Mixed-effects Polynomial Regression Models

chapter 5

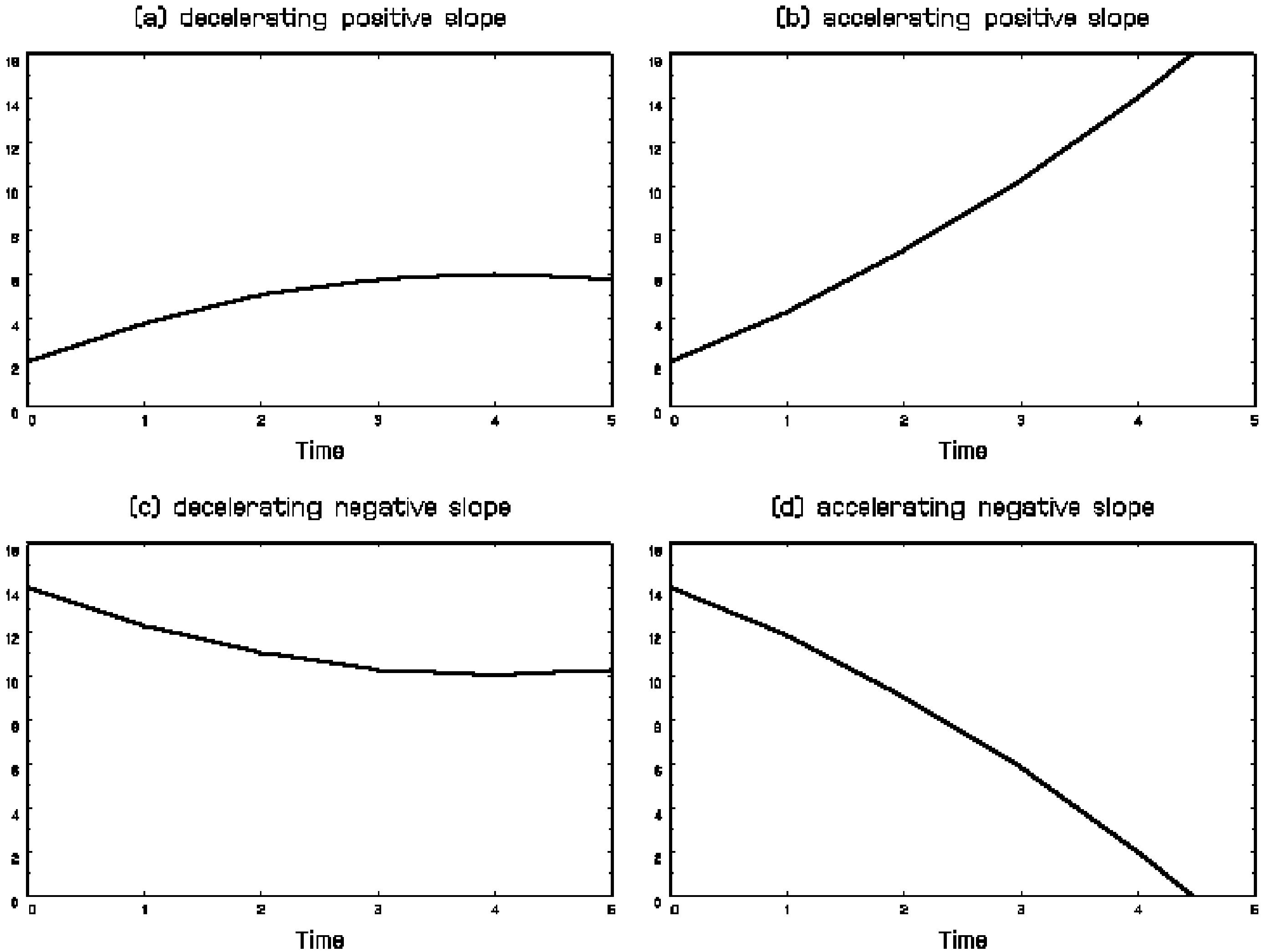

Figure 5.1 Various curvilinear models: (a) decelerating positive slope; (b) accelerating positive slope; (c) decelerating negative slope; (d) accelerating negative slope

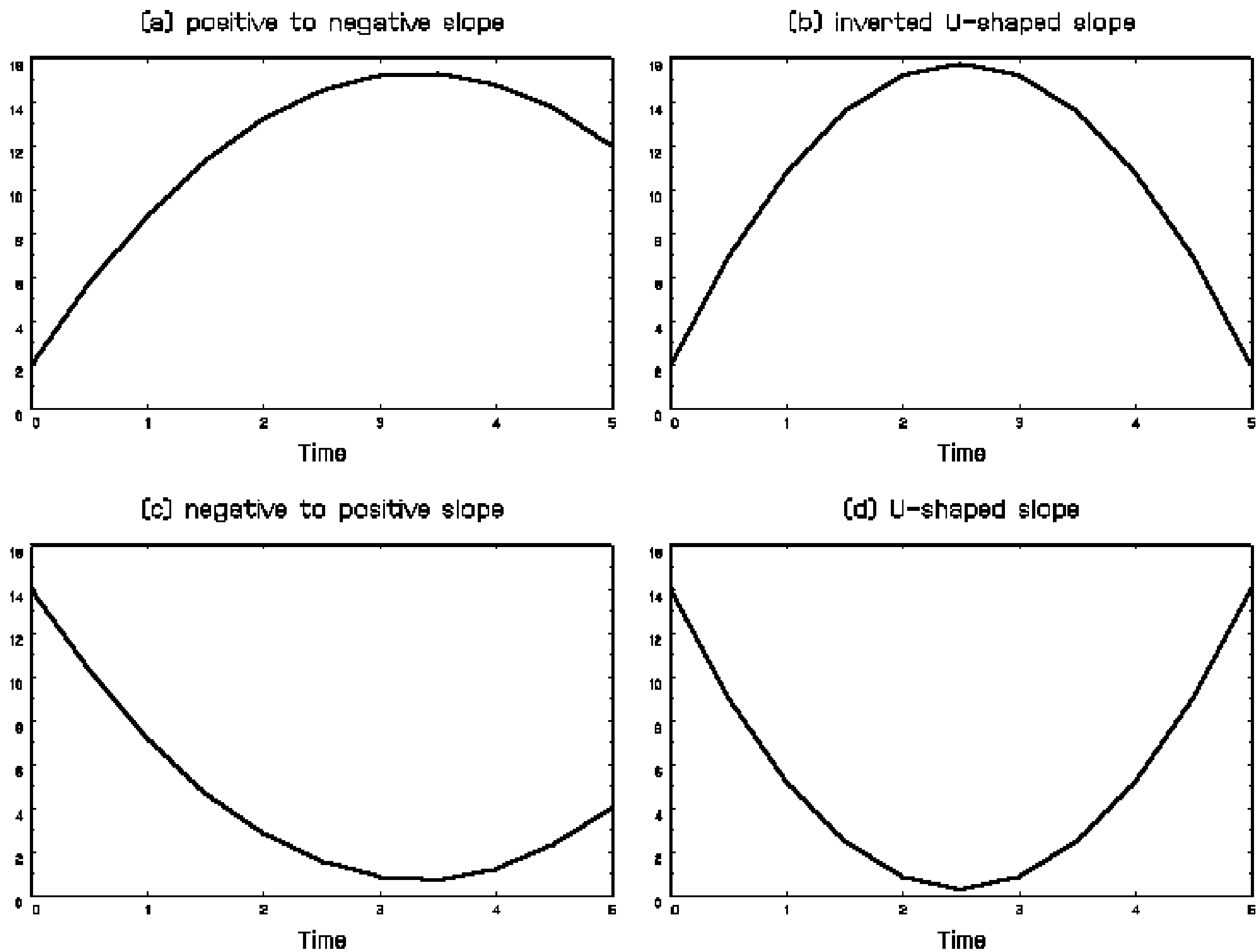

Figure 5.2 More curvilinear models:
(a) positive to negative slope ($\beta_0 = 2, \beta_1 = 8, \beta_2 = -1.2$);
(b) inverted U-shaped slope ($\beta_0 = 2, \beta_1 = 11, \beta_2 = -2.2$);
(c) negative to positive slope ($\beta_0 = 14, \beta_1 = -8, \beta_2 = 1.2$);
(d) U-shaped slope ($\beta_0 = 14, \beta_1 = -11, \beta_2 = 2.2$)

Expressing Time with Orthogonal Polynomials

Instead of

$$\boldsymbol{X}' = \boldsymbol{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix}$$

use

$$\boldsymbol{X}' = \boldsymbol{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -3 & -1 & 1 & 3 & 5 \\ 5 & -1 & -4 & -4 & -1 & 5 \end{bmatrix} \begin{array}{l} /\sqrt{6} \\ /\sqrt{70} \\ /\sqrt{84} \end{array}$$

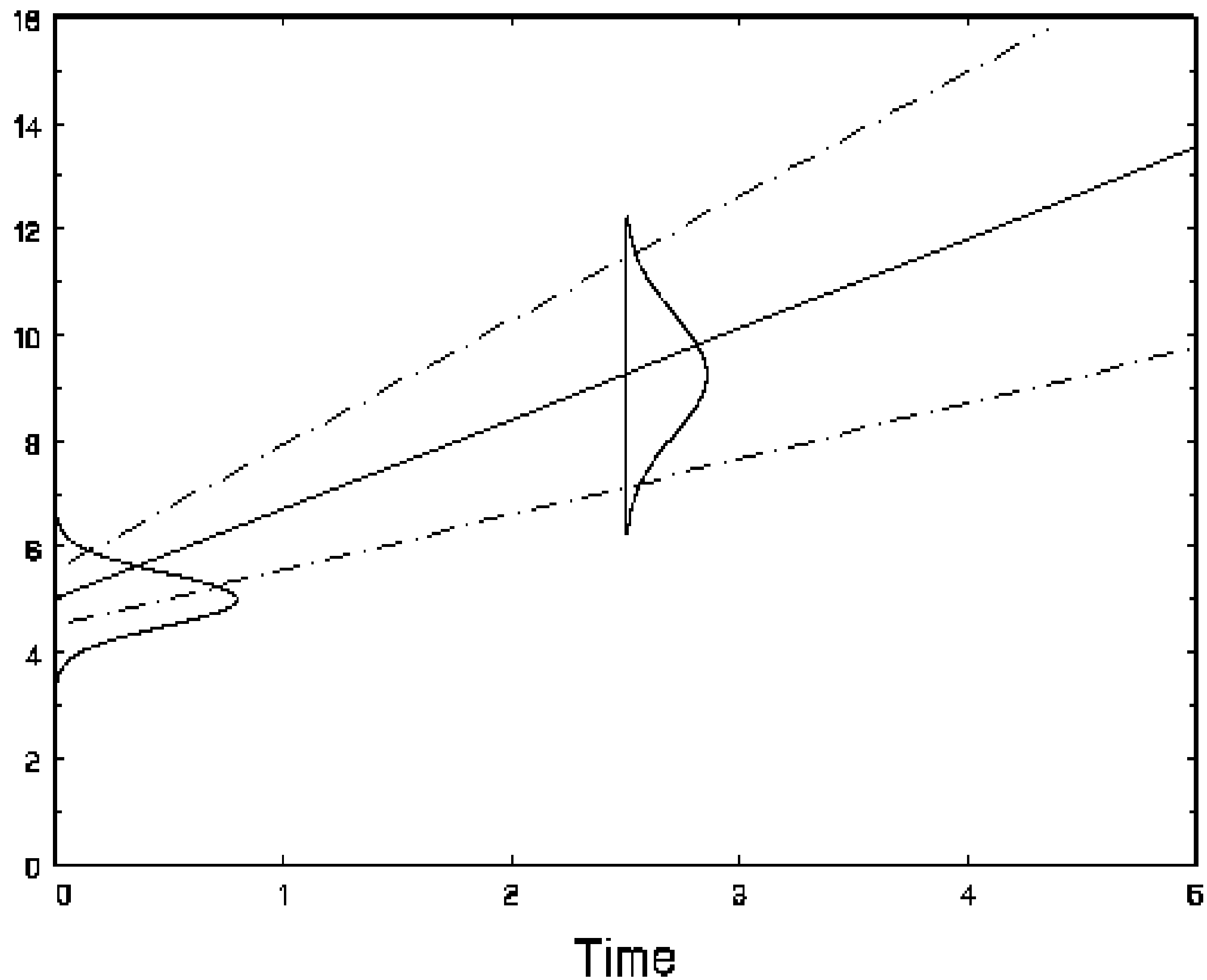

Figure 4.5 Intercept variance changes with coding of time

Top 10 reasons to use Orthogonal Polynomials

10 They look complicated, so it seems like you know what you're doing

9 With a name like orthogonal polynomials they have to be good

8 Good for practicing lessons learned from "hooked on phonics"

7 Decompose more quickly than (orthogonal?) polymers

6 Sound better than lame old releases from Polydor records

5 Less painful than a visit to the orthodonist

4 Unlike ortho, won't kill your lawn weeds

3 Might help you with a polygraph test

2 Great conversation topic for getting rid of unwanted "friends"

1 Because your instructor will give you an F otherwise

“Real” reasons to use Orthogonal Polynomials

- for balanced data, and CS structure, estimates of polynomial fixed effects $\boldsymbol{\beta}$ (*e.g.*, constant and linear) won’t change when higher-order polynomial terms (*e.g.*, quadratic and cubic) are added to the model
- in original scale, it gets increasingly difficult to estimate higher-degree terms (coefficients get smaller and smaller as the scale of X^p gets larger and larger)
- avoids high correlation between estimates, which can cause estimation problems
- provides comparison of importance of different polynomials
- intercept (and intercept-related parameters) represents the grand mean of Y

Orthogonal Polynomials via the Cholesky decomposition in 4 easy steps!

Suppose six equally-spaced timepoints

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

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

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

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 $(\boldsymbol{S'})^{-1}$

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

which yields the same orthogonal polynomial values as before

Orthogonal Polynomials via SAS

```
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];
```

Model Representations

Consider the matrix representation of the MRM:

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

- replace $\boldsymbol{X}$ with $\boldsymbol{X}(\boldsymbol{S}')^{-1}$ and $\boldsymbol{Z}$ with $\boldsymbol{Z}(\boldsymbol{S}')^{-1}$
- denote parameters in the orthogonal polynomial metric as $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}_i$ for the fixed and random effects, respectively

The reparameterized model is given as

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

Orthogonal Polynomial analysis of Reisby data

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		

$-2 \log L = 2207.64$

- log-likelihood value is identical to analysis in raw metric
- as before, only the constant and linear fixed-effect terms are significant (these terms also dominate in terms of magnitude), thus, the population trend is essentially linear
- 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 ($\hat{\sigma}^2_{\theta_2}$)
  - 71.7%, 23.7%, and 4.6%, respectively, of the sum of the estimated individual variance terms. At individual level, there is diminishing heterogeneity as the order of the polynomial increases
- a strong positive association between the constant and linear terms ($\hat{\sigma}^2_{\theta_1\theta_0} = 37.99$, expressed as a correlation $= .59$)
  - an individual's linear trend is positively associated with their average depression level

Can obtain same estimates of overall and individual trends as previous analysis (in raw metric)

$$\begin{aligned} \boldsymbol{\beta} &= (\boldsymbol{S}')^{-1}\boldsymbol{\gamma} \\ \boldsymbol{\upsilon}_i &= (\boldsymbol{S}')^{-1}\boldsymbol{\theta}_i \end{aligned}$$

e.g.,

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

$$= \begin{bmatrix} 0.4082 & -0.5976 & 0.5455 \\ 0 & 0.2390 & -0.8183 \\ 0 & 0 & 0.1637 \end{bmatrix} \begin{bmatrix} 43.24 \\ -9.94 \\ 0.31 \end{bmatrix} = \begin{bmatrix} 23.76 \\ -2.63 \\ 0.05 \end{bmatrix}$$

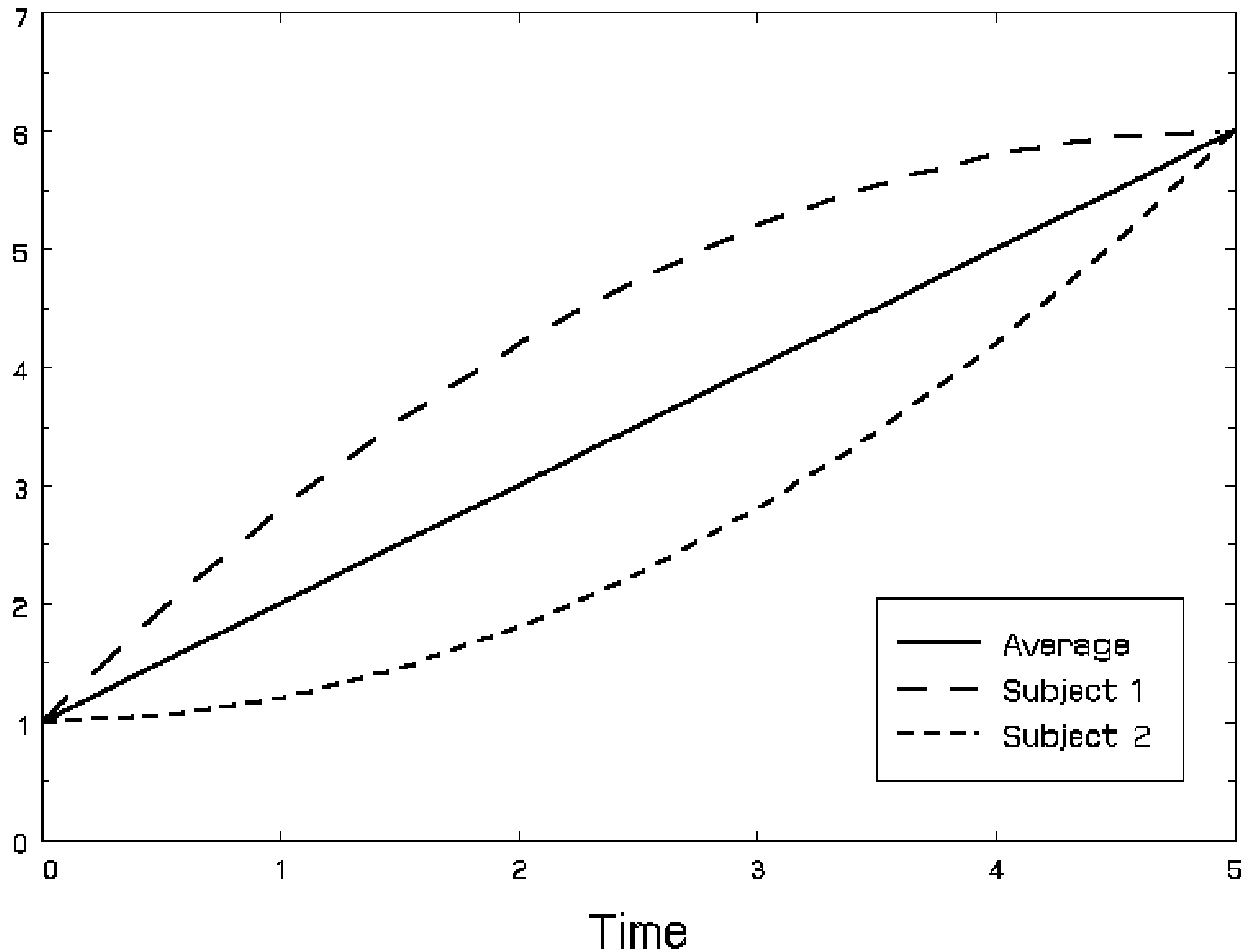

Figure 5.3 Average linear and individual quadratic trends

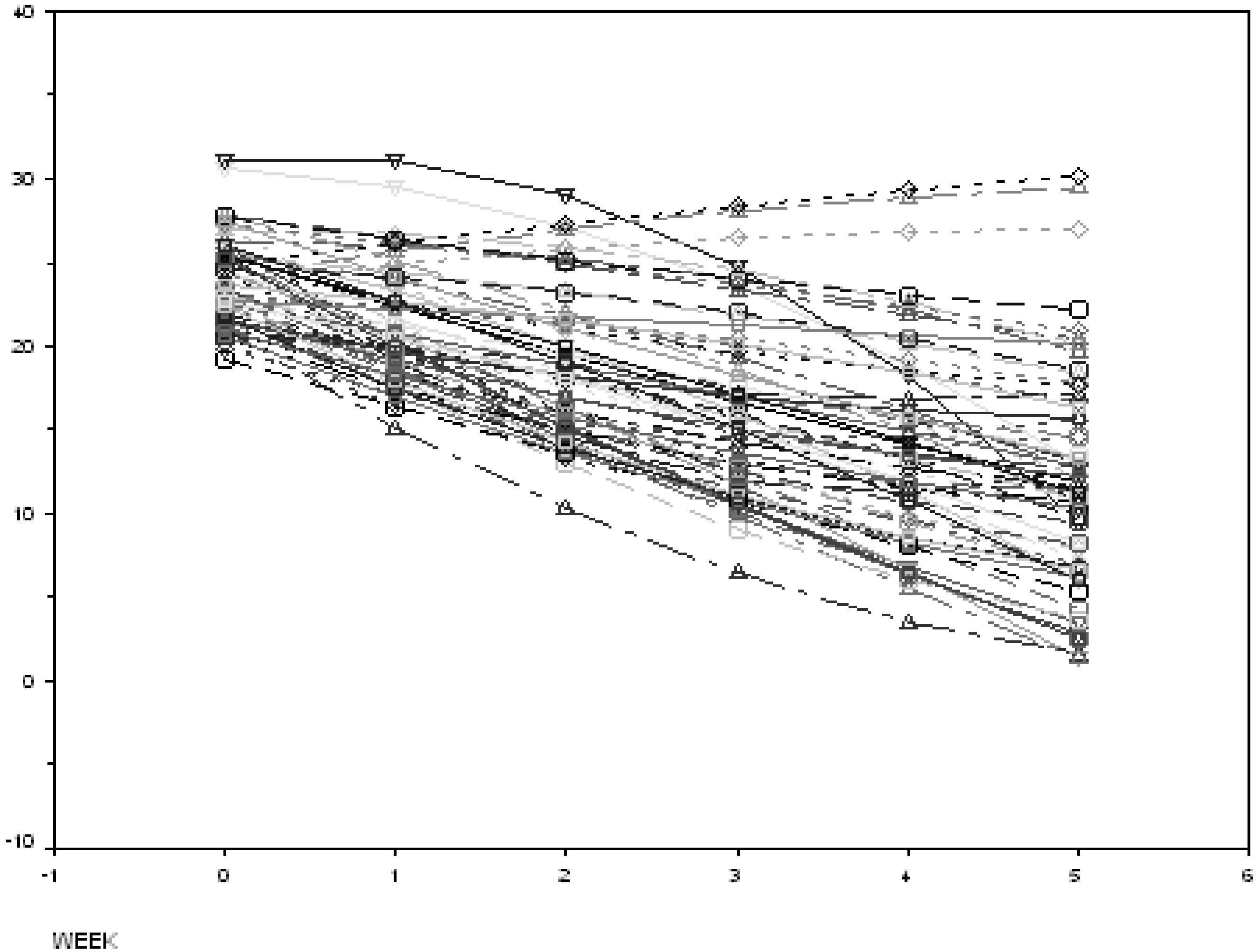

Figure 5.4 Reisby data: estimated curvilinear trends

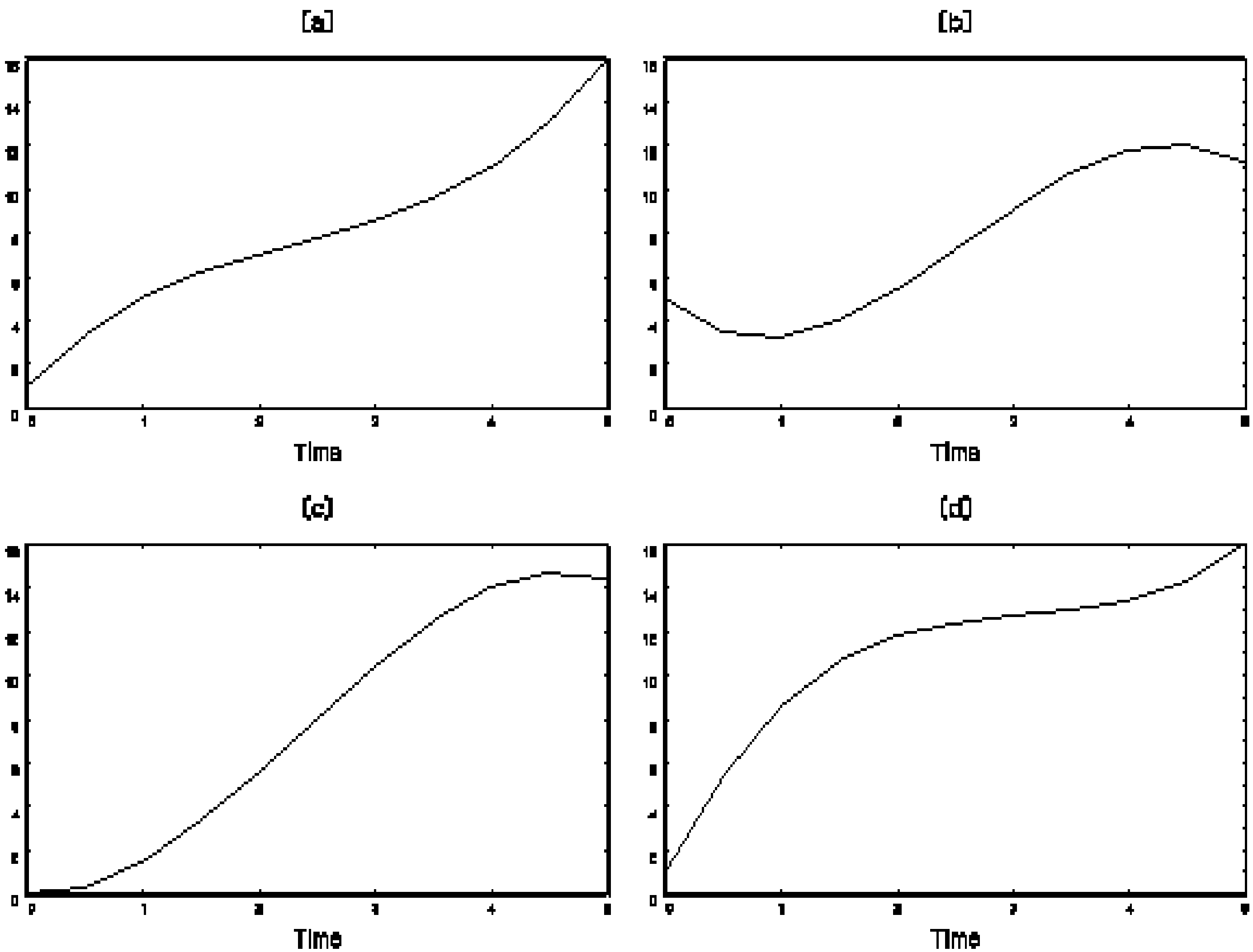

Figure 5.5 Cubic models of generally positive change across time:
(a) $y = 1 + 5.5t - 1.75t^2 + 0.25t^3$ (b) $y = 5 - 4.5t + 3.15t^2 - 0.40t^3$
(c) $y = 0 - 0.5t + 2.30t^2 - 0.325t^3$ (d) $y = 1 + 10.5t - 3.25t^2 + 0.35t^3$

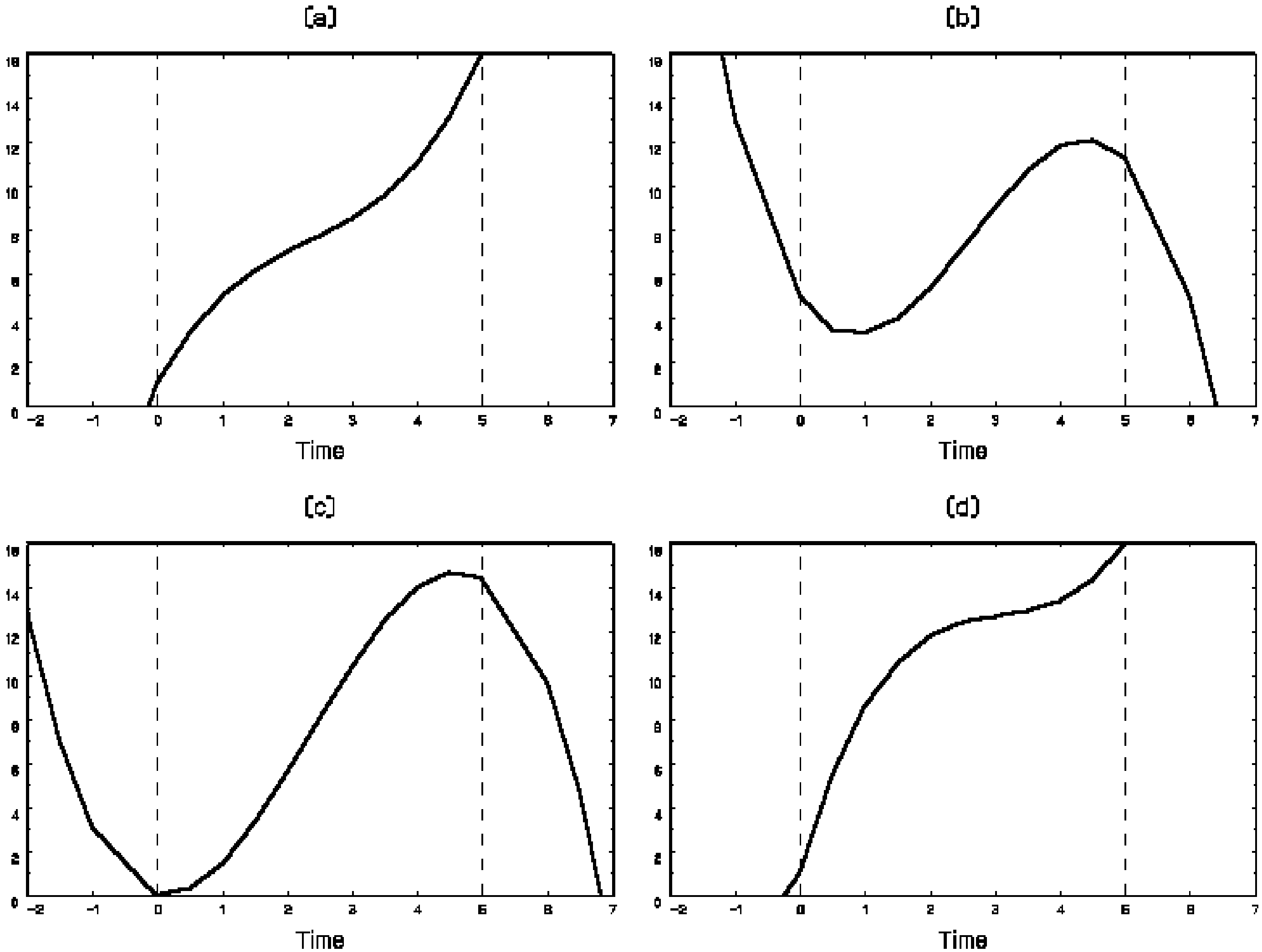

Figure 5.6 Extrapolation of cubic models across time:
(a) $y = 1 + 5.5t - 1.75t^2 + 0.25t^3$ (b) $y = 5 - 4.5t + 3.15t^2 - 0.40t^3$
(c) $y = 0 - 0.5t + 2.30t^2 - 0.325t^3$ (d) $y = 1 + 10.5t - 3.25t^2 + 0.35t^3$

Orthogonal Cubic Trend model - Reisby data

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		

$-2\log L = 2196.44$; $\chi^2_5 = 11.2, p < .05$ (even without dividing by 2)

- Relative to the previous model with quadratic trend, estimates for intercept, linear trend, and quadratic trends change very little
- Fixed effects reaffirm average linear response across time
- Variance estimates clearly diminish as order of polynomial increases (relative percentages of 68.7, 22.5, 5.7, 3.1)
- Positive association of constant and linear (higher average corresponds to higher slope)

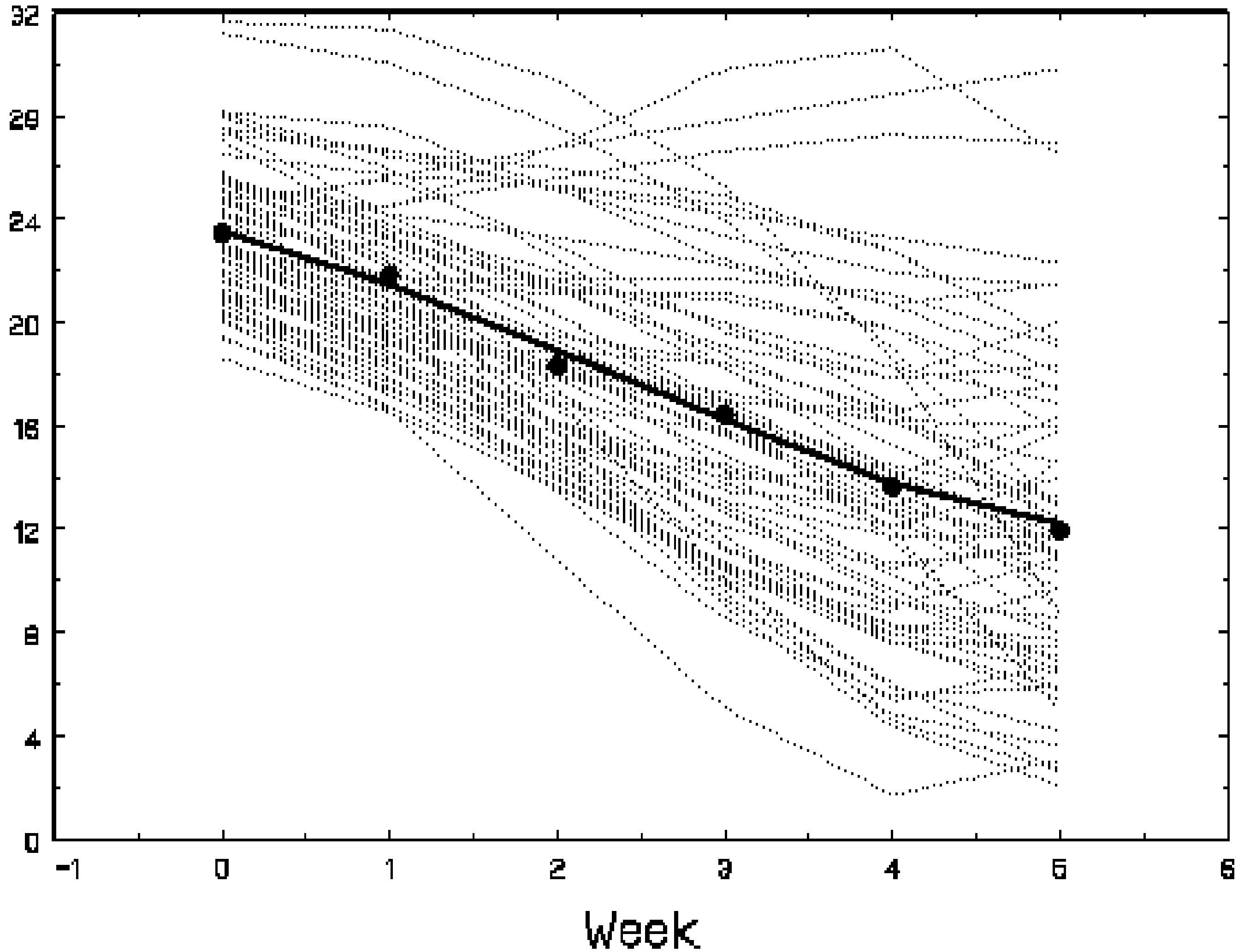

Figure 5.7 Reisby data: observed means (solid circles), estimated means (solid line), and estimated individual trends (dotted)

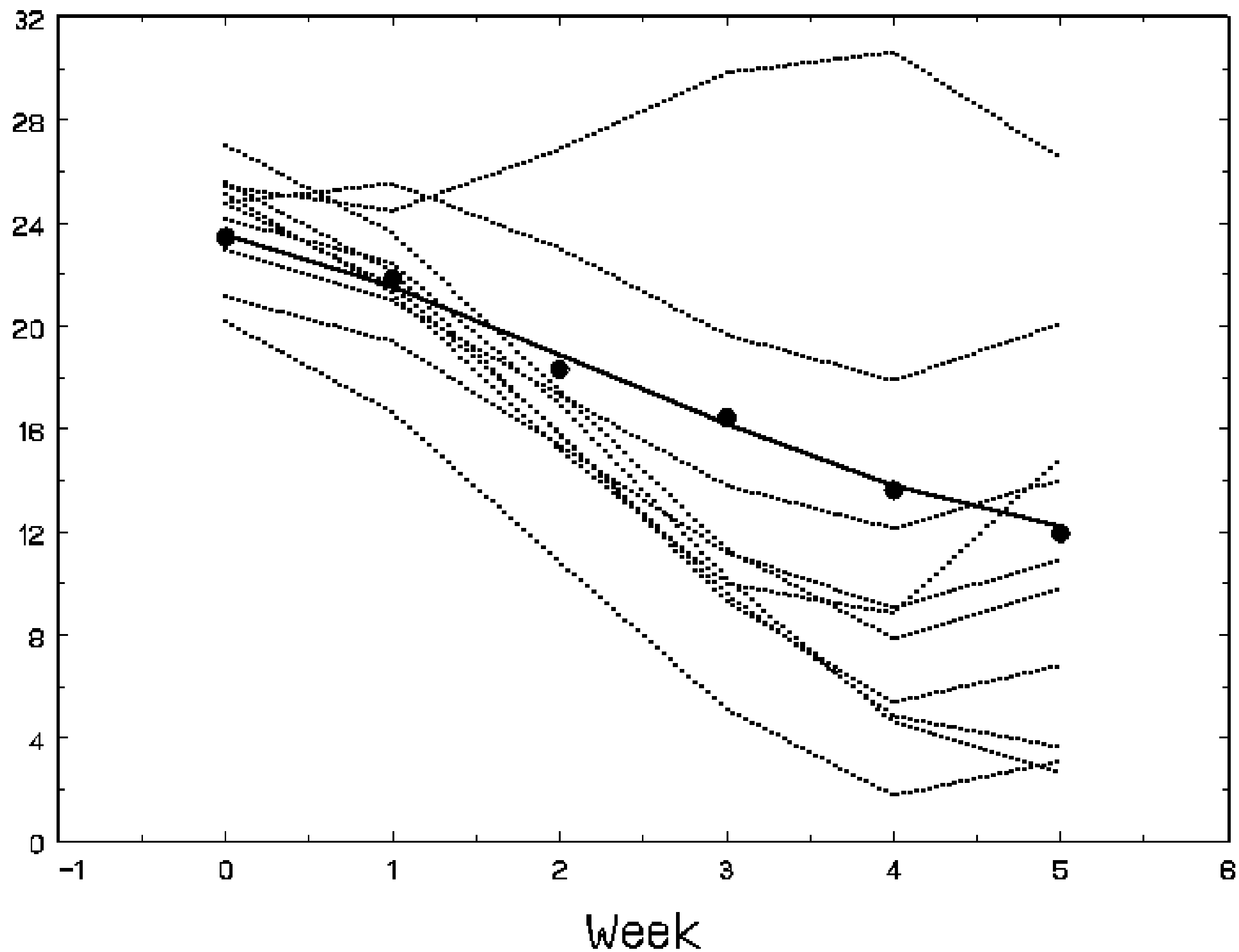

Figure 5.8 Reisby data: observed means (solid circles), estimated means (solid line), and estimated individual trends (dotted) for 10 subjects with largest cubic trend components

Observed and Estimated Standard Deviations

	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
Model-Based Estimates						
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

- fit of the observed standard deviations is incrementally better as the order of the polynomial is increased
- the cubic trend model does an excellent job of fitting the variation in the observed HDRS scores across time

Example 5a: Analysis of Riesby dataset. This handout contrasts quadratic trend models using the raw metric of week versus orthogonal polynomials. (SAS code and output) http://tigger.uic.edu/∼hedeker/riesquad.txt

Example 5b: Converting data from univariate to multivariate format, and getting the observed correlations and variance-covariance matrix of the repeated measures (SAS code and output)
http://www.uic.edu/classes/bstt/bstt513/riescov.txt

Example 5c: IML code that illustrates the calculation of estimated means and the variance-covariance matrix (SAS code and output)
http://tigger.uic.edu/∼hedeker/riesfit.txt