Chapter 5

Mixed-Effects Polynomial Regression Models for Continuous Outcomes

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 cross 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 non-linear 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:

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

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

$$y_{ij} = b_{0i} + (b_{1i} + b_{2i}t_{ij})t_{ij} + \varepsilon_{ij}$$

to point out that the "time effect" varies as a function of time. The level-2, between-subjects, model is now

$$\begin{aligned} b_{0i} &= \beta_0 + \upsilon_{0i} \\ b_{1i} &= \beta_1 + \upsilon_{1i} \\ b_{2i} &= \beta_2 + \upsilon_{2i} \end{aligned} \tag{5.2}$$

where β_0 is the intercept, β_1 is the average linear trend component, and β_2 is the average quadratic trend component. Similarly, υ_{0i} is the individual deviation from average intercept, and υ_{1i} and υ_{2i} represent the individual deviation the from average linear and quadratic trend components, respectively. Thus, this model allows curvilinearity at both the population (β_2) and individual (υ_{2i}) levels. In some situations it may be reasonable to restrict the curvilinear

part to the level-1 model only, in which case the final component of the level-2 model would simply be $b_{2i} = \beta_2$.

Figure 5.1 gives a sense of some of the possible time-trends that can be fit using curvilinear trend models.

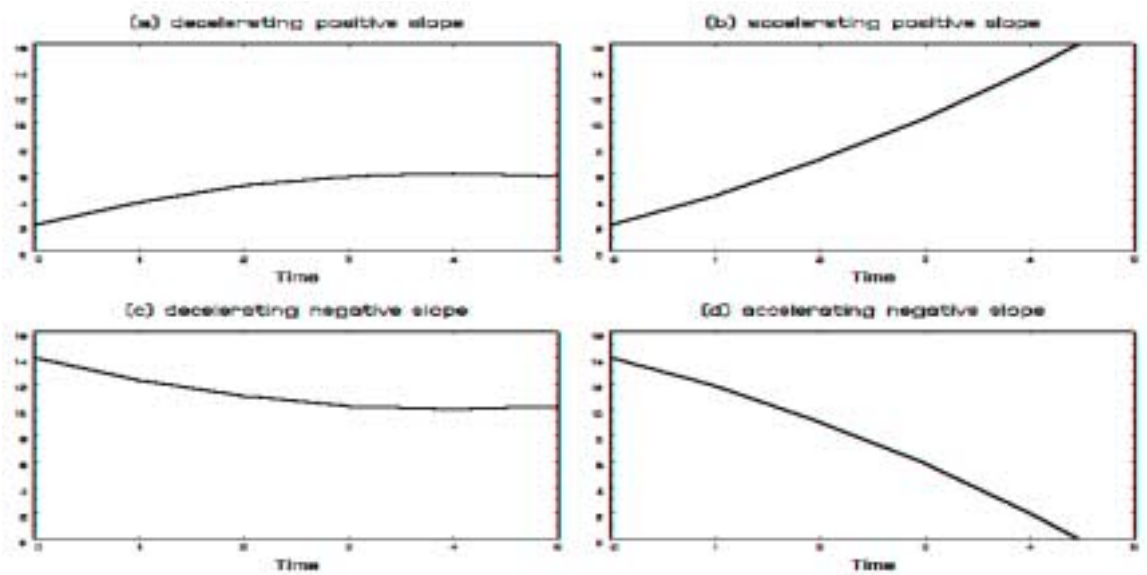

Figure 5.1. Various curvilinear models: (a) decelerating positive slope; (b) accelerating positive slope; (c) decelerating negative slope; (d) accelerating negative slope

In the top two graphs, positive trends are either decelerating (a) or accelerating (b). The linear component β_1 equals 2 for both, while the quadratic component β_2 is -.25 and .25, respectively. Similarly for the negative trends on the bottom, the linear component equals -2 for both, while the quadratic component is either .25 (decelerating negative) or -.25 (accelerating negative).

Notice that the decelerating positive trend in (a) is flat as time increases. Similarly, the negative trend in (c) flattens out. One can easily calculate the point at which a curvilinear trend "flattens out." For this, note that the derivative is

$$\frac{\partial y}{\partial t} = \beta_1 + 2\beta_2 t \ , \tag{5.3}$$

which equals 0 for $t = -\beta_1/(2\beta_2)$. So for our decelerating positive trend in (a) the flattening out point is when time equals 4. Similarly, the decelerating negative trend flattens out when time is 4. For proper interpretation of curvilinear models, it is very helpful to figure out where the flattening out point occurs, since beyond that point the trend reverses sign. That is, for (a) beyond week 4 the trend goes from being a positive one to a negative one.

This changing of signs can be useful in areas were the relationship across time is thought to be J- or U-shaped. Figure 5.2 presents some examples of curves of these types. As can be seen, a wide variety of relationships are possible simply by adding in the squared term for time in the model.

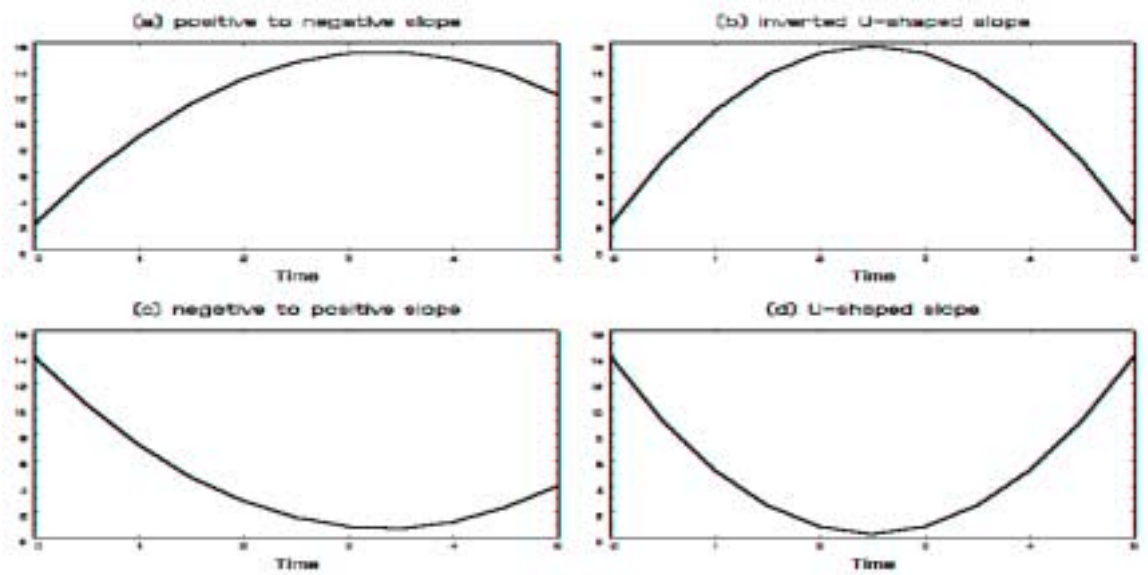

Figure 5.2. More curvilinear models:

(a) positive to negative slope ($\beta_0 = 2, \beta_1 = 8, \beta_2 = -1.2$);

(b) inverted U-shaped slope ($\beta_0 = 2, \beta_1 = 11, \beta_2 = -2.2$);

(c) negative to positive slope ($\beta_0 = 14, \beta_1 = -8, \beta_2 = 1.2$);

(d) U-shaped slope ($\beta_0 = 14, \beta_1 = -11, \beta_2 = 2.2$)

5.2.1 Curvilinear Trend Example

Turning again to the Reisby dataset, we will add a quadratic time trend to both the population and individual trends. Fitting this model yields the results given in Table 5.1.

Table 5.1

MRM results for level-1 model (5.1) and level-2 model (5.2).

Parameter	Estimate	SE	z	$p <$
β_0	23.76	0.55	43.04	.0001
β_1	-2.63	0.48	-5.50	.0001
β_2	0.05	0.09	0.58	.56
$\sigma^2_{\upsilon_0}$	10.44	3.59		
$\sigma_{\upsilon_0\upsilon_1}$	-0.92	2.41		
$\sigma^2_{\upsilon_1}$	6.64	2.76		
$\sigma_{\upsilon_0\upsilon_2}$	-0.11	0.42		
$\sigma_{\upsilon_1\upsilon_2}$	-0.94	0.49		
$\sigma^2_{\upsilon_2}$	0.19	0.09		
σ^2	10.52	1.11		

Note. $-2\log L = 2207.64$

Comparing this model to one without these quadratic terms (*i.e.*, the model listed in Table 5.5 of the previous chapter with $\beta_2 = \sigma^2_{\upsilon_2} = \sigma_{\upsilon_0\upsilon_2} = \sigma_{\upsilon_1\upsilon_2} = 0$) yields a deviance of 11.4, which is statistically significant on 4 degrees of freedom (even without an adjustment to the p-value for testing of variance terms). This is interesting given that the Wald test for β_2 is clearly non-significant. In fact, comparing the above model to one with $\sigma^2_{\upsilon_2} = \sigma_{\upsilon_0\upsilon_2} = \sigma_{\upsilon_1\upsilon_2} = 0$ (not shown) yields a deviance of 11.0. Nearly all of the improvement in model fit is through the inclusion of the quadratic term as a random effect, and not as a fixed effect. This suggests that although the trend across time is essentially linear at the population level, it is

curvilinear at the individual level. How is this possible? Consider Figure 5.3 which plots a hypothetical example with an average linear trend and curvilinear trends for two individuals.

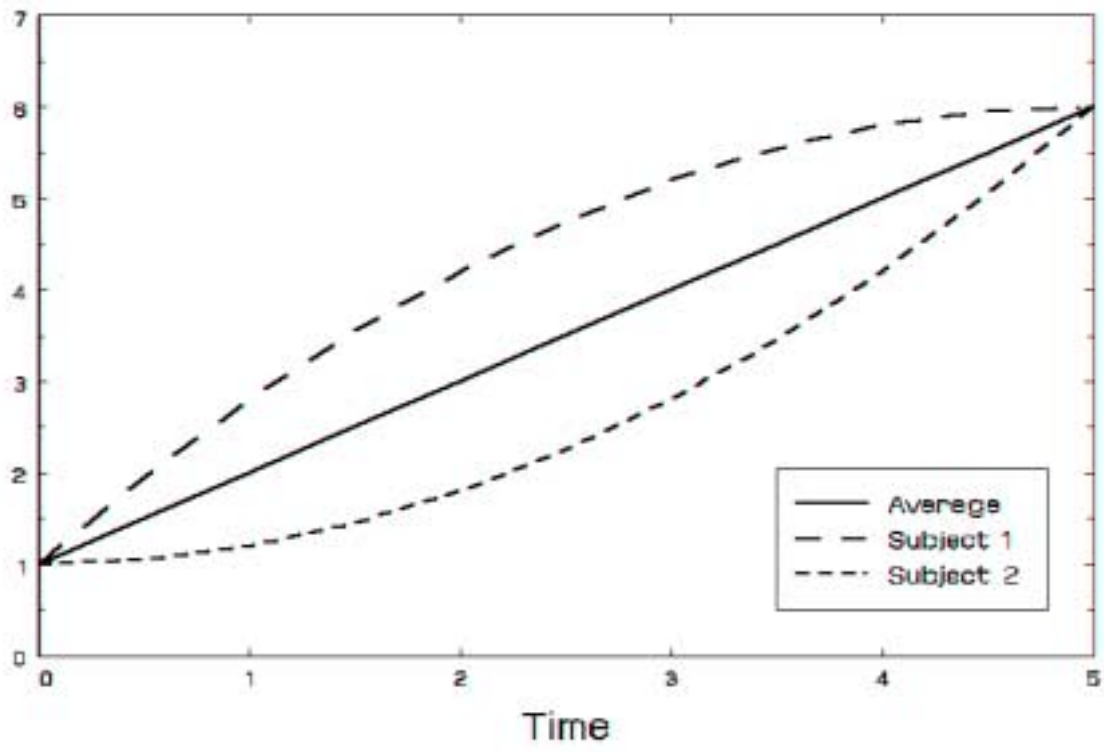

Figure 5.3. Average linear and individual quadratic trends.

Here, one individual has an accelerating positive trend (subject 2) and the other has a decelerating positive trend (subject 1). These two "cancel out" to yield an average positive trend that is strictly linear.

For the Reisby data, though the quadratic term is non-significant at the population level, it is instructive to calculate the point at which the decelerating negative trend flattens out. This yields:

$$\hat{t} \;=\; -\hat{\beta}_1/(2\hat{\beta}_2) \;=\; 1/2(2.63/0.05) \;=\; 26.3\;,$$

which is well beyond the time-frame of the study. Thus, the curvilinear component is clearly very slight (at the population level).

Figure 5.4 contains a plot of the individual trend estimates from this model. These are obtained by calculating $\hat{y}_{ij} = \hat{b}_{0i} + \hat{b}_{1i}t_{ij} + \hat{b}_{2i}t_{ij}^2$ for $t = 0, 1, \ldots, 5$, and then connecting the timepoint estimates for each individual. The plot makes apparent the wide heterogenity in trends across time, as well as the increasing variance in HD scores across time. Some individuals have accelerating downward trends suggesting a delay in the drug effect. Alternatively, others have decelerating downward trends, which are consistent with a leveling off of the drug effect. Some individuals even have positive trends indicating a worsening of their depressive symptoms across time. This is not too surprising given that antidepressants, like imipramine, are known to be ineffective for some patients. The figure is also interesting in showing that many of the individual trend lines are approximately linear. Thus, the improvement that the curvilinear model provides in describing change across time is perhaps modest.

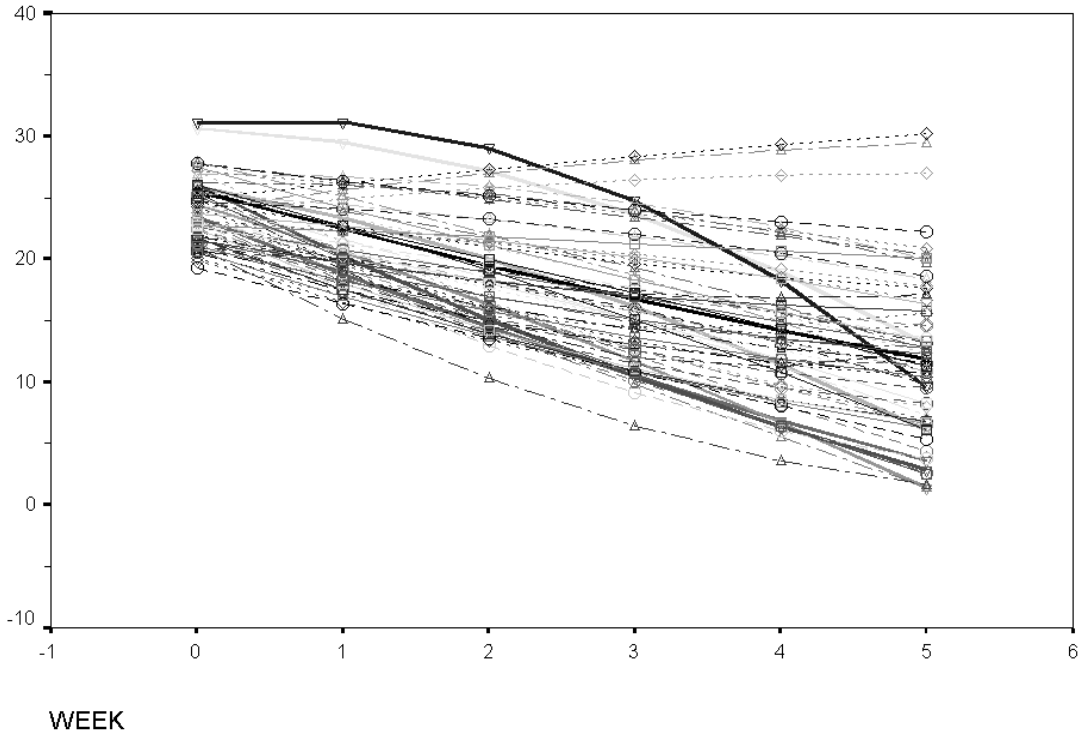

Figure 5.4. Reisby data: estimated curvilinear trends.

The model fit of the observed variance-covariance matrix of the repeated measures is obtained as:

$$\hat{V}(\boldsymbol{y}) \;=\; \boldsymbol{Z}\hat{\boldsymbol{\Sigma}}_{\upsilon}\boldsymbol{Z}' + \hat{\sigma}^2\boldsymbol{I}$$

$$= \begin{bmatrix} 20.96 & & & & & \\ 9.41 & 23.86 & & & & \\ 8.16 & 15.57 & 31.07 & & & \\ 6.68 & 16.08 & 23.11 & 38.31 & & \\ 4.98 & 14.88 & 23.26 & 30.12 & 45.98 & \\ 3.06 & 11.97 & 20.98 & 30.09 & 39.29 & 59.11 \end{bmatrix},$$

where

$$\boldsymbol{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix}, \quad \hat{\boldsymbol{\Sigma}}_v = \begin{bmatrix} 10.44 & -0.92 & -0.11 \\ -0.92 & 6.64 & -0.94 \\ -0.11 & -0.94 & 0.19 \end{bmatrix},$$

and $\hat{\sigma}^2 = 10.52$. By comparing this matrix with the observed variance-covariance matrix, presented in the previous chapter, one can see that the estimated variances are close to the observed, and the model is clearly picking up the pattern of diminishing covariance away from the diagonal and at the earlier timepoints. We will return to this issue regarding the the relative model fit of the observed (co)variances later in this chapter.

5.3 Orthogonal Polynomials

For trend models, it is often beneficial to represent the polynomials in orthogonal form [Bock, 1975]. One advantage is that it avoids collinearity problems that can result from using multiples of t (t^2, t^3, etc.) as regressors. To see this, consider a curvilinear trend model with three timepoints. Then t = 0, 1, and 2, while t^2 = 0, 1, and 4; these two variables are nearly perfectly correlated. To counter this, time is sometimes expressed in centered form, for example $(t - \bar{t})$ = -1, 0, and 1, and $(t - \bar{t})^2$ = 1, 0, and 1. If there are the

same number of observations at the three timepoints, this centering removes the correlation between the linear and quadratic trend components entirely. In the more usual situation of non-equal numbers of observations across time, this greatly diminishes the correlation between the polynomials. Another aspect of centering time is that the meaning of the model intercept changes. In the previous raw form of time, the intercept represented differences at the first timepoint (*i.e.*, when time $= 0$). Alternatively, in centered form, the model intercept represents differences at the midpoint of time.

An additional advantage of using orthogonal polynomials, over simply centering time, is that the polynomials are put on the same scale. Thus, their estimated coefficients can be compared in terms of their magnitude in the same way as standardized beta coefficients in ordinary regression analysis. Also, in the original scale it gets increasingly difficult to estimate the regression coefficients of higher-degree polynomial terms because the coefficients (and their standard errors) get smaller and smaller. This computational problem is removed by putting the polynomials on the same scale.

For equal time intervals, tables of orthogonal polynomials can be found in several statistics texts, for example in Pearson and Hartley [1976]. For example, for a study with six equally-spaced timepoints the orthogonal polynomials for the constant, linear, and quadratic trend components are given as:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -3 & -1 & 1 & 3 & 5 \\ 5 & -1 & -4 & -4 & -1 & 5 \end{bmatrix} \begin{matrix} /\sqrt{6} \\ /\sqrt{70} \\ /\sqrt{84} \end{matrix} .$$

The rows here represent the three orthogonal polynomials (constant, linear, and quadratic) and the columns are the six timepoints. Notice that these row vectors are independent of each

other, since the sum of inner products equals 0 in all cases (*e.g.*, $-5\times5+-3\times-1+\ldots+5\times5=0$). Also, by dividing the values by the square-root of the quantities on the right, which are simply the sum of squared values in a row, these polynomials have the same scale. Thus, these terms are simultaneously made independent of each other and standardized to the same (unit) scale. This holds exactly when the number of observations at each timepoint are equal, and approximately so in the more usual situation when they are unequal.

To use orthogonal polynomials in a MRM, variables representing the trend components must be created. This can be done using a series of `IF` statements in any statistical software program. For example, below is SAS code that can be used to create the orthogonal trend component variables for the first two timepoints, weeks 0 and 1, of this example with 6 timepoints (assuming a variable named `WEEK` that takes on values 0 to 5 in the dataset).

```
IF WEEK = 0 THEN DO;
       CONS = 1 / SQRT(6);
       LIN = -5 / SQRT(70);
       QUAD = 5 / SQRT(84);
END;
IF WEEK = 1 THEN DO;
       CONS = 1 / SQRT(6);
       LIN = -3 / SQRT(70);
       QUAD = -1 / SQRT(84);
END;
```

The above statements assign the orthogonal polynomial values that are in the first two columns of the matrix. Similar `IF` statements would be used to produce the orthogonal polynomial values for the remaning study timepoints (*i.e.*, the rightmost four columns of the matrix). The new variables `CONS`, `LIN`, and `QUAD` would then be entered into MRM analysis as regressors and random effects, with the caveat that an intercept should not be included since the `CONS` term has taken its place.

While looking up the values of orthogonal polynomials in statistical tables can be interesting (and possibly cure insomnia!), these can also be calculated directly. Bock [1975] describes a method for doing this utilizing the Cholesky factorization (*i.e.*, matrix square root) of a symmetric matrix. This factorization has many uses in statistical computing and so it is worth briefly going over. Specifically, the symmetric matrix $\boldsymbol{A}$ is factored as

$$\boldsymbol{A} = \boldsymbol{S}\boldsymbol{S}' \ ,$$

where $\boldsymbol{S}$ is the lower triangular Cholesky factor, and $\boldsymbol{S}'$ is its upper triangular counterpart (*i.e.*, the transpose of the lower triangular matrix, since $'$ denotes the transpose of a matrix). Continuing with our quest of directly obtaining orthogonal polynomials, denote the time matrix, including the intercept, as $\boldsymbol{T}$. For example, in our case of 6 timepoints and up to quadratic trend, we have:

$$\boldsymbol{T}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix} \ .$$

Then, the steps one takes to obtain the orthogonal polynomial matix are:

1. compute $\boldsymbol{T}'\boldsymbol{T}$, which yields a symmetric matrix.

$$\boldsymbol{T}'\boldsymbol{T} = \begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} .$$

2. obtain the Cholesky factor $\boldsymbol{S}$ of $\boldsymbol{T}'\boldsymbol{T}$, and express it in transpose form (*i.e.*, upper triangular).

$$\boldsymbol{S}' = \begin{bmatrix} 2.4495 & 6.1237 & 22.4537 \\ 0 & 4.1833 & 20.9165 \\ 0 & 0 & 6.1101 \end{bmatrix} .$$

3. obtain the inverse $(\boldsymbol{S}')^{-1}$.

$$(\boldsymbol{S}')^{-1} = \begin{bmatrix} 0.4082 & -0.5976 & 0.5455 \\ 0 & 0.2390 & -0.8183 \\ 0 & 0 & 0.1637 \end{bmatrix} .$$

4. multiply $\boldsymbol{T}$ by this inverse $(\boldsymbol{S}')^{-1}$.

$$\begin{aligned} \boldsymbol{T}(\boldsymbol{S}')^{-1} &= \begin{bmatrix} 0.4082 & -0.5976 & 0.5455 \\ 0.4082 & -0.3586 & -0.1091 \\ 0.4082 & -0.1195 & -0.4364 \\ 0.4082 & 0.1195 & -0.4364 \\ 0.4082 & 0.3586 & -0.1091 \\ 0.4082 & 0.5976 & 0.5455 \end{bmatrix} \\ &= \begin{bmatrix} 1/\sqrt{6} & -5/\sqrt{70} & 5/\sqrt{84} \\ 1/\sqrt{6} & -3/\sqrt{70} & -1/\sqrt{84} \\ 1/\sqrt{6} & -1/\sqrt{70} & -4/\sqrt{84} \\ 1/\sqrt{6} & 1/\sqrt{70} & -4/\sqrt{84} \\ 1/\sqrt{6} & 3/\sqrt{70} & -1/\sqrt{84} \\ 1/\sqrt{6} & 5/\sqrt{70} & 5/\sqrt{84} \end{bmatrix} , \end{aligned}$$

which yields the same orthogonal polynomial values as before.

Besides replicating the results found in statistical tables, this method has an important advantage in that it can be used to create orthogonal polynomials for cases of unequal time intervals simply by modifying the original $\boldsymbol{T}$ matrix. Notice also that to add a cubic term, a row containing the t^3 values would simply be added to the $\boldsymbol{T}'$ matrix. While it might seem difficult to implement, use of matrix algebra routines in statistical packages makes this procedure relatively easy. For instance, the SAS statements below can be used to perform the above calculations.

```
TITLE 'producing orthogonal polynomial matrix';

PROC IML;
 time = {  1  0   0    ,
           1  1   1    ,
           1  2   4    ,
           1  3   9    ,
           1  4  16    ,
           1  5  25  } ;
orthpoly = time*INV(ROOT(T(time)*time));

PRINT 'time matrix', time [FORMAT=8.4];

PRINT 'orthogonalized time matrix', orthpoly [FORMAT=8.4];
```

This code uses two built-in SAS matrix routines: `INV` which performs matrix inversion and `ROOT` which yields the transpose of the Cholesky factor (*i.e.*, the upper triangular matrix $\boldsymbol{S}'$). Additionally, matrix multiplication is performed using the $*$ operator, which is also the ordinary scalar multiplication operator. As can be seen, SAS allows these operations to be

performed all on one line, so that there is no need to save intermediate results.

5.3.1 Model Representations

To distinguish between the two representations of the MRM, the original time metric $\boldsymbol{T}$ and the orthogonal polynomial time metric $\boldsymbol{T}(\boldsymbol{S}')^{-1}$, it is helpful to modify the notation for the latter. Consider first the matrix representation of the MRM:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i \ , \tag{5.4}$$

where time is expressed (in $\boldsymbol{X}$ and $\boldsymbol{Z}$) in its original metric. The mean of the normally-distributed random effects $\boldsymbol{v}_i$ is $\mathbf{0}$ and the variance-covariance matrix is $\boldsymbol{\Sigma}_v$. Also, the errors $\boldsymbol{\varepsilon}_i$ are normally and independently distributed with $\mathbf{0}$ mean and variance $\sigma^2\boldsymbol{I}_{n_i}$. Thus, the dependent variable vector $\boldsymbol{y}_i$ has mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and variance-covariance matrix $\boldsymbol{Z}_i\boldsymbol{\Sigma}_v\boldsymbol{Z}_i'+\sigma^2\boldsymbol{I}_{n_i}$.

Now for the model with orthogonal polynomials for time (in both $\boldsymbol{X}$ and $\boldsymbol{Z}$), this requires a simple replacement of $\boldsymbol{X}$ with $\boldsymbol{X}(\boldsymbol{S}')^{-1}$ and $\boldsymbol{Z}$ with $\boldsymbol{Z}(\boldsymbol{S}')^{-1}$. Denote the parameters in the orthogonal polynomial metric as $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}_i$ for the fixed and random effects parameters, respectively. Then, the reparameterized model is given as:

$$\boldsymbol{y}_i = \boldsymbol{X}_i(\boldsymbol{S}')^{-1}\boldsymbol{\gamma} + \boldsymbol{Z}_i(\boldsymbol{S}')^{-1}\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i \ . \tag{5.5}$$

The random effects $\boldsymbol{\theta}_i$ are distributed normally with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}_\theta$. As a result, the dependent variable vector $\boldsymbol{y}_i$ has its mean given as $\boldsymbol{X}_i(\boldsymbol{S}')^{-1}\boldsymbol{\gamma}$

and variance-covariance as:

$$\begin{aligned} V(\boldsymbol{y}_i) &= (\boldsymbol{Z}_i(\boldsymbol{S}')^{-1})\boldsymbol{\Sigma}_\gamma(\boldsymbol{Z}_i(\boldsymbol{S}')^{-1})' + \sigma^2\boldsymbol{I}_{n_i} \\ & \quad \boldsymbol{Z}_i\left[(\boldsymbol{S}')^{-1}\boldsymbol{\Sigma}_\gamma\boldsymbol{S}^{-1}\right]\boldsymbol{Z}_i' + \sigma^2\boldsymbol{I}_{n_i} \end{aligned} \tag{5.6}$$

in the transformed metric.

5.3.2 Orthogonal Polynomial Trend Example

Here we refit the model presented in Table 5.1, except that orthogonal polynomials are used. The results of this analysis are given in Table 5.2.

Table 5.2

MRM results for orthogonal polynomial version of level-1 model (5.1) and level-2 model (5.2).

Parameter	Estimate	SE	z	$p <$
γ_0	43.24	1.37	31.61	.0001
γ_1	-9.94	0.86	-11.50	.0001
γ_2	0.31	0.54	0.58	.56
$\sigma^2_{\theta_0}$	111.91	21.60		
$\sigma_{\theta_0\theta_1}$	37.99	10.92		
$\sigma^2_{\theta_1}$	37.04	8.90		
$\sigma_{\theta_0\theta_2}$	-10.14	6.19		
$\sigma_{\theta_1\theta_2}$	-0.82	3.80		
$\sigma^2_{\theta_2}$	7.23	3.50		
σ^2	10.52	1.11		

Note. $-2\log L = 2207.64$

First, notice that the log-likelihood value is identical in Tables 5.1 and 5.2. Thus, the

two solutions are equivalent, one is simply a re-expressed version of the other. Comparing the regression coefficients, as before, we see that only the constant and linear terms are significant. These terms also dominate in terms of magnitude; not only is the quadratic term non-significant, it is negligible. Thus, at the population average level, the trend is essentially linear.

Turning to the variance estimates, we see that the estimated constant variance ($\hat{\sigma}^2_{\theta_0}$) is much larger than the estimated linear trend component ($\hat{\sigma}^2_{\theta_1}$), which is much larger than the estimated quadratic trend component ($\hat{\sigma}^2_{\theta_2}$). In terms of relative percentages, these three represent 71.7, 23.7, and 4.6, respectively, of the sum of the estimated individual variance terms. Thus, at the individual level there is heterogeneity in terms of all three components, but with diminishing return as the order of the polynomial increases. This analysis then quantifies what the plot of the empiricaly Bayes trends in Figure 5.4 depicts.

Inspection of the covariance terms reveals a strong positive association between the constant and linear terms ($\hat{\sigma}^2_{\theta_1\theta_0} = 37.99$, expressed as a correlation = .59). This seems to be in contrast with the results for this term from the previous analysis in Table 5.1, where there was a slight negative association between the intercept and linear terms ($\hat{\sigma}^2_{\upsilon_1\upsilon_0} = -.92$, expressed as a correlation = -.11). The reason for this apparent discrepancy is that in Table 5.1, the intercept represents the first timepoint, whereas the constant term in Table 5.2 represents the average across time. Thus, an individual's linear trend is both negatively associated with their baseline depression level and positively associated with their average depression level. Subjects with higher initial depression levels have more negative linear slopes, and as a result, lower average depression values.

5.3.3 Translating parameters

The parameters from a model using orthogonal polynomials can be directly related to those in the corresponding model that uses the original metric for time. One is simply a re-expressed or translated version of the other, with the matrix $\boldsymbol{S}'$ or $(\boldsymbol{S}')^{-1}$ serving as the translator, depending on the direction of the translation. To see the connection between the two, consider again the matrix representation of the MRM:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\upsilon}_i + \boldsymbol{\varepsilon}_i \tag{5.7}$$

in the original metric, and

$$\boldsymbol{y}_i = \boldsymbol{X}_i(\boldsymbol{S}')^{-1}\boldsymbol{\gamma} + \boldsymbol{Z}_i(\boldsymbol{S}')^{-1}\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i \tag{5.8}$$

in the orthogonal metric. Thus, the two sets of model parameters are related according to:

$$\boldsymbol{\beta} = (\boldsymbol{S}')^{-1}\boldsymbol{\gamma} \tag{5.9}$$

$$\boldsymbol{\upsilon}_i = (\boldsymbol{S}')^{-1}\boldsymbol{\theta}_i \tag{5.10}$$

The two representations of the variance-covariance matrix of the random effects are related in a similar way. From before, we have that

$$V(\boldsymbol{y}_i) = \boldsymbol{Z}_i\boldsymbol{\Sigma}_\upsilon\boldsymbol{Z}_i' + \sigma^2\boldsymbol{I}_{n_i} \tag{5.11}$$

in the original MRM, and

$$V(\boldsymbol{y}_i) = \boldsymbol{Z}_i\left[(\boldsymbol{S}')^{-1}\boldsymbol{\Sigma}_\gamma\boldsymbol{S}^{-1}\right]\boldsymbol{Z}_i' + \sigma^2\boldsymbol{I}_{n_i} \tag{5.12}$$

in the transformed metric. The relationship between the two versions of the variance-covariance matrix of the random effects is therefore:

$$\boldsymbol{\Sigma}_{\upsilon} = (\boldsymbol{S}')^{-1}\boldsymbol{\Sigma}_{\gamma}\boldsymbol{S}^{-1} \ . \tag{5.13}$$

As can be seen, $(\boldsymbol{S}')^{-1}$ is the matrix translating the orthogonal polynomial parameters to those in the original metric of time, and similarly $\boldsymbol{S}'$ translates the parameters from the original metric to the orthogonal polynomial metric, namely:

$$\boldsymbol{\gamma} = \boldsymbol{S}'\boldsymbol{\beta} \tag{5.14}$$

$$\boldsymbol{\theta}_i = \boldsymbol{S}'\boldsymbol{\upsilon}_i \tag{5.15}$$

$$\boldsymbol{\Sigma}_{\gamma} = \boldsymbol{S}'\boldsymbol{\Sigma}_{\upsilon}\boldsymbol{S} \tag{5.16}$$

These equations can be used to translate the estimates from one version of the model to the other simply by plugging in the parameter estimates that are obtained from an analysis. For example, one can derive the results from Table 5.1 (original time) based on those from Table 5.2 (orthogonal time) according to:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{S}')^{-1}\hat{\boldsymbol{\gamma}} \\ &= \begin{bmatrix} 0.4082 & -0.5976 & 0.5455 \\ 0 & 0.2390 & -0.8183 \\ 0 & 0 & 0.1637 \end{bmatrix} \begin{bmatrix} 43.24 \\ -9.94 \\ 0.31 \end{bmatrix} = \begin{bmatrix} 23.76 \\ -2.63 \\ 0.05 \end{bmatrix} \end{aligned} \tag{5.17}$$

To get standard errors of the parameter estimates a bit more work is required. For this, one must obtain the variance-covariance matrix of the parameter estimates. Note that the standard errors for a set of parameter estimates are equal to the square root of the diagonal

entries of this matrix. For the regression coefficients, denote these as $V(\hat{\boldsymbol{\beta}})$ and $V(\hat{\boldsymbol{\gamma}})$ for our two versions of the model. Both of these are symmetric $p \times p$ matrices, where p is the number of polynomial trend components (including the intercept). Then, the following relationships hold:

$$V(\hat{\boldsymbol{\beta}}) = (\boldsymbol{S}')^{-1}\, V(\hat{\boldsymbol{\gamma}})\, \boldsymbol{S}^{-1} \tag{5.18}$$

$$V(\hat{\boldsymbol{\gamma}}) = \boldsymbol{S}'\, V(\hat{\boldsymbol{\beta}})\, \boldsymbol{S} \tag{5.19}$$

Taking the square root of the diagonal elements of these matrices then provide the standard errors for these two versions of the regression coefficients.

From our orthogonal analysis in Table 5.2, we have

$$V(\hat{\boldsymbol{\gamma}}) = \begin{bmatrix} 1.8708 & 0.5823 & -0.1402 \\ 0.5823 & 0.7470 & 0.0214 \\ -0.1402 & 0.0214 & 0.2914 \end{bmatrix}$$

and so,

$$\begin{aligned} V(\hat{\boldsymbol{\beta}}) &= \begin{bmatrix} 0.4082 & -0.5976 & 0.5455 \\ 0 & 0.2390 & -0.8183 \\ 0 & 0 & 0.1637 \end{bmatrix} \begin{bmatrix} 1.8708 & 0.5823 & -0.1402 \\ 0.5823 & 0.7470 & 0.0214 \\ -0.1402 & 0.0214 & 0.2914 \end{bmatrix} \\ &\times \begin{bmatrix} 0.4082 & 0 & 0 \\ -0.5976 & 0.2390 & 0 \\ 0.5455 & -0.8183 & 0.1637 \end{bmatrix} \\ &= \begin{bmatrix} 0.3048 & -0.1199 & 0.0146 \\ -0.1199 & 0.2294 & -0.0382 \\ 0.0146 & -0.0382 & 0.0078 \end{bmatrix}. \end{aligned}$$

Taking the square root of the diagonal entries of this matrix yields 0.5521, 0.4790, and 0.0883, which agree with the reported standard errors for the fixed effects in Table 5.1.

To obtain the standard errors for the variance-covariance parameter estimates, it is somewhat more difficult though the logic is the same. Notice that the regression coefficients are $p \times 1$ parameter vectors, whereas $\boldsymbol{\Sigma}_\upsilon$ and $\boldsymbol{\Sigma}_\theta$ are $r \times r$ symmetric matrices of parameters (where r is the number of random effects). The first step is then to vectorize these parameter matrices. This can be done using results from McCulloch [1982]. Specifically, we can vectorize the relationship $\boldsymbol{\Sigma}_\theta = \boldsymbol{S}'\boldsymbol{\Sigma}_\upsilon\boldsymbol{S}$ as

$$\begin{aligned} \text{vec}\boldsymbol{\Sigma}_\theta &= \text{vec}(\boldsymbol{S}'\boldsymbol{\Sigma}_\upsilon\boldsymbol{S}) \\ &= (\boldsymbol{S}' \otimes \boldsymbol{S}')\text{vec}\boldsymbol{\Sigma}_\upsilon \end{aligned}$$

where $\text{vec}\boldsymbol{\Sigma}_\theta$ is the $r^2 \times 1$ vector that is obtained by stacking the r column vectors of the matrix $\boldsymbol{\Sigma}_\theta$ on top of each other (similarly for $\text{vec}\boldsymbol{\Sigma}_\upsilon$), and $\otimes$ represents the Kronecker product. For example, for a 3×3 matrix $\boldsymbol{\Sigma}$ the vec would be (using σ to represent either a variance or covariance parameter):

$$\text{vec}\boldsymbol{\Sigma} = (\sigma_{11}\,\sigma_{21}\,\sigma_{31}\,\sigma_{12}\,\sigma_{22}\,\sigma_{32}\,\sigma_{13}\,\sigma_{23}\,\sigma_{33})'$$

Since $\boldsymbol{\Sigma}_\upsilon$ and $\boldsymbol{\Sigma}_\theta$ are symmetric matrices, not all of the r^2 elements in $\text{vec}\boldsymbol{\Sigma}_\theta$ or $\text{vec}\boldsymbol{\Sigma}_\upsilon$ are unique; instead, each have $r \times (r+1)/2$ unique parameters. For this, McCulloch [1982] defines the "vech" as the $r^* \times 1$ vector containing only the unique elements of a symmetric matrix, where $r^* = r \times (r+1)/2$. The vech stacks the on or above diagonal elements of a square matrix on top of each other to form one column vector. Thus, for the same 3×3 matrix $\boldsymbol{\Sigma}$ considered above, the vech would be:

$$\text{vech}\boldsymbol{\Sigma} = (\sigma_{11}\,\sigma_{12}\,\sigma_{22}\,\sigma_{13}\,\sigma_{23}\,\sigma_{33})'$$

Furthermore, $\boldsymbol{G}$ is the $r^2 \times r^*$ transformation matrix that carries the vech into the vec, *i.e.*, $\text{vec}\boldsymbol{\Sigma} = \boldsymbol{G}\text{vech}\boldsymbol{\Sigma}$. For example, with $r = 3$, the form of the $\boldsymbol{G}$ matrix is:

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.20}$$

As a result, we can write:

$$\boldsymbol{G}\text{vech}\boldsymbol{\Sigma}_\theta = (\boldsymbol{S}' \otimes \boldsymbol{S}')\boldsymbol{G}\text{vech}\boldsymbol{\Sigma}_\upsilon$$

so that

$$\text{vech}\boldsymbol{\Sigma}_\theta = \boldsymbol{G}^+(\boldsymbol{S}' \otimes \boldsymbol{S}')\boldsymbol{G}\text{vech}\boldsymbol{\Sigma}_\upsilon \tag{5.21}$$

where $\boldsymbol{G}^+$ is the Moore-Penrose inverse of $\boldsymbol{G}$ with the property that $\text{vech}\boldsymbol{\Sigma} = \boldsymbol{G}^+\text{vec}\boldsymbol{\Sigma}$ [Magnus, 1988]. For $r = 3$, the $\boldsymbol{G}^+$ matrix is:

$$\boldsymbol{G}^+ = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.22}$$

One can verify that pre-multiplying the vec by $\boldsymbol{G}^+$ does yield the vech, and that that pre-multiplying the vech by $\boldsymbol{G}$ yields the vec for this 3×3 case.

Now we have the vector of r^* unique parameters in $\boldsymbol{\Sigma}_\theta$ expressed as a function of the r^* unique parameters in $\boldsymbol{\Sigma}_\upsilon$, pre-multiplied by the matrix product $\boldsymbol{G}^+(\boldsymbol{S}' \otimes \boldsymbol{S}')\boldsymbol{G}$. While this is a more complicated expression than that relating the regression coefficients (*i.e.*, $\boldsymbol{\gamma} = \boldsymbol{S}'\boldsymbol{\beta}$), it has the same basic form. As a result, to obtain the variance-covariance matrix of the r^* estimates in $\hat{\boldsymbol{\Sigma}}_\theta$, we get

$$V(\text{vech}\hat{\boldsymbol{\Sigma}}_\theta) = \boldsymbol{G}^+(\boldsymbol{S}' \otimes \boldsymbol{S}')\boldsymbol{G}\, V(\text{vech}\hat{\boldsymbol{\Sigma}}_\upsilon)\, \boldsymbol{G}'(\boldsymbol{S} \otimes \boldsymbol{S})\boldsymbol{G}^{+'} \,. \tag{5.23}$$

Notice that $V(\text{vech}\hat{\boldsymbol{\Sigma}}_\theta)$ is a $r^* \times r^*$ matrix, and that taking the square root of the diagonal elements of this matrix yields the standard errors for the r^* unique elements of $\hat{\boldsymbol{\Sigma}}_\theta$. In a similar way, we can obtain

$$V(\text{vech}\hat{\boldsymbol{\Sigma}}_\upsilon) = \boldsymbol{G}^+((\boldsymbol{S}')^{-1} \otimes (\boldsymbol{S}')^{-1})\boldsymbol{G}\, V(\text{vech}\hat{\boldsymbol{\Sigma}}_\theta)\, \boldsymbol{G}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{S}^{-1})\boldsymbol{G}^{+'} \,, \tag{5.24}$$

which again is a $r^* \times r^*$ matrix, which yields the standard errors for the r^* unique elements of $\hat{\boldsymbol{\Sigma}}_\upsilon$ by taking the square root of the diagonal elements of this matrix.

5.3.4 Higher-order Polynomial Models

Thus far we have only considered curvilinear or quadratic trend models. Adding higher order polynomials is sometimes of interest. In general if there are n timepoints than one can consider $n-1$ polynomials in terms of the mean structure, where the first-order polynomial is the linear, the second-order is the quadratic, etc. For example, for the Reisby dataset

with six timepoints, a model with up to the 5th-order polynomial could be fit in terms of the fixed effects. With equal numbers of observations per timepoint, such a model would fit the marginal timepoint means exactly because as many parameters (intercept plus $n-1$ polynomial terms) are estimated as timepoints (n). In the unbalanced case, the fit would be approximately equal.

How many polynomials can be fit in the random-effects part depends on the variation in the data and its correlational structure. The maximum would also be $n-1$ polynomials, though in this case it is not possible to also estimate the error variance. This is because the variance-covariance matrix of $\boldsymbol{y}$ is $n \times n$, and so a maximum of $n(n-1)/2$ variance-covariance parameters are estimable. Thus, a model with a random intercept plus $n-1$ random polynomial terms would represent a saturated model for the variance-covariance structure.

Often, interpretation of higher-order polynomials beyond say quadratic and cubic trends is complicated and there may be insufficient rationale to expect such trends anyway. Of course this depends on the problem, but often models with linear, quadratic, and possibly cubic polynomial trends do a reasonable job of balancing model fit and interpretation.

To augment our quadratic trend model to include a cubic trend, the t^3 (original or othogonal metric) is added to the within-subjects (level-1) model, namely,

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + b_{2i}t_{ij}^2 + b_{3i}t_{ij}^3 + \varepsilon_{ij} \quad . \tag{5.25}$$

Here, the additional term b_{3i} represents the subject's cubic trend component. The level-2,

between-subjects, model is now

$$
\begin{aligned}
b_{0i} &= \beta_0 + \upsilon_{0i} \\
b_{1i} &= \beta_1 + \upsilon_{1i} \\
b_{2i} &= \beta_2 + \upsilon_{2i} \\
b_{3i} &= \beta_3 + \upsilon_{2i} \qquad (5.26)
\end{aligned}
$$

where β_3 represents the average cubic change and υ_{3i} is the individual deviation from average cubic change. As in the quadratic trend model, in some cases it may be reasonable to restrict the cubic part to the level-1 model only (in which case the final component of the level-2 model would simply be $b_{3i} = \beta_3$).

Figure 5.5 illustrates some cubic curves, where the general trend is positive across time.

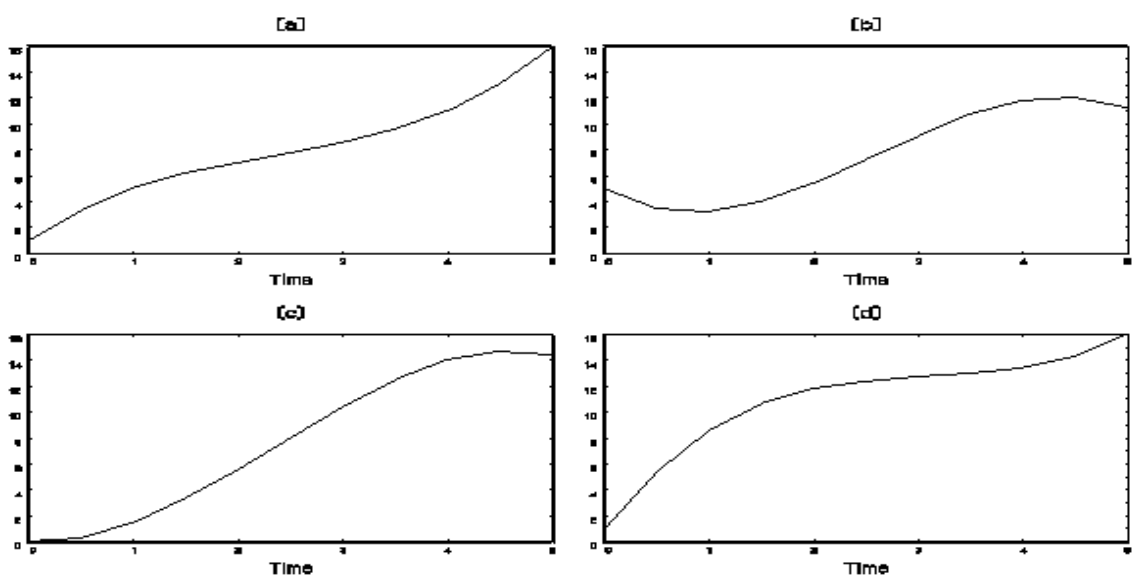

Figure 5.5. Cubic models of generally positive change across time:

(a) $y = 1 + 5.5t - 1.75t^2 + 0.25t^3$ (b) $y = 5 - 4.5t + 3.15t^2 - 0.40t^3$

(c) $y = 0 - 0.5t + 2.30t^2 - 0.325t^3$ (d) $y = 1 + 10.5t - 3.25t^2 + 0.35t^3$

As can be seen, a wide variety of curves are possible. One shape that is particularly useful is the s-shaped curve akin to Figure 5.5c. In this type of curve the response is first relatively flat, then rises, and finally levels off. Though such a non-linear relationship across time may be fitted best using a logistic response curve, the cubic model does a reasonable job of approximating this type of trend in some cases. Of course, one must keep in mind that these curves can yield nonsensical values of y if extrapolated in the x direction. Figure 5.6 illustrates these same cubic curves, except that now two extra weeks are added on to each side for time, so that now the range is -2 to 7 (the dashed lines in these figures represent the original timeframe of 0 to 5).

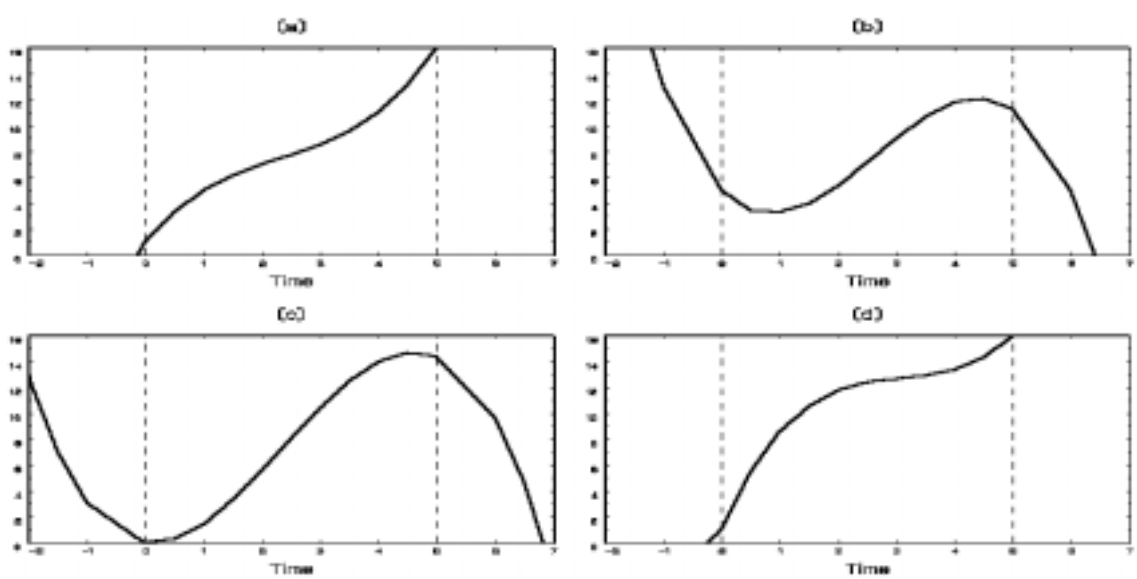

Figure 5.6. Extrapolation of cubic models across time:

(a) $y = 1 + 5.5t - 1.75t^2 + 0.25t^3$ (b) $y = 5 - 4.5t + 3.15t^2 - 0.40t^3$

(c) $y = 0 - 0.5t + 2.30t^2 - 0.325t^3$ (d) $y = 1 + 10.5t - 3.25t^2 + 0.35t^3$

Notice that these cubic trends don't level off at weeks 0 or 5, but instead keep going and going, yielding values of y that are (eventually) "off the map." Thus, it is a good idea to plot the estimated curve(s) from an analysis using higher-order polynomials to ensure that the interpretation is reasonable. Alternatively, though we won't consider them here, more general polynomial models with asymptotes are possible, and have recently been described in an MRM context by Jones [1996].

5.3.5 Cubic Trend Example

The Reisby dataset is fit using a orthogonal cubic trend model and the results of this analysis are given in Table 5.3.

Table 5.3

MRM results for orthogonal polynomial version of level-1 model (5.25) and level-2 model (5.26).

Parameter	Estimate	SE	z	$p <$
intercept γ_0	43.24	1.35	31.95	.0001
linear trend γ_1	-9.88	0.84	-11.70	.0001
quadratic trend γ_2	0.33	0.55	0.60	.55
cubic trend γ_3	0.62	0.48	1.29	.20
$\sigma^2_{\theta_0}$	110.91	21.08		
$\sigma_{\theta_0\theta_1}$	34.92	10.25		
$\sigma^2_{\theta_1}$	36.25	8.25		
$\sigma_{\theta_0\theta_2}$	-11.14	6.12		
$\sigma_{\theta_1\theta_2}$	0.26	3.69		
$\sigma^2_{\theta_2}$	9.24	3.58		
$\sigma_{\theta_0\theta_3}$	7.25	5.31		
$\sigma_{\theta_1\theta_3}$	-4.33	3.27		
$\sigma_{\theta_2\theta_3}$	4.03	2.16		
$\sigma^2_{\theta_3}$	5.04	2.85		
σ^2	8.92	1.14		

Note. $-2\log L = 2196.44$

Comparing the deviance from this model (2196.44) with that from the quadratic trend model (2207.64) yields a difference of $\chi^2 = 11.2$, which is significant at $p < .05$, even without an

adjustment of the p-value for the inclusion of variance terms in this global test of

$$H_0 : \gamma_3 = \sigma^2_{\theta_3} = \sigma_{\theta_0\theta_3} = \sigma_{\theta_1\theta_3} = \sigma_{\theta_2\theta_3} = 0 \ .$$

The estimates of the fixed effects again re-affirm the linearity of the response across time at the average level. Neither the quadratic nor the cubic terms are statistically significant. Relative to the model with only quadratic trend in Table 5.2, notice how little the estimates for the intercept, linear trend, and quadratice trend have changed. Again, if the data were complete across time these would not change at all, but because the data are not complete across time these trend components are not exactly independent of each other and so there are slight changes when an additional polynomial is added to the model.

Turning to the variance estimates, we see that these clearly diminish as the order of the polynomial is increased. In terms of relative percentages, the four trend components represent 68.7, 22.5, 5.7, and 3.1, respectively, of the sum of the estimated individual variance terms. Thus, over 90% of the individual heterogeneity is in terms of the constant (*i.e.*, average across time) and the linear trend component. Though statistically significant, the quadratic and cubic terms do not account for a great deal of the individual differences in trends across time.

This can also be seen from Figure 5.7 which presents the observed means, estimated means, and estimated individual trends based on the cubic model.

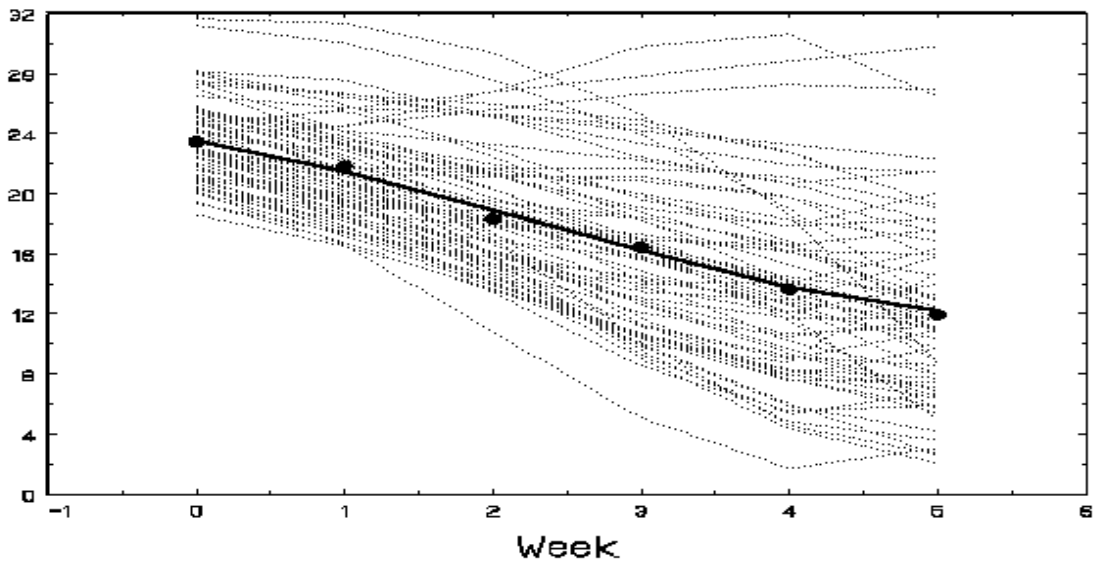

Figure 5.7. Reisby data: observed means (solid circles), estimated means (solid line), and estimated individual trends (dotted)

As has been seen before with this dataset, the observed means clearly descend linearly across time. Though the plot is a bit busy, one can discern a few individuals displaying non-linear trend across time, though the extent of the non-linearity does not appear to be that great. To zoom in on this a little more, Figure 5.8 presents the data for the ten subjects with largest estimates (in absolute value) of the cubic trend component $\hat{\gamma}_3 + \hat{\theta}_{i3}$.

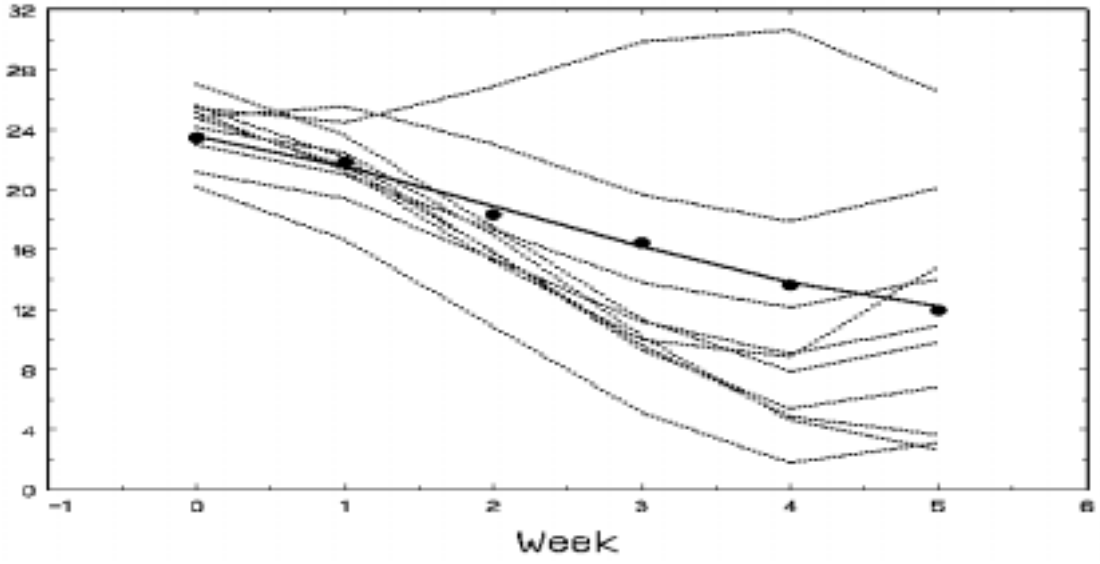

Figure 5.8. Reisby data: observed means (solid circles), estimated means (solid line), and estimated individual trends (dotted) for 10 subjects with largest cubic trend components

This plot does show a backwards S-shaped curve for at least a subset of individuals. For these subjects, the decrease in depression scores is initially not strong, then grows in the middle timepoints, and finally tapers off or increases slightly at the end. One subject even has increased depression scores following an elongated S-shaped curve. Thus, the cubic trend component does play a role in explaining some of the individual heterogeneity in trends across time, albeit a relatively modest role.

Finally, Table 5.4 presents the observed standard deviations across time, along with the

model-based estimates for the various MRMs fit to these data.

Table 5.4

Observed and estimated standard deviations across time

	week 0	week 1	week 2	week 3	week 4	week 5
observed	4.53	4.70	5.49	6.41	6.97	7.22
<u>model-based estimates</u>						
random intercept	5.93	5.93	5.93	5.93	5.93	5.93
random linear trend	4.98	4.91	5.24	5.92	6.84	7.91
random quadratic trend	4.58	4.88	5.57	6.19	6.78	7.69
random cubic trend	4.50	4.73	5.31	6.37	7.06	7.32

One can see that the fit of the observed standard deviations is incrementally better as the order of the polynomial is increased. In particular, comparing the random quadratic trend to the random cubic trend, one gets a feel for why the latter is preferred by the likelihood-ratio test; the cubic trend model does an excellent job of fitting the variation in the observed HD scores across time. It should be noted that each of these models includes all lower order polynomials as random effects as well. Thus, for example, the random quadratic model includes random intercept and linear trend components too.

The improved fit of the cubic model for the variance-covariance structure can also be seen in the (co)variance plot in Figure 5.9.

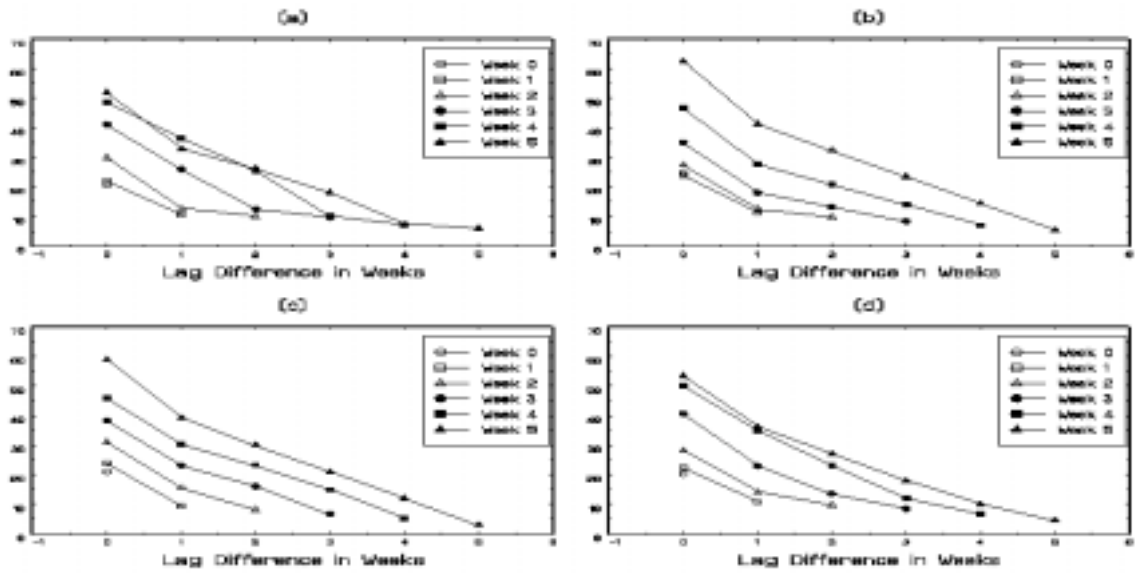

Figure 5.9. Reisby data: (a) observed (co)variances

(b) random linear trend model (co)variance estimates

(c) random quadratic trend model (co)variance estimates

(d) random cubic trend model (co)variance estimates

By comparing the different model estimates to the observed variances and covariances, one gets a sense of the improvement in model fit provided as the order of the polynomial is increased.

5.4 Summary

Polynomial trend models are considered in this chapter, in particular models with quadratic and cubic trend components. Since trends across time can be non-linear, the use of polynomials is a useful tool for longitudinal data analysis. It is important to remember that the model is still a linear one, since it is linear in terms of the regression coefficients. The polynomial regressors are what allow the linear model to fit some types of non-linear relationships. In some cases, a non-linear regression part is required, for example, for modeling of human stature across time [Bock and du Toit, 2004]. For interested readers, nonlinear mixed-effects regression models are considered extensively in Davidian and Giltinan [1995] and Vonesh and Chinchilli [1997].

Since the orthogonal polynomial representation greatly reduces any collinearity and scale differences in the regressors, it is computationally easier to obtain. For this reason, in cases where numerical difficulties are occurring with analyses using raw time values, investigators might consider using orthogonal polynomials instead. In this chapter we have considered the use of orthogonal polynomials in MRMs in detail, though they can be applied more generally in longitudinal data analysis models. For example, in later chapters we will present MRMs and GEE models for categorical outcomes, and orthogonal polynomials can be effectively applied within these types of models as well.