Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence
Judith D. Singer and John B. Willett

Print publication date: 2003
Print ISBN-13: 9780195152968
Published to Oxford Scholarship Online: September 2009
DOI: 10.1093/acprof:oso/9780195152968.001.0001

Treating TIME More Flexibly

Judith D. Singer
John B. Willett

DOI:10.1093/acprof:oso/9780195152968.003.0005

Abstract and Keywords

This chapter demonstrates how to apply the multilevel model to complex data sets. Section 5.1 begins by illustrating what to do when the number of waves is constant but their spacing is irregular. Section 5.2 illustrates what to do when the number of waves per person differs as well; it also discusses the problem of missing data, the most common source of imbalance in longitudinal work. Section 5.3 demonstrates how to include time-varying predictors in your data analysis. Section 5.4 concludes by discussing why and how to adopt alternative representations for the main effect of TIME.

Keywords: multilevel model, time, data analyses, missing data

> Change is a measure of time
>
> —Edwin Way Teale

All the illustrative longitudinal data sets in previous chapters share two structural features that simplify analysis. Each is: (1) balanced—everyone is assessed on the identical number of occasions; and (2) time-structured—each set of occasions is identical across individuals. Our analyses have also been limited in that we have used only: (1) time-invariant predictors that describe immutable characteristics of individuals or their environment (except for *TIME* itself); and (2) a representation of *TIME* that forces the level-1 individual growth parameters to represent "initial status" and "rate of change."

The multilevel model for change is far more flexible than these examples suggest. With little or no adjustment, you can use the same strategies to analyze more complex data sets. Not only can the waves of data be irregularly spaced, their number and spacing can vary across participants. Each individual can have his or her own data collection schedule and the number of waves can vary without limit from person to person. So, too, predictors of change can be time-invariant or time-varying, and the level-1 submodel can be parameterized in a variety of interesting ways.

In this chapter, we demonstrate how you can fit the multilevel model for change under these new conditions. We begin, in section 5.1, by illustrating what to do when the number of waves is constant but their spacing is irregular. In section 5.2, we illustrate what to do when the number of waves per person differs as well; we also discuss the problem of missing data, the most common source of imbalance in longitudinal work. In section 5.3, we demonstrate how to include time-varying predictors in your data analysis. We conclude, in section 5.4, by discussing why and how you can adopt alternative representations for the main effect of *TIME*.

(p.139) 5.1 Variably Spaced Measurement Occasions

Many researchers design their studies with the goal of assessing each individual on an identical set of occasions. In the tolerance data introduced in chapter 2, each participant was assessed five times, at ages 11, 12, 13, 14, and 15. In the early intervention data introduced in chapter 3 and the alcohol use data introduced in chapter 4, each participant was assessed three times: at ages 12, 24, and 36 months or ages 14, 15, and 16 years. The person-period data sets from these time-structured designs are elegantly balanced, with a temporal variable that has an identical cadence for everyone under study (like *AGE* in tables 2.1, and 3.1).

Yet sometimes, despite a valiant attempt to collect time-structured data, actual measurement occasions will differ. Variation often results from the realities of fieldwork and data collection. When investigating the psychological consequences of unemployment, for example, Ginexi, Howe, and Caplan (2000) designed a time-structured study with interviews scheduled at 1, 5, and 11 months after job loss. Once in the field, however, the interview times varied considerably around these targets, with increasing variability as the study went on. Although interview 1 was conducted between 2 and 61 days after job loss, interview 2 was conducted between 111 and 220 days, and interview 3 was conducted between 319 and 458 days. Ginexi and colleagues could have associated the respondents' outcomes with the *target* interview times, but they argue convincingly that the number of days since job loss is a better metric for the measurement of time. Each individual in their study, therefore, has a *unique* data collection schedule: 31, 150, and 356 days for person 1; 23, 162, and 401 days for person 2; and so on.

So, too, many researchers design their studies knowing full well that the measurement occasions may differ across participants. This is certainly true, for example, of those who use an *accelerated cohort* design in which an age-heterogeneous cohort of individuals is followed for a constant period of time. Because respondents initially vary in age, and *age*, not *wave*, is usually the appropriate metric for analysis (see the discussion of time metrics in section 1.3.2), observed measurement occasions will differ across individuals. This is actually what happened in the larger alcohol-use study from which the small data set in chapter 4 was excerpted. Not only were those 14-year-olds re-interviewed at ages 15 and 16, concurrent samples of 15- and 16-year-olds were re-interviewed at ages 16 and 17 and ages 17 and 18, respectively. The advantage of an accelerated cohort design is that you can model change over a longer temporal period (here, the five years between ages 14 and 18) using fewer waves of data. Unfortunately, under the usual conditions, the data sets **(p. 140)** are then sparser at the earliest and latest ages, which can complicate the specification of the level-1 submodel.

In this section, we show how you can use the methods of previous chapters to analyze data sets with variably spaced measurement occasions. All you need to deal with are some minor coding issues for the temporal predictor in the person-period data set; model specification, parameter estimation, and substantive interpretation proceeds as before. To illustrate just how simple the analysis can be, we begin by discussing data sets in which the *number* of waves is constant but their *spacing* varies. We discuss data sets in which the *number* of waves varies as well in section 5.2.

5.1.1 The Structure of Variably Spaced Data Sets

We illustrate how to analyze data sets with variably spaced measurement occasions using a small sample extracted from the Children of the National Longitudinal Study of Youth (CNLSY). The data set, comprising children's scores on the reading subtest of the Peabody Individual Achievement Test (PIAT), includes three waves of data for 89 African-American children. Each child was 6 years old in 1986, the first year of data collection. During the second wave of data collection, in 1988, these children were to be 8; during the third wave, in 1990, they were to be 10. We focus here on an unconditional growth model, not the inclusion of level-2 predictors, because this second aspect of analysis remains unchanged.

Table 5.1 presents excerpts from the person-period data set. Notice that its structure is virtually identical to all person-period data sets shown so far. The only difference is that it contains *three* temporal variables denoting the passage of time: *WAVE, AGE*, and *AGEGRP*. Although we will include only one of these in any given model, a distinctive feature of time-unstructured data sets is the possibility of multiple metrics for clocking time (often called metameters).

WAVE is the simplest but least analytically useful of the three. Although its values—1, 2, and 3—reflect the study's design, they have little substantive meaning when it comes to addressing the research question. Because *WAVE* does not identify the child's age at each occasion, nor does it capture the chronological distance between occasions, it cannot contribute to a meaningful level-1 submodel. We mention this issue explicitly because empirical researchers sometimes postulate individual growth models using design variables like *WAVE* (or year of data collection) even though other temporal predictors are generally more compelling.

AGE is a better predictor because it specifies the child's actual age (to the nearest month) on the day each test was administered. A child like **(p.141)**

Table 5.1: Excerpts from the person-period data set for the reading study

ID	WAVE	AGEGRP	AGE	PIAT
04	1	6.5	6.00	18
04	2	8.5	8.50	31
04	3	10.5	10.67	50
27	1	6.5	6.25	19
27	2	8.5	9.17	36
27	3	10.5	10.92	57
31	1	6.5	6.33	18
31	2	8.5	8.83	31
31	3	10.5	10.92	51
33	1	6.5	6.33	18
33	2	8.5	8.92	34
33	3	10.5	10.75	29
41	1	6.5	6.33	18
41	2	8.5	8.75	28
41	3	10.5	10.83	36
49	1	6.5	6.50	19
49	2	8.5	8.75	32

ID	WAVE	AGEGRP	AGE	PIAT
49	3	10.5	10.67	48
69	1	6.5	6.67	26
69	2	8.5	9.17	47
69	3	10.5	11.33	45
77	1	6.5	6.83	17
77	2	8.5	8.08	19
77	3	10.5	10.00	28
87	1	6.5	6.92	22
87	2	8.5	9.42	49
87	3	10.5	11.50	64
…	…	…	…	…

Note that *TIME* is clocked using three distinct variables: *WAVE, AGEGRP,* and *AGE*.

ID 04, who had just turned 6 at wave 1, has an *AGE* of 6.00 for that record; a child like *ID* 87, who would soon turn 7, has an *AGE* of 6.92. The average child is 6.5 years old at wave 1, as we would expect if births and testing occasions were randomly distributed. If data collection had proceeded according to plan, the average child would have been 8.5 and 10.5 years old at the next two waves. Not surprisingly, actual ages varied around these targets. By wave 2, the youngest child had just turned 8 while the oldest was well over 9. By wave 3, the youngest child had just turned 10 while the oldest was nearly 12. Like many longitudinal studies, the *CNLSY* suffers from "occasion creep"—over time, the temporal separation of **(p.142)** occasions widens as the actual ages exceed design projections. In this data set, the average child is 8.9 years in wave 2 and nearly 11 years in wave 3.

The third temporal variable, *AGEGRP*, is a time-structured predictor that is more substantively meaningful than the design variable *WAVE*. Its values indicate the child's "expected age" on each measurement occasion (6.5, 8.5, and 10.5). This time-structured predictor clocks time on a scale that is comparable numerically to the irregularly spaced predictor *AGE*. Adding *AGEGRP* to the person-period data set allows us to demonstrate that the characterization of a data set as time-structured or irregular can depend on nothing more than the *cadence* of the temporal predictor used to postulate a model. If we postulate our model using *AGEGRP*, the data set is time-structured; if we postulate a comparable model using *AGE*, it is not.

The multilevel model for change does not care if the individual-specific cadence of the level-1 predictor is identical for everyone or if it varies from case to case. Because we fit the model using the actual numeric values of the temporal predictor, spacing is irrelevant. We can postulate and fit a comparable model regardless of the variable's cadence. Of far greater importance is the choice of the functional form for the level-1 submodel. Should it represent linear change or a more complex shape for the individual growth trajectory? Might this decision depend upon the specific temporal predictor chosen for model building?

To address these questions, figure 5.1 presents empirical change plots with superimposed OLS linear change trajectories for 9 children. Each panel plots each child's *PIAT* scores twice, once for each temporal predictor. We use •'s and a dashed line when plotting by *AGE*; we use +'s and a solid line when plotting by *AGEGRP*. With just three waves of data—whichever temporal predictor we use—it is difficult to argue for anything but a linear change individual growth model.

If we can postulate a linear change individual growth model using either temporal predictor, which one should we use? As argued above, we prefer *AGE* because it provides more precise information about the child at the moment of testing. Why set this information aside just to use the equally spaced, but inevitably less accurate, *AGEGRP*. Yet this is what many researchers do when analyzing longitudinal data—indeed, it is what *we* did in chapters 3 and 4. There, instead of using the participant's precise ages, we used integers: 12, 18, and 24 months for the children in chapter 3; 14, 15, and 16 for the teenagers in chapter

4. Although the loss of precision may be small, as suggested by the close correspondence between the pairs of fitted OLS trajectories in each panel of figure 5.1, there are children for whom the differential is much larger. To investigate this question empirically, we fit two multilevel models for change to these data: one using *AGEGRP*, another using *AGE* as the **(p.143)**

(p.144) temporal predictor at level-1. Doing so allows us to demonstrate how to analyze irregularly spaced data sets *and* to illustrate the importance of assessing the merits for time empirically.

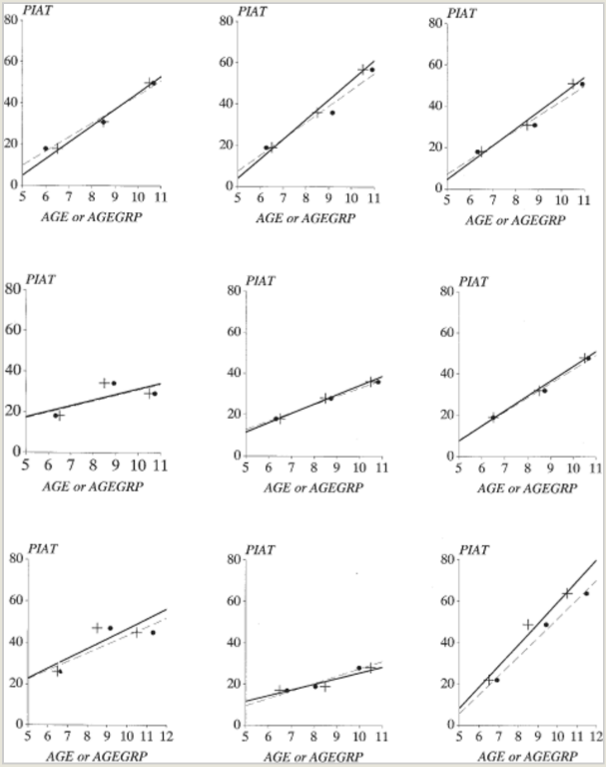

Figure 5.1. Comparing time-structured and time-unstructured representations of the effect of *TIME*. Empirical change plots with superimposed OLS trajectories for 9 participants in the reading study. The +’s and solid lines are for *TIME* clocked using the child’s *target* age at data collection; the •’s and dashed lines are for *TIME* clocked using each child’s observed age.

5.1.2 Postulating and Fitting Multilevel Models with Variably Spaced Waves of Data

Regardless of which temporal representation we use, we postulate, fit, and interpret the multilevel model for change using the same strategies. Adapting the general specification of an unconditional growth model in equations 4.9a and 4.9b, let Y_{ij} be child i’s *PIAT* score on occasion j and $TIME_{ij}$ represent either temporal variable:

(5.1a)

$$Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \varepsilon_{ij}$$
$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$
$$\pi_{1i} = \gamma_{10} + \zeta_{1i},$$

where

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ and } \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right). \tag{5.1b}$$

If we center both *AGE* and *AGEGRP* on age 6.5—the average child's age at wave 1—the parameters have the usual interpretations. In the population from which this sample was drawn, γ_{00} represents the average child's true initial status (at age 6.5); γ_{10} represents the average child's annual rate of true change between ages 6 and 11; σ_ε^2 summarizes the within-child scatter around his or her own true change trajectory; and σ_0^2 and σ_1^2 summarize the between-child variability in initial status and annual rates of change.

Use of a generic representation $TIME_{ij}$ in the level-1 growth model (instead of a specific representation like *AGE* – 6.5 or *AGEGRP* – 6.5) yields these interpretations. We can postulate the same model for either predictor because $TIME_{ij}$ includes subscripts that are both person-specific (*i*) and time-specific (*j*). If *TIME* represents *AGEGRP* – 6.5, the data set is time structured; if we use *AGE* – 6.5, it is not. From a data-analytic perspective, you just specify the relevant temporal representation to your statistical software. From an interpretive perspective, the distinction is moot.

Table 5.2 presents the results of fitting these two unconditional growth models to these data: the first uses *AGEGRP* – 6.5; the second uses *AGE* – 6.5. Each was fit using full ML in SAS PROC MIXED. The parameter estimates for initial status, $\hat{\gamma}_{00}$, are virtually identical—21.16 and 21.06—as are those for the within-child variance, σ_ε^2: 27.04 and 27.45. But the similarities stop there. For the slope parameter, γ_{10}, the estimated growth **(p.145)**

Table 5.2: Results of using alternative representations for the main effect of *TIME* (*n* = 89) when fitting an unconditional growth model to the CNLSY reading data

			Predictor representing TIME	
		Parameter	AGEGRP – 6.5	AGE – 6.5
Fixed Effects				
Initial status, π_{0i}	Intercept	γ_{00}	21.1629***	21.0608***
			(0.6143)	(0.5593)
Rate of Change, π_{1i}	Intercept	γ_{10}	5.0309***	4.5400***
			(0.2956)	(0.2606)
Variance Components				
Level-1:	within-person	σ_ε^2	27.04***	27.45***
Level-2:	In initial status	σ_0^2	11.05*	5.11
	In rate of change	σ_1^2	4.40***	3.30***
Goodness-of-fit				
	Deviance		1819.8	1803.9
	AIC		1831.9	1815.9
	BIC		1846.9	1830.8

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

The first model treats the data set as time-structured by using the predictor (*AGEGRP* – 6.5); the second model treats the data set as time-unstructured by using each child's actual age at each assessment, (*AGE* – 6.5).

Note. SAS Proc Mixed, Full ML. Also note that the covariance component, σ_{01}, is estimated, but not displayed.

rate is half a point larger in a model with *AGEGRP* – 6.5 (5.03 vs. 4.54). This cumulates to a two-point differential in PIAT scores over the four years under study. So, too, the two level-2 variance components are much larger for a model with *AGEGRP* – 6.5. Why are these estimates larger when we treat the data set as time-structured, using *AGEGRP* – 6.5 as our level-1 predictor, than when we treat it as irregular, using *AGE* – 6.5? We obtain a larger fixed effect for linear growth because *AGEGRP* associates the data for waves 2 and 3 with earlier ages (8.5 and 10.5) than observed. If we amortize the same gain over a shorter time period, the slope must be steeper. We obtain larger estimated variance components because the model with the time-structured predictor fits less well—there is more unexplained variation in initial status and growth rates—than when we associate each child's data with his or her age at testing. In other words, treating this unstructured data set as though it is time-structured introduces error into the analysis—error that we can reduce by using the child's age at testing as the temporal predictor.

We conclude that the model with *AGEGRP* as the level-1 temporal **(p.146)** predictor fits less well than the model with *AGE*. With the former representation, the slope is inappropriately larger—inaccurately implying more rapid gains—and there is more unexplained variation in initial status and rates of change. The superiority of the model with *AGE* as the temporal predictor is supported by its smaller AIC and BIC statistics. The bottom line: never "force" an unstructured data set to be structured. If you have several metrics for tracking time—and you often will—investigate the possibility of alternative temporal specifications. Your first choice, especially if tied to design, not substance, may not always be the best.

5.2 Varying Numbers of Measurement Occasions

Once you allow the spacing of waves to vary across individuals, it is a small leap to allow their *number* to vary as well. Statisticians say that such data sets are *unbalanced*. As you would expect, balance facilitates analysis: models can be parameterized more easily, random effects can be estimated more precisely, and computer algorithms will converge more rapidly.

Yet a major advantage of the multilevel model for change is that it is easily fit to unbalanced data. Unlike approaches such as repeated measures analysis of variance, with the multilevel modeling of change it is straightforward to analyze data sets with varying numbers of waves of data. To illustrate the general approach, we begin, in section 5.2.1, by introducing a new data set in which the number of waves per person varies widely, from 1 to 13. We extend this discussion in section 5.2.2, by discussing implementation and estimation problems that can arise when data are unbalanced. We conclude, in section 5.2.3, by discussing potential causes of imbalance—especially missing data—and how they can affect statistical analysis.

5.2.1 Analyzing Data Sets in Which the Number of Waves per Person Varies

Murnane, Boudett, and Willett (1999) used data from the National Longitudinal Survey of Youth (NLSY) to track the labor-market experiences of male high school dropouts. Like many large panel studies, the NLSY poses a variety of design complications: (1) at the first wave of data collection, the men varied in age from 14 to 17; (2) some subsequent waves were separated by one year, others by two; (3) each wave's interviews were conducted at different times during the calendar year; and (4) respondents could describe more than one job at each interview. Person-specific schooling and employment patterns posed further problems. Not only could respondents drop out of school at different times and enter the **(p.147)**

Table 5.3: Excerpts from the person-period data set for the high school dropout wage study

ID	EXPER	LNW	BLACK	HGC	UERATE
206	1.874	2.028	0	10	9.200
206	2.814	2.297	0	10	11.000
206	4.314	2.482	0	10	6.295
332	0.125	1.630	0	8	7.100
332	1.625	1.476	0	8	9.600
332	2.413	1.804	0	8	7.200
332	3.393	1.439	0	8	6.195
332	4.470	1.748	0	8	5.595
332	5.178	1.526	0	8	4.595
332	6.082	2.044	0	8	4.295
332	7.043	2.179	0	8	3.395
332	8.197	2.186	0	8	4.395
332	9.092	4.035	0	8	6.695
1028	0.004	0.872	1	8	9.300
1028	0.035	0.903	1	8	7.400
1028	0.515	1.389	1	8	7.300
1028	1.483	2.324	1	8	7.400

ID	EXPER	LNW	BLACK	HGC	UERATE
1028	2.141	1.484	1	8	6.295
1028	3.161	1.705	1	8	5.895
1028	4.103	2.343	1	8	6.900

labor force at different times, they also changed jobs at different times. To track wages on a common temporal scale, Murnane and colleagues decided to clock time from each respondent's first day of work. This allows each hourly wage to be associated with a temporally appropriate point in the respondent's labor force history. The resulting data set has an unusual temporal schedule, varying not only in spacing but length.
Table 5.3 presents excerpts from the person-period data set. To adjust for inflation, each hourly wage is expressed in constant 1990 dollars. To address the skewness commonly found in wage data and to linearize the individual wage trajectories, we analyze the natural logarithm of wages, *LNW*. Then, to express this outcome on its original scale, we take antilogs (e.g., $e^{(2.028)}$ = \$7.60 per hour).

The temporal variable *EXPER* identifies the specific moment—to the nearest day—in each man's labor force history associated with each observed value of *LNW*. Notice the variability in the number and spacing of waves. Dropout 206 has three waves, for jobs held at 1.874, 2.814 and 4.314 years of experience after labor force entry. Dropout 332 has 10 waves, the first for a job held immediately after entering the labor force, the others for jobs held approximately every subsequent year. Dropout **(p.148)** 1028 has 7 waves; the first three describe the first six months of work (at 0.004, 0.035, and 0.515 years). Across the full sample, 77 men have 1 or 2 waves of data, 82 have 3 or 4, 166 have 5 or 6, 226 have 7 or 8, 240 have 9 or 10, and 97 have more than 10. The earliest wave describes someone's first day of work; the latest describes a job held 13 years later.

This is the first data set we have presented in which the number of waves of data varies across individuals. Some men even have fewer than three waves—less than the minimum articulated in previous chapters. A major advantage of the multilevel model for change is that everyone can participate in the estimation, regardless of how many waves he contributes to the data set. Even the 38 men with just 1 wave of data and the 39 with just 2 waves are included in the estimation. Although they provide less, or no, information about within-person variation—and hence do not contribute to variance component estimation—they can still contribute to the estimation of fixed effects where appropriate. Ultimately, each person's fitted trajectory is based on a combination of his: (1) observed trajectory, and (2) a model-based trajectory determined by the values of the predictors.

You need no special procedures to fit a multilevel model for change to unbalanced data. All you need do is specify the model appropriately to your statistical software. As long as the person-period data set includes enough people with enough waves of data for the numeric algorithms to converge, you will encounter no difficulties. If the data set is severely unbalanced, or if too many people have too few waves for the complexity of your hypothesized model, problems may arise in the estimation. For now, we continue with this data set, which includes so many people with so many waves that estimation is straightforward. We discuss strategies for identifying and resolving estimation problems in section 5.2.2.

Table 5.4 presents the results of fitting three multilevel models for change to the wage data, using full ML in SAS PROC MIXED. First examine the results for Model A, the unconditional growth model. The positive and statistically significant fixed effect for *EXPER* indicates that inflation-adjusted wages rise over time. Because the outcome, *LNW*, is expressed on a logarithmic scale, its parameter estimate, $\hat{\gamma}_{10}$, is not a linear growth rate. As in regular regression, however, transformation facilitates interpretation. If an outcome in a linear relationship, *Y*, is expressed as a natural logarithm and $\hat{\gamma}_{10}$, is the regression coefficient for a predictor *X*, then 100 ($e^{(\hat{\gamma}_{10})}$ – 1) is the *percentage change* in *Y* per unit difference in *X*. Because *EXPER* is calibrated in years, this transformation yields an annual percentage growth rate in wages. Computing $100(e^{(0.0457)} – 1) = 4.7$, we estimate that the average high school dropout's inflation-adjusted hourly wages rise by 4.7% with each year of labor force participation.

(p.149)

Table 5.4: Results of fitting a taxonomy of multilevel models for change to the high school dropout wage data (*n* = 888)

		Parameter	Model A	Model B	Model C
Fixed Effects					
Initial status, π_{0i}	Intercept	γ_{00}	1.7156***	1.7171***	1.7215***
			(0.0108)	(0.0125)	(0.0107)
	(*HGC* – 9)	γ_{01}		0.0349***	0.0384***
				(0.0079)	(0.0064)
	BLACK	γ_{02}		0.0154	
				(0.0239)	
Rate of change, π_{1i}	Intercept	γ_{10}	0.0457***	0.0493***	0.0489***
			(0.0023)	(0.0026)	(0.0025)
	(*HGC* – 9)	γ_{11}		0.0013	
				(0.0017)	
	BLACK	γ_{12}		−0.0182**	−0.0161***
				(0.0055)	(0.0045)
Variance Components					
Level-1:	within-person	σ_{ε}^{2}	0.0951***	0.0952***	0.0952***
Level-2:	In initial status	σ_{0}^{2}	0.0543***	0.0518***	0.0518***
	In rate of change	σ_{1}^{2}	0.0017***	0.0016***	0.0016***

	Parameter	Model A	Model B	Model C
Goodness-of-fit				
	Deviance	4921.4	4873.8	4874.7
	AIC	4933.4	4893.8	4890.7
	BIC	4962.1	4941.7	4929.0

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

Model A is an unconditional growth model; Model B includes the effects of highest grade completed (*HGC* – 9) and race (*BLACK*) on both initial status and rate of change; Model C is a reduced model in which (*HGC* – 9) predicts only initial status and *BLACK* predicts only rate of change.

Note: SAS Proc Mixed, Full ML. Also note that the covariance component, σ_{01}, is estimated, but not displayed.

After specifying a suitable individual growth model, you add level-2 predictors in the usual way. The statistically significant variance components in Model A, for both initial status and rate of change, suggest the wisdom of this action. Models B and C examine the effects of two predictors: (1) the race/ethnicity of the dropout; and (b) the highest grade he completed before dropping out. Although the sample includes 438 Whites, 246 African Americans, and 204 Latinos, analyses not shown here suggest that we cannot distinguish statistically between the trajectories of Latino and White dropouts. For this reason, these models include just one race/ethnicity predictor (*BLACK*). Highest grade completed, *HGC*, is a continuous variable that ranges from 6th through 12th grade, with an average of 8.8 and a standard deviation of 1.4. To facilitate **(p. 150)** interpretation, our analyses use a rescaled version, *HGC* – 9, which centers *HGC* around this substantively meaningful value near the sample mean (see section 4.5.4 for a discussion of centering).

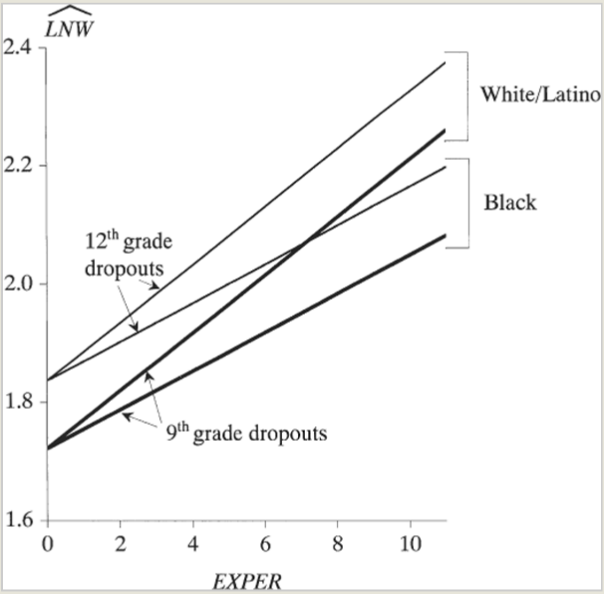

Figure 5.2. Displaying the results of a fitted multilevel model for change. Log wage trajectories from Model C of table 5.4 for four prototypical dropouts: Blacks and Whites/Latinos who dropped out in 9th and 12th grades.

Model B of Table 5.4 associates each predictor with initial status and rate of change. The estimated fixed effects suggest that *HGC* – 9 is related only to initial status while *BLACK* is related only to the rate of change. We therefore fit Model C, whose level-2 submodels reflect this observation. The fixed effect for *HGC* – 9 on initial status tells us that dropouts who stay in school longer earn higher wages on labor force entry ($\hat{\gamma}_{01}$ = 0.0384, $p < .001$), as we might expect because they are likely to have more skills than peers who left school earlier. The fixed effect for *BLACK* on rate of change tells us that, in contrast to Whites and Latinos, the wages of Black males increase less rapidly with labor force experience ($\hat{\gamma}_{12}$ = –0.0161, $p < .001$). The statistically significant level-2 variance components indicate the presence of additional unpredicted interindividual variation in both initial status and rate of change. In sections 5.3.3 and 6.1.2, we add other predictors that explain some of this remaining variation.

Figure 5.2 summarizes the effects in Model C by displaying wage trajectories for four prototypical dropouts: Blacks and Whites/Latinos who dropped out in 9th and 12th grades. We obtained these trajectories using the same two-stage process presented in section 4.5.3. We first substituted the two values of *BLACK* (0 and 1) into Model C and then substituted in two prototypical values of *HGC* – 9 (0 and 3, to correspond to 9 and 12 years of education). The plots document the large and statistically significant effects of education and race on the wage trajectories. The longer a prospective dropout stays in school, the higher his wages on labor force entry. But race plays an important role, not on initial wages but on the rate of change. Although the average Black dropout initially earns an **(p.151)** hourly wage indistinguishable from the average White or Latino dropout, his annual percentage increase is lower. Controlling for highest grade completed, the average annual percentage increase is $100(e^{(0.0489)} - 1) = 5.0\%$ for Whites and Latinos in comparison to $100(e^{(0.0328)} - 1) = 3.3\%$ for Blacks. Over time, this race differential overwhelms the initial advantage of remaining in school. Beyond 7 years of labor force participation, a Black male who left school in 12th grade earns a lower hourly wage than a White or Latino male who left in 9th.

5.2.2 Practical Problems That May Arise When Analyzing Unbalanced Data Sets

We encountered no problems when fitting models to the unbalanced data in section 5.2.1. The most complex model (C) converged in just three iterations and we could estimate every parameter in the model. But if your data set is severely unbalanced, or if too few people have enough waves of data, computer iterative algorithms may not converge and you may be unable to estimate one or more variance components.

Why does imbalance affect the estimation of variance components but not fixed effects? No matter how unbalanced the person-period data set, the estimation of fixed effects is generally no more difficult than the estimation of regression coefficients in a regular linear model. To demonstrate why, let us begin with a multilevel model—for simplicity, an unconditional growth model—expressed in composite form:

(5.2a) $$Y_{ij} = [\gamma_{00} + \gamma_{10} TIME_{ij}] + [\zeta_{0i} + \zeta_{1i} TIME_{ij} + \varepsilon_{ij}].$$

If we re-express the composite error term in the second set of brackets as: $\varepsilon^*_{ij} = [\zeta_{0i} + \zeta_{1i} TIME_{ij} + \varepsilon_{ij}]$, we obtain an equivalent representation of equation 5.2a:

(5.2b) $$Y_{ij} = \gamma_{00} + \gamma_{10} TIME_{ij} + \varepsilon^*_{ij}.$$

Equation 5.2b resembles a standard regression model, with γ’s instead of β’s and ε^*_{ij} instead of ε_{ij} The difference is that we do not assume that the composite residuals ε^*_{ij} are independent and normally distributed with mean 0 and variance $\sigma^2_{\varepsilon^*}$. Instead we assume that their constituents—ζ_{0i},ζ_{1i}, and ε_{ij}—follow the assumptions:

$$\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right) \text{ and } \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right).$$

It is these complex assumptions—about the variance components—that complicate estimation.

Now consider the following thought experiment. Suppose we are willing **(p. 152)** to make a simplifying assumption about the composite residuals, declaring them to be independent and normally distributed: $\varepsilon_{ij}^* \sim N(0, \sigma_{\varepsilon^*}^2)$. This is tantamount to assuming that both level-2 residuals, ζ_{0i} and ζ_{1i}, are always 0, as would be their associated variance components (i.e., both σ_0^2 and σ_1^2 are also 0). In the language of multilevel modeling, we would be *fixing* the intercept and rate of change, making them constant across individuals. Whether each person contributed one wave or many, estimation of the two fixed effects and the one variance component would then become a standard regression problem. All you would need are a sufficient number of distinct values of $TIME_{ij}$ in the person-period data set—enough distinct points in a plot of Y_{ij} vs. $TIME_{ij}$—to identify the level-1 submodel's functional form. In a time-structured data set, this plot would be composed of vertical stripes, one for each measurement occasion. This is why you would need at least three waves of data—the stripes would lie at just those three occasions. In an unstructured data set, the variable spacing of waves makes it easier to estimate fixed effects because the data points are more separated "horizontally." This allows you to relax the data minimum per person—allowing some people to have fewer than three waves—as long as you have enough distinct values of $TIME_{ij}$ to estimate the fixed effects.

If we are unwilling to make these simplifying assumptions—and we generally are—estimation of variance components can be difficult if too many people have too few waves. Variability in the spacing of waves helps, but may not resolve the problem. Estimation of variance components requires that enough people have sufficient data to allow quantification of within-person residual variation—variation in the residuals over and above the fixed effects. If too many people have too little data, you will be unable to quantify this residual variability.

When does the numeric task become so difficult that the variance components cannot be estimated? We offer no rules because so many issues are involved, including the degree of imbalance, the complexity of the model, the number of people with few vs. many waves, and the inclusion of time-varying predictors (discussed in section 5.3). Suffice it to say that when imbalance is severe enough, numeric computer algorithms can produce theoretically impossible values or fail to converge. Each statistical software program has its own way of informing the user of a problem; once discovered, we recommend that you be proactive and not automatically accept the default “solution” your program offers. Below, we discuss each of the two major estimation problems.

Boundary Constraints

Many population parameters have *boundary constraints*—limits beyond which they cannot theoretically lie. Like variances and correlation **(p.153)** coefficients, the variance/covariance components in the multilevel model have clear boundaries: (1) a variance component cannot be negative; and (2) a covariance component, expressed in correlation form, must lie between −1 and +1. Because of the complexity of the estimation task—especially with unbalanced data—as well as the iterative nature of the computational algorithms, multilevel modeling programs occasionally generate parameter estimates that reach, or lie outside, these limits. When this happens, the program may output the implausible estimate or its boundary value (e.g., it might set a variance component to 0).

How will you know if you have encountered a boundary constraint? The warning signs differ across programs. If you use SAS PROC MIXED, the program log will note that “the G matrix [the variance-covariance matrix for the variance components] is not positive definite.” By default, SAS sets the offending estimate to its boundary value. MLwiN does not provide a note; instead, it sets the offending estimate, and all associated estimates, to boundary values. If your output indicates that an estimate is exactly 0, you have likely encountered a boundary constraint. HLM will provide you with a warning message and modify its computational algorithm to avoid the problem. With all software, one clue that you may be approaching a “boundary” is if you find you need an excessive number of iterations to reach convergence.

We recommend that you never let a computer program arbitrarily make important decisions like these. Regardless of which program you use, you should be proactive about boundary constraints. Overspecification of the model’s stochastic portion is the usual cause; model simplification is generally the cure. A practical solution is to compare alternative models that remove one, or more, offending random effects systematically until the model can be fit. This strategy, known as *fixing* a predictor’s effect, usually resolves the problems.

We illustrate this approach using a small data set purposefully selected from the larger wage data set just analyzed. We constructed this sample for pedagogic purposes, hoping to create such extreme imbalance that boundary constraints would arise. This new data set is composed of the 124 men who had three or fewer waves of wage data: 47 men have three waves, 39 have two, and 38 have only one. The earliest value of *EXPER* is 0.002; the latest is 7.768. This data set is *not* a random sample of the original group.

Table 5.5 presents the results of fitting three models to this smaller data set; each is based upon Model C, the “final” model of table 5.4. As before, each was fit using ML in SAS PROC MIXED. In the first model, which is identical to Model C, the estimated variance component for linear growth, $\hat{\sigma}_1^2$, is exactly 0. This is a standard sign of a boundary **(p.154)**

Table 5.5: Comparison of three alternative approaches to fitting Model C of table 5.4 to a severely unbalanced subset of the high school dropout wage data (n = 124)

		Parameter	A Default method	B Removing boundary constraints	C Fixing rates of change
Fixed Effects					
Initial status, π_{0i}	Intercept	γ_{00}	1.7373***	—	1.7373***
			(0.0476)		(0.0483)
	(*HGC* – 9)	γ_{01}	0.0462~	—	0.0458~
			(0.0245)		(0.0245)
Rate of	Intercept	γ_{10}	0.0516*	—	0.0518*
change, π_{1i}			(0.0211)		(0.0209)
	BLACK	γ_{12}	−0.0596~	—	−0.0601~
			(0.0348)		(0.0346)
Variance Components					
Level-1:	Within-person	σ_ε^2	0.1150***	0.1374***	0.1148***
Level-2:	In initial status	σ_0^2	0.0818**	0.0267	0.0842***
	In rate of change	σ_1^2	0.0000	−0.0072	—
Goodness-of-fit					
	Deviance		283.9	—	283.9

Parameter		A Default method	B Removing boundary constraints	C Fixing rates of change
AIC		297.9	—	295.9
BIC		317.6	—	312.8

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

Model A uses the default option in SAS PROC MIXED; Model B removes boundary constraints for the variance components; Model C removes the level-2 residual for rate of change, eliminating the associated variance component (as well as the associated covariance component).

Note. SAS Proc Mixed, Full ML. Also note that the covariance component, σ_{01}, is estimated where appropriate, but not displayed.

problem, used by both SAS PROC MIXED and MLwiN. Estimates of 0 are always suspicious; here they indicate that the algorithm has encountered a boundary constraint. (Note that SAS allows the associated covariance component to be non-zero, whereas MLwiN would also set that term to 0.)

Model B in table 5.5 represents our dogged attempt to fit the specified model to data. To do so, we invoke a software option that relaxes the default boundary constraint permitting us to obtain a negative variance component. When analyzing severely unbalanced data, eliminating automatic fix-ups can help identify problems with boundary constraints. Unfortunately, in this case, the iterative algorithm does not converge (a different problem that we will soon discuss). Nevertheless, notice that the estimated variance component for rate of change at the last iteration is **(p.155)** negative—a logical impossibility. This, too, is another sign suggesting the need for model simplification.

Model C in table 5.5 constrains the variance component for the linear growth rate, and its associated covariance component, to be 0. Notice that the deviance statistic for this model is identical to that of the first, suggesting the wisdom of fixing this parameter. This model fits no worse and involves fewer parameters (as reflected by the superior AIC and BIC statistics). This means that with this data set—which is *not* a random sample from the original—we cannot confirm the existence of any systematic residual variation in the slopes of the wage trajectories beyond the modest effect of *BLACK* shown in the final column of table 5.5.

Nonconvergence

As discussed in section 4.3, all multilevel modeling programs implement iterative numeric algorithms for model fitting. These algorithms compare fit criteria (such as the log-likelihood statistic) across successive iterations and declare convergence when the change in the fit criterion is sufficiently “small.” Although the user can determine how small is “small enough,” all programs have a default criterion, generally an arbitrarily small proportional change. When the criterion is met, the algorithm *converges* (i.e., stops iterating). If the criterion cannot be met in a large number of iterations, estimates should be treated with suspicion.

How many iterations are needed to achieve convergence? If your data set is highly structured and your model simple, convergence takes just a few iterations, well within the default values set by most programs. With unbalanced data sets and complex models, convergence can take hundreds or thousands of iterations although the algorithms in specialized packages (e.g., HLM and MLwiN) usually converge more rapidly than those in multipurpose programs (e.g., SAS PROC MIXED).

For every model you fit—but especially for models fit to unbalanced data—be sure to check that the algorithm has converged. In complex problems, the program’s default limits on the maximum number of iterations may be too low to reach convergence. All packages allow you to increase this limit. If the algorithm still does not converge, sequentially increase the limit until it does. Some programs allow you to facilitate this search by providing “starting values” for the variance and covariance components.

No matter how many iterations you permit and no matter how much prior information you provide, there will be times when the algorithm will not converge. Nonconvergence can result from many factors, but two common causes are poorly specified models and insufficient data; their **(p.156)** combination can be deadly. If you need an extremely large number of iterations to fit a model to data, closely examine the variance components and determine whether you have sufficient information to warrant allowing level-2 residuals for both initial status and rates of change. (If you are fitting nonlinear models using the methods of chapter 6, scrutinize other variance components as well.) Remember that any given data set contains a finite amount of information. You can postulate a complex model, but it is not always possible to fit that model to the available data.

We conclude by noting that other problems besides boundary constraints can cause nonconvergence. One problem, easily remedied, is a variable’s scale. If an outcome’s values are too small, the variance components will be smaller still; this can cause nonconvergence via rounding error issues. Simple multiplication of the outcome by 100, 1000, or another factor of 10 can usually ameliorate this difficulty. Predictor scaling can also cause problems but usually you want to adjust its metric in the *opposite* direction. For a temporal predictor, for example, you might move from a briefer time unit to a longer one (from days to months or months to years) so as to increase the growth rate’s magnitude. These kinds of transformations have only cosmetic effects on your essential findings. (They will change the value of the log likelihood and associated statistics, but leave the results of tests unaffected.)

5.2.3 Distinguishing among Different Types of Missingness

No discussion of imbalance is complete without a complementary discussion of its underlying source. Although some researchers build imbalance into their design, most imbalance is unplanned, owing to scheduling problems, missed appointments, attrition, and data processing errors. Further imbalance accrues if individuals who miss a wave of data collection subsequently return to the sample. For example, although the NLSY has a low annual attrition rate—less than 5% of the original sample initially leave in each of the first 13 years—many participants miss one or two waves. In their exhaustive study of NSLY attrition, MaCurdy, Mroz, and Gritz (1998) find many differences among persisters, dropouts, and returnees. Of relevance for the wage analyses just presented are the findings that attrition is higher for both the unemployed and men who once earned high wages.

Unplanned imbalance, especially when it stems from attrition or other potentially systematic sources, may invalidate your inferences. The issue is not the technical ability to fit a model but rather a substantive question about credible generalization. To probe the issues, statisticians frame **(p.157)** the problem, not in terms of imbalance, but rather in terms of *missing data*. When you fit a multilevel model for change, you implicitly assume that each person's observed records are a random sample of data from his or her underlying true growth trajectory. If your design is sound and has no built-in bias, and everyone is assessed on every planned occasion, your observed data will meet this assumption. If one or more individuals are not assessed on one or more occasions, your observed data may not meet this assumption. In this case, your parameter estimates may be biased and your generalizations incorrect.

Notice that we use the word "may," not "will," throughout the previous paragraph. This is because missingness, in and of itself, is not necessarily problematic. It all depends upon what statisticians call the *type of missingness*. In seminal work on this topic, Little (1995), refining earlier work with Rubin (Little & Rubin, 1987), distinguishes among three types of missingness: (1) *missing completely at random* (MCAR); (2) covariate-dependent dropout (CDD); and (3) *missing at random* (MAR) (see also Schafer, 1997). As Laird (1988) demonstrates, we can validly generalize the results of fitting a multilevel model for change under all three of these missingness conditions, which she groups together under rubric *ignorable nonresponse*.

When we say that data are MCAR, we argue that the observed values are a random sample of all the values that could have been observed (according to plan), had there been no missing data. Because time-invariant predictors are usually measured when a study begins, their values are rarely missing. As a result, when a multilevel model includes no time-varying predictors, the only predictor that can be missing is *TIME* itself (when a planned measurement occasion is missed). This means that longitudinal data are MCAR if the probability of assessment on any occasion is independent of: (1) the particular time; (2) the values of the substantive predictors; and (3) the values of the outcome (which are, by definition, unobserved). For the NLSY wage data just analyzed, we can make a case for the MCAR assumption if the probability of providing wage data at any point in time is independent of the particular moment in that individual's labor force history, all other predictors, and the unobserved wage. There cannot be particular moments when a man would be unlikely to grant an interview, as would be the case if men were unwilling to do so on specific days (which seems unlikely). But missingness must also not vary systematically by an individual's wage or other potentially unobserved characteristics. MaCurdy and colleagues (1998) convincingly demonstrate that these latter two conditions are implausible for the NLSY.

The conclusion that the MCAR assumption is untenable for the NLSY **(p.158)** data is unsurprising as this assumption is especially restrictive—wonderful when met, but rarely so. Covariate dependent dropout (CDD) is a less restrictive assumption that permits associations between the probability of missingness and observed predictor values ("covariates"). Data can be CDD even if the probability of missingness is systematically related to either *TIME* or observed substantive predictors. For the NLSY wage data, we can argue for the validity of the CDD assumption even if there are particular moments when men are unlikely to grant interviews. Missingness can also vary by either race or highest grade completed (our two observed predictors). By including these observed predictors in the multilevel model, we deflect the possibility of bias, allowing appropriate generalization of empirical results.

The major difficulty in establishing the tenability of the MCAR and CDD assumptions is the requirement of demonstrating that the probability of missingness at any point in time is unrelated to the contemporaneous value of the associated outcome. Because this outcome is unobserved, you cannot provide empirical support as you lack the very data you need. Only a substantive argument and thought experiment will do. Any potential relationship between the unobserved outcome and the probability of missingness invalidates these assumptions. For example, if men with particularly high or low wages are less likely to participate in an NLSY interview, we cannot support either assumption. As this hypothesis is both tenable and likely, we cannot defend either assumption for the NLSY wage data (nor for many other longitudinal data sets).

Fortunately, there is an even less restrictive type of missingness—more common in longitudinal research—that still permits valid generalization of the multilevel model for change: the MAR assumption. When data are MAR, the probability of missingness can depend upon *any* observed data, for either the predictors or any outcome values. It cannot, however, depend upon any *un*observed value of either any predictor or the outcome. So if we are willing to argue that the probability of missingness in the NLSY depends only upon observed predictor values (that is, *BLACK* and *HGC*) *and* wage data, we can make a case for the MAR assumption. The allowance for dependence upon observed outcome data can account for a multitude of sins, often supporting the credibility of the MAR assumption even when MCAR and CDD assumptions seem far-fetched.

As general as it seems, you should not accept the MAR assumption without scrutiny. Greenland and Finkle (1995) examine this assumption in cross-sectional research and suggest that even it can be difficult to meet. To illustrate their point, they argue that someone's unwillingness to answer a question about sexual preference (i.e., heterosexual vs. homosexual) **(p.159)** is likely correlated with his or her true sexual preference. We agree, but believe that there are many times when an individual's outcome values will adequately reflect such concerns. Yet even this assertion can be untrue. For example, a recovering alcoholic's willingness to continue participating in a study about abstinence is likely related to his or her ability to stay sober on each occasion. Such a systematic pattern—even if impossible to prove—invalidates the MAR assumption.

In practice, the burden of evaluating the tenability of these missingness assumptions rests with you. Any type of ignorable missingness permits valid inference; you just need to determine which seems most credible for your project. We suggest that you act as your own harshest critic—better you than the reviewers! As MAR is the least restrictive assumption, it provides the acid test. The key question is whether it is safe to assume that the probability of missingness is unrelated to unobserved concurrent outcomes (conditional on all observed outcomes). For the NLSY wage data, we can invent two plausible scenarios that undermine this assumption: If men are less likely to be interviewed at a particular wave if, at that time, they are earning especially: (1) *high* wages—because they might be less willing to take the time off from work to participate; or (2) *low* wages—because they might be less willing to reveal these low values to an interviewer. Because current wages (even unobserved) are strongly correlated with past and future wages, however, these risks are likely minimal. We therefore conclude that they are unlikely to be a major source of missingness for these data, supporting the credibility of the MAR assumption.[1]

If you cannot invoke one of these three missingness assumptions, you will need to add corrections to the multilevel model for change. Two different strategies are currently used: *selection models* and *pattern mixture models*. Under the selection approach, you build one statistical model for the "complete" data and a second model for the selection process that gave rise to the missingness. Under the pattern mixture approach, you identify a small number of missingness patterns and then fit a multilevel model stratified by these patterns. For further information, we direct your attention to the excellent papers by Hedeker and Gibbons (1997), Little (1995), and Little and Yau (1998).

5.3 Time-Varying Predictors

A time-varying predictor is a variable whose *values* may differ over time. Unlike their time-invariant cousins, which record an individual's static status, time-varying predictors record an individual's potentially differing **(p.160)** status on each associated measurement occasion. Some time-varying predictors have values that change naturally; others have values that change by design.

In their four-year study of how teen employment affects the amount of time adolescents spend with their families, Shanahan, Elder, Burchinal, and Conger (1996) examined the effects of three time-varying predictors: (1) the average number of hours worked per week; (2) the total amount of money earned per year; and (3) whether earnings were used for nonleisure activities (e.g., schoolbooks or savings). At age 12½, the average adolescent spent 16.3 hours per week with his or her family; over time, this amount declined at an average annual rate of 1.2 hours per week. Teen employment had both positive and negative effects. Although teens who made more money experienced steeper declines than peers who made less, those who spent some earnings on nonleisure activities or who worked especially long hours spent *more* time, on average, with their families (although their rates of decline were no shallower). The authors conclude that: “adolescent work constitutes a potentially positive source of social development, although this depends on how its multiple dimensions—earnings, spending patterns, [and] hours …—fit with the adolescent’s broader life course” (p. 2198).

In this section, we demonstrate how you can include time-varying predictors in the multilevel model for change. We begin, in section 5.3.1, by showing how to parameterize, interpret, and graphically display a model that includes a time-varying predictor’s main effect. In section 5.3.2, we allow the *effect* of a time-varying predictor to vary over time. In section 5.3.3, we discuss how to recenter time-varying predictors so as to facilitate interpretation. We conclude, in section 5.3.4, with some words of caution. Having described the analytic opportunities that time-varying predictors afford, we raise complex conceptual issues that can compromise your ability to draw clear convincing conclusions.

5.3.1 Including the Main Effect of a Time-Varying Predictor

Conceptually, you need no special strategies to include the main effect of a time-varying predictor in a multilevel model for change. The key to understanding why this is so lies in the *structure* of the person-period data set. Because each predictor—whether time-invariant or time-varying—has its own value on *each* occasion, it matters little whether these values vary across each person’s multiple records. A time-invariant predictor’s values remain constant; a time-varying predictor’s values vary. There is nothing more complex to it than that.

(p.161)

Table 5.6: Excerpts from the person-period data set for the unemployment study

ID	MONTHS	CES-D	UNEMP
7589	1.3142	36	1
7589	5.0924	40	1
7589	11.7947	39	1
55697	1.3471	7	1
55697	5.7823	4	1
65641	0.3285	32	1

ID	MONTHS	CES-D	UNEMP
65641	4.1068	9	0
65641	10.9405	10	0
65441	1.0842	27	1
65441	4.6982	15	1
65441	11.2690	7	0
53782	0.4271	22	1
53782	4.2382	15	0
53782	11.0719	21	1

We illustrate the general approach using data from Ginexi and colleagues' (2000) study of the effects of unemployment on depressive symptoms (mentioned briefly in section 5.1). By recruiting 254 participants from local unemployment offices, the researchers were able to interview individuals soon after job loss (within the first 2 months). Follow-up interviews were conducted between 3 and 8 months and 10 and 16 months after job loss. Each time, participants completed the Center for Epidemiologic Studies' Depression (CES-D) scale (Radloff, 1977), which asks them to rate, on a four-point scale, the frequency with which they experience each of 20 depressive symptoms. CES-D scores can vary from a low of 0 for someone with no symptoms to a high of 80 for someone in serious distress.

Just over half the sample (n = 132) was unemployed at every interview. Others had a variety of re-employment patterns: 62 were always working after the first interview; 41 were still unemployed at the second interview but working by the third; 19 were working by the second interview but unemployed at the third. We investigate the effect of unemployment using the time-varying predictor, *UNEMP*. As shown in the person-period data set in table 5.6, *UNEMP* represents individual *i*'s unemployment status at each measurement occasion. Because subjects 7589 and 55697 were consistently unemployed, their values of *UNEMP* are consistently 1. Because the unemployment status of the remaining cases *changed*, their values of *UNEMP* change as well: subject 65641 was working at both the second and third interviews (pattern 1-0-0); subject 65441 was working by the third (pattern 1-1-0); and subject 53782 was working at the second **(p. 162)** interview but unemployed again by the third (pattern 1-0-1). For any individual, *UNEMP* can be either 0 or 1 at each measurement occasion except the first (because, by design, everyone was initially unemployed).

We begin, as usual, with an unconditional growth model without substantive predictors:

$$Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \varepsilon_{ij}$$
$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$
(5.3a) $$\pi_{1i} = \gamma_{10} + \zeta_{1i},$$
where
(5.3b)
$$\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right) \text{ and } \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right).$$

Model A of table 5.7 presents the results of fitting this model to data, where $TIME_{ij}$ indicates the number of months (to the nearest day) between the date of interview *j* for person *i* and his date of initial unemployment. On the first day of job loss ($TIME_{ij} = 0$), we estimate that the average person has a non-zero CES-D score of 17.67 ($p < .01$); overt ime, this level declines linearly at a rate of 0.42 per month ($p < .001$). The variance components for both initial status and rates of change are statistically significant, suggesting the wisdom of exploring the effects of person-specific predictors.

Using a Composite Specification

Because many respondents eventually find work, the unconditional growth model likely tells an incomplete story. If employment alleviates depressive symptoms, might the reemployment of half the sample explain the observed decline? If you exclusively use level-1/level-2 representations, you may have difficulty postulating a model that addresses this question. In particular, it may not be clear where—in which model—the time-varying predictor should appear. So far, person-specific variables have appeared in level-2 submodels as predictors of level-1 growth parameters. Although you might therefore conclude that substantive predictors must always appear at level-2, this conclusion would be incorrect!

The easiest way of understanding how to include a time-varying predictor is to use the composite specification of the multilevel model. It is not that we cannot include a time-varying predictor in a model written using a level-1/level-2 specification (we will soon show how to do so), but rather that it is easier to learn how these predictors' effects operate and what types of models you might fit, if you start here.

We begin with the composite specification for the unconditional **(p.163)**

Table 5.7: Results of fitting a taxonomy of multilevel models for change to the unemployment data (*n* = 254)

		Parameter	Model A	Model B	Model C	Model D
Fixed Effects						
Composite	Intercept	γ_{00}	17.6694**	12.6656***	9.6167***	11.2666***
model	(initial status)		(0.7756)	(1.2421)	(1.8893)	(0.7690)
	TIME	γ_{10}	−0.4220***	−0.2020*	0.1620	
	(rate of change)		(0.0830)	(0.0933)	(0.1937)	
	UNEMP	γ_{20}		5.1113***	8.5291***	6.8795***
				(0.9888)	(1.8779)	(0.9133)
	UNEMP by	γ_{30}			−0.4652*	−0.3254**
	TIME				(0.2172)	(0.1105)
Variance Components						
Level−1:	Within-person	σ_{ε}^{2}	68.85***	62.39***	62.03***	62.43***
Level−2:	In intercept	σ_{0}^{2}	86.85***	93.52***	93.71***	41 52***
	In rate of Change	σ_{1}^{2}	0.36*	0.46**	0.45**	—
	In *UNEMP*	σ_{2}^{2}	—	—	—	40.45*
	In *UNEMP* by *TIME*	σ_{3}^{2}	—	—	—	0.71**
Goodness-of-fit						

	Parameter	Model A	Model B	Model C	Model D
Deviance		5133.1	5107.6	5103.0	5093.6
AIC		5145.1	5121.6	5119.7	5113.6
BIC		5166.3	5146.4	5147.3	5148.9

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

These models predict depression scores (on the *CES-D*) in the months following unemployment as a function of the time-varying predictor *UNEMP*. Model A is an unconditional growth model (see equation 5.4). Model B adds the main effect of *UNEMP* as a fixed effect (see equation 5.5); Model C also adds the interaction between *UNEMP* and linear *TIME* (see equation 5.7). Model D allows *UNEMP* to have both fixed and random effects (see equation 5.10). Notice that we have changed the order in which the fixed effects appear to correspond to the composite specification of the model.

Note: Full ML, SAS Proc Mixed. Also note the models include all associated covariance parameters, which we do not display to conserve space.

growth model, formed by substituting the second and third equations in equation 5.3a into the first:

(5.4) $$Y_{ij} = [\gamma_{00} + \gamma_{10}TIME_{ij}] + [\zeta_{0i} + \zeta_{1i}TIME_{ij} + \varepsilon_{ij}].$$

As in chapter 4, we use brackets to distinguish the model's fixed and stochastic portions. Because the fixed portion in the first bracket resembles a standard regression model, we can add the main effect of the time-varying predictor, *UNEMP*, by writing: **(p.164)**

(5.5)
$$Y_{ij} = [\gamma_{00} + \gamma_{10}TIME_{ij} + \gamma_{20}UNEMP_{ij}] + [\zeta_{0i} + \zeta_{1i}TIME_{ij} + \varepsilon_{ij}].$$

The two subscripts on *UNEMP* signify its time-varying nature. In writing equation 5.5, we assume that individual *i*'s value of *Y* at time *j* depends upon: (1) the number of months of since job loss (*TIME*); (2) his or her contemporaneous value of *UNEMP*; and (3) three person-specific residuals, ζ_{0i}, ζ_{1i}, and ε_{ij}.

What does this model imply about the time-varying predictor's main effect? Because the fixed effects, the γ's, are essentially regression parameters, we can interpret them using standard conventions:

- γ_{10} is the population average monthly rate of change in CES-D scores, controlling for unemployment status.
- γ_{20} is the population average difference, over time, in CES-D scores between the unemployed and employed.

The intercept, γ_{00}, refers to a logical impossibility: someone who is employed (*UNEMP* = 0) on the first day of job loss (*TIME* = 0). As in regular regression, an intercept can fall outside the range of the data (or theoretical possibility) without undermining the validity of the remaining parameters.

We can delve further into the model's assumptions by examining figure 5.3, which presents four average population trajectories implied by the model. As in figure 3.4, we obtained these trajectories by substituting in specific values for the substantive predictor(s). But because *UNEMP* is time-varying, we substitute in *time-varying patterns* not constant values. Since everyone was initially unemployed, *UNEMP* can take on one of four distinct patterns: (1) 1 1 1, for someone consistently unemployed; (2) 1 0 0, for someone who soon finds a job and remains employed; (3) 1 1 0, for someone who remains unemployed for a while but eventually finds a job; and (4) 1 0 1, for someone who soon finds a job only to lose it. Each pattern yields a different population trajectory, as shown in figure 5.3.

The unbroken trajectory in the upper left panel represents the predicted change in depressive symptoms for people who remain unemployed during the study. Because their values of *UNEMP* do not change, their implied average trajectory is linear. In displaying this single line, we do not mean to suggest that everyone who is consistently unemployed follows this line. The person-specific residuals, ζ_{0i} and ζ_{1i}, allow different individuals to have unique intercepts and slopes. But every true trajectory for someone who is consistently unemployed is linear, regardless of its level or slope.

The remaining trajectories in figure 5.3 reflect different patterns of temporal variation in *UNEMP*. Unlike the population trajectories in previous **(p.165)**

(p.166) chapters, these are *discontinuous*. Discontinuity is a direct consequence of *UNEMP*'s dichotomous time-varying nature. The upper right panel, for the 1 0 0 pattern, presents a hypothesized population trajectory for someone who finds a job at 5 months and remain employed. The lower left panel, for the 1 1 0 pattern, presents a hypothesized trajectory for someone who finds a job at 10 months and remains employed. The lower right panel, for the 1 0 1 pattern, presents a hypothesized trajectory for someone who finds a job at 5 months only to lose it at 10.

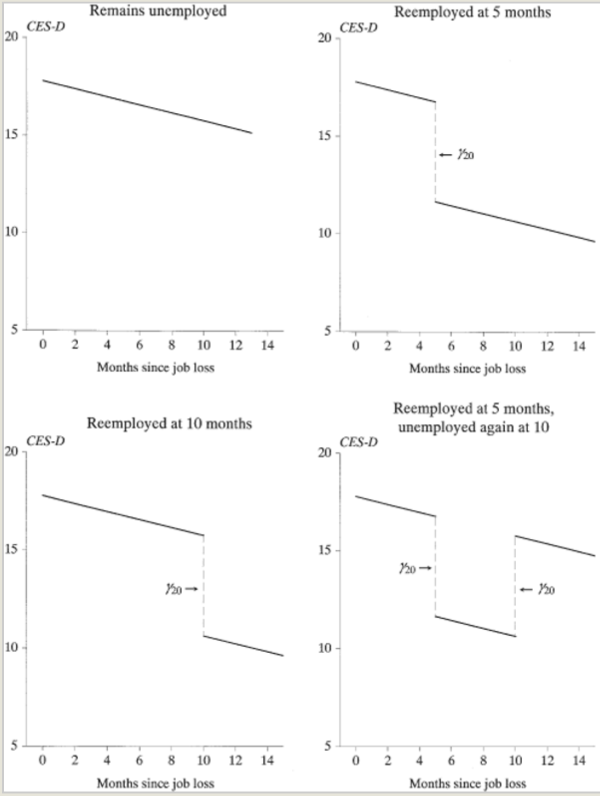

Figure 5.3. Identifying a suitable level-1 model for a time-varying predictor. Four average population trajectories implied by equation 5.5 for the effects of time-varying unemployment (*UNEMP*) on CES-D scores. In each panel, the magnitude of the effect of unemployment remains constant (at γ_{20}), but because *UNEMP* is time-varying, the model implies different population average trajectories corresponding to alternative patterns of unemployment and reemployment.

In offering these hypothetical trajectories, we must mention two caveats. First, although we link the upper and lower segments in each panel using dashed lines, our model implies only the solid portions. We use the dashed lines to emphasize that a change in unemployment status is associated with a switch in trajectory. Second, these few trajectories are not the only ones implied by the model. As in the first panel, person-specific residuals—ζ_{0i} and ζ_{1i}—suggest the existence of many other discontinuous trajectories, each with its own intercept and slope. But because the model constrains the effect of *UNEMP* to be constant, the *gap* between trajectories—for any individual—will be identical, at γ_{20}, the parameter associated with *UNEMP*. (We relax this assumption in section 5.3.2.)

Model B of table 5.7 presents the results of fitting this model to data. The parameter estimate for *TIME*, $\hat{\gamma}_{10}$, suggests that the monthly rate of decline in CES-D, while still statistically significant, has been cut in half (to 0.20 from 0.42 in Model A). This suggests that reemployment explains some of the observed decline in CES-D scores. This conclusion is reinforced by: (1) the large statistically significant effect of *UNEMP*—the average CES-D score is 5.11 points higher ($p < .001$) among the unemployed; and (2) the poorer fit of Model A in comparison to Model B—the difference in deviance statistics is 25.5 on the addition of one parameter ($p < .001$) and the AIC and BIC statistics are much lower as well. (We discuss the variance components later in this section.)

The left panel of figure 5.4 displays prototypical trajectories for Model B. Rather than present many different discontinuous trajectories reflecting the wide variety of transition times for *UNEMP*, we present just two continuous trajectories: the upper one for someone consistently unemployed; the lower one for someone consistently employed after 3.5 months. Displaying only two trajectories reduces clutter and highlights the most extreme contrasts possible. Because of this study’s design, we start the fitted trajectory for *UNEMP* = 0 at 3.5 months, the earliest time when a participant could be interviewed while working. To illustrate what would happen were we to extrapolate this trajectory back to *TIME* = 0, we include the dashed line. Because the model includes only the main effect of *UNEMP*, the two fitted trajectories are constrained to be parallel.

(p.167)

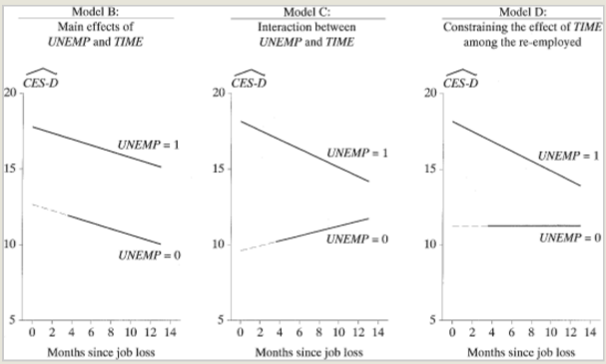

Figure 5.4. Displaying the results of fitted multilevel models for change that include a time-varying predictor. Prototypical trajectories from three models presented in table 5.7: Model B—the main effect of *UNEMP* and *TIME*, Model C—the interaction between *UNEMP* and *TIME*, and Model D—which

How do these two fitted trajectories display the main effect of unemployment status in Model B? Had the study followed just two static groups—the consistently unemployed and the consistently employed—these two trajectories would be the only ones implied by the model. But because *UNEMP* is time-varying, Model B implies the existence of many more depression trajectories, one for each possible *pattern* of unemployment/employment. Where are these additional trajectories? We find it helpful to think of the extremes shown as a conceptual *envelope* encompassing all discontinuous trajectories implied by the model. If *UNEMP* remains constant, an individual stays on one depression trajectory; if *UNEMP* changes, an individual shifts trajectories. As everyone in this study is unemployed at the first interview, everyone begins on the top trajectory. Those who find new jobs drop to the lower trajectory. Those who remain employed stay there. Those who lose their new jobs return to the upper trajectory. Conceptually, envision many dashed vertical lines running from the upper trajectory to the bottom (and back again) for individuals who change employment status. The set of these trajectories, which fall within the envelope shown, represent the complete set of prototypes implied by the model.

constrains the effect of *TIME* to be 0 among the reemployed.

(p.168) ***Using a Level-1/Level-2 Specification***

Having included a time-varying predictor under the composite specification, we now show how you can specify the identical model using a level-1/level-2 specification. This representation provides further insight into how time-varying predictors' effects operate; it also allows you to include time-varying predictors using software packages (e.g., HLM) that require a level-1/level-2 specification of the multilevel model for change.

To derive the level-1/level-2 specification that corresponds to a given composite specification, you proceed backwards. In other words, just as we can substitute level-2 submodels into a level-1 submodel to form a composite specification, so, too, can we *decompose* a composite model into its constituent level-1 and level-2 parts. Because the time-specific subscript *j* can appear only in a level-1 model, all time-varying predictors must appear in at level-1. We therefore write the level-1 submodel for the composite main effects model in equation 5.5 as:

(5.6a) $$Y_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \pi_{2i}UNEMP_{ij} + \varepsilon_{ij}.$$

Person-specific predictors that vary over time appear at level-1, not level-2. If you have no time-invariant predictors, as here, the accompanying level-2 models are brief:

$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$
$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$
(5.6b) $$\pi_{2i} = \gamma_{20}.$$

You can verify that substituting these level-2 models into the level-1 model in equation 5.6a yields the composite specification in equation 5.5. To add the effects of time-invariant predictors, you include them, as usual, in the level-2 submodels.

Notice that the third equation in equation 5.6b, for π_{2i}, the parameter for *UNEMP*, includes no level-2 residual. All the multilevel models fit so far have invoked a similar constraint—that the effect of a person-specific predictor is constant across population members. Time-invariant predictors require this assumption because they have no *within*-person variation to allow for a level-2 residual. But for time-varying predictors we could easily modify the last model in equation 5.6b to be:

(5.6c) $$\pi_{2i} = \gamma_{20} + \zeta_{2i}.$$

This allows the effect of *UNEMP* to vary randomly across individuals in the population. Adding this residual relaxes the assumption that **(p.169)** the gap between postulated trajectories in figure 5.3 is constant. To fit the new model to data, we revise the distributional assumptions for the residuals as presented in equation 5.3b. Commonly, we expand the assumption of multivariate normality to include all three level-2 residuals:

(5.6d) $$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ and } \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \\ \zeta_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix} \right).$$

Notice that in adding one extra residual, ζ_{2i} we add three extra variance components: σ_2^2, σ_{20} and σ_{21}.

Just because we *can* add these terms to our model does not mean that we should. Before doing so, we must decide whether the additional parameters are: (1) necessary; and (2) estimable using the available data. To address the first issue, consider whether the effect of employment on CES-D scores, controlling for time, *should* vary randomly across individuals. Before answering yes, remember that we are talking about *random* variation. If we expect the effect of unemployment to vary *systematically* across people, we can add substantive predictors that reflect this hypothesis. The question here is whether we should go further and add a residual that allows the effect of *UNEMP* to vary randomly. To be sure, much of our caution stems from concerns about the second point—the ability to estimate the additional parameters. With three (and sometimes fewer) measurement occasions per person, we often lack sufficient data to estimate additional variance components. Indeed, if we attempt to fit this more elaborate model, we encounter boundary constraints (as described in section 5.2.2). We therefore suggest that you resist the temptation to automatically allow the effects of time-varying predictors to vary at level-2 unless you have good reason, and sufficient data, to do so. (We will soon do so in section 5.3.2.)

As your models become more complex, we offer some practical advice (born of the consequences of the failure to follow it). When including time-varying predictors, we suggest that you write out the entire model before specifying your choice to a computer package. We suggest this extra step because it is not always obvious which random effects to include. In equation 5.6b, for example, the level-2 submodels require the first two parameters to be random and the third to be fixed. In other words, to fit this model you must use what appears to be an *inconsistent* set of level-2 submodels. As in many aspects of longitudinal analysis, the default or “standard” specifications may not yield the model *you* want to fit.

(p.170) ***Time-Varying Predictors and Variance Components***

In section 4.5.2, we discussed how the magnitude of variance components generally change on the inclusion of time-invariant predictors: (1) the level-1 variance component, σ_ε^2, remains relatively stable because time-invariant predictors cannot explain much within-person variation; but (2) the level-2 variance components, σ_0^2 and σ_1^2, will decline if the time-invariant predictors "explain" some of the between-person variation in initial status or rates of change, respectively. Time-varying predictors, in contrast, can affect all three variance components because they vary both within- *and* between-persons. And although you can interpret a decrease in the magnitude of the level-1 variance component, changes in level-2 variance components may not be meaningful, as we now show.

The general principles can be illustrated simply using Models A and B in table 5.7. Adding *UNEMP* to the unconditional growth model (Model A) reduces the magnitude of the within-person variance component, σ_ε^2, by 9.4% (from 68.85 to 62.39). Using strategies from section 4.4.3, equation 4.13, we conclude that time-varying unemployment status explains just over 9% of the variation in CES-D scores. This interpretation is straightforward because the time-varying predictor is added to the level-1 model, reducing the magnitude of the level-1 residual, ε_{ij}.

But ascribing meaning to observed changes in the level-2 variance components σ_0^2 and σ_1^2 can be nearly impossible. As we move from Model A to B both estimates *increase*! Although we alluded to this possibility in section 4.4.3, this is first example in which we observe such a pattern. The explanation for this seeming paradox—that changes in level-2 variance components do not assess the effects of time-varying predictors—lies in the associated level-1 submodel. When you add a time-varying predictor, as either a main effect or an interaction, you *change* the meaning of the individual growth parameters because:

- The intercept parameter, π_{0i}, now refers to the value of the outcome when *all* level-1 predictors, not only *TIME* but also the time-varying predictor, are zero.
- The slope parameter, π_{1i}, is now a *conditional* rate of change, controlling for the effects of the time-varying predictor.

Altering the population quantity that each parameter represents alters the meaning of the associated level-2 variance component. Hence, it makes no sense to compare the magnitude of these variance components across successive models.

This means that you must rely on changes in the time-varying predictors fixed effects, and associated goodness-of-fit statistics, when deciding **(p.171)** whether to retain a time-varying predictor in your model. As tempting as it is to compute the percentage reduction in a variance component associated with the inclusion of a time-varying predictor, there is no consistently meaningful way of doing so.

5.3.2 Allowing the Effect of a Time-Varying Predictor to Vary over Time

Might unemployment status also affect the trajectory's slope? In previous chapters, we initially associated predictors with both initial status *and* rates of change. Yet because Model B includes only the *main* effects of *TIME* and *UNEMP*, the trajectories are constrained to be parallel.

There are many ways to specify a model in which the trajectories' slopes vary by unemployment status. The easiest approach, and the one we suggest you begin with, is to add the cross-product—here, between *UNEMP* and *TIME*—to the main effects model:

(5.7)
$$Y_{ij} = [\gamma_{00} + \gamma_{10} TIME_{ij} + \gamma_{20} UNEMP_{ij} + \gamma_{30} UNEMP_{ij} \times TIME_{ij}] \\ + [\zeta_{0i} + \zeta_{1i} TIME_{ij} + \varepsilon_{ij}].$$

Notice the close resemblance between this and the composite model that includes an interaction between a time-*invariant* predictor and *TIME* (shown in equation 4.3). The differences between the two are purely cosmetic: (1) the substantive predictor (here *UNEMP* and there *COA*) has an additional subscript *j* to indicate that it is time-varying; and (2) different subscripts reference the relevant fixed effects (the γ's).

Model C of table 5.7 presents the results of fitting this model to data. The interaction between *TIME* and *UNEMP* is statistically significant ($\hat{\gamma}_{30} = -0.46,\ p < .05$). As with all interactions, we can interpret this effect in two ways: (1) the effect of unemployment status on CES-D scores varies over time; and (2) the rate of change in CES-D scores over time differs by unemployment status. Rather than delve into these interpretations, we draw your attention to the prototypical trajectories for this model displayed in the middle panel of figure 5.4. Here we find an unexpected pattern: while CES-D scores decline among the unemployed, the *reverse* is found among the re-employed—their CES-D scores appear to increase! The parameter estimate for the main effect of *TIME*, $\hat{\gamma}_{10}$ suggests why we observe this anomaly—it is not statistically significant (it is even smaller than its standard error, 0.19). Although we estimate a non-zero rate of change among the re-employed, we might have obtained this estimate even if the true rate of change in the population was zero.

This suggests that it might be wise to constrain the trajectory among the re-employed to be flat, with a slope of 0, while allowing the trajectory **(p.172)** among the unemployed to decline over time. Were we fitting a standard regression model, we might achieve this goal by removing the main effect of *TIME*:

(5.8)
$$Y_{ij} = [\gamma_{00} + \gamma_{20}UNEMP_{ij} + \gamma_{30}UNEMP_{ij} \times TIME_{ij}] + [\zeta_{0i} + \zeta_{1i}TIME_{ij} + \varepsilon_{ij}].$$

Had we fit this model to data and obtained fitted trajectories by unemployment status we would find: when *UNEMP* = 0, $\hat{Y}_{ij} = \hat{\gamma}_{00}$, when *UNEMP* = 1, $\hat{Y}_{ij} = (\hat{\gamma}_{00} + \hat{\gamma}_{20}) + \hat{\gamma}_{30}TIME_{ij}$.

This model's structural portion yields trajectories with the desired properties: (1) for the employed, we would have a flat line at level $\hat{\gamma}_{00}$; and (2) for the unemployed, we would have a slanted line, with intercept $\hat{\gamma}_{00} + \hat{\gamma}_{20}$ and slope $\hat{\gamma}_{30}$.

We do not fit this model, however, because of the lack of congruence between its structural and stochastic portions. Comparing the elements in the two sets of brackets in equation 5.8, notice that the model includes: (1) a random effect for *TIME*, ζ_{1i}, but no corresponding main effect (we removed γ_{10} from the model when we removed the main effect of *TIME*); and (2) a fixed effect for the *UNEMP* by *TIME* interaction (γ_{30}) and no corresponding random effect. We therefore postulate an alternative model in which the fixed and random effects are better aligned:

$$\begin{aligned} Y_{ij} &= [\gamma_{00} + \gamma_{20}UNEMP_{ij} + \gamma_{30}UNEMP_{ij} \times TIME_{ij}] \\ &\quad + [\zeta_{0i} + \zeta_{3i}UNEMP_{ij} \times TIME_{ij} + \varepsilon_{ij}]. \end{aligned}$$
(5.9)

Notice that the interaction term, *UNEMP* by *TIME*, appears as both a fixed and a random effect. But when we attempt to fit the model in equation 5.9 to data, we find that its AIC and BIC statistics are larger (worse) than that of Model C (we cannot conduct a formal test because this model is not fully nested within the other, nor do we present the results in table 5.7).

It might appear, then, that Model C is preferable. But before reaching this conclusion, we revisit a question raised in the previous section: Should the effect of *UNEMP* be constant across the population? When we previously attempted to allow this effect to vary randomly (by augmenting Model B, which included the main effect of *TIME*) we could not fit the model to data. But having constrained the model's structural portion so that the trajectory among the re-employed is flat, we notice an inconsistency in equation 5.8: it allows the intercept among the employed, γ_{00}, to vary randomly (through the inclusion of the residual, ζ_{0i}) but not the increment to this intercept associated with unemployment, γ_{20} (there is no corresponding residual, ζ_{2i}). Why should we allow the flat level of the trajectory among the re-employed to vary and *not* **(p.173)** allow the increment to this flat level (which yields the the intercept among the unemployed) to vary randomly as well? Perhaps the fit of the model in equation 5.9 is poorer than Model C because of this unrealistically stringent constraint on the random effects.

We address this supposition by fitting Model D:

$$Y_{ij} = [\gamma_{00} + \gamma_{20}UNEMP_{ij} + \gamma_{30}UNEMP_{ij} \times TIME_{ij}]$$
$$+ [\zeta_{0i} + \zeta_{2i}UNEMP_{ij} + \zeta_{3i}UNEMP_{ij} \times TIME_{ij} + \varepsilon_{ij}], \quad (5.10)$$

which allows each fixed effect to have an associated random effect. The results of fitting this model are shown in the final column of table 5.7 and are graphed in the right panel of figure 5.4. Immediately upon layoff, the average unemployed person in the population has a CES-D score of 18.15 (=11.27 + 6.88). Over time, as they acclimate to their new status, the average unemployed person's CES-D scores decline at a rate of –0.33 per month ($p < .01$). CES-D scores among those who find a job are lower (by as much as 6.88 if the job is found immediately after layoff or as little as 2.97 if 12 months later (14.24 – 11.27). Once a formerly unemployed individual finds a job and keeps it, we find no evidence of systematic change in CES-D scores over time. We believe that this model provides a more realistic representation of the patterns of change in CES-D scores over time than Model C. Not only is it substantively compelling, its AIC statistic is superior (and its BIC nearly equivalent) even though it includes several additional parameters (the extra variance components shown in table 5.7 as well as the extra covariance components not shown).

We hope that this example illustrates how you can test important hypotheses about time-varying predictors' effects and investigate even more ways in which outcomes might change over time (here, how CES-D scores change not just with time but also re-employment). As we will show in chapter 6, the ability to include time-varying predictors opens up a world of analytic opportunities. Not only can level-1 individual growth models be smooth and linear, they can also be discontinuous and curvilinear. This allows us to postulate and fit level-1 submodels that better reflect our hypotheses about the population processes that give rise to sample data and assess the tenability of such hypotheses with data. But to adequately build a foundation for pursuing those types of analyses, we must consider other issues that arise when working with time-varying predictors, and we do so by beginning with issues of centering.

5.3.3 Recentering Time-Varying Predictors

In chapter 4, when discussing interpretation of parameters associated with time-invariant predictors, we introduced the practice of recentering: **(p.174)** subtracting a constant from a predictor's values to alter its parameter's meaning. In some analyses, we subtracted a predictor's overall sample mean (known as *grand-mean centering*); in others, we subtracted a substantively interesting value (such as 9 for *highest grade completed*). We now describe similar strategies you can use with time-varying predictors.

To concretize the discussion, let us return to the wage data for high school dropouts summarized in table 5.4. We can express Model C in composite form by writing: $Y_{ij} = [\gamma_{00} + \gamma_{10}\,TIME_{ij} + \gamma_{01}(HGC_i - 9) + \gamma_{12}\,BLACK_i \times TIME_{ij}] + [\zeta_{0i} + \zeta_{1i}\,TIME_{ij} + \varepsilon_{ij}]$. As did the original researchers, we now introduce the possibility that wages might be affected by a time-varying predictor, *UERATE*, the unemployment rate in the local geographic area:

(5.11)

$$Y_{ij} = [\gamma_{00} + \gamma_{10}TIME_{ij} + \gamma_{01}(HGC_i - 9) + \gamma_{12}BLACK_i \times TIME_{ij} + \gamma_{20}UERATE_{ij}] + [\zeta_{0i} + \zeta_{1i}TIME_{ij} + \varepsilon_{ij}].$$

We restrict attention to the main effect of *UERATE* because extensive analysis suggests that its effect on log wages does not vary over time.

Adapting recentering strategies outlined in section 4.5.4 for time-invariant predictors, we could include *UERATE* in several different ways, each using one of the following:

- Its raw values
- Deviations around its grand mean *in the person-period data set* (7.73)
- Deviations from another meaningful constant (say, 6, 7 or 8, common unemployment rates during the time period under study)

Each strategy would lead to virtually identical conclusions. Were we to fit the model in equation 5.11 using each, we would find identical parameter estimates, standard errors, and goodness-of-fit statistics *with just one exception*: for the intercept, γ_{00}. Inspecting equation 5.11 clarifies why this is so. As in regression, adding a main effect does not alter the meaning of the model's remaining parameters. If *UERATE* is expressed on its raw scale, γ_{00} estimates the average log wage on the first day of work (*EXPER* = 0) for a black male who dropped out in ninth grade (*HGC* – 9 = 0) and who lives in an area with no unemployment (*UERATE* = 0). If *UERATE* is grand-mean centered, γ_{00} estimates the average log-wage for a comparable male who lives in an area with an "*average*" unemployment rate. But because this "average" would be computed in the person-period data set, in which both the measurement occasions and number of waves vary across people, it may not be particularly meaningful.

(p.175)

Table 5.8: Results of adding three alternative representations of the time-varying predictor for local area unemployment rate (*UERATE*) to Model C of table 5.4 for the high school dropout wage data (*n* = 888)

		Parameter	Model A: centered at 7	Model B: within person centering	Model C: time-1 centered
Fixed Effects					
Initial status, π_{0i}	Intercept	γ_{00}	1.7490***	1.8743***	1.8693***
			(0.0114)	(0.0295)	(0.0260)
	(*HGC* – 9)	γ_{01}	0.0400***	0.0402***	0.0399***
			(0.0064)	(0.0064)	(0.0064)
	UERATE	γ_{20}	−0.0120***	−0.0177***	−0.0162***
			(0.0018)	(0.0035)	(0.0027)
	Deviation of	γ_{30}		−0.0099***	−0.0103***
	UERATE from			(0.0021)	(0.0019)
	centering value				
Rate of change, π_{1i}	Intercept	γ_{10}	0.0441***	0.0451*	0.0448***
			(0.0026)	(0.0027)	(0.0026)
	BLACK	γ_{12}	−0.0182***	−0.0189***	−0.0183***
			(0.0045)	(0.0045)	(0.0045)
Variance Components					
Level−1:	within-person	σ_{ε}^{2}	0.0948***	0.0948***	0.0948***

		Parameter	Model A: centered at 7	Model B: within person centering	Model C: time-1 centered
Level−2:	In initial status	σ_0^2	0.0506***	0.0510***	0.0503***
	In rate of change	σ_1^2	0.0016***	0.0016***	0.0016***
Goodness-of-fit					
	Deviance		4830.5	4827.0	4825.8
	AIC		4848.5	4847.0	4845.8
	BIC		4891.6	4894.9	4893.7

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

Model A adds (*UERATE* – 7); Model B centers *UERATE* at each person's mean; Model C centers *UERATE* around each person's value of *UERATE* at his first measurement occasion.

Note: SAS Proc Mixed, Full ML. Also note that the covariance component, σ_{01}, is estimated, but not displayed.

We therefore often prefer recentering time-varying predictors *not* around the grand-mean but rather around a substantively meaningful constant—here, say 7. This allows γ_{00} to describe the average log-wage for someone whose local area has a 7% unemployment rate. The results of fitting this last model appear in the first column of table 5.8. As in section 5.2.1, we can interpret this parameter estimate by computing $100(e^{(-0.0120)} - 1) = -1.2$. We conclude that each one-percentage point difference in local area unemployment rate is associated with wages that are 1.2 percent lower.

(p.176) Given that centering has so little effect on model interpretation, you may wonder why we raise this issue. We do so for three reasons: (1) the topic receives much attention in the multilevel literature (see, e.g., Kreft et al., 1995; Hofmann & Gavin, 1998); (2) some computer programs tempt analysts into recentering their predictors through the availability of simple toggle switches on an interactive menu; and (3) there are still other meaningful ways of recentering. Not only can you recenter around a *single* constant, you can recenter around *multiple constants*, one per person. It is this approach, also known as within-context or group-mean centering, to which we now turn.

The general idea behind within-context centering is simple: instead of representing a time-varying predictor using a single variable, decompose the predictor into multiple constituent variables, which, taken together, separately identify specific sources of variation in the outcome. Of the many ways of decomposing a time-varying predictor, two deserve special mention:

- *Within-person centering*: include the *average* unemployment rate for individual *i*, $\overline{UERATE}_{i0}$, as well as the deviation of each period's rate from this average, $(UERATE_{ij} - \overline{UERATE}_{i0})$.
- *Time-1 centering*: include *time-1*'s unemployment rate for individual *i*, $UERATE_{i1}$, as well as the deviation of each subsequent rate from this original value, $(UERATE_{ij} - UERATE_{i1})$.

Within-context centering provides *multiple* ways of representing a time varying predictor. Under within-person centering, you include a time-invariant *average* value and deviations from that average; under time-1 centering, you include the time-invariant *initial* value and deviations from that starting point. In both cases, as well as in the many other possible versions of within-context centering, the goal is to represent the predictor in a way that provides greater insight into its effects. (Of course, within-person centering raises interpretive problems of endogeneity, discussed in the following section.)

The last two columns of table 5.8 present the results of fitting the multilevel model for change with *UERATE* centered within-person (Model B) and around time-1 (Model C). Each contributes a particular insight into the negative effect of local unemployment on dropouts' wages. Model B reveals an association between wages and two aspects of the unemployment: (1) its average over time

—the lower the average rate, the lower the wage; and (2) its relative magnitude, at each point in time, in comparison to this average. Model C demonstrates that wages are also associated with two other aspects of the time-varying unemployment rates: (1) their *initial* value, when the dropout first enters the labor force; and (2) the **(p.177)** *increment or decrement*, at each subsequent point in time, from that initial value. Is either of these centered options clearly superior to the raw variable representation? Given that we cannot compare deviance statistics (because no model is nested within any other), comparison of AIC and BIC statistics suggests that all three are roughly comparable, with BIC giving the nod to Model A and AIC the nod to Model C.

These strategies for representing the effect of a time-varying predictor are hardly the only options. We offer them primarily in the hope that they will stimulate your thinking about substantively interesting ways of representing predictors' effects. We find routine recommendations to always, or never, center unconstructive. We prefer instead to recommend that you think carefully about which representations might provide the greatest insight into the phenomenon you are studying.

5.3.4 An Important Caveat: The Problem of Reciprocal Causation

Most researchers get very excited by the possibility that a statistical model could represent the relationship between changing characteristics of individuals and their environments, on the one hand, and individual outcomes on the other. We now dampen this enthusiasm by highlighting interpretive difficulties that time-varying predictors can present. The problem, known generally as *reciprocal causation* or *endogeneity*, is the familiar "chicken and egg" cliché: if X is correlated with Y, can you conclude that X *causes* Y or is it possible that Y *causes* X?

Many, but not all, time-varying predictors are subject to these problems. To help identify which are most susceptible, we classify time-varying predictors into four groups: *defined, ancillary, contextual*, and *internal*[2]. In the context of individual growth modeling, classification is based on the degree to which a predictor's values at time t_{ij} are: (1) assignable a priori; and (2) potentially influenced by the study participant's contemporaneous outcome. The more "control" a study participant has over his or her predictor values, the more clouded your inferences.

A time-varying predictor is *defined* if, in advance of data collection, its values are predetermined for everyone under study. Defined predictors are impervious to issues of reciprocal causation because no one—not the study participants nor the researchers—can alter their values. Most defined predictors are themselves functions of time. All representations of *TIME* are defined because their values depend solely on a record's time-period. Time-varying predictors that reflect other periodic aspects of time—such as season (fall, summer, etc) or anniversary

(anniversary month, nonanniversary month)—are defined because once the metric **(p.178)** for time is chosen, so, too, are their values. Predictors whose values are set by an external schedule are also defined. If Ginexi and colleagues (2000) added a variable representing each person’s time-varying unemployment benefits, its values would be defined because payments reflect a uniform schedule. Similarly, when comparing the efficacy of time-varying drug tapering regimens in a randomized trial, an individual’s dosage is defined if the researcher determines the entire dosing schedule *a priori*. Different people may take different doses at different times, but if the schedule is predetermined, the predictor is defined.

A time-varying predictor is *ancillary* if its values cannot be influenced by study participants because they are determined by a stochastic process totally external to them. We use the term “stochastic process” to emphasize that, unlike a defined predictor, an ancillary predictor can behave erratically over time. Ancillary predictors are impervious to issues of reciprocal causation because no one involved in the study directly affects their values. Most ancillary predictors assess potentially changing characteristics of the physical or social environment in which respondents live. In his study of marital dissolution, for example, South (1995) divided the United States into 382 local marriage markets and used census data to create a time-varying predictor assessing the availability of spousal alternatives in each market. His *availability index* contrasted the number of unmarried persons “locally available” to the respondent with the number of unmarried persons “locally available” to the respondent’s spouse. As no respondent could be part of the local marriage market (because all were married), this predictor is ancillary. If some *were* part of the local market (as they would be in a study of marital *initiation*), this predictor would be approximately ancillary because: (1) the contribution of any individual to the index would be negligible (given that the smallest marriage market included over a half million people); and (2) few individuals move to a particular area because of the availability of spousal alternatives. Following this logic, the local area unemployment rate just used in the high school dropout wage analysis is approximately ancillary. Other ancillary predictors include weather (Young, Meaden, Fogg, Cherin, & Eastman, 1997) and treatment, if randomly assigned.

A *contextual* time-varying predictor also describes an “external” stochastic process, but the connection between units is closer—between husbands and wives, parents and children, teachers and students, employers and employees. Because of this proximity, contextual predictors can be influenced by an individual’s contemporaneous outcome values; if so, they are susceptible to issues of reciprocal causation. To assess whether reciprocal causation is a problem, you must analyze the particular situation. For example, in their 30-year study of the effects of parental divorce **(p.179)** on mental health, Cherlin, Chase-Lansdale and McRae (1998) included time-varying predictors denoting whether children had experienced a parental divorce during four developmental phases: 7–10, 11–15, 16–22, and 23–33. These contextual time-varying predictors are unlikely to create interpretive problems because it is doubtful that someone's level of emotional problems would influence either the occurrence or the timing of a parental divorce. But in their three-year study of the link between the quality of childcare centers and children's early cognitive and language development, Burchinal et al. (2000) face a thornier problem. Because parents may *choose* particular childcare centers precisely because they emphasize particular skills, observed links between center quality and child development may be due to a link between development and quality, not quality and development. If such criticisms seem reasonable, we suggest that you treat a contextual time-varying predictor as if it were internal, and address issues of reciprocal causation in ways we now describe.

Internal time-varying predictors describe an individual's potentially changeable status over time. Some describe *psychological* states (mood or satisfaction), while others describe *physical* states (respiratory function, blood levels), *social* states (married/unmarried, working/unemployed), or other personal attributes. In their four-year study of adolescent smoking, for example, Killen, Robinson, Haydel, et al. (1997) annually assessed dozens of internal predictors ranging from counts of the number of friends who smoke and the frequency of drinking to the adolescent's height and weight. And in their four-year study of conduct disorder in boys, Lahey, McBurnett, Loeber, & Hart (1995) collected annual data on receipt of various kinds of psychological treatment, both in-patient and out-patient, medication and talk therapy.

Internal time-varying predictors raise serious interpretive dilemmas. Isn't it reasonable to argue, for example, that as teens start smoking, they increase the number of friends who smoke, increase their frequency of drinking, and lose weight? So, too, isn't it possible that as a child's behavior worsens a parent may be more likely to initiate psychotherapy? Although the causal link may be from predictor to outcome, it may also run the opposite way. Some readers may believe that longitudinal data—and the associated statistical models—should resolve such concerns. But resolution of the directional arrow is more difficult. As long as a model links *contemporaneous* information about time-varying predictors and outcomes, we effectively convert a longitudinal problem into a cross-sectional one, fully burdened by questions of reciprocal causation.

Given the conceptual appeal of internal and contextual time-varying predictors, what should you do? We have two concrete recommendations. **(p.180)** First, use theory as a guide, play your own harshest critic, and determine whether your inferences are clouded by reciprocal causation. Second, if your data allow, consider coding time-varying predictors so that their values in each record in the person-period data set refer to a *previous* point in chronological time. After all, there is nothing about the multilevel model for change that requires contemporaneous data coding. Most researchers use contemporaneous values by default. Yet it is often more logical to link *prior* status on a predictor with current status on an outcome.

For example, in their study of conduct disorder (CD) in boys, Lahey and colleagues (1995) carefully describe three ways they coded the effect of time-varying predictors representing treatment:

> In each case, the treatment was considered to be present in a given year if that form of treatment had been provided during all or part of the *previous 12 months* (emphasis added).… In addition, the analyses of treatment were repeated using the cumulative number of years that the treatment had been received as the time-varying covariate to determine whether the accumulated number of years of treatment influenced the number of CD symptoms in each year. Finally, a 1-year time-lagged analysis was conducted to look at the effect of treatment on the number of CD symptoms in the following year. (p. 90)

By linking each year's outcomes to prior treatment data, the researchers diminish the possibility that their findings are clouded by reciprocal causation. So, too, by carefully describing several alternative coding strategies, each of which describes a predictor constructed from the prior year's data, the researchers appear more credible and thoughtful in their work.

How might we respond to questions about reciprocal causation in Ginexi and colleagues' (2000) study of the link between unemployment and depression? A critic might argue that individuals whose CES-D scores decline over time are more likely to find jobs than peers whose levels remain stable or perhaps increase. If so, the observed link between re-employment and CES-D scores might result from the effects of CES-D on employment, not employment on CES-D. To rebut this criticism, we emphasize that the re-employment predictor indicates whether the person is *currently* employed at each subsequent interview. As a result, the moment of re-employment is temporally prior to the collection of CES-D scores. This design feature helps ameliorate the possibility that the observed relationship between unemployment and depression is a result of reciprocal causation. Had the CES-D and re-employment data been collected simultaneously, it would have been more difficult to marshal this argument.

(p.181) Our message is simple: just because you can establish a link between a time-varying predictor and a time-varying outcome does not guarantee that the link is causal. While longitudinal data can help resolve issues of temporal ordering, the inclusion of a time-varying predictor can muddy the very issues the longitudinal models were intended to address. Moreover, as we will show in the second half of this book, issues of reciprocal causation can be even thornier when studying event occurrence because the links between outcomes and predictors are often more subtle than the examples just presented suggest. This is not to say you should not include time-varying predictors in your models. Rather, it is to say that you must recognize the issues that such predictors raise and not naively assume that longitudinal data alone will resolve the problem of reciprocal causation.

5.4 Recentering the Effect of *TIME*

TIME is the fundamental time-varying predictor. It therefore makes sense that if recentering a substantive time-varying predictor can produce interpretive advantages, so, too, should recentering *TIME*. In this section, we discuss an array of alternative recentering strategies, each yielding a different set of level-1 individual growth parameters designed to address related, but slightly different, research questions.

So far, we have tended to recenter *TIME* so that the level-1 intercept, π_{0i}, represents individual *i*'s true *initial status*. Of course, the moment corresponding to someone's "initial status" is context specific—it might be a particular chronological age in one study (e.g., age 3, 6.5, or 13) or the occurrence of a precipitating event in another (e.g., entry into or exit from the labor force). In selecting a sensible starting point, we seek an early moment, ideally during the period of data collection, inherently meaningful for the process under study. This strategy yields level-2 submodels in which all parameters are directly and intrinsically interpretable, and it ensures that the value of *TIME* associated with the intercept, π_{0i}, falls within *TIME*'s observed range. Not coincidentally, this approach also yields a level-1 submodel that reflects everyday intuition about intercepts as a trajectory's conceptual "starting point."

Although compelling, this approach is hardly sacrosanct. Once you are comfortable with model specification and parameter interpretation, a world of alternatives opens up. We illustrate some options using data from Tomarken, Shelton, Elkins, and Anderson's (1997) randomized trial evaluating the effectiveness of supplemental antidepressant medication for individuals with major depression. The study began with an overnight **(p.182)**

Table 5.9: Alternative coding strategies for *TIME* in the antidepressant trial

WAVE	DAY	READING	TIME OF DAY	TIME	(TIME – 3.33)	(TIME – 6.67)
1	0	8 A.M.	0.00	0.00	−3.33	−6.67
2	0	3 P.M.	0.33	0.33	−3.00	−6.33
3	0	10 P.M.	0.67	0.67	−2.67	−6.00
4	1	8 A.M.	0.00	1.00	−2.33	−5.67
5	1	3 P.M.	0.33	1.33	−2.00	−5.33
6	1	10 P.M.	0.67	1.67	−1.67	−5.00
…						
11	3	3 P.M.	0.33	3.33	0.00	−3.33
…						
16	5	8 A.M.	0.00	5.00	1.67	−1.67
17	5	3 P.M.	0.33	5.33	2.00	−1.33
18	5	10 P.M.	0.67	5.67	2.33	−1.00
19	6	8 A.M.	0.00	6.00	2.67	−0.67
20	6	3 P.M.	0.33	6.33	3.00	−0.33
21	6	10 P.M.	0.67	6.67	3.33	0.00

hospital stay for 73 men and women who were already being treated with a nonpharmacological therapy that included bouts of sleep deprivation. During the pre-intervention night, the researchers prevented each participant from obtaining any sleep. The next day, each person was sent home with a week's worth of pills (placebo or treatment), a package of mood diaries (which use a five-point scale to assess positive and negative moods), and an electronic pager. Three times a day—at 8 A.M., 3 P.M., and 10 P.M.—during the next month, respondents were electronically paged and reminded to fill out a mood diary. Here we analyze the first week's data, focusing on the participants' positive moods. With full compliance, each person would have 21 assessments. Although two people were recalcitrant (producing only 2 and 12 readings), everyone else was compliant, filling out at least 16 forms.

Table 5.9 presents seven variables that represent related, but distinct, ways of clocking time. The simplest, *WAVE*, counts from 1 to 21; although great for data processing, its cadence has little intuitive meaning because few of us divide our weeks into 21 conceptual components. *DAY*, although coarse, has great intuitive appeal, but it does not distinguish among morning, afternoon, and evening readings. One way to capture this finer information is to add a second temporal variable, such as *READING* or *TIME OF DAY*. Although the metric of the former makes it difficult to analyze, the metric of the latter is easily understood: 0 for morning readings; 0.33 for afternoon readings; 0.67 for evening readings. (We could **(p.183)** also use a 24-hour clock and assign values that were not equidistant.) Another way to distinguish within-day readings is to create a single variable that combines both aspects of time. The next three variables, *TIME, TIME* – 3.33, and *TIME* – 6.67, achieve this goal. The first, *TIME*, operates like our previous temporal variables—it is centered on initial status. The others are linear transformations of *TIME*: one centered on 3.33, the study's *midpoint*, and the other centered on 6.67, the study's *final* wave.

Having created these alternative variables, we could now specify a separate set of models for each. Instead of proceeding in this tedious fashion, let us write a general model that uses a generic temporal variable (*T*) whose values are centered around a generic constant (*c*):

(5.12a) $$Y_{ij} = \pi_{0i} + \pi_{1i}\left(T_{ij} - c\right) + \varepsilon_{ij}.$$

We can then write companion level-2 models for the effect of treatment:

$$\pi_{0i} = \gamma_{00} + \gamma_{01}TREAT_i + \zeta_{0i}$$

(5.12b) $$\pi_{1i} = \gamma_{10} + \gamma_{11}TREAT_i + \zeta_{1i}$$

and invoke standard normal theory assumptions for the residuals. This same model can be used for most of the temporal variables in table 5.9 (except those that distinguish only between within-day readings).

Table 5.10 presents the results of fitting this general model using the three different temporal variables, *TIME, TIME* – 3.33, and *TIME* – 6.67. Begin with the initial status representation of *TIME*. Because we cannot reject null hypotheses for either linear change or treatment, we conclude that: (1) on average, there is no linear trend in positive moods over time in the placebo group ($\hat{\gamma}_{10} = -2.42,\ n.s.$); and (2) when the study began, the groups were indistinguishable ($\hat{\gamma}_{01} = -3.11,\ n.s.$) as randomization would have us expect. The statistically significant coefficient for the effect of *TREAT* on linear change ($\hat{\gamma}_{11} = 5.54,\ p < 0.05$) indicates that the trajectories' slopes differ. The prototypical trajectories in figure 5.5 illustrate these findings. On average, the two groups are indistinguishable initially, but over time, the positive mood scores of the treatment group increase while those of the control group decline. The statistically significant variance components for the intercept $(\hat{\sigma}_0^2 = 2111.33,\ p < .001)$ and linear change $(\sigma_1^2 = 63.74,\ p < .001)$ indicate that that substantial variation in these parameters has yet to be explained.

What happens as we move the centering constant from 0 (initial status), to 3.33 (the study's midpoint), to 6.67 (the study's endpoint)? As expected, some estimates remain identical, while others change. The general principle is simple: parameters related to the *slope* remain stable while those related to the *intercept* differ. On the stable side, we obtain **(p.184)**

Table 5.10: Results of using alternative representations for the main effect of *TIME* when evaluating the effect of treatment on the positive mood scores in the antidepressant trial (*n* = 73)

			Temporal predictor in level−1 model		
		Parameter	TIME	(TIME – 3.33)	(TIME – 6.67)
Fixed Effects					
Level−1	Intercept	γ_{00}	167.46***	159.40***	151.34***
intercept, π_{0i}			(9.33)	(8.76)	(11.54)
	TREAT	γ_{01}	−3.11	15.35	33.80*
			(12.33)	(11.54)	(15.16)
Rate of	Intercept	γ_{10}	−2.42	−2.42	−2.42
Change, π_{1i}			(1.73)	(1.73)	(1.73)
	TREAT	γ_{11}	5.54*	5.54*	5.54*
			(2.28)	(2.28)	(2.28)
Variance Components					
Level−1:	within-person	σ_ε^2	1229.93***	1229.93***	1229.93***
Level−2:	In level-1 intercept	σ_0^2	2111.33***	2008.72***	3322.45***
	In rate of change	σ_1^2	63.74***	63.74***	63.74***
	Covariance	σ_{01}	−121.62*	90.83	303.28***
Goodness-of-fit					

	Temporal predictor in level−1 model		
Parameter	TIME	(TIME – 3.33)	(TIME – 6.67)
Deviance	12680.5	12680.5	12680.5
AIC	12696.5	12696.5	12696.5
BIC	12714.8	12714.8	12714.8

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

TIME is centered around initial status, middle status, and final status.

Note: Full ML, SAS PROC MIXED.

identical estimates for the linear rate of change in the placebo group ($\hat{\gamma}_{10} = -2.42,\ n.s.$) and the effect of treatment on that rate ($\hat{\gamma}_{11} = 5.54,\ p < 0.05$). So, too, we obtain identical estimates for the residual variance in the rate of change ($\hat{\sigma}_1^2 = 63.74,\ p < .001$) and the within-person residual variance ($\hat{\sigma}_\varepsilon^2 = 1229.93$). And, most important, the deviance, AIC and BIC statistics remain unchanged because these models *are* structurally identical.

Where these models differ is in the location of their trajectories' *anchors*, around their starting point, midpoint, or endpoint. Because the intercepts refer to these anchors, each model tests a different set of hypotheses about them. If we change *c*, we change the anchors, which changes the estimates and their interpretations. In terms of the general model in equations 5.12a and 5.12b, γ_{00} assesses the elevation of the population average change trajectory at time *c*; γ_{01} assesses the differential elevation of this trajectory at time *c* between groups; σ_0^2 assesses the population variance in true status at time *c*; and σ_{01} assesses the population **(p.185)** covariance between true status at time *c* and the per-unit rate of change in *Y*.

Although general statements like these are awkward, choice of a suitable centering constant can create simple, even elegant, interpretations. If we choose *c* to be 3.33, this study's midpoint, the intercept parameters assess effects at midweek. Because the treatment is still nonsignificant ($\hat{\gamma}_{01} = 15.35,\ n.s.$), we conclude that the average elevation of the two trajectories remains indistinguishable at this time. If we choose *c* to be 6.67, this study's endpoint, the intercept parameters assess effects at week's end. Doing so yields an important finding:

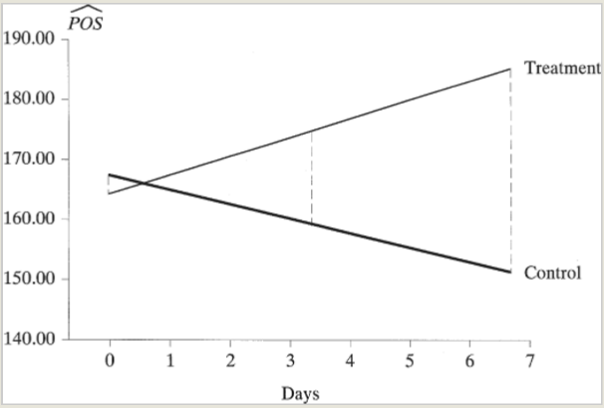

Figure 5.5. Understanding the consequences of rescaling the effect of *TIME*. Prototypical trajectories for individuals by *TREATMENT* status in the antidepressant experiment. The dashed vertical lines reflect the magnitude of the effect of *TREATMENT* if time is centered at the study's beginning (0), midpoint (3.33), and endpoint (6.67).

Instead of reinforcing the expected nonsignificant early differences between groups, we now find a statistically significant treatment effect ($\hat{\gamma}_{01} = 33.80,\ p < .05$). After a week of antidepressant therapy, the positive mood score for the average member of the treatment group differs from that of the average member of the control group.

How can changing the centering constant for *TIME* have such a profound impact, especially since the fundamental model is unchanged? The dashed vertical lines in the prototypical plots in figure 5.5 provide an explanation. In adopting a particular centering constant, we cause the resultant estimates to describe the trajectories' behavior at that specific point in time. Changing the trajectory's anchor changes the location of the focal comparison. Of course, you could conduct *post hoc* tests of these contrasts (using methods of section 4.7) and obtain identical results. But **(p.186)** when doing data analysis, it is sometimes easier to establish level-1 parameters that automatically yield readymade tests for hypotheses of greatest interest. We urge you to identify a scale for *TIME* that creates a level-1 submodel with directly interpretable parameters. Initial status often works well, but there are alternatives. The midpoint option is especially useful when *total study duration* has intrinsic meaning; the endpoint option is especially useful when *final status* is of special concern.

Statistical considerations can also suggest the need to recenter *TIME*. As shown in table 5.10, a change in center can change the interpretation, and hence values, of selected random effects. Of particular note is the effect that a recentering can have on σ_{01}, the covariance between a level-1 model's intercept and slope. Not only can a recentering affect this parameter's magnitude, it can also affect its sign. In these data, the covariance between intercept and slope parameters moves from – 121.62 to 90.83 to 303.28 as the centering constant changes. These covariances (and their associated variances) imply correlation coefficients of –0.33, 0.25, and 0.66, respectively. As you might imagine, were we to choose an even larger centering constant, outside the range of the data, it would be possible to find oneself specifying a model in which the correlation between parameters is close to 1.00. As Rogosa and Willett (1985) demonstrate, you can always alter the correlation between the level-1 growth parameters simply by changing the centering constant.

Understanding that the correlation between level-1 individual growth parameters can change through a change of centering constants has important analytic consequences. Recall that in section 5.2.2, we alluded to the possibility that you might encounter boundary constraints if you attempted to fit a model in which the correlation between intercept and slope is so high that iterative algorithms may not converge and you cannot find stable estimates. We now introduce the possibility that the correlation between true intercept and true slope can be so high as to preclude model fitting. When this happens, recentering *TIME* can sometimes ameliorate your problem.

There is yet another reason you might recenter time: it can sometimes lead to a simpler level-1 model. For this to work, you must ask yourself: Is there a centering constant that might totally eliminate the need for an explicit intercept parameter? If so, you could decrease the number of parameters needed to effectively characterize the process under study. This is precisely what happened in the work of Huttenlocher, Haight, Bryk, Seltzer, and Lyons (1991). Using a sample of 22 infants and toddlers, the researchers had data on the size of children's vocabularies at up to six measurement occasions between 12 and 26 months. Reasoning that there must be an age at which we expect children to have *no* words, **(p.187)** the researchers centered *TIME* on several early values, such as 9, 10, 11, and 12 months. In their analyses, they found that centering around age 12 months allowed them to eliminate the intercept parameter in their level-1 submodel, thereby dramatically simplifying their analyses.

We conclude by noting that there are other scales for *TIME* that alter not only a level-1 submodel's intercept but also its slope. It is possible, for example, to specify a model that uses neither a traditional intercept nor slope, but rather parameters representing initial and final status. To do so, you need to create two new temporal predictors, one to register each feature, and eliminate the stand-alone intercept term.

To fit a multilevel model for change in which the level-1 individual growth parameters refer to initial and final status, we write:

(5.13a)

$$Y_{ij} = \pi_{0i}\left(\frac{max\ time - TIME_{ij}}{max\ time - min\ time}\right) + \pi_{1i}\left(\frac{TIME_{ij} - min\ time}{max\ time - min\ time}\right) + \varepsilon_{ij}.$$

In the context of the antidepressant medication trial, in which the earliest measurement is at time 0 and the latest at time 6.67, we have:

$$Y_{ij} = \pi_{0i}\left(\frac{6.67 - TIME_{ij}}{6.67}\right) + \pi_{1i}\left(\frac{TIME_{ij}}{6.67}\right) + \varepsilon_{ij}.$$

Although it may not appear so, this model is identical to the other linear growth models; it is just that its parameters have new interpretations. This is true despite the fact that equation 5.13a contains no classical "intercept" term and *TIME* appears twice in two different predictors.

To see how the individual growth parameters in this model represent individual *i*'s initial and final status, substitute the minimum and maximum values for *TIME* (0 and 6.67) and simplify. When *TIME* = 0, we are describing someone's initial status. At this moment, the second term of equation 5.13a falls out and the first term becomes π_{0i} so that individual *i*'s initial status is $\pi_{0i} + \varepsilon_{ij}$ Similarly, when *TIME* = 6.67, we are describing someone's final status. At this moment, the first term of equation 5.13a falls out and the second term becomes π_{1i} so that individual *i*'s final status is $\pi_{1i} + \varepsilon_{ij}$.

We can then specify standard level-2 submodels—for example:

$$\pi_{0i} = \gamma_{00} + \gamma_{01} TREAT_i + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11} TREAT_i + \zeta_{1i} \quad (5.13b)$$

and invoke standard normal theory assumptions about the residuals. When we fit this model to data, we find the same deviance statistic we found before—12,680.5—reinforcing the observation that this model is identical to the three linear models in table 5.10. And when it comes to **(p.188)** the parameter estimates, notice the similarity between these and selected results in table 5.10:

$$\hat{\pi}_{0i} = 167.46 - 3.11 TREAT_i$$

$$\hat{\pi}_{1i} = 151.34 + 33.80 TREAT_i.$$

The first model provides estimates of initial status in the control group (167.46) and the differential in initial status in the treatment group (−3.11). The second model provides estimates of final status in the control group (151.34) and the differential in final status in the treatment group (33.80).

This unusual parameterization allows you to address questions about initial and final status simultaneously. Simultaneous investigation of these questions is superior to a piecemeal approach based on separate analyses of the first and last wave. Not only do you save considerable time and effort, you increase statistical power by using all the longitudinal data, even those collected at intermediate points in time.

Notes:

(1.) A fundamental advantage of the multilevel model for change is that it can be used under all three missingness assumptions. This stands in contrast to other longitudinal methods (including the generalized estimating equations (GEE) approach of Diggle, Liang, and Zeger (1994), which requires the MCAR assumption).

(2.) The terms defined, ancillary, and internal were first used by Kalbfleisch and Prentice (1980). The fourth category—contextual—builds upon ideas in Blossfeld and Rohwer (1995) and Lancaster (1990).