CHAPTER 4

MIXED-EFFECTS REGRESSION MODELS FOR CONTINUOUS OUTCOMES

4.1 INTRODUCTION

The previous chapters have considered traditional analysis of variance methods for longitudinal data analysis. Unfortunately, these traditional methods are of limited use because of restrictive assumptions concerning missing data across time and the variance–covariance structure of the repeated measures. The univariate "mixed-model" analysis of variance assumes that the variances and covariances of the dependent variable across time are equal (*i.e.*, compound symmetry). Alternatively, the multivariate analysis of variance for repeated measures only includes subjects with complete data across time. Also, these procedures focus on estimation of group trends across time and provide little help in understanding about how specific individuals change across time. For these and other reasons, mixed-effects regression models (MRMs) have become popular for modeling longitudinal data.

Variants of MRMs have been developed and described under a variety of names: random-effects models [Laird and Ware, 1982], variance component models [Dempster et al., 1981], multilevel models [Goldstein, 1995], hierarchical linear models [Raudenbush and Bryk, 2002], two-stage models [Bock, 1989], random coefficient models [de Leeuw and Kreft, 1986], mixed models [Longford, 1987; Wolfinger, 1993], empirical Bayes models [Hui and Berger, 1983; Strenio et al., 1983], and random regression models [Bock, 1983a,b; Gibbons et al., 1988, 1993]. A basic characteristic of these models is the inclusion of random subject effects into regression models in order to account for the influence of subjects on their repeated observations. These random subject effects thus describe each person's trend across time, and explain the correlational structure of the longitudinal data.

Longitudinal Data Analysis. By Donald Hedeker and Robert D. Gibbons

Additionally, they indicate the degree of subject variation that exists in the population of subjects.

There are several features that make MRMs especially useful in longitudinal research. First, subjects are not assumed to be measured on the same number of timepoints, thus, subjects with incomplete data across time are included in the analysis. The ability to include subjects with incomplete data across time is an important advantage relative to procedures that require complete data across time because (a) by including all data, the analysis has increased statistical power, and (b) complete-case analysis may suffer from biases to the extent that subjects with complete data are not representative of the larger population of subjects. Because time is treated as a continuous variable in MRMs, subjects do not have to be measured at the same timepoints. This is useful for analysis of longitudinal studies where follow-up times are not uniform across all subjects. Both time-invariant and time-varying covariates can be included in the model. Thus, changes in the outcome variable may be due to both stable characteristics of the subject (*e.g.*, their gender or race) as well as characteristics that change across time (*e.g.*, life-events). Finally, whereas traditional approaches estimate average change (across time) in a population, MRM can also estimate change for each subject. These estimates of individual change across time can be particularly useful in longitudinal studies where a proportion of subjects exhibit change across time that deviates from the average trend.

Applications of MRMs are steadily increasing and can be found in many different fields, including studies on alcohol [Curran et al., 1997], smoking [Niaura et al., 2002], HIV/AIDS [Gallagher et al., 1997], drug abuse [Carroll et al., 1994; Halikas et al., 1997], psychiatry [Elkin et al., 1995; Serretti et al., 2000], and child development [Huttenlocher et al., 1991; Campbell and Hedeker, 2001], to name a few. Not only do these articles illustrate the wide applicability of MRMs, they also give a sense of how MRM results are typically reported in the various literatures.

This chapter will focus on describing MRMs for continuous outcomes in a very practical way. We will first illustrate how MRMs can be seen as an extension of an ordinary linear regression model. Starting with a simple linear regression model, the model will slowly be extended and described, in order to guide the reader going from familiar to less familiar territory. Following the descriptions of the statistical models, several MRM analyses will be presented using a longitudinal psychiatric dataset. These analyses will illustrate many of the key features of MRMs for longitudinal data analysis. For further illustration, readers are reminded that they can download the dataset and program files to replicate the analyses in this chapter.[1]

4.2 A SIMPLE LINEAR REGRESSION MODEL

To introduce MRMs, consider a simple linear regression model for the measurement y of individual i ($i = 1, 2, \ldots, N$ subjects) on occasion j ($j = 1, 2, \ldots, n_i$ occasions):

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij}. \tag{4.1}$$

Ignoring subscripts, this model represents the regression of the outcome variable y on the independent variable time (denoted t). The subscripts keep track of the particulars of the data, namely whose observation it is (subscript i) and when was this observation made

[1] http://www.uic.edu/~hedeker/long.html

(the subscript j). The independent variable t gives a value to the level of time, and may represent time in weeks, months, etc. Since y and t carry both i and j subscripts, both the outcome variable and the time variable are allowed to vary by individuals and occasions.

In linear regression models, like (4.1), the errors ε_{ij} are assumed to be normally and *independently* distributed in the population with zero mean and common variance σ^2. This independence assumption makes the model given in equation (4.1) an unreasonable one for longitudinal data. This is because the outcomes y are observed repeatedly from the same individuals, and so it is much more reasonable to assume that errors within an individual are correlated to some degree. Furthermore, the above model posits that the change across time is the same for all individuals since the model parameters (β_0, the intercept or initial level, and β_1, the linear change across time) do not vary by individuals. For both of these reasons, it is useful to add individual-specific effects into the model that will account for the data dependency and describe differential time trends for different individuals. This is precisely what MRMs do. The essential point is that MRMs therefore can be viewed as augmented linear regression models.

4.3 RANDOM INTERCEPT MRM

A simple extension of the regression model given in (4.1) to allow for the influence of each individual on their repeated outcomes is provided by

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + \varepsilon_{ij}, \tag{4.2}$$

where v_{0i} represents the influence of individual i on his/her repeated observations. Notice that if individuals have no influence on their repeated outcomes, then all of the v_{0i} terms would equal 0. However, it is more likely that subjects will have positive or negative influences on their longitudinal data, and so the v_{0i} terms will deviate from 0.

To better reflect how this model characterizes an individual's influence on their observations, it is helpful to represent the model in a hierarchical or multilevel form [Goldstein, 1995; Raudenbush and Bryk, 2002]. For this, it is partitioned into the following within-subjects (or level-1) model,

$$y_{ij} = b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}, \tag{4.3}$$

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

$$\begin{aligned} b_{0i} &= \beta_0 + v_{0i}, \\ b_{1i} &= \beta_1. \end{aligned} \tag{4.4}$$

Here, the level-1 model indicates that individual i's response at time j is influenced by his/her initial level b_{0i} and time trend, or slope, b_{1i}. The level-2 model indicates that individual i's initial level is determined by the population initial level β_0, plus a unique contribution for that individual v_{0i}. Thus, each individual has their own distinct initial level. Conversely, the present model indicates that each individual's slope is the same; all are equal to the population slope β_1. Another way to think about it is that each person's trend line is parallel to the population trend determined by β_0 and β_1. The difference

between each individual's trend and the population trend is v_{0i}, which is constant across time.

The between-subjects, or level-2, model is sometimes referred to as a "slopes as outcomes" model [Burstein, 1980]. The hierarchical representation shows that just as within-subjects (level-1) covariates can be included in the model to explain variation in level-1 outcomes (y_{ij}), between-subjects (level-2) covariates can be included to explain variation in level-2 outcomes (the subject's intercept b_{0i} and slope b_{1i}). Note that combining the within- and between–subjects models (4.3) and (4.4) yields the previous single-equation model (4.2).

Since individuals in a sample are typically thought to be representative of a larger population of individuals, the individual-specific effects v_{0i} are treated as random effects. That is, v_{0i} are considered to be representative of a distribution of individual effects in the population. The most common form for this population distribution is the normal distribution with mean 0 and variance σ_v^2. In the model given by equation (4.2), the errors ε_{ij} are now assumed to be normally and *conditionally independently* distributed in the population with zero mean and common variance σ^2. Conditional independence here means conditional on the random individual-specific effects v_{0i}. Since the errors now have an influence due to individuals removed from them, this conditional independence assumption is much more reasonable than the ordinary independence assumption associated with (4.1).

As mentioned, individuals deviate from the regression of y on t in a parallel manner in this model (since there is only one subject effect v_{0i}). Thus, it is sometimes referred to as a random-intercept model, with each v_{0i} indicating how individual i deviates from the population trend. Figure 4.1 represents this model graphically.

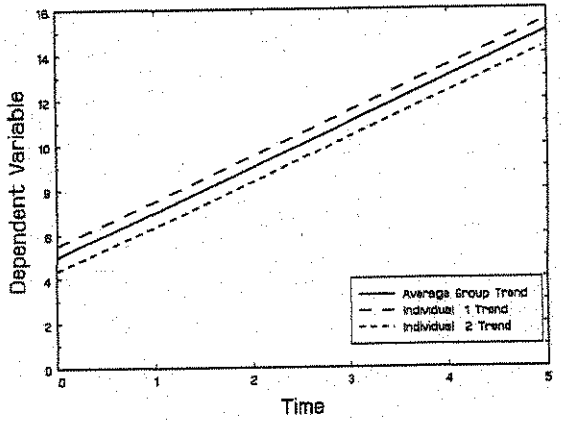

Figure 4.1. Random-intercept MRM.

In this figure, the solid line represents the population average trend, which is based on β_0 and β_1. Also depicted are two individual trends, one below and one above the population

(average) trend. For a given sample there are N such lines, one for each individual. The variance term σ_v^2 represents the spread of these lines. If σ_v^2 is near-zero, then the individual lines would not deviate much from the population trend. In this case, individuals do not exhibit much heterogeneity in their change across time. Alternatively as individuals differ from the population trend, the lines move away from the population trend line and σ_v^2 increases. In this case, there is more individual heterogeneity in time trends.

4.3.1 Incomplete Data Across Time

The occasions range from $j = 1$ to n_i in the model specification, with each person being measured on n_i timepoints. Since n carries the i subscript, each subject may vary in terms of the number of measured occasions. Furthermore, there are no restrictions on the number of observations per individual, subjects who are missing at a given timepoint are not excluded from the analysis. Also, since the time variable t carries the i subscript, subjects can be measured on different occasions. It is even possible that there are no "common" timepoints or "waves" for measurements; each individual could be measured on an individualized schedule. The underlying assumption of the model is that the data that are available for a given individual are representative of how that individual deviates from the population trend across the timeframe of the study.

Chapter 14 will discuss missing data issues more thoroughly. For now, a few points are worth mentioning. As Laird [1988] points out, MRMs for longitudinal data using maximum likelihood (ML) estimation provide valid statistical tests in the presence of ignorable nonresponse. By ignorable nonresponse, it is meant that the probability of nonresponse can depend on observed covariates (*e.g.*, time) *and* observed values of the dependent variable from the subjects with missing data. The notion here is that if missingness is related to observed performance (*i.e.*, observed values of the dependent variable), in addition to other observable subject characteristics (*i.e.*, observed covariates), then MRMs provide valid statistical inferences for the model parameters. This is a very useful result, because many instances of missing data can be assumed to be related to observed performance or other subject characteristics. Thus, MRMs provide an attractive method for dealing with incomplete longitudinal data.

In considering missing data and whether they are ignorable or not, a related issue is the distinction between attrition (*i.e.*, subjects dropping out of the study and not returning) and sporadic or intermittent missing data (*i.e.*, subjects with missing data between observed timepoints). It may very well be that these arise from distinct processes and so, for example, it might be more plausible to assume ignorable missingness for intermittent missing data rather than attrition. Chapter 14 will discuss these issues in greater detail and also describe methods of analysis not requiring ignorable nonresponse.

4.3.2 Compound Symmetry and Intraclass Correlation

The random intercept model implies a compound symmetry assumption for the variances and covariances of the longitudinal data. That is, both the variances and covariances across time are assumed to be the same, namely,

$$\begin{aligned} V(y_{ij}) &= \sigma_v^2 + \sigma^2, \\ C(y_{ij}, y_{ij'}) &= \sigma_v^2, \qquad \text{where } j \neq j'. \end{aligned} \tag{4.5}$$

Expressing the covariance as a correlation yields the *intraclass correlation*, which is the ratio of the individual variance σ_v^2 to the total variance $\sigma^2 + \sigma_v^2$. This coefficient represents the degree of association of the longitudinal data within subjects, and specifically indicates the proportion of variance in the data attributable to individuals.

As an important caveat, it should be noted that a random intercept model that includes autocorrelated errors, as described in Chapter 7, provides a variance–covariance structure that is more general than the above compound symmetry structure. So, to be precise, it is a random intercept model with independent errors that implies compound symmetry.

4.3.3 Inference

Hypothesis testing for the fixed-effects parameters (*i.e.*, β) generally involves the so-called "Wald test" [Wald, 1943], which uses the ratio of parameter estimate to its standard error to determine statistical significance. These test statistics (*i.e.*, Z = ratio of the parameter estimate to its standard error) are compared to a standard normal frequency table to test the null hypothesis that the parameter equals 0. Alternatively, these Z-statistics are sometimes squared, in which case the resulting test statistic is distributed as chi-square on one degree of freedom. In either case, the p-values are identical.

For the variance and covariance terms, there are concerns in using the standard errors in constructing Wald test statistics particularly when the population variance is thought to be near zero and the number of subjects is small [Bryk and Raudenbush, 1992]. This is because variance parameters are bounded; they cannot be less than zero and so using the standard normal for the sampling distribution is not reasonable. As a result, in this text we will not include the Wald tests for variance and covariance terms.

For nested models, the likelihood ratio test can be used to perform uni- or multi-parameter hypothesis tests. For this, one compares the model deviance values (*i.e.*, $-2 \log L$) to a chi-square distribution, where the degrees of freedom equals the number of parameters set equal to zero in the more restrictive model. It should be noted that while use of the likelihood ratio test for fixed effects is not problematic, for variance and covariance terms this test also suffers from the variance boundary problem mentioned above [Verbeke and Molenberghs, 2000]. Based on simulation studies it can be shown that the likelihood ratio test is too conservative (for testing null hypotheses about variance and covariance parameters), namely, it does not reject the null hypothesis often enough. This would then lead 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 dividing the p-value obtained from the likelihood ratio test (of variance and covariance parameters) by two.

4.3.4 Psychiatric Dataset

Throughout this and later chapters, we will consider data from a psychiatric study described in Reisby et al. [1977]. This study focused on the longitudinal relationship between imipramine (IMI) and desipramine (DMI) plasma levels and clinical response in 66 depressed inpatients. Imipramine is the prototypic drug in the series of compounds known as tricyclic antidepressants, and is commonly prescribed for the treatment of major depression [Seiden and Dykstra, 1977]. Since imipramine biotransforms into the active metabolite desmethylimipramine (or desipramine), measurement of desipramine was also done in this study. Major depression is often classified in terms of two types. The first type, nonendogenous or reactive depression, is associated with some tragic life event such as the death of a close friend or family member, whereas the second type, endogenous depression, is not a result of any specific event and appears to occur spontaneously. It is sometimes held that antidepressant medications are more effective for endogenous depression [Willner, 1985]. In this sample, 29 patients were classified as nonendogenous and the remaining 37 patients were deemed to be endogenous.

The study design was as follows. Following a placebo period of 1 week, patients received 225 mg/day doses of imipramine for four weeks. In this study, subjects were rated with the Hamilton Depression Rating Scale (HDRS) [Hamilton, 1960] twice during the baseline placebo week (at the start and end of this week) as well as at the end of each of the four treatment weeks of the study. These HDRS scores represent the dependent variable that is measured across time. Higher scores on the HDRS represent higher levels of depression and lower scores indicate less depression. Plasma level measurements of both IMI and its metabolite DMI were made at the end of each week; these will be treated as time-varying covariates. The sex and age of each patient was recorded and a diagnosis of endogenous or nonendogenous depression was made for each patient. These time-invariant (*i.e.*, individual-level) variables are all potential covariates, though our analyses will only focus on diagnosis.

Although the total number of subjects in this study was 66, the number of subjects with all measures at each of the weeks fluctuated: 61 at week 0 (start of placebo week), 63 at week 1 (end of placebo week), 65 at week 2 (end of first drug treatment week), 65 at week 3 (end of second drug treatment week), 63 at week 4 (end of third drug treatment week), and 58 at week 5 (end of fourth drug treatment week). Of the 66 subjects, only 46 had complete data at all timepoints. Thus, complete-case analysis under repeated measures MANOVA, for example, would discard approximately one-third of the dataset. MRM, alternatively, uses the data that are available from all 66 subjects.

Table 4.1 presents observed HDRS means, standard deviations, and sample sizes across the six study timepoints. Because the HDRS means are decreasing across time, there appears to be consistent improvement across time. Additionally, it is clear that the standard deviations are increasing across time. There is more spread in HDRS scores as time goes by. This is reasonable because some patients likely improved across time, to varying degrees, while others did not.

Table 4.1. Observed HDRS Means, Standard Deviations (sd), and n Across Time

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5
Mean	23.44	21.84	18.31	16.42	13.62	11.95
sd	4.53	4.70	5.49	6.42	6.97	7.22
n	61	63	65	65	63	58

Correlations, both pairwise and listwise, of the repeated HDRS outcomes are given in Table 4.2. Note that pairwise correlations are calculated based on all available data for a given pair of variables (*e.g.*, all subjects with week 0 and week 1 measurements are included in the calculation of the week 0 and week 1 correlation), whereas listwise correlations requires subjects to have complete data on all variables (*i.e.*, only subjects with data at all six timepoints are included). The correlations follow the commonly seen pattern of diminishing in value as one goes further away from the diagonal. This is true for both

the pairwise and listwise correlations, which are similar. Also, within a given time lag, it appears that the association may increase across time. For example, for the time lag of one week, the pairwise correlation equals .49 at the beginning (*i.e.*, week 0 and week 1 correlation) and increases to .65 at the end (*i.e.*, week 4 and week 5 correlation). Taken together, these data do not appear to satisfy a compound symmetry assumption of equal variances and covariances across time.

Table 4.2. Observed HDRS Correlations: **Listwise (n = 46)** and *Pairwise (46 ≤ n ≤ 66)*

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5
Week 0	1.0	*.49*	*.41*	*.33*	*.23*	*.18*
Week 1	**.49**	1.0	*.49*	*.41*	*.31*	*.22*
Week 2	**.42**	**.49**	1.0	*.74*	*.67*	*.46*
Week 3	**.44**	**.51**	**.73**	1.0	*.82*	*.57*
Week 4	**.30**	**.35**	**.68**	**.78**	1.0	*.65*
Week 5	**.22**	**.23**	**.53**	**.62**	**.72**	1.0

Figure 4.2 presents the so-called "spaghetti plot" of the data. This plot is obtained by constructing a scatterplot of the HDRS scores by time, and then connecting the dots of each individual's data across time. This plot is useful for assessing overall aspects of the data. For example, the plot above suggests that there is a general linear decline in the HDRS scores across time, though clearly there is considerable individual heterogeneity in this. Also, the plot clearly shows the increasing variance in HDRS scores across time.

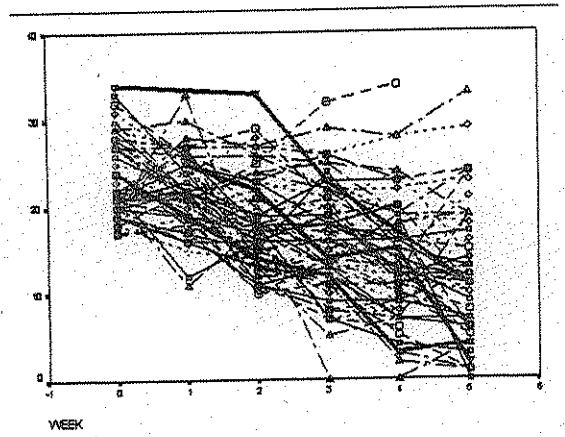

Figure 4.2. Reisby data: Spaghetti plot of observed data.

4.3.5 Random Intercept Model Example

The first model fit to these data corresponds to within-subjects model (4.3) and between-subjects model (4.4). This is the random-intercept model with only time as a regressor, where time is treated using incremental values from 0 to 5. Though the descriptive statistics above indicate that the compound symmetry assumption of this model is very dubious, we will fit this simple model here as a starting point in our examination of these data. Table 4.3 presents the results from this analysis using maximum likelihood (ML) estimation.

Table 4.3. MRM Results for Level-1 Model (4.3) and Level-2 Model (4.4)

Parameter	Estimate	SE	Z	$p <$
β_0	23.55	0.64	36.80	.0001
β_1	-2.38	0.14	-17.00	.0001
$\sigma^2_{y_0}$	16.15	3.41		
σ^2	19.04	1.53		

Note. $-2 \log L = 2285.19$. SE = standard error.

Focusing first on the estimated regression parameters, this model indicates that patients start off, on average, with a HDRS score of 23.55 and change by -2.38 points each week. Lower scores on the HDRS reflect less depression, so patients are improving across time by a little over 2 points per week. Both the intercept and slope are statistically significant ($p < .0001$) in this analysis. The intercept being significant is not particularly meaningful; it just indicates that HDRS scores are different than zero at baseline. However, because the slope is significant, the rate of improvement is significantly different from zero based on this analysis. On average, patients are improving across time.

The model estimates can be used to generate estimated values of the mean HDRS scores across time. Specifically, $\hat{y} = 23.552 - 2.376$ Week. These are displayed in Table 4.4 along with the observed means.

Table 4.4. Observed and Estimated Means

	Week					
	0	1	2	3	4	5
Observed	23.44	21.84	18.31	16.42	13.62	11.95
Estimated	23.55	21.18	18.80	16.42	14.05	11.67

Comparing the estimated to observed means in Table 4.4 indicates excellent model fit of these marginal means. Thus, overall it appears that the change across time in HDRS scores is linear. In their report, Reisby et al. [1977] classified patients into three groups based on their final HDRS scores: responders had scores below 8, partial responders were between 8 and 15, and nonresponders had final HDRS scores above 15. By this criteria, the estimated average trend is in the partial response range at the final timepoint (*i.e.*, = 11.67). For a more quantitative assessment of model fit, the interested reader is referred to Kaplan and

George [1998], which describes use of econometric forecasting statistics to assess various forms of fit between observed and estimated means.

The model fit of the variances and covariances can also be examined. Here, the estimated variance, which is assumed to be constant over time, is 16.15+19.04=35.19, or expressed as a standard deviation it yields 5.93. Since the observed standard deviations displayed in Table 4.1 clearly increase across time, this estimate of constant variance is an oversimplification. Turning to the correlations of the repeated measures, the intraclass correlation here equals $r = 16.15/(16.15+19.04) = .46$, which indicates that 46% of the unexplained variance in HDRS scores (*i.e.*, that part of the HDRS scores not explained by the linear effect of week) is at the individual level. Thus, subjects display considerable heterogeneity in depression levels. Comparing this value of .46 to the correlation matrix in Table 4.2 again suggests that this model is an oversimplification; while the average of the correlations might be approximately .46, there is considerable variation in these correlations and so assuming that they are all the same does not appear to be reasonable.

As a final point of comparison, note that performing an OLS simple linear regression on these data, as described in Section 4.2 ignoring the clustering of observations within subjects, yields $\hat{\beta}_0 = 23.60$ (SE = .55), $\beta_1 = -2.41$ (SE = .18), and $\hat{\sigma}^2 = 35.40$ (SE = 2.59) [a similar analysis using ML estimation yields the same regression results and $\hat{\sigma}^2 = 35.21$ with SE = 2.57]. Thus, our results from the random-intercept model are in very close agreement with these, though of course the standard errors are considerably different. What is very interesting to note is that what the ordinary regression model lumped together into error variance (35.40 or 35.21), the random-intercept model separates into within-subjects and between-subjects variances (19.04 and 16.15, respectively). This illustrates a golden rule of statistics: One statistician's error term is another's career!

4.4 RANDOM INTERCEPT AND TREND MRM

For longitudinal data, the random intercept model is often too simplistic for a number of reasons. First, it is unlikely that the rate of change across time is the same for all individuals. It is more likely that individuals differ in their time trends; not everyone changes at the same rate. Furthermore, the compound symmetry assumption of the random intercept model is usually untenable for most longitudinal data. In general, measurements at points close in time tend to be more highly correlated than measurements further separated in time. Also, in many studies, subjects are more similar at baseline, and they grow at different rates across time. Thus, it is natural to expect that variability will increase over time.

For these reasons, a more realistic MRM allows both the intercept and time trend to vary by individuals. For this, the level-1 model is as before in (4.3), but the level-2 model is augmented as

$$\begin{aligned} b_{0i} &= \beta_0 + \upsilon_{0i}, \\ b_{1i} &= \beta_1 + \upsilon_{1i}. \end{aligned} \tag{4.6}$$

In this model, β_0 is the overall population intercept, β_1 is the overall population slope, υ_{0i} is the intercept deviation for subject i, and υ_{1i} is the slope deviation for subject i. As before, ε_{ij} is an independent error term distributed normally with mean 0 and variance σ^2. The assumption regarding the independence of the errors is one of conditional independence, that is, they are independent conditional on υ_{0i} and υ_{1i}. With two random individual-specific effects, the population distribution of intercept and slope deviations is assumed to be bivariate normal $\mathcal{N}(0, \Sigma_\upsilon)$, with the random-effects variance–covariance matrix as

$$\Sigma_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0\upsilon_1} \\ \sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_1} \end{bmatrix}.$$

This model can be thought of as a personal trend or change model since it represents the measurements of y as a function of time, both at the individual (υ_{0i} and υ_{1i}) and population (β_0 and β_1) levels. The intercept parameters indicate the starting point, and the slope parameters indicate the degree of change over time. The population intercept and slope parameters represent the overall (population) trend, while the individual parameters express how subjects deviate from the population trend.

Figure 4.3 represents this model graphically. Again, the figure represents the population trend with the solid line and the trends from two individuals, who now deviate both in terms of the intercept and slope. Because the slope varies for individuals, this model allows the possibility that some individuals do not change across time, while others can exhibit dramatic change. The population trend is the average across the individuals and the variance terms indicate how much heterogeneity there is in the population. Specifically, the variance term $\sigma^2_{\upsilon_0}$ indicates how much spread there is around the population intercept, and $\sigma^2_{\upsilon_1}$ represents the spread in slopes. To the degree that each individual's deviation from the population trend is only due to random error, these variance terms will approach zero. Alternatively, as each individual's deviation from the population trend is nonrandom, but characterized by the individual trend parameters υ_{0i} and υ_{1i} as being nonzero, these variance terms will increase from zero. Additionally, the covariance term, $\sigma_{\upsilon_0\upsilon_1}$, represents the degree to which the individual intercept and slope parameters covary. For example, a positive covariance term would suggest that individuals with higher initial values have greater positive slopes, while a negative covariance would suggest the opposite.

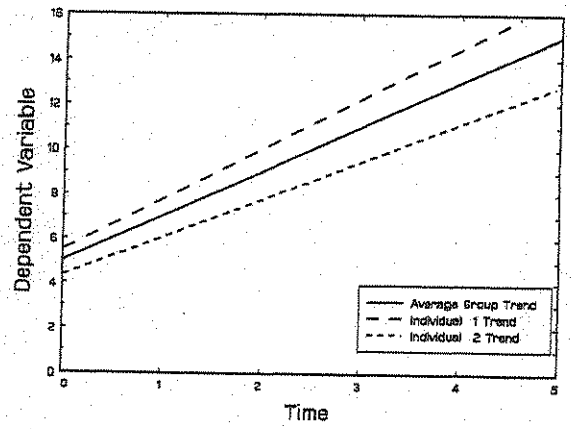

Figure 4.3. Random intercept and trend MRM.

4.4.1 Random Intercept and Trend Example

Continuing with our psychiatric example, we will fit the within-subjects model (4.3) and between-subjects model (4.6). As before, time is treated using incremental values from 0 to 5. The ML results are presented in Table 4.5.

Table 4.5. MRM Results for Level-1 Model (4.3) and Level-2 Model (4.6)

Parameter	Estimate	SE	Z	$p <$
β_0	23.58	0.55	43.22	.0001
β_1	−2.38	0.21	−11.39	.0001
$\sigma^2_{v_0}$	12.63	3.53		
$\sigma_{v_0 v_1}$	−1.42	1.04		
$\sigma^2_{v_1}$	2.08	0.52		
σ^2	12.22	1.12		

Note. $-2 \log L = 2219.04$. SE = standard error.

The results for the regression coefficients are very similar to the previous random-intercept analysis of these data. This model indicates that patients start off, on average, with a HDRS score of 23.58 and change by −2.38 points each week. As before, the effect of time is significant, and so we can conclude that the rate of improvement is significantly different from zero in this study.

For the variance and covariance terms, statistical significance is not indicated in the table because of the problem with use of the Wald test for these parameters (discussed in Section 4.3.3). However, the magnitude of the estimates does reveal the degree of individual heterogeneity in both the intercepts and slopes. For example, while the average intercept in the population is estimated to be 23.58, the estimated population standard deviation for the intercept is 3.55 ($= \sqrt{12.63}$). Similarly, the average population slope is −2.38, but the estimated population standard deviation for the slope equals 1.44 ($= \sqrt{2.08}$), and so approximately 95% of subjects in the population are expected to have slopes in the interval $-2.38 \pm (1.96 \times 1.44) = -5.20$ to .44. That the interval includes positive slopes reflects the fact that not all subjects improve across time. Thus, there is considerable heterogeneity in terms of patients' initial level of depression and in their change across time. Finally, the covariance between the intercept and linear trend is negative; expressed as a correlation it equals −.28, which is moderate in size. This suggests that patients who are initially more depressed (*i.e.*, greater intercepts) improve at a greater rate (*i.e.*, more pronounced negative slopes). An alternative explanation, though, is that of a floor effect due to the HDRS rating scale. Simply put, patients with less depressed initial scores have a more limited range of lower scores than those with higher initial scores.

An interesting question, at this point, is whether the between-subjects model in equation (4.6) is necessary over that in equation (4.4). In other words, is the assumption of compound symmetry rejected or not. Because these are nested models, they can be compared using a likelihood ratio test, albeit with the caveat that because the testing involves variance terms the p-value obtained from the likelihood ratio test should be divided by two. In the present case, the difference in model deviance values equals $2285.19 - 2219.04 = 66.15$ on 2 degrees of freedom (the 2 degrees of freedom are for the addition of the slope variance and the slope-intercept covariance), and so the p-value is less than .001 regardless. Thus, there is clear evidence that the assumption of compound symmetry is rejected.

Finally, empirical Bayes estimates of the individual random effects, $\hat{b}_{0i}$ and $\hat{b}_{1i}$, are often of interest. The derivation of these estimates will be described later. For now, these are plotted in Figure 4.4. The dashed lines indicate the estimated population intercepts and slopes. Thus, $\hat{v}_{0i}$ is represented by the vertical distance between a point and the horizontal line, while $\hat{v}_{1i}$ is represented by the horizontal distance between a point and the vertical line. This scatterplot reveals the wide range of observed intercepts and slopes in this sample. In particular, there are some patients who are very depressed initially but who improve to a great degree (upper left-hand corner). Similarly, there are some patients who show little or no improvement over time (towards the right side).

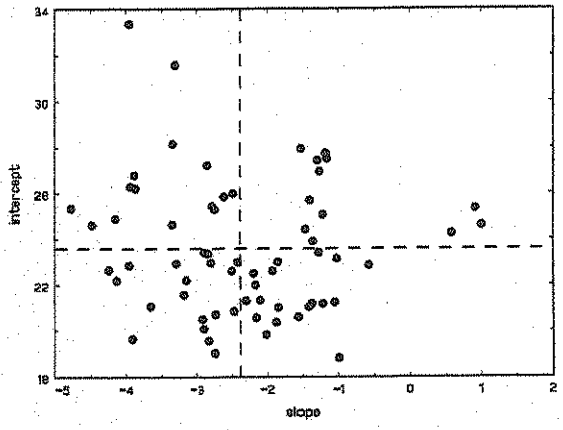

Figure 4.4. Reisby data: Estimated random effects.

4.4.2 Coding of Time

The coding of the time variable t has implications for the interpretation of the model parameters. For example, as in our Reisby analysis, t can start with the value 0 for baseline and be incremented according to the measurement timeline (*e.g.*, 1, 2, 3, 4, and 5 for the weekly follow-ups). In this formulation, the intercept parameters (β_0, v_{0i}, and $\sigma^2_{v_0}$) characterize aspects of the baseline timepoint. Alternatively, t can be expressed in centered form, where the average of time is subtracted from each time value (*e.g.*, −2.5, −1.5, −.5, .5, 1.5, 2.5). In this case, the meaning of the intercept parameters changes to reflect aspects about the midpoint of time, and not the baseline timepoint.

Figure 4.5 represents how the "intercept" variance $\sigma^2_{v_0}$ can change dramatically between baseline and centered codings of time. In the former, $\sigma^2_{v_0}$ represents the degree of individual heterogeneity at time 0, whereas in the latter it would represent heterogeneity at week 2.5 (the center of time). The figure portrays the average trend across time (the solid line)

and two individual trends (the dot-dashed lines); the latter are meant to reflect the range of trends in the population of individuals (for an actual dataset there would be N such individual trends). The normal distributions on the figure represent the degree of individual heterogeneity at time equal to 0 and 2.5. Notice that the spread in the normal distribution, and thus $\sigma^2_{\upsilon_0}$, is much greater when time equals its midpoint relative to its baseline value. Thus, analysis of the same dataset would yield very different estimates of $\sigma^2_{\upsilon_0}$ if baseline-incremented versus centered coding of time was used.

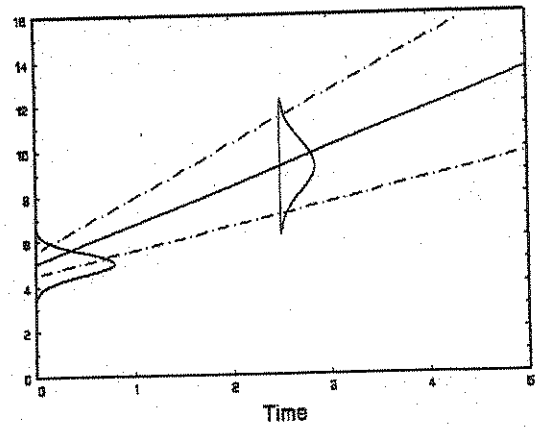

Figure 4.5. Intercept variance changes with coding of time.

As yet another coding choice, sometimes substantive interest focuses on the end of the measurement timeline. Here, time could be coded as -5, -4, -3, -2, -1, and 0 (in this case with six timepoints), so that the intercept parameters reflect aspects of the final timepoint. The choice of which representation to use often depends on ease of interpretation and the hypotheses of interest. It is important, though, that careful consideration is given to what is treated as "zero time," because all of the intercept parameters are effected by this choice. In particular, the zero time value should, in general, be one that is within the range of the observed data. Otherwise the intercept parameters represent extrapolations (in time) of the data. For example, a common mistake is to code time sequentially as 1, 2, 3, ..., n. In such a model, with time included, the intercept parameters represent a timepoint that is one unit *before* the first study timepoint. Clearly, one is typically not interested in trying to estimate what happened before the study began. For the interested reader, Biesanz et al. [2004], Horwitz et al. [1990], and Mehta and West [2000] describe this issue of zero time, and related issues, in great detail.

4.4.2.1 Example To illustrate the effect that the coding of time has, we reran the random intercept and trend model using a centered version of week, namely weekc = week $-$ 2.5. The ML results for this analysis are presented in Table 4.6. Comparing these results to those in Table 4.5 illustrates the effect that the coding of time has. First, notice that the deviance values and slope estimates are identical. This includes both the estimate of average slope $\hat{\beta}_1$ and the heterogeneity in slopes $\hat{\sigma}^2_{\upsilon_1}$. The slope estimates haven't changed because the scale of the time variable has not changed. However, the parameters involving the location of the time variable have changed because the origin of this variable has shifted. Specifically, the estimate of the intercept is now 17.63, which corresponds to the average HDRS level when time equals 2.5 weeks (at the center of time). This intercept estimate is less than its counterpart in Table 4.5 because individuals are, on average, improving across time. Likewise, the intercept variance estimate of 18.52 reflects the degree of individual heterogeneity at 2.5 weeks. This value is greater than its counterpart in Table 4.5, indicating that the phenomenon illustrated in Figure 4.5 is occurring for these data. Namely, subjects are more alike at the beginning of the study than at the middle of the study.

Table 4.6. MRM Results for Level-1 Model (4.3) and Level-2 Model (4.6) with Centered Week

Parameter	Estimate	SE	Z	$p <$
β_0	17.63	0.56	31.47	.0001
β_1	-2.38	0.21	-11.39	.0001
$\sigma^2_{\upsilon_0}$	18.52	3.62		
$\sigma_{\upsilon_0\upsilon_1}$	3.78	1.08		
$\sigma^2_{\upsilon_1}$	2.08	0.52		
σ^2	12.22	1.12		

Note. $-2 \log L = 2219.04$. SE = standard error.

Finally, the covariance has not only changed values, but signs as well. How could this happen? Easier than one might at first imagine. Notice that from Table 4.5, the interpretation of the negative covariance would be that subjects with higher initial HDRS values have more negative slopes across time. Thus, subjects who are very depressed initially improve at a greater rate than those who are not so depressed to begin with. Turning to the centered results in Table 4.6, the positive covariance suggests that subjects with higher mid-study HDRS values have less negative (or more positive) slopes across time. In other words, subjects who are more depressed at mid-study are those that have improved less than subjects with lower mid-study depression levels. Clearly both of these interpretations are reasonable and consistent with each other; Figure 4.6 illustrates the spaghetti plot of two subjects following these patterns. The two dashed horizontal lines in the figure indicate the mean HDRS at week 0 (23.6) and week 2.5 (17.6). As the figure shows, the subject with the above average HDRS value at week 0 has the more negative slope, while the subject with the above average HDRS value at week 2.5 has the less negative (or more positive) slope. Thus, this example has highlighted the fact that correct model interpretation depends on an understanding of the coding of the variables in the analysis.

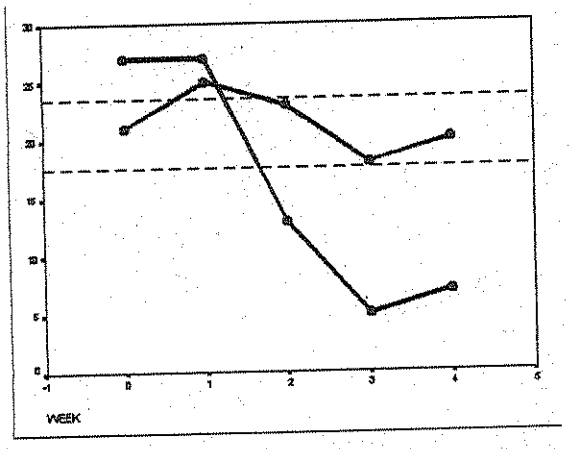

Figure 4.6. Spaghetti plot of two subjects supporting differential interpretation of intercept slope covariance.

4.4.3 Effect of Diagnosis on Time Trends

At this point, it may be interesting to examine whether we can explain some of the heterogeneity in intercepts and slopes, depicted earlier in Figure 4.4, in terms of particular subject characteristics. For example, in this study it may be that a subject's diagnosis (endogenous versus nonendogenous depression) is related to their initial depression level and change across time. Preparing for this analysis, note the observed HDRS means across time stratified by diagnostic group in Table 4.7.

Table 4.7. Observed HDRS Means and n Across Time Stratified by Group

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5
Endogenous	24.0	23.0	19.3	17.3	14.5	12.6
n	33	34	37	36	34	31
Nonendogenous	22.8	20.5	17.0	15.3	12.6	11.2
n	28	29	28	29	29	27

As the means indicate, both groups are clearly improving across time, though the endogenous group is consistently higher than the nonendogenous group. To explore this, we will augment the level-2 model to include a covariate DX which equals 0 if the patient's diagnosis is nonendogenous (NE) and equals 1 if the patient is endogenous (E). This variable

enters the level-2 model rather than the level-1 model because it varies only with subjects (i) and not with time (j).

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_2 DX_i + v_{0i} \\ b_{1i} &= \beta_1 + \beta_3 DX_i + v_{1i}\,. \end{aligned} \tag{4.7}$$

Here, β_0 represents the average week 0 HDRS level for NE patients, and β_1 the average HDRS weekly improvement for NE patients. Similarly, β_2 represents the average week 0 HDRS difference for E patients (relative to NE patients) and β_3 the average difference in HDRS weekly improvement rates for E patients (relative to NE patients). Thus, β_3 represents the diagnosis by time interaction, indicating the degree to which the time trends vary by diagnostic group. In this augmented model, v_{0i} is the individual's deviation from their diagnostic group intercept and v_{1i} is the individual's deviation from their diagnostic group slope. To the degree that the variable DX is useful in explaining intercept and slope variation, these individual deviations and their corresponding variances, ($\sigma^2_{v_0}$ and $\sigma^2_{v_1}$), will be reduced. Results for this model are listed in Table 4.8.

A likelihood ratio test comparing this model to the previous one can be used to test the null hypothesis that the diagnosis-related effects (*i.e.*, β_2 and β_3) are zero. This yields $X^2_2 = 2219.04 - 2214.94 = 4.1$, which is not statistically significant. Inspection of the estimates in Table 4.8 reveals a marginally significant difference in terms of their initial scores, with endogenous patients about 2 points higher, and absolutely no difference in their trends across time. This is also borne out if one compares the variance estimates from Tables 4.5 and 4.8. Notice that the intercept variance has diminished slightly from to 12.63 to 11.64, as a result of the marginally significant intercept difference, whereas the slope variance is the same. Taken together, there is no real evidence that the two diagnostic groups differ in terms of their HDRS scores across time.

Table 4.8. MRM Results for Level-1 Model (4.3) and Level-2 Model (4.7)

Parameter	Estimate	SE	Z	$p <$
NE intercept β_0	22.48	0.79	28.30	.0001
NE slope β_1	−2.37	0.31	−7.59	.0001
E intercept difference β_2	1.99	1.07	1.86	.063
E slope difference β_3	−0.03	0.42	−0.06	.95
$\sigma^2_{v_0}$	11.64	3.53		
$\sigma_{v_0 v_1}$	−1.40	1.00		
$\sigma^2_{v_1}$	2.08	0.50		
σ^2	12.22	1.11		

Note. $-2\log L = 2214.94$.

Figure 4.7 illustrates the observed and estimated trends for these two groups. For the latter, these are simply computed as $22.48 - 2.37\,Week$ for the nonendogenous group, and $(22.48 + 1.99) + (-2.37 - 0.03)\,Week$ for the endogenous group. The figure helps to illustrate the conclusions of the analysis. As can be seen, there is only a marginal difference between the two groups that is consistent across time. Also, the observed and estimated means are in close agreement for both groups.

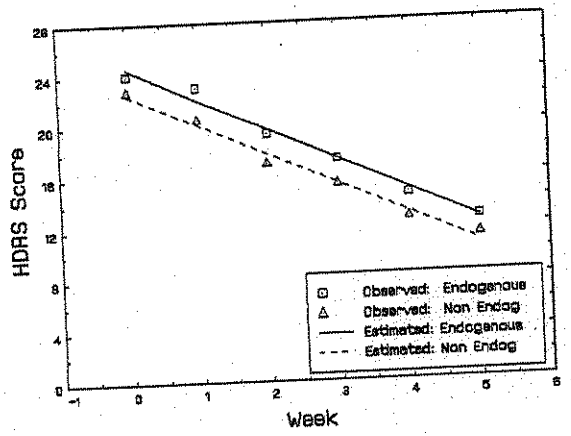

Figure 4.7. Reisby data: Estimated and observed means across time and diagnostic groups.

4.5 MATRIX FORMULATION

A more compact representation of the model is afforded using matrices and vectors. This formulation is particularly useful in summarizing statistical aspects of the model. For this, the MRM for the $n_i \times 1$ response vector $\boldsymbol{y}$ for individual i can be written as

$$\underset{n_i\times 1}{\boldsymbol{y}_i} = \underset{n_i\times p}{\boldsymbol{X}_i}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n_i\times r}{\boldsymbol{Z}_i}\ \underset{r\times 1}{\boldsymbol{\upsilon}_i} + \underset{n_i\times 1}{\boldsymbol{\varepsilon}_i} \tag{4.8}$$

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, $\boldsymbol{Z}_i$ is the $n_i \times r$ design matrix for the random effects, $\boldsymbol{\upsilon}_i$ is the $r \times 1$ vector of random individual effects, and $\boldsymbol{\varepsilon}_i$ is the $n_i \times 1$ error vector.

For example, in the random intercept and slope MRM just considered, we would have

$$\boldsymbol{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \cdots \\ \cdots \\ y_{in_i} \end{bmatrix} \quad \text{and} \quad \boldsymbol{X}_i = \boldsymbol{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \cdots & \cdots \\ \cdots & \cdots \\ 1 & t_{in_i} \end{bmatrix}$$

for the data matrices, and

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\upsilon}_i = \begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix}$$

for the population and individual trend parameter vectors, respectively. For the model including diagnosis and the diagnosis by time interaction, the data matrix for $\boldsymbol{X}$ would be changed to

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & t_{i1} & DX_i & DX_i \times t_{i1} \\ 1 & t_{i2} & DX_i & DX_i \times t_{i2} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 1 & t_{in_i} & DX_i & DX_i \times t_{in_i} \end{bmatrix},$$

while the $\boldsymbol{Z}$ matrix would be the same. The matrix representation does not really distinguish between person-varying (level-2) and time-varying (level-1) covariates, it's all $\boldsymbol{X}$. Thus, again, the multilevel or hierarchical representation of the model into the level-1 and level-2 submodels might give the impression that several models are simultaneously being estimated, but actually there is only one model (that is broken apart to aid in our interpretation of that model).

The distributional assumptions about the random effects and errors are

$$\begin{aligned} \boldsymbol{\varepsilon}_i &\sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_{n_i}), \\ \boldsymbol{\upsilon}_i &\sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\upsilon}). \end{aligned}$$

As a result, it can be shown that the observations $\boldsymbol{y}_i$ and random effects $\boldsymbol{\upsilon}_i$ have the joint multivariate normal distribution:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{\upsilon}_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{\Sigma}_{\upsilon}\boldsymbol{Z}_i' + \sigma^2\boldsymbol{I}_{n_i} & \boldsymbol{Z}_i\boldsymbol{\Sigma}_{\upsilon} \\ \boldsymbol{\Sigma}_{\upsilon}\boldsymbol{Z}_i' & \boldsymbol{\Sigma}_{\upsilon} \end{bmatrix} \right). \tag{4.9}$$

Using results from multivariate statistics, it can be further shown that the mean of the posterior distribution of $\boldsymbol{\upsilon}_i$, given $\boldsymbol{y}_i$, yields the formula for the empirical Bayes estimates of the random effects,

$$\hat{\boldsymbol{\upsilon}}_i = \left[\boldsymbol{Z}_i'(\sigma^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{Z}_i + \boldsymbol{\Sigma}_{\upsilon}^{-1}\right]^{-1} \boldsymbol{Z}_i'(\sigma^2\boldsymbol{I}_{n_i})^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}). \tag{4.10}$$

Similarly, the corresponding posterior covariance matrix is given by

$$\boldsymbol{\Sigma}_{\upsilon|\boldsymbol{y}_i} = \left[\boldsymbol{Z}_i'(\sigma^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{Z}_i + \boldsymbol{\Sigma}_{\upsilon}^{-1}\right]^{-1}. \tag{4.11}$$

Note that the variance–covariance matrix of the repeated measures $\boldsymbol{y}$ is of the form

$$V(\boldsymbol{y}_i) = \boldsymbol{Z}_i\boldsymbol{\Sigma}_{\upsilon}\boldsymbol{Z}_i' + \sigma^2\boldsymbol{I}_{n_i}. \tag{4.12}$$

As an example of what this equation implies, consider a model including a random intercept and time trend, $r = 2$, and a subject with three timepoints, $n = 3$. Then, their random-effects design matrix might be

$$\boldsymbol{Z}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix},$$

and so the variance–covariance matrix equals

$$\sigma^2 I_{n_i} + \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma^2_{\upsilon_0} + \sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_0} + 2\sigma_{\upsilon_0\upsilon_1} \\ \sigma^2_{\upsilon_0} + \sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_0} + 2\sigma_{\upsilon_0\upsilon_1} + \sigma^2_{\upsilon_1} & \sigma^2_{\upsilon_0} + 3\sigma_{\upsilon_0\upsilon_1} + 2\sigma^2_{\upsilon_1} \\ \sigma^2_{\upsilon_0} + 2\sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_0} + 3\sigma_{\upsilon_0\upsilon_1} + 2\sigma^2_{\upsilon_1} & \sigma^2_{\upsilon_0} + 4\sigma_{\upsilon_0\upsilon_1} + 4\sigma^2_{\upsilon_1} \end{bmatrix}.$$

Notice that this structure, which is based on a model with a random intercept and time trend, allows the variances and covariances to change across time. For example, if $\sigma_{\upsilon_0\upsilon_1}$ is positive, then clearly the variance increases across time. Diminishing variance across time is also possible if, for example, $-2\sigma_{\upsilon_0\upsilon_1} > \sigma^2_{\upsilon_1}$. Other patterns are possible depending on the values of these variance and covariance parameters.

Models with more than random intercepts and linear trends are also possible, as are models that allow autocorrelated errors, that is $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2\boldsymbol{\Omega}_i)$; these will be described in subsequent chapters. For now, note that by including both multiple random effects, and possibly autocorrelated errors, a wide range of variance–covariance structures for the repeated measures is possible. This flexibility is in sharp contrast to the traditional ANOVA models which assume either a compound symmetry structure (univariate ANOVA) or a totally general structure (MANOVA). Typically, compound symmetry is too restrictive and a general structure is not parsimonious. MRMs, alternatively, provide these two and everything in between, and so allow efficient modeling of the variance–covariance structure of the repeated measures. More discussion about this is included in Chapters 6 and 7.

4.5.1 Fit of Variance–Covariance Matrix

For the Reisby dataset considered in this chapter, it is of interest to consider model fit of the variances and covariances associated with the repeated outcomes. Below is the observed variance–covariance matrix for the six study timepoints. These are calculated based on the pairwise data for the covariances and the available data for each of the variances.

$$V(\boldsymbol{y}) = \begin{bmatrix} 20.55 & & & & & \\ 10.50 & 22.07 & & & & \\ 10.20 & 12.74 & 30.09 & & & \\ 9.69 & 12.43 & 25.96 & 41.15 & & \\ 7.17 & 10.10 & 25.56 & 36.54 & 48.59 & \\ 6.02 & 7.39 & 18.25 & 26.31 & 32.93 & 52.12 \end{bmatrix}.$$

As noted above, the mixed model formulates that the variance–covariance matrix of the repeated measures follow the equation $V(\boldsymbol{y}_i) = \boldsymbol{Z}_i\boldsymbol{\Sigma}_\upsilon\boldsymbol{Z}_i' + \sigma^2\boldsymbol{I}_{n_i}$. This is specifically the variance–covariance matrix of the repeated measures given the model covariates $\boldsymbol{X}$. If the only covariates in $\boldsymbol{X}$ are time trends (*e.g.*, week and week squared) and if the model fits the observed marginal timepoint means well, then we can compare the estimated variance–covariance matrix given by the model to the observed variance–covariance matrix. For the Reisby dataset, the random intercept and trend model did fit the observed timepoint means well, and so using the estimates in Table 4.5, namely $\hat{\sigma}^2 = 12.22$ and

$$\hat{\boldsymbol{\Sigma}}_\upsilon = \begin{bmatrix} 12.63 & -1.42 \\ -1.42 & 2.08 \end{bmatrix},$$

with the design matrix of the random effects

$$\boldsymbol{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix},$$

yields

$$\hat{V}(\boldsymbol{y}) = \boldsymbol{Z}\hat{\boldsymbol{\Sigma}}_\upsilon\boldsymbol{Z}' + \hat{\sigma}^2\boldsymbol{I} = \begin{bmatrix} 24.85 & & & & & \\ 11.21 & 24.08 & & & & \\ 9.79 & 12.52 & 27.48 & & & \\ 8.37 & 13.18 & 18.00 & 35.03 & & \\ 6.95 & 13.84 & 20.73 & 27.63 & 46.74 & \\ 5.53 & 14.50 & 23.47 & 32.44 & 41.41 & 62.60 \end{bmatrix}$$

as the estimated variance–covariance matrix. Given that this variance–covariance matrix of 21 elements is represented by only 4 parameter estimates, the fit appears reasonably good. The model is clearly picking up on the increasing variance across time and the diminishing covariance away from the diagonal.

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 example, in the covariance plot, lag 0 would correspond to the variance of the repeated measures at the different study timepoints, lag 1 would correspond to the covariances of the repeated measures one unit of time apart (*e.g.*, week 0 and week 1 covariance, week 1 and week 2 covariance, . . ., week 4 and 5 covariance), lag 2 would correspond to covariances 2 units of time apart (*e.g.*, week 0 and week 2 covariance, week 1 and week 3 covariance, . . ., week 3 and 5 covariance), etc. Similarly for the correlation plot, with the exception that there is no lag 0 in the plot because these would all be correlations of 1.

Figures 4.8 and 4.9 show the observed covariance and correlation plots. Note the plot of the variances at lag = 0 in Figure 4.8; these increase steadily from week 0 to week 5. Similarly, examining the covariances or correlations show that, for a given lag, the level of association generally increases across timepoints. These plots make very clear that a compound symmetric structure (*i.e.*, a random intercept model) of equal variances and covariances would not fit well.

Instead, it is of interest to see the graphs for the random intercept and trend model of Table 4.5. These are provided in Figures 4.10 and 4.11. These plots suggest reasonable model fit of the variances and covariances, though as one would expect the patterns are more systematic than the actual data. The model is emphasizing the increasing (co)variance values within a lag to a greater extent than the observed data. While in Chapters 6 and 7 we will describe more statistical tools that can be used for model selection regarding the variance–covariance structure, these plots can be quite useful in getting a "feel" for this.

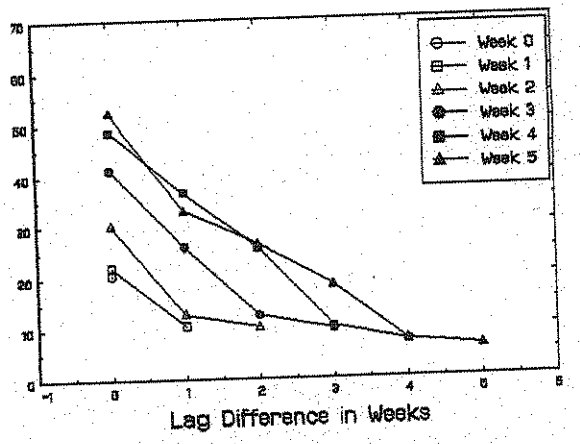

Figure 4.8. Reisby data: Observed covariance plot.

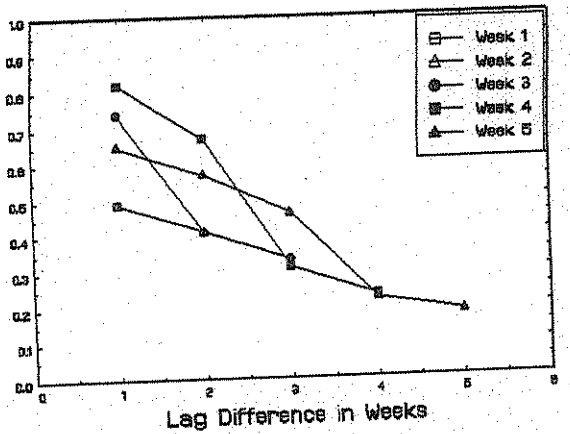

Figure 4.9. Reisby data: Observed correlation plot.

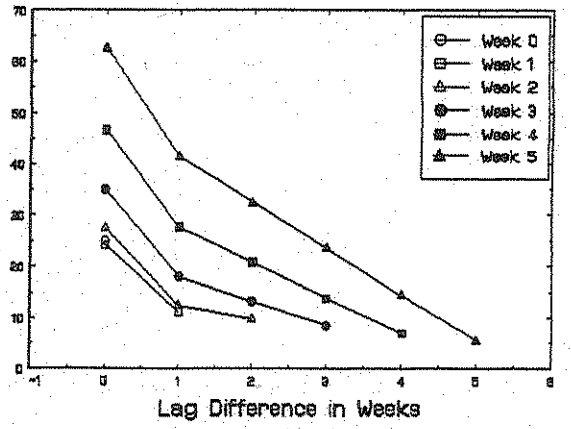

Figure 4.10. Reisby data: Random intercept and trend model estimated covariance plot.

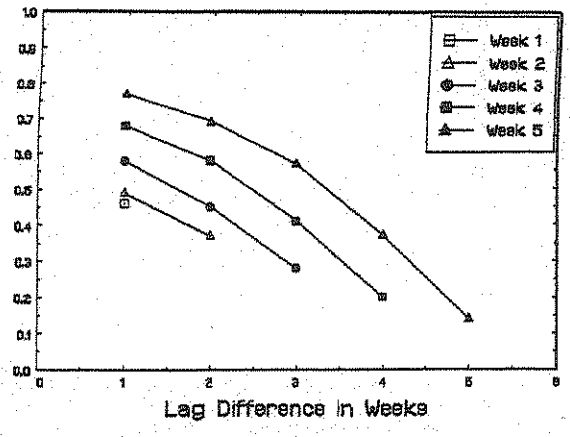

Figure 4.11. Reisby data: Random intercept and trend model estimated correlation plot.

4.5.2 Model with Time-Varying Covariates

In this section, we examine the effects of the time-varying drug plasma levels IMI and DMI. Since an inspection of the data indicated that the magnitude of these measurements varied greatly between individuals (from 4 to 312 μg/L for IMI and from 1 to 740 μg/L for DMI), a log transformation is used for these covariates. This helps to ensure that the estimated regression coefficients are not unduly influenced by extreme values on these covariates.

Also, these variables, ln IMI and ln DMI, are expressed in grand-mean centered form so that the model intercept represents HDRS scores for patients with average drug levels. To obtain the grand-mean centered versions of these variables, the variable's sample mean is subtracted from each observation. For notational simplicity in the model equations, I_{ij} and D_{ij} will represent the grand-mean centered versions of ln IMI and ln DMI, respectively, in what follows. Also, whereas the previous models considered HDRS outcomes from weeks 0 to 5, the models of this section only include HDRS outcome data from weeks 2 to 5. This is because the drug plasma levels are not available at the first two timepoints of the study (*i.e.*, week 0, or baseline, and week 1, or the end of the drug-washout period). While MRM does allow incomplete data across time, data must be complete within a given timepoint (in terms of both the dependent variable and covariates) for that timepoint to be included in the analysis. Thus, the analyses that follow are for the four week period following the drug-washout period with t_{ij} coded as 0, 1, 2, and 3 for these four respective timepoints. As a result, the intercept represents HDRS scores for week 2 of the study (*i.e.*, when $t_{ij} = 0$).

The first level-1 model is given by

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + b_{2i}I_{ij} + b_{3i}D_{ij} + \varepsilon_{ij}, \tag{4.13}$$

where, b_{0i} is the week 2 HDRS level for patient i under average levels of both ln IMI and ln DMI, b_{1i} is the weekly change in HDRS for patient i, b_{2i} is the patient's change in HDRS due to ln IMI, and b_{3i} is the change in HDRS due to ln DMI. The between-subjects model is given as

$$\begin{aligned} b_{0i} &= \beta_0 + \upsilon_{0i}, \\ b_{1i} &= \beta_1 + \upsilon_{1i}, \\ b_{2i} &= \beta_2, \\ b_{3i} &= \beta_3, \end{aligned} \tag{4.14}$$

where β_0 is the average week 2 HDRS level for patients with average ln IMI and ln DMI values, β_1 is the average HDRS weekly change, β_2 is the average HDRS difference for a unit change in ln IMI, and β_3 is the average HDRS difference for a unit change in ln DMI. Also, υ_{0i} is the individual intercept deviation, and υ_{1i} is the individual slope deviation. Notice that the level-2 model indicates that the drug effects could also be treated as random. This would be accomplished by adding υ_{2i} and υ_{3i} to the model, and would allow individual variation in terms of the drug level effect on HDRS scores. Given that antidepressants like IMI and DMI are not effective for all individuals, it is plausible that the drug levels are more strongly related to changes in depression for some individuals, whereas for others they are less so. Similarly, one could add individual-level covariates (*e.g.*, endogenous/nonendogenous group) into the models for b_{2i} and b_{3i} to examine whether the drug effects vary with individual-level covariates. Again, it is feasible that the drug effects on outcome are stronger for endogenous than nonendogenous patients. An example of an MRM allowing such individual variation in relationships is described by Hedeker et al. [1996]. Fitting the present model yields the results given in Table 4.9.

It is interesting to note that neither of the drug levels seems to be significantly related to the depression scores across time. However, note that the model given in (4.13) specifies that a person's drug level is related to their depression score at that same timepoint. It might be more plausible to instead posit that a person's drug level is related to their *change* in

Table 4.9. MRM Results for Level-1 Model (4.13) and Level-2 Model (4.14)

Parameter	Estimate	SE	Z	$p <$
intercept β_0	18.17	0.71	25.70	.0001
time slope β_1	−2.03	0.28	−7.15	.0001
ln IMI β_2	0.60	0.85	0.71	.48
ln DMI β_3	−1.20	0.63	−1.90	.06
$\sigma^2_{\upsilon_0}$	24.83	5.79		
$\sigma_{\upsilon_0\upsilon_1}$	−0.72	1.74		
$\sigma^2_{\upsilon_1}$	2.73	0.95		
σ^2	10.46	1.37		

Note. $-2\log L = 1502.5$.

depression score, or improvement, at that same timepoint. For this, the following alternative level-1 model is considered:

$$(y_{ij} - y_{i0}) = b_{0i} + b_{1i}t_{ij} + b_{2i}I_{ij} + b_{3i}D_{ij} + \varepsilon_{ij}, \tag{4.15}$$

where y_{i0} is the individual's HDRS score at baseline (or at week 1 for those few subjects with a missing baseline score). This yields the results presented in Table 4.10.

Table 4.10. MRM Results for Level-1 Model (4.15) and Level-2 Model (4.14)

Parameter	Estimate	SE	Z	$p <$
intercept β_0	−5.18	0.66	−7.87	.0001
slope β_1	−1.97	0.29	−6.90	.0001
ln IMI β_2	0.63	0.82	0.77	ns
ln DMI β_3	−1.97	0.60	−3.26	.0014
$\sigma^2_{\upsilon_0}$	20.50	5.13		
$\sigma_{\upsilon_0\upsilon_1}$	0.84	1.61		
$\sigma^2_{\upsilon_1}$	2.78	0.97		
σ^2	10.53	1.39		

Note. $-2\log L = 1498.8$.

Interestingly, now the effect of DMI, the metabolite of IMI, is highly significant and negative. Thus, greater DMI values are associated with greater improvement (*i.e.*, more negative HDRS change scores). However, the parent drug IMI is not significantly related to HDRS change scores and in fact its coefficient is positive. It's important to remember that the model estimates the IMI effect controlling for the DMI effect, and vice versa. These two drug levels are moderately correlated with each other (r = .18, .23, .22, and .18 for the four respective timepoints) and so the results above are not necessarily indicative of the marginal relationships of each drug with depression scores.

Correlations of the drug plasma levels with the HDRS scores, both raw and expressed as change scores, are given in Table 4.11. These bear out the fact that the drug levels are much more associated with the HDRS change scores than the actual scores. These correlations also show the greater association between HDRS change scores and DMI, rather than IMI, drug levels.

Table 4.11. Correlation Between HDRS Scores and Plasma Levels (Natural Log Units)

Drug	Week 2	Week 3	Week 4	Week 5
		HDRS total score		
ln IMI	−0.034	−0.038	−0.003	−0.189
ln DMI	−0.177	−0.075	−0.246	−0.293
		HDRS change from baseline		
ln IMI	−0.049	−0.106	−0.046	−0.240
ln DMI	−0.366	−0.281	−0.363	−0.361

4.5.2.1 *Within and Between-Subjects Effects for Time-Varying Covariates*

When time-varying covariates are included in a MRM, as in the manner of the last analysis, an assumption is made that the between and within-subjects effects of these variables are equal. To see this, express the time-varying covariates I_{ij} and D_{ij} as

$$\begin{aligned} I_{ij} &= \overline{I}_i + (I_{ij} - \overline{I}_i) \\ D_{ij} &= \overline{D}_i + (D_{ij} - \overline{D}_i) \end{aligned}$$

where $\overline{I}_i$ and $\overline{D}_i$ are the means of these two time-varying covariates computed for each individual. Thus, the first term following the equality represents the individual's mean on the time-varying covariate (*i.e.*, a between-subjects variable) and the second term represents the individual's deviation around their mean (*i.e.*, a within-subjects variable). Figure 4.12 shows a plot, considering the exact same data points, illustrating the difference between (a) a purely between-subjects effect versus (b) a purely within-subjects effect.

Focusing first on (a), notice that for these three subjects there is no within-subjects effect of the covariate, since the value of the dependent variable is constant within subjects. There is, however, a large between-subjects effect indicating that y increases as the subject average on x increases. Turning attention to (b), one can see a large within-subjects effect of the covariate for the two subjects. For a given subject, y increases as the value of x increases. However, there is no between-subjects effect of x in (b) since the mean of y is identical for these two subjects.

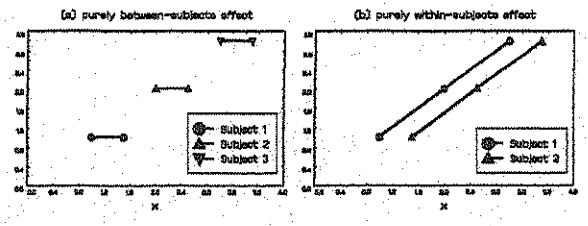

Figure 4.12. Time-varying covariate effects: (a) Purely between-subjects and (b) purely within-subjects.

To separate the within- and between-subjects effects of time-varying covariates one can include both the subject's average $\overline{x}_i$ and the subject's time-varying deviation $x_{ij} - \overline{x}_i$ into the model. In the present example including both of these terms into the MRM yields

$$(y_{ij} - y_{i0}) = b_{0i} + b_1 t_{ij} + b_{2i}(I_{ij} - \overline{I}_i) + b_{3i}(D_{ij} - \overline{D}_i) + \epsilon_{ij}, \tag{4.16}$$

and

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_4 \overline{I}_i + \beta_5 \overline{D}_i + \upsilon_{0i}, \\ b_{1i} &= \beta_1 + \upsilon_{1i}, \\ b_{2i} &= \beta_2, \\ b_{3i} &= \beta_3 \end{aligned} \tag{4.17}$$

for the level-1 and level-2 models. Thus, the total effect of IMI, for example,

$$\beta_2(I_{ij} - \overline{I}_i) + \beta_4 \overline{I}_i$$

is partitioned into its within- and between-subjects effects (*i.e.*, β_2 and β_4, respectively). The between-subjects part indicates the degree to which the individual's average drug level is related to their average depression level, averaging across time. In other words, it may be that subjects with consistently high drug levels have consistently low depression scores. Alternatively, the within-subjects component represents the degree to which variation in an individual's drug level is associated with a change in their depression scores (*i.e.*, a within-subject change). Thus, it may be that a higher relative drug level for an individual is associated with a lower relative depression score for that individual at a particular timepoint. If these two are equal ($\beta_2 = \beta_4$), then the IMI effect is

$$\beta_2(I_{ij} - \overline{I}_i) + \beta_2 \overline{I}_i \; = \; \beta_2 I_{ij},$$

which is exactly what was used in the last analysis. Thus, we implicitly assumed that the within- and between-subjects effects of these two drug levels were the same in the previous analysis. This assumption can be tested by comparing the previous model with the more general model of (4.16) and (4.17). Table 4.12 includes the results of this latter analysis.

Table 4.12. MRM Results for Level-1 Model (4.16) and Level-2 Model (4.17)

Parameter	Estimate	SE	Z	$p <$
intercept β_0	−5.09	0.66	−7.71	.0001
slope β_1	−2.02	0.29	−6.94	.0001
within ln IMI β_2	2.44	1.46	1.68	.10
within ln DMI β_3	−1.80	1.00	−1.80	.075
between ln IMI β_4	−0.31	1.00	−0.31	ns
between ln DMI β_5	−2.37	0.80	−2.97	.004
$\sigma^2_{v_0}$	20.32	5.10		
$\sigma_{v_0 v_1}$	0.50	1.64		
$\sigma^2_{v_1}$	2.83	0.97		
σ^2	10.38	1.36		

Note. $-2\log L = 1495.8$.

Comparing the two models yields a likelihood ratio test statistic of $X^2_2 = 3.0$, which is not statistically significant; the assumption of homogeneity of the between- and within-subjects regressions is not rejected. Inspecting the estimated coefficients for DMI supports this: −1.8 and −2.4 for the within- and between-subjects effects, respectively. Conversely, the estimates for IMI are very different, and even of opposite sign. However, neither is statistically significant and the standard errors for these two IMI estimates are quite large. In conclusion, for these data, there is not sufficient evidence to reject the assumption of equality in the within- and between-subjects effects for these two drug levels.

4.5.2.2 Time Interactions with Time-Varying Covariates

In some cases, it can be of substantive interest to examine whether there are interactions between a time-varying covariate and time. For example, one might posit that the relationship between the time-varying covariate and the outcome either increases or decreases across time. This is clearly plausible in the present example since the effectiveness of antidepressants is not thought to be immediate, but instead to develop over time [Reisby et al., 1977]. In other words, the drug plasma levels might be minimally related or unrelated to the depression outcome initially, with effects emerging across time. Thus, it is of interest to examine the degree to which the effects of the time-varying drug plasma levels on the change in depression scores vary across time. To explore this possibility, the level-1 model can be augmented to include the time interactions, namely,

$$(y_{ij} - y_{i0}) = b_{0i} + b_{1i}t_{ij} + b_{2i}I_{ij} + b_{3i}D_{ij} + b_{4i}(I_{ij} \times t_{ij}) + b_{5i}(D_{ij} \times t_{ij}) + \varepsilon_{ij}, \quad (4.18)$$

with the accompanying level-2 model,

$$\begin{aligned} b_{0i} &= \beta_0 + v_{0i}, \\ b_{1i} &= \beta_1 + v_{1i}, \\ b_{2i} &= \beta_2, \\ b_{3i} &= \beta_3, \\ b_{4i} &= \beta_4, \\ b_{5i} &= \beta_5. \end{aligned} \quad (4.19)$$

To correctly interpret the model parameters it is important to remember that the drug levels have been grand-mean centered, that the week variable equals 0 for the second week of the study, and that interpretation of the "main effects" is altered when interactions are present (*i.e.*, they represent the effect of the variable when the interacting variable equals 0). Thus, in this model, β_0 represents the average week 2 HDRS change score for patients with average drug levels, β_1 is the average weekly change in HDRS change scores for patients with average drug levels, β_2 is the HDRS change-score difference for a unit change of ln IMI at week 2, and β_3 represents the HDRS change-score difference per unit change of ln DMI at week 2. One can think of β_2 as the regression slope corresponding to the plot of HDRS change scores versus ln IMI levels considering week 2 data only (with the caveat that this regression slope is really a partial regression slope adjusting for the other drug level). Similar comments apply for interpreting β_3 in terms of ln DMI. Turning to the interactions β_4 and β_5, these indicate the per-week change in the drug effects on the HDRS change scores. In terms of the plot analogy, these interactions correspond to the change in (partial) regression slopes associated with separate weekly plots of HDRS change scores versus drug levels as one goes across the weeks. In other words, how does the slope for a given drug vary across time. Finally, v_{0i} represents the individual intercept deviation and v_{1i} is the individual slope (*i.e.*, time) deviation. Table 4.13 lists the results of this analysis.

Table 4.13. MRM Results for Level-1 Model (4.18) and Level-2 Model (4.19)

Parameter	Estimate	SE	Z	$p <$
intercept β_0	−5.12	0.65	−7.82	.0001
slope (*i.e.*, time) β_1	−1.94	0.28	−7.04	.0001
ln IMI β_2	0.40	0.87	0.46	ns
ln DMI β_3	−1.51	0.62	−2.43	.017
ln IMI by time β_4	0.16	0.41	0.39	ns
ln DMI by time β_5	−0.90	0.34	−2.65	.01
$\sigma^2_{v_0}$	20.24	5.05		
$\sigma_{v_0 v_1}$	0.99	1.52		
$\sigma^2_{v_1}$	2.50	0.91		
σ^2	10.35	1.36		

Note. $-2\log L = 1492.0$.

Comparing this model to the one without drug by time interaction (*i.e.*, from Table 4.10) yields a likelihood ratio test statistic of $X^2_2 = 6.8$, which is statistically significant at the .05 level. Thus, there is evidence that the drug effects on depression do vary across time. Inspecting the estimates and their test statistics in Table 4.13 reveals that it is ln DMI, and not ln IMI, that is interacting significantly with time. Specifically, ln DMI has an initial week 2 effect that is significant ($p < .017$), indicating that higher levels of ln

DMI are associated with greater improvement on the HDRS scale at this timepoint, and this beneficial effect of ln DMI gets more pronounced across time ($p < .01$). Concretely, for a one-unit change in ln DMI at week 2 (*i.e.*, when time is coded 0) the estimate is a 1.51 point reduction on the HDRS change score, whereas by the last timepoint (*i.e.*, when time is coded 3) it is a 4.21 point reduction ($1.51 + 3 \times .9$).

At first glance, it might seem a bit unusual that the ln DMI by time interaction is so highly significant given the reported correlations in Table 4.11. To better understand this, consider the simple linear regression slopes that are obtained from regressing HDRS change scores on ln DMI values at each of the four timepoints separately: these are -2.081, -2.195, -3.370, and -3.3765, respectively. These regression slopes provide clearer evidence of the ln DMI by time interaction, as they increase (in absolute value) more dramatically across time than the analogous correlations in Table 4.11. Why do these two sets of descriptive statistics suggest different conclusions? Remembering that the correlation is essentially a scale-free representation of the slope (*i.e.*, $r = \hat{\beta}\, s_x/s_y$), it is clear that the scales of the dependent and independent variable play a role here. Interestingly, the scale of these two go in opposite directions across time; the standard deviations of the HDRS change scores increase (5.38, 6.51, 7.35, and 7.88 across the four timepoints), whereas the standard deviations of the ln DMI values decrease (.95, .84, .79, and .76 across these same four timepoints). Thus, the metric for the slopes across time is very different (*i.e.*, the ratio of standard deviations s_x/s_y equals .18, .13, .11, and .10, respectively) which explains why the simple slopes and correlations are not in such close agreement, and why the significant ln DMI by time interaction of the MRM is a bit at odds with the apparent consistent pattern of the correlations across time. As this final MRM and the descriptive statistics make clear, it is the scale-dependent slope of ln DMI (*i.e.*, how much change in depression is associated with a unit change in the natural log of this blood level) that is increasing across time, and not the scale-free association.

4.6 ESTIMATION

Estimation of MRMs generally uses a combination of two complementary methods [Laird and Ware, 1982; Bock, 1989]. For the random individual effects υ_i, empirical Bayes (EB) methods are used, while maximum (marginal) likelihood (ML) or restricted (or residual) maximum likelihood (REML [McCulloch and Searle, 2001]) methods are used for estimation of variance parameters, σ^2 and $\boldsymbol{\Sigma}_\upsilon$, and regression coefficients $\boldsymbol{\beta}$. For general MRMs, the solution is fairly complex and requires iterative algorithms like Newton–Raphson or Fisher-scoring procedures. Essentially, these algorithms continue iterating through the data until all parameter estimates are changing by a very small degree, at which point the estimation process is said to have converged. In the following, we will present the formulas in the case of a random-intercept MRM to illustrate the essential ideas and show connections with estimation in ordinary (*i.e.*, fixed-effects) regression models. More complete details for specific models are presented in the subsequent chapters.

EB estimates of individual effects are sometimes termed EAP ("Expected *a Posteriori*") estimates, since they are derived as the mean of the posterior distribution of υ, given $\boldsymbol{y}_i$. Denoting the EB estimate of υ_i as $\tilde{\upsilon}_i$ to distinguish it from subsequent ML estimates, and given the model assumptions, we get the following EB estimator of individual parameters:

$$\tilde{\upsilon}_i = \rho_{n_i n_i} \frac{1}{n_i} \mathbf{1}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) = \rho_{n_i n_i} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta}), \qquad (4.20)$$

where $\boldsymbol{x}_{ij}$ is the vector of regressors for a individual i at time j, and $\rho_{n_i n_i}$ is equivalent to the Spearman–Brown reliability formula [Guilford, 1954], given as $\rho_{nn} = nr/[1 + (n-1)r]$ with r as the intraclass correlation.

A property of EB estimation is that $\tilde{\upsilon}_i$ is a function of both the individual's data and the empirical prior distribution specified for υ_i. As information about an individual increases (*i.e.*, the reliability $\rho_{n_i n_i}$ increases toward 1), by either increasing data interdependency within the subject (increasing r) and/or increasing sample size (n_i), the EB estimate approaches the average deviation (across time) for that individual, $(\sum_{j=1}^{n_i} y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})/n_i$. Note that this latter formula yields the OLS estimator of the individual effect. Alternatively, as information about an individual decreases (*i.e.*, $\rho_{n_i n_i}$ decreases toward 0), by decreasing data interdependency within the subject and/or decreasing sample size, the EB estimate approaches the posited mean of the empirical prior distribution of υ_i, namely 0. Thus, given r, if a subject has few measurements, then the EB estimate will be smaller (in absolute value) than the corresponding OLS estimate. Alternatively, if the subject has many measurements across time, then the EB and OLS estimates would be very similar. Because of this, the EB estimates are said to be *shrunken to the mean*, where the mean of the random effects equals zero in the population. The degree of shrinkage depends on the number of measurements an individual has. An important advantage of EB estimates relative to OLS estimates is that they are not as prone to the undue influence of outliers.

Figure 4.13 shows a plot of the EB estimates versus their OLS counterparts for a random-intercept model of the Reisby data. Notice that the spread of the OLS estimates is greater than the spread of the EB estimates, and that the points do not fall exactly on the 45-degree line. If the EB and OLS estimates were exactly the same all of the points would be on the 45-degree line. Instead, the points suggest a line that is slightly more horizontal (*i.e.*, toward zero), which results from the shrinkage of the EB estimates toward zero.

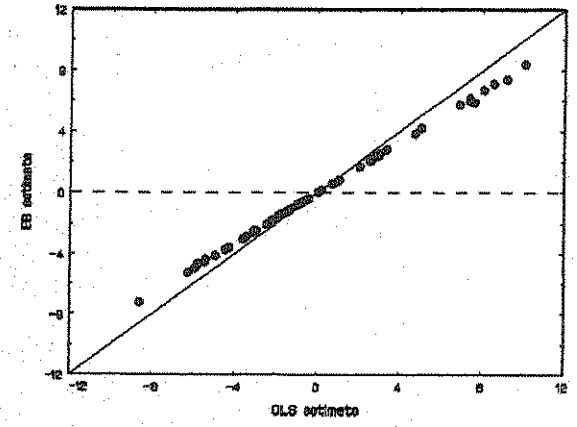

Figure 4.13. Reisby data: EB versus OLS estimates of subject effects.

In addition to the EB estimate of the posterior mean, the variance of the posterior distribution of υ is given as

$$\sigma^2_{\upsilon|y_i} = \sigma^2_\upsilon(1 - \rho_{n_i n_i}). \tag{4.21}$$

Again, the form reveals the nature of this EB estimator of the posterior variance: As information about the individual increases, the posterior variance becomes a fraction of the empirical prior variance (σ^2_υ), while as information about the individual decreases, this variance approaches the empirical prior variance.

To estimate covariate effects β and variance parameters σ^2_υ and σ^2, ML estimation can be used. The ML estimation procedure is more fully presented in the appendix and is described using two numerical algorithms: the EM algorithm solution [Dempster et al., 1981] and the Fisher scoring solution [Longford, 1987]. From the EM algorithm solution, we can see how the random-intercept MRM can be viewed as a generalization of the ordinary multiple linear regression model. Namely, the following equations are used in the iterative EM algorithm solution,

$$\hat{\beta} = \left[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X}_i\right]^{-1}\left[\sum_{i=1}^{N} \boldsymbol{X}_i'(\boldsymbol{y}_i - \mathbf{1}_i\tilde{\upsilon}_i)\right], \tag{4.22}$$

$$\hat{\sigma}^2_\upsilon = \frac{1}{N}\sum_{i}^{N}\tilde{\upsilon}_i^2 + \sigma^2_{\upsilon|y_i}, \tag{4.23}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i}^{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\beta} - \mathbf{1}_i\tilde{\upsilon}_i)'(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\beta} - \mathbf{1}_i\tilde{\upsilon}_i) + n_i\sigma^2_{\upsilon|y_i} \tag{4.24}$$

with the solution proceeding by iterating between EB equations (4.20) and (4.21) and ML equations (4.22)–(4.24) until convergence. Note that as estimates of the individual effects $\tilde{\upsilon}_i$ and variances $\sigma^2_{\upsilon|y_i}$ approach zero, the subject variance estimate (σ^2_υ) approaches zero, and the equations for regression coefficients β and error variance σ^2 approach the maximum likelihood solution of these parameters in the usual fixed-effects regression model, namely,

$$\hat{\beta} = \left[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X}_i\right]^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{y}_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\beta})'(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\beta}).$$

Thus, as the dependency of the data within individuals decreases, the solution approaches the (ML) solution for an ordinary multiple linear regression model.

The formula for the random effect variance (4.23) reveals an interesting connection between this population variance and the sample variance of the EB estimates. Notice that this equation can be written as

$$\hat{\sigma}^2_\upsilon = \frac{1}{N}\sum_{i}^{N}\tilde{\upsilon}_i^2 + \frac{1}{N}\sum_{i}^{N}\sigma^2_{\upsilon|y_i}. \tag{4.25}$$

Because the mean of the random effects is approximately zero in the sample, the first term after the equality is essentially the (ML) estimate of the sample variance of the EB estimates, while the second term is the average of the posterior variances. Thus, the estimate of the population variance of the random effects $\hat{\sigma}^2_\upsilon$ will always exceed the sample variance of the EB random effects (except for the trivial case where all of the posterior variances equal 0).

4.6.1 ML Bias in Estimation of Variance Parameters

It can be shown that the ML estimates of variance parameters in MRMs are biased downwards (*i.e.*, they are too small). This is also true of the ML estimate of the error variance in ordinary multiple regression, and the equations for this parameter illustrate the point well. Note that the ML estimate of the error variance equals

$$\sigma^2 = \frac{SSE}{N}, \tag{4.26}$$

where SSE is the sum of squared errors and N is the sample size. The unbiased ordinary least square (OLS) estimate of the same parameter is

$$\sigma^2 = \frac{SSE}{N - p - 1}, \tag{4.27}$$

where p is the number of regressors. These equations make clear that this bias is negligible if $N - p$ is relatively large (say, over 100), but can be of concern when $N - p$ is not large. While ML estimates of the fixed effects are not greatly affected by this, their standard errors can be different if the variance parameters are downwardly biased. Again, the standard errors will be too small under ML estimation, though the difference is negligible if $N - p$ is relatively large.

To correct for this bias, the MRM parameters can be estimated by REML [Patterson and Thompson, 1971]). Clearly, REML estimates are preferred over ML estimates, however there is one important consideration to keep in mind. Because REML adjusts the likelihood for the number of covariates in a model, one cannot use REML likelihood ratio tests for comparing models with different covariates. ML likelihood ratio tests do not have this limitation. Because of this, and because the sample size for the datasets in this text are relatively large, we will present ML estimates unless otherwise noted.

4.7 SUMMARY

As this chapter has demonstrated, MRMs are useful for analyzing longitudinal data. MRMs allow for the presence of missing data, irregularly-spaced measurements across time, time-varying and invariant covariates, accommodation of individual-specific deviations from the average time trend, and estimation of the population variance associated with these individual effects. Perhaps the most popular feature of MRMs is their treatment of missing data. As has been illustrated, subjects are not assumed to be measured at the same number of timepoints. Since there are no restrictions on the number of observations per individual, subjects who are missing at a given interview wave are not excluded from the analysis. The assumption of the model is that the available data for a given subject are representative of that subject's deviation from the average trends across time (which are estimated based on the whole sample). It is important to note that the missing data are not imputed in estimation of the MRM; rather the model parameters are estimated using all available data. Further treatment of missing data in longitudinal studies is described in Chapter 14.

Statistical software to perform MRM analysis has proliferated, especially for continuous outcomes. Most of the major statistical packages now include procedures for estimating continuous MRMs. Additionally, there are several independent programs that were specifically designed for MRM analysis; these include HLM [Raudenbush et al., 2000a], MLwiN [Goldstein et al., 1998], and MIXREG [Hedeker and Gibbons, 1996b], to mention a few. Review articles comparing some of these software programs include van der Leeden et al. [1996] and de Leeuw and Kreft [2001]. Rather than present software examples here, the website for this book (`http://www.uic.edu/~hedeker/long.html`) includes syntax files for the analyses presented in this chapter for some of these software programs.

This chapter has focused on the modeling aspects of MRM without a great deal of discussion on parameter estimation. Because these models are more complex than ordinary fixed-effects regression models, it is sometimes the case that the iterative procedures used for estimation of an MRM do not converge to a solution. If this occurs, it is often because the model is overly complex, relative to the data being used to estimate it, and so model simplification is necessary. Although it is not always apparent why a particular model does not converge, building models in a sequential piecewise manner can help to isolate where troubles occur.

In the example, repeated observations were observed nested within individuals. In the terminology of multilevel analysis [Goldstein, 1995] and hierarchical linear models [Raudenbush and Bryk, 2002], this is termed a two-level data structure with individuals representing level-2 and the nested repeated observations level-1. The models that we have presented are thus referred to as two-level models. Individuals themselves, though, are often observed clustered within some higher-level unit, for example, a classroom, clinic, or worksite. Cross-sectional clustered data can also be considered as two-level data, with the clusters representing level-2 and the clustered subjects level-1. Modeling of such clustered data is described in detail in several texts [Goldstein, 1995; Hox, 2002; Kreft and de Leeuw, 1998; Longford, 1993; Raudenbush and Bryk, 2002; Snijders and Bosker, 1999]. In some studies, subjects are clustered and also repeatedly measured, resulting in three-levels of data: the cluster (level-3), individual (level-2), and repeated observation (level-1). Analysis of three-level data is also described in some of the aforementioned texts, and will be treated in Chapter 13.

In this chapter, we have considered MRMs for a single longitudinal response process in which individual members of the population can systematically deviate from the overall process; however, there is a single overall process from which individuals may deviate. In some cases, this assumption may be unreasonable. In such cases, it may be more plausible to assume that there are two or more different processes in the population, and individual subjects deviate from a particular response process. From the statistical point of view, in such cases, the population may be better represented by a mixture of temporal response processes rather than a single process. Furthermore, within each component of the temporal response distribution or "latent class," the effects of treatment or other fixed-effects in the model may vary. If the data confirm the existence of multiple latent classes, each with a different temporal response process and possibly different treatment related effects, we may then be able to target specific treatments to specific subjects, based on pre-treatment characteristics that are predictive of latent class membership. Several authors have described MRMs where the random effects are drawn from either a latent class or a mixture distribution in the population [Verbeke and Lesaffre, 1996; Muth´en and Shedden, 1999; McCulloch et al., 2002; Xu and Hedeker, 2002]. These articles make clear that these extensions of MRMs can be quite useful in some contexts.

CHAPTER 5

MIXED-EFFECTS POLYNOMIAL REGRESSION MODELS

5.1 INTRODUCTION

In many situations, it is too simplistic to assume that the change across time is linear. For example, it may be that the outcome changes across time in a curvilinear manner. A curvilinear trend would allow a leveling off or accelerating of the change across time. This is clearly plausible in many situations, but especially for rating scale data, like that considered in the last chapter, where ceiling and floor effects can easily occur. In this chapter we will explore the use of polynomial trend models, illustrating how they can be used to model particular kinds of nonlinear relationships across time both at the individual and population levels.

5.2 CURVILINEAR TREND MODEL

To begin, consider the following curvilinear trend model that is obtained by adding a quadratic, or squared, term to the level-1 model:

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + b_{2i}t_{ij}^2 + \varepsilon_{ij}. \tag{5.1}$$

Here, b_{0i} is the intercept for subject i, b_{1i} is the linear trend component for subject i, and b_{2i} is the quadratic trend component for subject i. Notice that this model can also be written as