

where $\sigma_{\text{True Slope}}^2$ is the population variance of the true rate of change and $CSST$ is the corrected sum-of-squares-time, now common across individuals (Willett, 1988). Because $\sigma_{\text{True Slope}}^2$ appears in both the numerator and denominator, it plays a central role in determining reliability. If everyone is growing at the same true rate, all true growth trajectories will be parallel and there will be no variability in the true rate of change across people. When this happens, both $\sigma_{\text{True Slope}}^2$ and the reliability of change will be 0, no matter how precisely the individual change is measured. Ironically, this means that the OLS slope can be a very precise yet completely unreliable measure of change. If there are large differences in the true rate of change across people, the true growth trajectories will crisscross considerably. When this happens, $\sigma_{\text{True Slope}}^2$ will be large, dominating both numerator and denominator, and the reliability of the OLS slope will tend to 1, regardless of its precision. This means that the OLS slope can be an imprecise yet reliable measure of change. The conclusion: you can be fooled about the quality of your change measurement if you use reliability as your sole criterion.

We can also use equation 2.2 to reinforce our earlier conclusions about longitudinal research design. First, for a given level of interindividual difference in true change in the population, the reliability of the OLS slope depends solely on the residual variance. Once again, the better the quality of your outcome measurement, the better the reliability with which change can be measured because at least part of the residual variance is simply measurement error. Second, reliability can be improved through design, by manipulating the number and spacing of the measurement occasions. Anything that you can do to increase corrected sum-of-squares time, $CSST$, will help. As you add waves of data or move the existing waves further away from the center of the data collection period, the reliability with which change can be measured will improve.

3

Introducing the Multilevel Model for Change

When you're finished changing, you're finished

—Benjamin Franklin

In this chapter, we introduce the multilevel model for change, demonstrating how it allows us to address within-person and between-person questions about change simultaneously. Although there are several ways of writing the statistical model, here we adopt a simple and common approach that has much substantive appeal. We specify the multilevel model for change by simultaneously postulating a pair of subsidiary models—a level-1 submodel that describes how each person changes over time, and a level-2 model that describes how these changes differ across people (Bryk & Raudenbush, 1987; Rogosa & Willett, 1985).

We begin, in section 3.1, by briefly reviewing the rationale and purpose of statistical models in general and the multilevel model for change in particular. We then introduce the level-1 model for individual change (section 3.2) and the level-2 model for interindividual heterogeneity in change (section 3.3). In section 3.4, we provide an initial foray into the world of estimation, introducing the method of maximum likelihood. (We discuss other methods of estimation in subsequent chapters.) We close, in sections 3.5 and 3.6, by illustrating how the resultant parameter estimates can be interpreted and how key hypotheses can be tested.

We do not intend this chapter to present a complete and general account of the multilevel model for change. Our goal is to provide a single “worked” example—from beginning to end—that illustrates all the steps you must go through when specifying the model, fitting it to data, and interpreting its results. We proceed in this way because we believe it is easier to learn about the model by first walking through a simple, but complete, analysis in a constrained, yet realistic, context. This minimizes notational and analytic complexity and lets us focus on interpretation and

understanding. As a result, this chapter is limited to: (1) a linear change model for individual growth; (2) a time-structured data set in which everyone shares an identical data collection schedule; (3) an evaluation of the impact of a single dichotomous time-invariant predictor; and (4) the use of one piece of dedicated statistical software, HLM. In subsequent chapters, we extend this basic model in many ways, generalizing it to situations in which growth is curvilinear or discontinuous; the timing, spacing, and number of waves of data differ across individuals; interest centers on the effects of many predictors, both discrete and continuous, time-invariant and time-varying; distributional assumptions differ; and other methods of estimation and statistical software are used.

3.1 What Is the Purpose of the Multilevel Model for Change?

Even though you have surely fit many types of statistical models in your data analytic career, experience tells us that when researchers get caught up in a novel and complex analysis, they often need to be reminded just what a statistical model is and what it is not. So before presenting the multilevel model for change itself, we briefly review the purpose of statistical models.

Statistical models are mathematical representations of population behavior; they describe salient features of the hypothesized process of interest among individuals in the target population. When you use a particular statistical model to analyze a particular set of data, you implicitly declare that *this* population model gave rise to *these* sample data. Statistical models are not statements about sample behavior; they are statements about the *population process* that generated the data.

To provide explicit statements about population processes, statistical models are expressed using parameters—intercepts, slopes, variances, and so on—that represent specific population quantities of interest. Were you to use the following simple linear regression model to represent the relationship between infant birth weight (in pounds) and neurological functioning on a single occasion in a cross-sectional data set (with the usual notation) $NEURO_i = \beta_0 + \beta_1(BWGT_i - 3) + \varepsilon_i$, you would be declaring implicitly that, in the population from which your sample was drawn: (1) β_0 is an unknown intercept parameter that represents the expected level of neurological functioning for a three-pound newborn; and (2) β_1 is an unknown slope parameter that represents the expected difference in functioning between newborns whose birth weights differ by one pound. Even an analysis as simple as a one-sample *t*-test invokes a statis-

tical model expressed in terms of an unknown population parameter: the population mean, μ . In conducting this test, you use sample data to evaluate the evidence concerning μ 's value: Is μ equal to zero (or some other prespecified value)? Analyses may differ in form and function, but a statistical model underpins every inference.

In whatever context, having postulated a statistical model, you then fit the model to sample data and estimate the population parameters' unknown values. Most methods of estimation provide a measure of "goodness-of-fit"—such as an R^2 statistic or a residual variance—that quantifies the correspondence between the fitted model and sample data. If the model fits well, you can use the estimated parameter values to draw conclusions about the direction and magnitude of hypothesized effects in the population. Were you to fit the simple linear regression model just specified above, and find that $\widehat{NEURO}_i = 80 + 5(BWGT_i - 3)$, you would be able to predict that an average three-pound newborn has a functional level of 80 and that functional levels are five points higher for each extra pound at birth. Hypothesis tests and confidence intervals could then be used to make inferences from the sample back to the population.

The simple regression model above is designed for cross-sectional data. What kind of statistical model is needed to represent change processes in longitudinal data? Clearly, we seek a model that embodies two types of research questions: level-1 questions about *within-person change* and level-2 questions about *between-person differences in change*. If the hypothetical study of neurological functioning just described were longitudinal, we might ask: (1) How does each child's neurological functioning change over time? and (2) Do children's trajectories of change vary by birth weight? The distinction between the within-person and the between-person questions is more than cosmetic—it provides the core rationale for specifying a statistical model for change. It suggests that a model for change must include components at two levels: (1) a level-1 submodel that describes how individuals change over time; and (2) a level-2 submodel that describes how these changes vary across individuals. Taken together, these two components form what is known as a multilevel statistical model (Bryk & Raudenbush, 1987; Rogosa & Willett, 1985).

In this chapter, we develop and explain the multilevel model for change using an example of three waves of data collected by Burchinal and colleagues (1997). As part of a larger study of the effects of early intervention on child development, these researchers tracked the cognitive performance of 103 African-American infants born into low-income families. When the children were 6 months old, approximately half ($n = 58$) were randomly assigned to participate in an intensive early intervention program designed to enhance their cognitive functioning; the other

Table 3.1: Excerpts from the person-period data set for the early intervention study

<i>ID</i>	<i>AGE</i>	<i>COG</i>	<i>PROGRAM</i>
68	1.0	103	1
68	1.5	119	1
68	2.0	96	1
70	1.0	106	1
70	1.5	107	1
70	2.0	96	1
71	1.0	112	1
71	1.5	86	1
71	2.0	73	1
72	1.0	100	1
72	1.5	93	1
72	2.0	87	1
...
902	1.0	119	0
902	1.5	93	0
902	2.0	99	0
904	1.0	112	0
904	1.5	98	0
904	2.0	79	0
906	1.0	89	0
906	1.5	66	0
906	2.0	81	0
908	1.0	117	0
908	1.5	90	0
908	2.0	76	0
...

half ($n = 45$) received no intervention and constituted a control group. Each child was assessed 12 times between ages 6 and 96 months. Here, we examine the effects of program participation on changes in cognitive performance as measured by a nationally normed test administered three times, at ages 12, 18, and 24 months.

Table 3.1 presents illustrative entries from the person-period data set for this example. Each child has three records, one per wave of data collection. Each record contains four variables: (1) *ID*; (2) *AGE*, the child's age (in years) at each assessment (1.0, 1.5, or 2.0); (3) *COG*, the child's cognitive performance score at that age; and (4) *PROGRAM*, a dichotomy that describes whether the child participated in the early intervention program. Because children remained in their group for the duration of data collection, this predictor is time-invariant. Notice that all eight empirical growth records in table 3.1 suggest a decline in cognitive per-

formance over time. As a result, although we might wish that we would be determining whether program participants experience a faster rate of *growth*, it appears that we will actually be determining whether they experience a slower rate of *decline*.

3.2 The Level-1 Submodel for Individual Change

The *level-1* component of the multilevel model, also known as the *individual growth model*, represents the change we expect each member of the population to experience during the time period under study. In the current example, the level-1 submodel represents the individual change in cognitive performance that we hypothesize will occur during each child's second year of life.

Whatever level-1 submodel we specify, we must believe that the observed data could reasonably have come from a population in which the model is functioning. To align expectations with reality, we usually precede level-1 submodel specification with visual inspection of the empirical growth plots (although purists might question the wisdom of "peeking"). Figure 3.1 presents empirical growth plots of *COG vs AGE* for the 8 children whose data appear in table 3.1. We also examined plots for the 95 other children in the sample but we do not present them here, to conserve space. The plots reinforce our perception of declining cognitive performance over time. For some, the decline appears smooth and systematic (subjects 71, 72, 904, 908); for others, it appears scattered and irregular (subjects 68, 70, 902, 906).

When examining empirical growth plots like these, with an eye toward ultimate model specification, we ask global questions such as: What type of population individual growth model might have generated these sample data? Should it be linear or curvilinear with age? Smooth or jagged? Continuous or disjoint? As discussed in chapter 2, try and look beyond inevitable sample zigs and zags because plots of observed data confound information on true change with the effects of random error. In these plots, for example, the slight nonlinearity with age for subjects 68, 70, 902, 906, and 908 might be due to the imprecision of the cognitive assessment. Often, and especially when you have few waves of data, it is difficult to argue for anything except a linear-change individual-growth model. So when we determine which trajectory to select for modeling change, we often err on the side of parsimony and postulate a simple linear model.¹

Adopting an individual growth model in which change is a linear function of *AGE*, we write the level-1 submodel as:

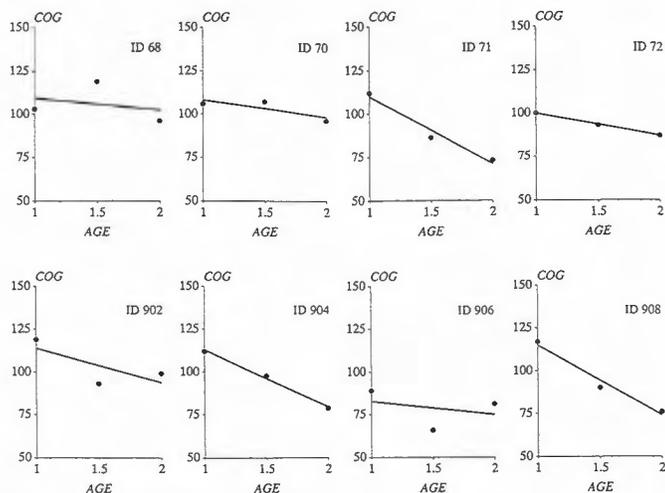


Figure 3.1. Identifying a suitable functional form for the level-1 submodel. Empirical growth plots with superimposed OLS trajectories for 8 participants in the early intervention study.

$$Y_{ij} = [\pi_{0i} + \pi_{1i}(AGE_{ij} - 1)] + [\varepsilon_{ij}]. \quad (3.1)$$

In postulating this submodel, we assert that, in the population from which this sample was drawn, Y_{ij} , the value of *COG* for child i at time j , is a linear function of his or her age on that occasion (AGE_{ij}). This model assumes that a straight line adequately represents each person's true change over time and that any deviations from linearity observed in sample data result from random measurement error (ε_{ij}).

Equation 3.1 uses two subscripts, i and j , to identify individuals and occasions, respectively. For these data, i runs from 1 through 103 (for the 103 children) and j runs from 1 through 3 (for the three waves of data). Although everyone in this data set was assessed on the same three occasions (ages 1.0, 1.5, and 2.0), the level-1 submodel in equation 3.1 is not limited in application to *time-structured* designs. The identical submodel could be used for data sets in which the timing and spacing of waves differs across people.² For now, we work with this time-structured

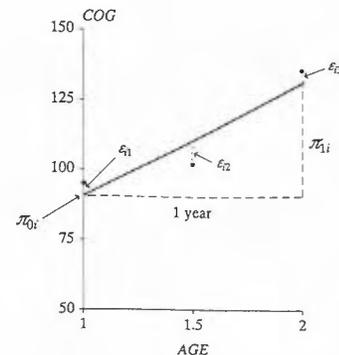


Figure 3.2. Understanding the structural and stochastic features of the level-1 individual growth model. Mapping the model in equation 3.1 onto imaginary data for child i , an arbitrarily selected member of the population.

example; in chapter 5, we extend our presentation to data sets in which data collection schedules vary across people.

In writing equation 3.1, we use brackets to distinguish two parts of the submodel: the *structural* part (in the first set of brackets) and the *stochastic* part (in the second). This distinction parallels the classical psychometric distinction between “true scores” and “measurement error,” but as we discuss below, its implications are much broader.

3.2.1 The Structural Part of the Level-1 Submodel

The structural part of the level-1 submodel embodies our hypotheses about the shape of each person's *true trajectory of change* over time. Equation 3.1 stipulates that this trajectory is linear with age and has *individual growth parameters* π_{0i} and π_{1i} that characterize its shape for the i th child in the population. Harkening back to section 2.2.2, these individual growth parameters are the population parameters that lie beneath the individual intercepts and slopes obtained when we fit OLS-estimated individual change trajectories in our exploratory analyses.

To clarify what the individual growth model says about the population, examine figure 3.2, which maps the model onto imaginary data for an arbitrarily selected member of the population, child i . First notice the intercept. Because we specify the level-1 submodel using the predictor ($AGE-I$), the intercept, π_{0i} , represents child i 's true cognitive performance at age 1. We concretize this interpretation in figure 3.2 by showing that the child's hypothesized trajectory intersects the Y axis at π_{0i} . Because we hypothesize that each child in the population has his or her own

intercept, this growth parameter includes the subscript i . Child 1's intercept is π_{01} , child 2's intercept is π_{02} , and so on.

Notice that equation 3.1 uses a special representation for the predictor, *AGE*. We used a similar approach in chapter 2, when we subtracted 11 from each adolescent's age before fitting exploratory OLS change trajectories to the tolerance data. This practice, known as *centering*, facilitates parameter interpretation. By using (*AGE-1*) as a level-1 predictor, instead of *AGE*, the intercept in equation 3.1 represents child i 's true value of Y at age 1. Had we simply used *AGE* as a level-1 predictor, with no centering, π_{0i} would represent child i 's true value of Y at age 0, an age that precedes the onset of data collection. This representation is less attractive because: (1) we would be predicting beyond the data's temporal limits; and (2) we don't know whether the trajectory extends back to birth linearly with age.

As you become adept at positing level-1 submodels, you will find that it is wise to consider empirical and interpretive issues like these when choosing the scale of your temporal predictor. In section 5.4, we explore other temporal representations, including those in which we center time on its *middle* and *final* values. The approach we adopt here—centering time on the first wave of data collection—is usually a good way to start. Aligning π_{0i} with the first wave of data collection allows us to interpret its value using simple nomenclature: it is child i 's true *initial status*. If π_{0i} is large, child i has a high true initial status; if π_{0i} is small, child i has low true initial status. We summarize this interpretation in the first row of the top panel of table 3.2, which defines all parameters in equation 3.1.

The second parameter in equation 3.1, π_{1i} , represents the *slope* of the postulated individual change trajectory. The slope is the most important parameter in a level-1 linear change submodel because it represents the rate at which individual i changes over time. Because *AGE* is clocked in years, π_{1i} represents child i 's true annual rate of change. We represent this parameter in figure 3.2 using the right triangle whose hypotenuse is the child's hypothesized trajectory. During the single year under study in our example—as child i goes from age 1 to 2—the trajectory rises by π_{1i} . Because we hypothesize that each individual in the population has his (or her) own rate of change, this growth parameter is subscripted by i . Child 1's rate of change is π_{11} , child 2's rate of change is π_{12} , and so on. If π_{1i} is positive, child i 's true outcome increases over time; if π_{1i} is negative, child i 's true outcome decreases over time (this latter case prevails in our example).

In specifying a level-1 submodel that attempts to describe everyone (all the i 's) in the population, we implicitly assume that all the true individual change trajectories have a common algebraic form. But we do not assume that everyone has the same exact trajectory. Because each person

Table 3.2: Definition and interpretation of parameters in the multilevel model for change

	Symbol	Definition	Illustrative interpretation
Level-1 Model (See Equation 3.1)			
<i>Individual growth parameters</i>	π_{0i}	<i>Intercept</i> of the true change trajectory for individual i in the population.	Individual i 's true value of <i>COG</i> at age 1 (i.e., his <i>true initial status</i>).
	π_{1i}	<i>Slope</i> of the true change trajectory for individual i in the population.	Individual i 's yearly rate of change in true <i>COG</i> (i.e., his <i>true annual rate of change</i>).
<i>Variance component</i>	σ_i^2	<i>Level-1 residual variance</i> across all occasions of measurement, for individual i in the population.	Summarizes the net (vertical) scatter of the observed data around individual i 's hypothesized change trajectory.
Level-2 Model (See Equation 3.3)			
<i>Fixed effects</i>	γ_{00}	Population average of the level-1 intercepts, π_{0i} , for individuals with a level-2 predictor value of 0.	Population average true initial status for nonparticipants.
	γ_{01}	Population average difference in level-1 intercept, π_{0i} , for a 1-unit difference in the level-2 predictor.	Difference in population average true initial status between participants and nonparticipants.
	γ_{10}	Population average of the level-1 slopes, π_{1i} , for individuals with a level-2 predictor value of 0.	Population average annual rate of true change for nonparticipants.
	γ_{11}	Population average difference in level-1 slope, π_{1i} , for a 1-unit difference in the level-2 predictor.	Difference in population average annual rate of true change between participants and nonparticipants.
<i>Variance components</i>	σ_0^2	Level-2 residual variance in true intercept, π_{0i} , across all individuals in the population.	Population residual variance of true initial status, controlling for program participation.
	σ_1^2	<i>Level-2 residual variance in true slope</i> , π_{1i} , across all individuals in the population.	Population residual variance of true rate of change, controlling for program participation.
	σ_{01}	Level-2 residual covariance between true intercept, π_{0i} , and true slope, π_{1i} , across all individuals in the population.	Population residual covariance between true initial status and true annual rate of change, controlling for program participation.

has his or her own individual growth parameters (intercepts and slopes), different people can have their own distinct change trajectories.

Positing a level-1 submodel allows us to distinguish the trajectories of different people using just their individual growth parameters. This leap is the cornerstone of individual growth modeling because it means that we can study interindividual differences in change by studying interindividual variation in the growth parameters. Imagine a population in which each member dips into a well of possible individual growth parameter values and selects a pair—a personal intercept and a slope. These values then determine his or her true change trajectory. Statistically, we say that each person has drawn his or her individual growth parameter values from an underlying bivariate distribution of intercepts and slopes. Because each individual draws his or her coefficients from an unknown *random* distribution of parameters, statisticians often call the multilevel model for change a *random coefficients model*.

3.2.2 The Stochastic Part of the Level-1 Submodel

The *stochastic* part of the level-1 submodel appears in the second set of brackets on the right-hand side of equation 3.1. Composed of just one term, the stochastic part represents the effect of random error, ϵ_{ij} , associated with the measurement of individual i on occasion j . The level-1 errors appear in figure 3.2 as ϵ_{i1} , ϵ_{i2} and ϵ_{i3} . Each person's *true* change trajectory is determined by the structural component of the submodel. But each person's *observed* change trajectory also reflects the measurement errors. Our level-1 submodel accounts for these perturbations—the differences between the true and observed trajectories—by including random errors: ϵ_{i1} for individual i 's first measurement occasion, ϵ_{i2} for individual i 's second measurement occasion, and so on.

Psychometricians consider random errors a natural consequence of measurement fallibility and the vicissitudes of data collection. We think it wise to be less specific, labeling the ϵ_{ij} as *level-1 residuals*. For these data, each residual represents that part of child i 's value of *COG* at time j not predicted by his or her age. We adopt this vaguer interpretation because we know that we can reduce the magnitude of the level-1 residuals by introducing selected time-varying predictors other than *AGE* into the level-1 submodel (as we show in section 5.3). This suggests that the stochastic part of the level-1 submodel is not just measurement error.

Regardless of how you conceptualize the level-1 errors, one thing is incontrovertible: they are *unobserved*. In ultimately fitting the level-1 submodel to data, we must invoke assumptions about the distribution of the level-1 residuals, from occasion to occasion and from person to person.

Traditional OLS regression invokes “classical” assumptions: that residuals are independently and identically distributed, with homoscedastic variance across occasions and individuals. This implies that, regardless of individual and occasion, each error is drawn independently from an underlying distribution with zero mean and an unknown residual variance. Often, we also stipulate the form of the underlying distribution, usually claiming normality. When we do, we can embody our assumptions about the level-1 residuals, ϵ_{ij} , by writing:

$$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \quad (3.2)$$

where the symbol \sim means “is distributed as,” N stands for a normal distribution, and the first element in parentheses identifies the distribution's mean (here, 0) and the second element identifies its variance (here, σ_ϵ^2). As documented in table 3.2, the residual variance parameter σ_ϵ^2 captures the scatter of the level-1 residuals around each person's true change trajectory.

Of course, classical assumptions like these may be less credible in longitudinal data. When individuals change, their level-1 error structure may be more complex. Each person's level-1 residuals may be autocorrelated and heteroscedastic over time, not independent as equation 3.2 stipulates. Because the same person is measured on several occasions, any unexplained person-specific time-invariant effect in the residuals will create a correlation across occasions. So, too, the outcome may have a different precision (and reliability) for individuals at different times, perhaps being more suitable at some occasions than at others. When this happens, the error variance may differ over time and the level-1 residuals will be heteroscedastic over occasions within person. How does the multilevel model for change account for these possibilities? Although this is an important question, we cannot address it fully without further technical work. We therefore delay addressing the issues of residual autocorrelation and heteroscedasticity until chapter 4, where we show, in section 4.2, how the full multilevel model for change accommodates automatically for certain kinds of complex error structure. Later, in chapter 8, we go further and demonstrate how using covariance structure analysis to conduct analyses of change lets you hypothesize, implement, and evaluate other alternative error structures.

3.2.3 Relating the Level-1 Submodel to the OLS Exploratory Methods of Chapter 2

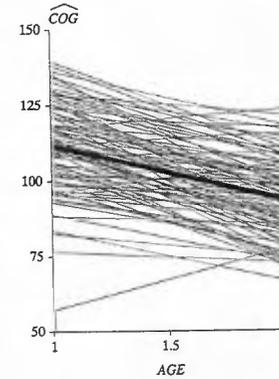
The exploratory OLS-fitted trajectories of section 2.2.2 may now make more sense. Although they are not fully efficient because they do not

properly exploit all the information present in longitudinal data, they do provide invaluable insights into the functioning of the hypothesized individual growth model. The top panel of figure 3.3 presents the results of using OLS methods to fit the level-1 submodel in equation 3.1 to the data for all 103 children (regressing *COG* on $(AGE-1)$, separately by *ID*). The bottom panel presents stem and leaf displays for three summary statistics from these models: the fitted intercepts, the fitted slopes, and the estimated residual variances.

For most children, cognitive performance declines over time. For some, the decline is rapid; for others, less so. Few children show any improvement. Each fitted intercept estimates that child's true initial status; each fitted slope estimates that child's true annual rate of change during the second year of life. The fitted intercepts are centered near 110; the fitted slopes are centered near -10 . This suggests that at age 1, the average child has a true cognitive level slightly above the national norm (of 100 for this test). Over time, however, most children decline (we estimate that only 7 improve).

The stem-and-leaf displays in the bottom left panel of figure 3.3 reveal great heterogeneity in fitted intercept and slope across children in the sample and suggest that not all children have identical trajectories of change. Of course, you must be cautious when interpreting the interindividual heterogeneity in change trajectories evident in figure 3.3. The between-person variation in the estimated change trajectories that you observe is necessarily inflated over the underlying interindividual variability in the unknown true change trajectories because the fitted trajectories, having been estimated from observed data, are *fallible* representations of true change. The actual variability in underlying true change will always be somewhat less than what you observe in exploratory analysis, with the magnitude of the difference depending on the quality of your outcome measurement and the efficacy of your hypothesized individual growth model.

The skewed distribution of residual variances in the bottom right panel of figure 3.3 suggests great variation in the *quality* of the OLS summaries across children (we expect the distribution of these statistics to be skewed, as they are "squared" quantities and are therefore bounded by zero below). When the residual variance is near 0, as it is for many children, the fitted trajectories are reasonable summaries of the observed data for those children. When the residual variance is larger, as it often is here, the fitted trajectories are poorer summaries: the observed values of *COG* are further away from the fitted lines, making the magnitude of the estimated level-1 residuals, and therefore the residual variance, large.



<u>Fitted initial status</u>	<u>Fitted rate of change</u>	<u>Residual variance</u>
14. 0	2. 0	46 8
13* 5588	1*	44
13. 00134	1. 0	42
12* 5556778999	0* 79	40 00
12. 02233344	0. 134	38
11* 55667777888889	-0* 4444332	36 8
11. 00011112222233334444	-0. 99998888777765	34
10* 55666688999	-1* 433322211000	32 3
10. 001222244	-1. 988887766655	30
9* 6666677799	-2* 4432211110000	28 4
9. 344	-2. 999977776655	26 7
8* 89	-3* 443322100000	24 1444
8. 34	-3. 987	22 8
7* 7	-4* 443111	20
7.		18 3
6*		16 00011
6.		14
5*		12 21
		10 44433
		8 1118886666
		6 77744
		4 333844
		2 0444488883338888888
		0 000011122233334444444466668111114447

Figure 3.3. Observed variation in fitted OLS trajectories. Fitted OLS trajectories for participants in the early intervention study as well as stem and leaf displays for fitted initial status, fitted rate of change, residual variance.

3.3 The Level-2 Submodel for Systematic Interindividual Differences in Change

The level-2 submodel codifies the relationship between interindividual differences in the change trajectories and time-invariant characteristics of the individual. The ability to formulate this relationship using a

level-2 submodel stems from the realization that adoption of a common level-1 submodel forces people to differ only in the values of their individual growth parameters. When we use a level-1 linear change model, people can differ only in their intercepts and slopes. This allows us to recast vague questions about the relationship between “change” and predictors as specific questions about the relationship between the individual growth parameters and predictors.

Like all statistical models, the level-2 submodel describes hypothesized population processes, not sample behavior. But insights gleaned from sample data can often provide valuable insight into model formulation. In this spirit, examine the top panel of figure 3.4, which separately plots fitted OLS trajectories according to the child’s program participation (program participants in the right panel, nonparticipants in the left). The average change trajectory for each group is shown in bold. Program participants tend to have higher scores at age 1 and decline less precipitously over time. This suggests that their intercepts are higher but their slopes are shallower. Also note the substantial interindividual heterogeneity *within* groups. Not all participants have higher intercepts than nonparticipants; not all nonparticipants have steeper slopes. Our level-2 model must simultaneously account for both the general patterns (here, the between-group differences in intercepts and slopes) *and* interindividual heterogeneity in patterns within groups.

What kind population model might have given rise to these patterns? The preceding discussion suggests four specific features for the level-2 submodel. First, its outcomes must be the individual growth parameters (here, π_{0i} and π_{1i} from equation 3.1). As in regular regression, where we model the population distribution of a random variable by making it an outcome, here, where we model the population distribution of the individual growth parameters, they, too, must be the outcomes. Second, the level-2 submodel must be written in separate parts, one for each level-1 growth parameter. When we use a linear change individual growth model at level-1 (as in equation 3.1), we need two level-2 submodels: one for the intercept, π_{0i} , another for the slope, π_{1i} . Third, each part must specify a relationship between an individual growth parameter and the predictor (here, *PROGRAM*). As you move across the panels in the top of figure 3.4, the value of the predictor, *PROGRAM*, shifts from 0 to 1. This suggests that each level-2 model should ascribe differences in either π_{0i} or π_{1i} to *PROGRAM* just as in a regular regression model. Fourth, each model must allow individuals who share common predictor values to vary in their individual change trajectories. This means that each level-2 submodel must allow for stochastic variation in the individual growth parameters.

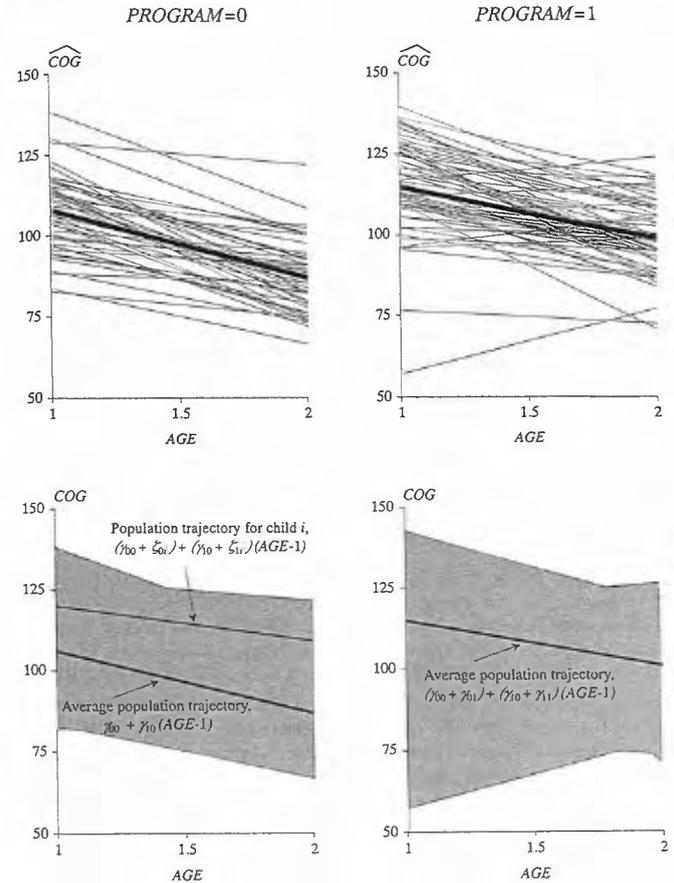


Figure 3.4. Understanding the structural and stochastic features of the level-2 submodel for inter-individual differences in change. Top panel presents fitted OLS trajectories separately by levels of the predictor *PROGRAM*. Bottom panel maps the model in equation 3.3 onto imaginary data for an arbitrary child i and the average population trajectory. The shaded portion in each of the lower panels is designed to suggest the existence of many distinct population trajectories for different children.

These considerations lead us to postulate the following level-2 submodel for these data:

$$\begin{aligned}\pi_{0i} &= \gamma_{00} + \gamma_{01}PROGRAM_i + \zeta_{0i} \\ \pi_{1i} &= \gamma_{10} + \gamma_{11}PROGRAM_i + \zeta_{1i}.\end{aligned}\quad (3.3)$$

Like all level-2 submodels, equation 3.3 has more than one component, each resembling a regular regression model. Taken together, the two components treat the intercept (π_{0i}) and the slope (π_{1i}) of an individual's growth trajectory as level-2 outcomes that may be associated with the predictor, *PROGRAM*. Each component also has its own residual—here, ζ_{0i} and ζ_{1i} —that permits the level-1 parameters (the π 's) of one person to differ stochastically from those of others.

Although not yet apparent, the two components of this level-2 submodel have *seven* population parameters: the four regression parameters (the γ 's) shown in equation 3.3 and three residual variance/covariance parameters we will soon define. All are estimated when we fit the multilevel model for change to data. We list, label, and define these parameters in the second section of table 3.2 and illustrate their action in the bottom panel of figure 3.4. We discuss their interpretation below.

3.3.1 Structural Components of the Level-2 Submodel

The structural parts of the level-2 submodel contain four level-2 parameters— γ_{00} , γ_{01} , γ_{10} , and γ_{11} —known collectively as the *fixed effects*. The fixed effects capture systematic interindividual differences in change trajectory according to values of the level-2 predictor(s). In equation 3.3, two of the fixed effects, γ_{00} and γ_{10} , are level-2 intercepts; two, γ_{01} and γ_{11} , are level-2 slopes. As in regular regression, the slopes are of greater interest because they represent the effect of predictors (here, the effect of *PROGRAM*) on the individual growth parameters. You can interpret the level-2 parameters much as you do regular regression coefficients, except that you must remember that they describe variation in “outcomes” that are themselves level-1 individual growth parameters.

The easiest way to unravel the meaning of the level-2 fixed effects is to identify a *prototypical individual* distinguished by particular predictor values, substitute those values into the level-2 submodel, and examine the consequences. To derive the postulated level-2 submodel for a prototypical nonparticipant, for example, we set *PROGRAM* to 0 in both parts of equation 3.3 to find: when *PROGRAM* = 0, $\pi_{0i} = \gamma_{00} + \zeta_{0i}$ and $\pi_{1i} = \gamma_{10} + \zeta_{1i}$. This model hypothesizes that, in the population of nonparticipants, the values of initial status and annual rate of change, π_{0i} and π_{1i} , are centered around the level-2 parameters γ_{00} and γ_{10} . γ_{00} represents the average true initial

status (cognitive score at age 1); γ_{10} represents the average true annual rate of change. By fitting the multilevel model for change to data and estimating these parameters, we address the question: What is the average true trajectory of change in the population for children who did not receive the early intervention program? The lower left panel of figure 3.5 depicts this average population trajectory. Its intercept is γ_{00} ; its slope is γ_{10} .

We repeat this process for program participants by setting *PROGRAM* to 1: in this case, $\pi_{0i} = (\gamma_{00} + \gamma_{01}) + \zeta_{0i}$ and $\pi_{1i} = (\gamma_{10} + \gamma_{11}) + \zeta_{1i}$. In the population of program participants, the values of initial status and annual rate of change, π_{0i} and π_{1i} , are centered around $(\gamma_{00} + \gamma_{01})$ and $(\gamma_{10} + \gamma_{11})$. Comparing these centers to those for nonparticipants illustrates that the level-2 parameters γ_{01} and γ_{11} capture the effects of *PROGRAM*. γ_{01} represents the hypothesized difference in average true initial status between groups; γ_{11} represents the hypothesized difference in average true annual rate of change. This allows us to think of the level-2 slopes, γ_{01} and γ_{11} , as “shifts” associated with program participation. The lower right panel of figure 3.4 depicts these shifts. If γ_{01} and γ_{11} are non-zero, the average population trajectories in the two groups differ; if they are both 0, they do not. These two level-2 slope parameters therefore address the question: What is the difference in the average trajectory of true change associated with program participation?

3.3.2 Stochastic Components of the Level-2 Submodel

Each part of the level-2 submodel contains a residual that allows the value of each person's growth parameters to be scattered around the relevant population averages. These residuals, ζ_{0i} and ζ_{1i} in equation 3.3, represent those portions of the level-2 outcomes—the individual growth parameters—that remain “unexplained” by the level-2 predictor(s). As is true for most residuals, we are interested less in their specific values than in their population variances and covariance, which we label σ_0^2 , σ_1^2 , and σ_{01} . You should know that labeling conventions for these population variances vary considerably across authors and statistical packages. For example, Raudenbush and Bryk (2002) label them τ_{00} , τ_{11} , and τ_{01} , while Goldstein (1995) labels them $\sigma_{\epsilon_0}^2$, $\sigma_{\epsilon_1}^2$, and $\sigma_{\epsilon_{01}}$.

If child *i* is a member of the population of nonparticipants, *PROGRAM* takes on the value 0 and the level-2 residuals in equation 3.3 represent deviations between his or her true initial status and annual rate of change from the population average intercept and slope for nonparticipants (γ_{00} and γ_{10}). We display a trajectory for this prototypical child in the lower left panel of figure 3.4. The trajectory begins at a true initial status of ($\gamma_{00} + \zeta_{0i}$) and has a (declining) true annual rate of change of ($\gamma_{10} + \zeta_{1i}$).

Trajectories for other children can be constructed similarly by combining parameters γ_{00} and γ_{10} with other child-specific residuals. The shaded area in this panel is designed to suggest the existence of many different true trajectories, one for each nonparticipant in the population (if they could be fully enumerated). Similarly, if child i is a member of the population of participants, *PROGRAM* takes on the value 1 and the level-2 residuals in equation 3.3 represent deviations between his true initial status and annual rate of change and the population average intercept and slope for participants ($\gamma_{00} + \gamma_{01}$) and ($\gamma_{10} + \gamma_{11}$). To illustrate the heterogeneity in change for this group, the lower right panel of figure 3.4 also includes a shaded area.

Because the level-2 residuals represent deviations between the individual growth parameters and their respective population averages, their variances, σ_0^2 and σ_1^2 , summarize the population variation in true individual intercept and slope around these averages. Because they describe those portions of the intercepts and slopes *left over* after accounting for the effect(s) of the model's predictor(s), they are actually *conditional* residual variances. Conditional on the presence of the model's predictors, σ_0^2 represents the population residual variance in true initial status and σ_1^2 represents the population residual variance in true annual rate of change. These variance parameters allow us to address the question: How much heterogeneity in true change remains after accounting for the effects of program participation?

When we posit a level-2 submodel, we also allow for a possible association between individual initial status and individual rates of change. Children who begin at a higher level may have higher (or lower) rates of change. To account for this possibility, we permit the level-2 residuals to be correlated. Since ζ_{0i} and ζ_{1i} represent the deviations of the individual growth parameters from their population averages, their population covariance summarizes the association between true individual intercepts and slopes. Again because of their conditional nature, the population covariance of the level-2 residuals, σ_{01} , summarizes the magnitude and direction of the association between true initial status and true annual rate of change, controlling for program participation. This parameter allows us to address the question: Controlling for program participation, are true initial status and true rate of change related?

To fit the multilevel model for change to data, we must make some assumptions about the level-2 residuals (just as we did for the level-1 residuals in equation 3.2). But because we have two level-2 residuals, we describe their underlying behavior using a *bivariate distribution*. The standard assumption is that the two level-2 residuals, ζ_{0i} and ζ_{1i} , are bivariate normal with mean 0, unknown variances, σ_0^2 and σ_1^2 , and unknown

covariance, σ_{01} . We can express these assumptions compactly using matrix notation by writing:

$$\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right). \quad (3.4)$$

Matrix notation greatly simplifies the way in which we codify the model's assumptions. In broad outline, we interpret equation 3.4 in the same way we interpret the assumptions about the level-1 residuals in equation 3.2. The first matrix on the right of the equals sign in parentheses specifies the bivariate distribution's mean vector; here, we assume it to be 0 for each residual (as usual). The second matrix specifies the bivariate distribution's variance-covariance matrix, also known as the *level-2 error covariance matrix* because it captures the covariation among the level-2 residuals (or errors). Two variances, σ_0^2 and σ_1^2 , appear along the diagonal, the covariance, σ_{01} , appears on the off-diagonal. Because the covariance between ζ_{0i} and ζ_{1i} is the same as the covariance between ζ_{1i} and ζ_{0i} , the off-diagonal elements are identical—that is, $\sigma_{01} = \sigma_{10}$. The complete set of residual variances and covariances—both the level-2 error variance-covariance matrix and the level-1 residual variance, σ_ϵ^2 —is known collectively as the model's *variance components*.

3.4 Fitting the Multilevel Model for Change to Data

Until the widespread availability of software for fitting multilevel models, researchers used ad hoc strategies like those presented in chapter 2 to analyze longitudinal data: they fitted individual growth trajectories in separate within-person OLS-regression analyses and then they regressed the individual growth parameter estimates obtained on selected level-2 predictors (Willett, 1989). But as previously discussed, this approach has at least two flaws: (1) it ignores information about the individual growth parameter estimates' precision, even though we know that it varies (as seen in the varying residual variances in the bottom panel of figure 3.3); and (2) it replaces *true* individual growth parameters—the real outcomes in a level-2 submodel—with their fallible estimates. The level-2 submodels do not describe the relationship between the parameter *estimates* and predictors, but between the parameters' *true values* and predictors.

Beginning in the 1980s, several teams of statisticians began developing specialized software for fitting the multilevel model for change to data. By the early 1990s, four major packages were widely used: HLM (Bryk, Raudenbush, & Congdon, 1988), MLn (Rasbash & Woodhouse, 1995), GENMOD (Mason, Anderson, & Hayat, 1988), and VARCL (Longford, 1993).

Although the latter two are no longer supported, HLM (Raudenbush, Bryk, Cheong, & Congdon, 2001, available from <http://www.ssicentral.com>) and MLwiN (Goldstein, 1998, available from <http://multilevel.ioe.ac.uk>) continue to be modified, expanded, and upgraded regularly to handle an increasing variety of multilevel models. Several multipurpose software packages have also added multilevel routines, including SAS PROC MIXED and PROC NL MIXED (SAS Institute, 2001, <http://www.sas.com>), the STATA "xt" routines, such as xtreg (Stata, 2001, <http://www.stata.com>), and SPLUS' NLME library (Pinheiro & Bates, 2001, available from <http://cm.bell-labs.com/cm/ms/departments/sia/project/nlme/>). So, too, teams of statisticians continue to develop new specialty programs including BUGS (Gilks, Richardson, & Spiegelhalter, 1996, available from <http://www.mrcbsu.cam.ac.uk/bugs>) and MIXREG (Hedeker & Gibbons, 1996; available from <http://www.uic.edu/~hedeker>).

As this list suggests, you have a wide and growing array of model fitting options in the investigation of change. We ourselves have no vested interest in any particular software program and do not promote any one above the others. All have their strengths, and we use many of them in our research and in this book. At their core, each program does the same job: it fits the multilevel model for change to data and provides parameter estimates, measures of precision, diagnostics, and so on. There is also some evidence that all the different packages produce the same, or similar, answers to a given problem (Kreft & de Leeuw, 1990). So, in one sense, it does not matter which program you choose. But the packages do differ in many important ways including the "look and feel" of their interfaces, their ways of entering and preprocessing data, their model specification process, their estimation methods, their strategies for hypothesis testing, and the provision of diagnostics. These differences may lead you to decide that one piece of software is especially convenient for your work.

For now, we focus on one particular method of estimation—*maximum likelihood*—as implemented in one program, HLM (Raudenbush, Bryk, Cheong, & Congdon, 2001). In subsequent chapters, we describe other methods of estimation and we apply other statistical software, allowing us to provide advice and compare the competing approaches and packages.

3.4.1 The Advantages of Maximum Likelihood Estimation

The method of maximum likelihood (ML) is currently the most popular approach to statistical estimation. Its popularity results, in part, from its excellent performance in large random samples from well-defined target

populations. As sample size increases, ML estimates have three desirable properties: (1) they are *asymptotically unbiased (consistent)*—they converge on the unknown true values of population parameters; (2) they are *asymptotically normally distributed*—their sampling distributions are approximately normal with known variance; and (3) they are *asymptotically efficient*—their standard errors are smaller than those derived by other methods. Another advantage is that any function of ML estimates is also an ML estimate. This means that predicted growth trajectories (constructed from ML estimates of initial status and rates of change) are ML estimates of the true trajectories. All else being equal, statisticians prefer estimates that are consistent and efficient, that make use of well-established normal theory, and that can generate decent estimates of more complex quantities. Hence the appeal of ML methods.

Notice that the attractive properties of ML estimates are *asymptotic*. This means that in practice—in any actual analysis of a real sample—the properties hold only *approximately*. In large samples, they are likely to hold; in small samples, they may not.⁵ To enjoy these advantages, you need a relatively large sample, and the question, how large is large, has no simple answer. Although 10 is certainly small and 100,000 is certainly large, no one can say definitively how large is large enough. In cross-sectional work, Long (1997), for example, recommends a minimum of 100 individuals and he labels sample sizes of 500 "adequate." For a general multilevel model, Snijders and Bosker (1999) consider samples of 30 or more large. Although "rules of thumb" like these provide broad guidelines, we tend to distrust them. The answer to the question "How large?" differs by context, by the particularities of different types of ML estimation, by features of the data, and by the requirements of the tests conducted. Instead we simply offer practical advice: if you use ML methods in "small" samples, treat *p*-values and confidence intervals circumspectly.

Derivation of computational formulas for ML estimation is beyond our scope or intent here. Below, we offer a heuristic explanation of what happens when you use ML methods to fit a multilevel model for change. Our goal is to lay the conceptual foundation for future chapters by explaining why ML estimates make sense and why they have such useful properties. Readers interested in mathematical details should consult Raudenbush and Bryk (2002), Goldstein (1995), or Longford (1993).

3.4.2 Using Maximum Likelihood Methods to Fit a Multilevel Model

Conceptually, maximum likelihood estimates are those guesses for the values of the unknown population parameters that maximize the

probability of observing a particular sample of data. In the early intervention study, they are those estimates of the fixed effects and variance components that make it most likely we would have observed the specific patterns of change found for these 103 children.

To derive an ML estimate for a population parameter, a statistician must first construct a *likelihood function*—an expression that describes the probability of observing the sample data as a function of the model's unknown parameters. Then, he, she, or more accurately, a computer, numerically examines the relative performance of potentially competing estimates until those that maximize the likelihood function are found. The likelihood function for the early intervention data is a function of the probability that we would observe the particular temporal pattern of COG values found in the person-period data set. We seek estimates of the fixed effects and variance components whose values maximize the probability of observing this specific pattern.

All likelihood functions are expressed as the product of probabilities (or probability densities). For cross-sectional data, each sample member usually contributes just one term, related to the probability that *that* person has his or her observed data. But because longitudinal data consist of several observations, one per measurement occasion, each person contributes several terms to the likelihood function, which contains as many terms as there are records in the person-period data set.

The particular term that each person contributes on each occasion depends on the specification and assumptions of the hypothesized model. The multilevel model contains structural parts (as shown in, for example, in equations 3.1 and 3.3) and stochastic parts (whose behavior is described in equations 3.2 and 3.4). The structural portion describes the true outcome value for person i on occasion j for his or her particular predictor values. It depends on the unknown values of the fixed effects. The stochastic portion—the level-1 and level-2 residuals—introduce an element of randomness into the proceedings, scattering the observations for person i on occasion j from the structurally specified value.

To derive a maximum likelihood estimate, we must also make assumptions about the *distribution* of the residuals. We have already stated assumptions in equation 3.2 for the level-1 residual, ε_{ij} , and in equation 3.4 for the two-level-2 residuals, ζ_{0i} and ζ_{1i} . Each is assumed to be normally distributed with mean 0; ε_{ij} has unknown variance, σ_{ε}^2 ; ζ_{0i} and ζ_{1i} have unknown variances, σ_0^2 and σ_1^2 , and covariance, σ_{01} . We also assume that the level-2 residuals are independent of the level-1 residual and that all residuals are independent of the model's predictors.

Given a model and its underlying assumptions, a statistician can write

a mathematical expression for the distribution, or *probability density*, of the outcome. This expression has a mean determined by the model's structural parts and a variance determined by its stochastic parts. As a probability density function, it also describes the likelihood that a person with particular values of the predictors—only *PROGRAM* in equation 3.3—could have particular outcome values using a set of unknown fixed effects and variance components whose values we would like to estimate. That is, it also contains the actual data values observed for that person on that occasion.

It is a short step from here to the full sample likelihood, which we reach by exploiting the well-known multiplicative property of independent probabilities. If you toss one coin, there is a probability of .5 that it will turn up heads. If you independently toss two coins, the probability that each will turn up heads is still .5. But taken together, the probability that you will obtain two heads is only .25 (.5 \times .5). If you independently toss three coins, the probability of three heads declines to 0.125 (.5 \times .5 \times .5). Statisticians use this principle to create a full sample likelihood from the separate person-period likelihoods just developed. First they write down the value of the probability density of the outcome for each person in the data set on every occasion, thereby describing the likelihood that he or she obtained his or her particular value of the outcome on that occasion. Then they multiply these terms together, yielding an expression for the likelihood of simultaneously observing *all* the data in the person-period data set. Because each person-period likelihood is a function of the data and the unknown parameters, so is their product the full sample likelihood.

To find ML estimates of the unknown population parameters, we identify those values of the unknown parameters that maximize this product of probabilities. Conceptually, imagine a computer trying out billions of alternative estimates, multiplying them together as specified in the sample likelihood function to yield a numeric value for the likelihood, and comparing those numeric values across all of the billions of tries until those estimates that yield the maximum value of the likelihood function are found. These would be the maximum likelihood estimates for this particular problem.

Of course, an enormous numerical search like this is daunting, even with fast computers. Calculus can facilitate the search, but it cannot eliminate the difficulty of working with the products of probability densities that make up the sample likelihood function. To facilitate the search, statisticians use a simple strategy: instead of finding those values of the unknown parameters that maximize the likelihood function, they find those that maximize its logarithm. Working with this new function, known

as the *log-likelihood function*, sacrifices nothing because the values that maximize it also maximize the raw likelihood function. The transformation to logarithms simplifies the intensive numerical calculations involved because (1) the logarithm of a product is a *sum* of the separate logarithms, and (2) the logarithm of a term raised to a power is the power multiplied by the logarithm of the term. And so, since the sample likelihood contains both multiplicative and exponentiated terms, the logarithmic transformation moves the numerical maximization into a more tractable sphere, computationally speaking.

Although simpler than maximizing the likelihood function itself, maximizing the log-likelihood function also involves iteration. All software programs that provide ML estimates for the multilevel model for change use an iterative procedure. To begin, the program generates reasonable "starting" values for all model parameters, usually by applying something like the OLS methods we just rejected in chapter 2! In successive iterations, the program gradually refines these estimates as it searches for the log-likelihood function's maximum. When this search converges—and the difference between successive estimates is trivially small—the resultant estimates are output. If the algorithm does not converge (and this happens more often than you might like), you must repeat the search allowing more iterations or you must improve your model specification. (We discuss these issues in section 5.2.2.)

Once the ML estimates are found, it is relatively easy for a computer to estimate their associated sampling variation in the form of *asymptotic standard errors* (*ase*). We use the adjective "asymptotic" because, as noted earlier, ML standard errors are accurate only in large samples. Like any standard error, the *ase* measures the precision with which an estimate has been obtained—the smaller the *ase*, the more precise the estimate.

We now use maximum likelihood methods to fit the multilevel model in equations 3.1 and 3.3 to the early intervention data. Table 3.3 presents results obtained using the HLM software.⁴ We first discuss the estimated fixed effects in the first four rows; in section 3.6, we discuss the estimated variance components shown in the next four rows.

3.5 Examining Estimated Fixed Effects

Empirical researchers usually conduct hypothesis tests before scrutinizing parameter estimates to determine whether an estimate warrants inspection. If an estimate is consistent with a null hypothesis of no population effect, it is unwise to interpret its direction or magnitude.

Table 3.3: Results of fitting a multilevel model for change to the early intervention data ($n = 103$)

	Parameter	Estimate	ase	z	
Fixed Effects					
Initial status, π_{0i}	Intercept	γ_{00}	107.84***	2.04	52.97
	PROGRAM	γ_{01}	6.85*	2.71	2.53
Rate of change, π_{1i}	Intercept	γ_{10}	-21.13***	1.89	-11.18
	PROGRAM	γ_{11}	5.27*	2.52	2.09
Variance Components					
Level 1:	Within-person, ϵ_{ij}	σ_{ϵ}^2	74.24***	10.34	7.17
Level 2:	In initial status, ζ_{0i}	σ_0^2	124.64***	27.38	4.55
	In rate of change, ζ_{1i}	σ_1^2	12.29	30.50	0.40
	Covariance between ζ_{0i} and ζ_{1i}	σ_{01}	-36.41	22.74	-1.60

- $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

This model predicts cognitive functioning between ages 1 and 2 years as a function of (AGE-1) (at level-1) and PROGRAM (at level-2).

Note: Full ML, HLM.

Although we agree that it is wise to test hypotheses before interpreting parameters, here we reverse this sequence for pedagogic reasons, discussing interpretation in section 3.5.1 and testing in section 3.5.2. Experience convinces us that when learning a new statistical method, it is easier to understand what you are doing if you interpret parameters first and conduct tests second. This sequence emphasizes conceptual understanding over up-or-down decisions about "statistical significance" and ensures that you understand the hypotheses you test.

3.5.1 Interpreting Estimated Fixed Effects

The fixed effects parameters of the level-2 submodel—the γ 's of equation 3.3—quantify the effects of predictors on the individual change trajectories. In our example, they quantify the relationship between the individual growth parameters and program participation. We interpret these estimates much as we do any regression coefficient, with one key difference: the level-2 "outcomes" that these fixed effects describe are the level-1 individual growth parameters themselves.

Until you are comfortable directly interpreting the output from software programs, we strongly recommend that you take the time to actually write down the structural portion of the fitted model before attempting to interpret the fixed effects. Although some software programs facilitate the linkage between model and estimates through

structured displays (e.g., MlwiN), others (e.g., SAS PROC MIXED) use somewhat esoteric conventions for labeling output. Substituting estimates $\hat{\gamma}$ in table 3.3 into the level-2 submodel in equation 3.3, we have:

$$\begin{aligned}\hat{\pi}_{0i} &= 107.84 + 6.85PROGRAM_i \\ \hat{\pi}_{1i} &= -21.13 + 5.27PROGRAM_i\end{aligned}\quad (3.5)$$

The first part of the fitted submodel describes the effects of *PROGRAM* on initial status; the second part describes its effects on the annual rates of change.

Begin with the first part of the fitted submodel, for initial status. In the population from which this sample was drawn, we estimate the true initial status (*COG* at age 1) for the average nonparticipant to be 107.84; for the average participant, we estimate that it is 6.85 points higher (114.69). The means of both groups are higher than national norms (100 for this test). The age 1 performance of participants is 6.85 points higher than that of nonparticipants. Before concluding that this differential in initial status casts doubt on the randomization mechanism, remember that the intervention started *before* the first wave of data collection, when the children were already 6 months old. This modest seven-point elevation in initial status may reflect early treatment gains attained between ages 6 months and 1 year.

Next, examine the second part of the fitted submodel, for the annual rate of change. In the population from which this sample was drawn, we estimate the true annual rate of change for the average nonparticipant to be -21.13 ; for the average participant, we estimate it to be 5.27 points higher (-15.86). The average nonparticipant dropped over 20 points during the second year of life; the average participant dropped over 15. The cognitive functioning of both groups of children declines over time. As we suspected when we initially examined these data, the intervention slows the rate of decline.

Another way of interpreting fixed effects is to plot fitted trajectories for prototypical individuals. Even in a simple analysis like this, which involves just one dichotomous predictor, we find it invaluable to inspect prototypical trajectories visually. For this particular multilevel model, only two prototypes are possible: a program participant (*PROGRAM* = 1) and a nonparticipant (*PROGRAM* = 0). Substituting these values into equation 3.5 yields the estimated initial status and annual growth rates for each:

$$\begin{aligned}\text{When } PROGRAM = 0: & \hat{\pi}_{0i} = 107.84 + 6.85(0) = 107.84 \\ & \hat{\pi}_{1i} = -21.13 + 5.27(0) = -21.13 \\ \text{When } PROGRAM = 1: & \hat{\pi}_{0i} = 107.84 + 6.85(1) = 114.69 \\ & \hat{\pi}_{1i} = -21.13 + 5.27(1) = -15.86.\end{aligned}$$

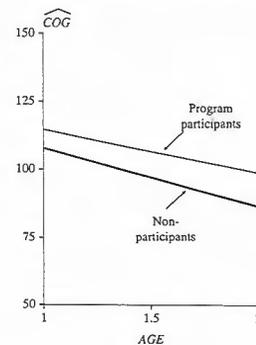


Figure 3.5. Displaying the results of a fitted multilevel model for change. Prototypical trajectories for an average program participant and nonparticipant in the early intervention data.

We use these estimates to plot the fitted individual change trajectories in figure 3.5. These plots reinforce the numeric conclusions just articulated. In comparison to nonparticipants, the average participant has a higher score at age 1 and a slower annual rate of decline.

3.5.2 Single Parameter Tests for the Fixed Effects

As in regular regression, you can conduct a hypothesis test on each fixed effect (each γ) using a single parameter test. Although you can equate the parameter value to any pre-specified value in your hypothesis test, most commonly you examine the null hypothesis that, controlling for all other predictors in the model, the population value of the parameter is 0, $H_0: \gamma = 0$, against the two-sided alternative that it is not, $H_1: \gamma \neq 0$. When you use ML methods, this test's properties are known only asymptotically (for exceptions, see note 3). You test this hypothesis for each fixed effect by computing the familiar z -statistic:

$$z = \frac{\hat{\gamma}}{ase(\hat{\gamma})}. \quad (3.7)$$

Most multilevel modeling programs provide z -statistics; if not, you can easily compute them by hand. However, care is needed because there is much looseness and inconsistency in output labels; terms like z -statistic, z -ratio, quasi- t -statistic, t -statistic, and t -ratio, which are not the same, are

used interchangeably. In HLM, the package we used here, this statistic is labeled a “*t*-ratio.” Most programs also output either an associated *p*-value or confidence interval to facilitate testing.⁵

Table 3.3 presents *z*-statistics (column 6) and approximate *p*-values (as superscripts in column 4) for testing hypotheses about the fixed effects. We reject all four null hypotheses, suggesting that each parameter plays a role in the story of the program’s effect on children’s cognitive development. In rejecting (at the .001 level) the null hypotheses for the two level-2 intercepts, γ_{00} and γ_{10} , we conclude that the average nonparticipant had a non-zero cognitive score at age 1 (hardly surprising!) which declined over time. In rejecting (at the .05 level) the null hypotheses for the two level-2 slopes, γ_{01} and γ_{11} , we conclude that differences between program participants and nonparticipants—in both initial status and annual rates of change—are statistically significant.

3.6 Examining Estimated Variance Components

Estimated variance and covariance components are trickier to interpret as their numeric values have little absolute meaning and there are no graphic aids to fall back on. Interpretation for a single fitted model is especially difficult as you lack benchmarks for evaluating the components’ magnitudes. This increases the utility of hypothesis testing, for at least the tests provide some benchmark (against the null value of 0) for comparison.

3.6.1 Interpreting Estimated Variance Components

Variance components assess the amount of outcome variability left—at either level-1 or level-2—after fitting the multilevel model. The level-1 residual variance, σ_e^2 , summarizes the population variability in an average person’s outcome values around his or her own true change trajectory. Its estimate for these data is 74.24, a number that is difficult to evaluate in absolute terms. In chapter 4, we provide strategies making relative comparisons to residual variances in other models.

The level-2 variance components summarize the between-person variability in change trajectories that remains after controlling for predictors (here, *PROGRAM*). Using the matrix notation of equation 3.4, we write:

$$\begin{bmatrix} 124.64 & -36.41 \\ -36.41 & 12.29 \end{bmatrix}$$

Because hypothesis tests, discussed below, reveal that only one of these elements, σ_0^2 , is significantly different from 0, it is the only parameter we

discuss here. But because we have no point of comparison, it is difficult to say whether its value, 124.64, is small or large. All we can say is that it quantifies the amount of residual variation in true initial status remaining after we control for program participation.

3.6.2 Single Parameter Tests for the Variance Components

Tests for variance components evaluate whether there is any remaining *residual* outcome variation that could potentially be explained by other predictors. The level of the particular variance component—either level-1 or level-2—dictates the type of predictor that might be added. In general, all the tests are similar in that they assess the evidence concerning the null hypothesis that the parameter’s population value is 0, $H_0: \sigma^2 = 0$, against the alternative that it is not, $H_1: \sigma^2 \neq 0$.

There are two very different methods for conducting these hypothesis tests. In this chapter, we offer the simpler approach—the *single parameter test*. Some programs provide this test as a *z*-statistic—the ratio of the estimated variance component to its asymptotic standard error. Others offer the identical test by squaring the *z*-statistic and labeling it a χ^2 statistic on one degree of freedom. The appeal of a single parameter hypothesis test is simple. Even if you fit just one statistical model, as we have here, you can garner some insight into the variance components’ relative values—at least in comparison to 0.

Unfortunately, statisticians disagree as to the nature, form, and effectiveness of these tests. Miller (1986), Raudenbush and Bryk (2002), and others have long questioned their utility because of their sensitivity to departures from normality. Longford (1999) describes their sensitivity to sample size and imbalance (unequal numbers of observations per person) and argues that they are so misleading that they should be abandoned completely. Because they can be useful for quick, albeit imprecise, assessment, we suggest you examine them only with extreme caution. In section 4.6, we present a superior method for testing hypotheses about variance components, an approach whose use we normally recommend.

Table 3.3 presents single-parameter hypothesis tests for the model’s four variance/covariance components. The first three test the null hypothesis that the population variance of the level-1 residuals, σ_e^2 , is 0, that the population variance of the level-2 residuals for initial status, σ_0^2 , is 0 and that the population variance of the level-2 residuals for the annual rate of change, σ_1^2 , is 0. The last tests whether the covariance between the level-2 residuals for initial status and annual rates of change, σ_{01} , is 0, indicating whether true initial status and true annual rate of

change are correlated, after participation in the intervention program is accounted for.

For these data, we reject only two of these null hypotheses (each at the .001 level). The test for the level-1 residual, on σ_{ϵ}^2 , suggests the existence of additional outcome variation at level-1, which may be predictable. To explain some of this remaining within-person variation, we might add suitable time-varying predictors such as the number of books in the child's home or the amount of parent-child interaction to the level-1 submodel.

The test for the level-2 residual for initial status, on σ_{0i}^2 , suggests the existence of additional variation in true initial status, π_{0i} , after accounting for the effects of program participation. This again suggests the need for additional predictors, but because this is a level-2 variance component (describing residual variation in true initial status), we would consider adding both time-invariant *and* time-varying predictors to the multilevel model.

We cannot reject the null hypotheses for the two remaining variance components. Failure to reject the null hypothesis for σ_1^2 indicates that *PROGRAM* explains all the potentially predictable variation between children in their true annual rates of change. Failure to reject the null hypothesis for σ_{01} indicates that the intercepts and slopes of the individual true change trajectories are uncorrelated—that there is no association between true initial status and true annual rates of change (once the effects of *PROGRAM* are removed). As we discuss in subsequent chapters, the results of these two tests might lead us to drop the second level-2 residual, ζ_{1i} , from our model, for neither its variance nor covariance with ζ_{0i} is significantly different from 0.

Doing Data Analysis with the Multilevel Model for Change

We are restless because of incessant change, but we would be frightened if change were stopped.

—Lyman Bryson

In chapter 3, we used a pair of linked statistical models to establish the multilevel model for change. Within this representation, a level-1 submodel describes how each person changes over time and a level-2 submodel relates interindividual differences in change to predictors. To introduce these ideas in a simple context, we focused on just one method of estimation (maximum likelihood), one predictor (a dichotomy), and a single multilevel model for change.

We now delve deeper into the specification, estimation, and interpretation of the multilevel model for change. Following introduction of a new data set (section 4.1), we present a *composite* formulation of the model that combines the level-1 and level-2 submodels together into a single equation (section 4.2). The new composite model leads naturally to consideration of alternative methods of estimation (section 4.3). Not only do we describe two new methods—*generalized least squares* (GLS) and *iterative generalized least squares* (IGLS)—within each, we distinguish further between two types of approaches, the *full* and the *restricted*.

The remainder of the chapter focuses on real-world issues of data analysis. Our goal is to help you learn how to articulate and implement a coherent approach to model fitting. In section 4.4, we present two “standard” multilevel models for change that you should always fit initially in any analysis—the *unconditional means* model and the *unconditional growth* model—and we discuss how they provide invaluable baselines for subsequent comparison. In section 4.5, we discuss strategies for adding time-invariant predictors to the multilevel model for change. We then discuss methods for testing complex hypotheses (sections 4.6 and 4.7) and examining model assumptions and residuals (section 4.8). We conclude,