Longitudinal Data Analysis Examples with Random Coefficient Models

David Rogosa   Stanford University
Hilary Saner   RAND Corporation

Abstract

Longitudinal panel data examples are used to illustrate estimation methods for individual growth curve models. These examples constitute one of the basic multilevel analysis settings, and they are used to illustrate issues and concerns in the application of hierarchical modeling estimation methods, specifically the widely-advertised HLM procedures of Bryk and Raudenbush. One main expository purpose is to "demystify" these kind of analyses by showing equivalences with simpler approaches. Perhaps more importantly, these equivalences indicate useful data analytic checks and diagnostics to supplement the multilevel estimation procedures. In addition, we recommend the general use of standardized canonical examples for the checking and exposition of the various multilevel procedures; as part of this effort, methods for the construction of longitudinal data examples with known structure are described.

Keywords: longitudinal data analysis; hierarchical linear models

Longitudinal Data Analysis Examples with Random Coefficient Models

I. Preamble.

This paper attempts to give a thorough treatment of one small, but prominent, example in multilevel analysis--individual growth curve analyses of longitudinal panel data. The history of this paper began with a presentation on longitudinal data analysis (Rogosa, 1989) at the October 1989 conference, "Best Methods for Analyzing Change" at USC; one section of that presentation (prepared with Hilary Saner) compared results from simpler longitudinal methods with those from the HLM program of Bryk and Raudenbush. One intended audience for this paper is users, past and future, of the HLM program; other relevant audiences include those interested in longitudinal data analysis (whether or not using hierarchical methods), and some parts may be useful to developers of hierarchical modeling estimation methods.

The individual growth curve models for longitudinal panel data are one common example of random coefficient models, and applications of empirical Bayes methods for estimation of these models date back at least to Fearn (1975) and Hui and Berger (1983). The general model building strategy which underlies our treatment of longitudinal panel data consists of models for the separate individual processes coupled with representations of individual differences, by allowing, for example, the individual-unit model parameters to differ over individuals. A motto for this is "Everyone has their own model"--the key is that the individual unit model must have some degree of seriousness. The model-building strategy of starting with the individual unit model and then building in individual differences is important and applicable in many social-science settings. Some other examples of the use of collections of individual unit models are: Rogosa and Ghandour (1991) for observations of behavior; Holland (1988) for outcomes of experiments; Efron and Feldman (1991) for dose-response curves and compliance in medical field trials; and Rogosa (1991) for aptitude-treatment interaction research designs.

For the measurement of change and analysis of longitudinal panel data, the unifying principle is that useful measures of change or stability are based on collections of individual growth curves. The important contrast is with analysis of associations and covariance structures, such as path analysis or LISREL models, the failures of which for longitudinal data are exhaustively discussed in, for example, Rogosa (1987; 1988; 1993; in press). We wish to emphasize that our modeling strategy for longitudinal data is entirely consistent with the structure and aims of the estimation procedures developed under the hierarchical modeling framework. The point of departure here is not over whether the parameters estimated are at all meaningful (which is the crux of the problems with LISREL, etc.); our concern is with identifying the most useful and dependable methods for estimation and exposition.

2. Straight-line Growth Curve Formulation.

The structure of the longitudinal panel data examples is deliberately kept very simple for the following reasons: (1) to match with the lead example of the HLM manual (Bryk, Raudenbush, Seltzer, Congdon, 1989), the Rat data, and (2) because this structure is adequate to illustrate the key technical and expository issues. The within-unit model is a straight-line growth curve, the observables have the basic classical test theory measurement model, and the between-unit model has a single exogenous predictor (perfectly measured and most often with no missing data). Start with an attribute η, such as reading proficiency or social competence, which exhibits systematic change over time. For individual p in the population of individuals, denote the form of the growth curve in η for individual p as $\eta_p(t)$. A straight-line growth-curve is written as

$$\eta_p(t) = \eta_p(0) + \theta_p t \; . \qquad (1)$$

We can rewrite (1) using the centering parameter t^o (from Rogosa and Willett, 1985, Sec. 2) which specifies a center for the time metric; $t^o = -\sigma_{\eta(0)\theta}/\sigma^2_\theta$. The centering parameter t^o has the convenient property that θ and $\eta(t^o)$ are uncorrelated over the population of individuals. Then the straight-line growth model can be written in terms of the uncorrelated random variables $\eta(t^o)$ and θ in the form:

$$\eta_p(t) = \eta_p(t^o) + \theta_p(t - t^o) \; . \qquad (1')$$

The constant rate of change θ_p in this individual growth curve model is often the key parameter of interest in research questions about change. The parameters of the individual growth curves have a distribution over the population of individuals (often assumed to be Gaussian by default). The first two moments of the rate of change are written as μ_θ and σ^2_θ . Longitudinal data sets also commonly include at least one exogenous (background) characteristic, denoted as Z , which allows us to address additional research questions about systematic individual differences in growth (i.e. correlates of change) and also to examine possible improvements in estimating growth curve parameters. The relation, over individuals, of Z to the rate parameter θ, is summarized by the conditional expectation $E(\theta \mid Z)$, which is stated here as the simplest possible straight-line regression

$$E(\theta \mid Z) = \mu_\theta + \gamma\ (Z - \mu_Z) \quad , \qquad (2)$$

where the regression slope parameter γ for the exogenous variable could also be written as $\beta_{\theta Z}$. Equation (2) is an example of a "between-unit" model. A similar relation can be stated for the intercept (aka "base" variable) in Equation 1. In the case where there is no measured exogenous variable, this between-unit model is $E(\theta \mid Z) = \mu_\theta$ (see later discussion of the North Carolina data example).

Observables. Times of observation are $\{t_i\} = t_1 ,..., t_T$, which in these data analysis examples are the same for all p (except when observations at some t_i are missing for some p). From these discrete values of the times of observation, we then have values for the $\eta_p(t_i)$ for p = 1, ..., n. The completion of this set-up is the standard (oversimplified) statement that the observable Y is an imperfectly measured η, and the relation between Y and η is through the basic classical test theory model. $Y_p(t_i) = \eta_p(t_i) + \epsilon_i$ for $p = 1, \ldots, n$. For convenience, the observables for individual p are written as $Y_{1p}, \ldots, Y_{Tp}$. It is convenient to consider Z to be measured perfectly to conform to the common assumptions and especially to make γ a main parameter of interest in the between-unit model (i.e., not distorted by measurement error in Z).

3. Parameters of Interest

Although in applications we might argue that descriptive analyses of the individual trajectories, rates of improvement, etc. are of the greatest substantive value, for the purposes here we give undue emphasis to the estimation of variance components and model parameters. Some key quantities, which are also the focus in the presentation in Bryk and Raudenbush (1987, pp 151-4; see also 1992, Chap. 6), are represented by the following parameters:

a. The first two moments of the rate of change over the population of individuals, μ_θ and σ^2_θ, which for the straight-line growth model address questions about typical rates of change and heterogeneity (individual differences) in rates of change.

b. The reliability of the growth curve estimates of θ_p, which for psychometric purposes addresses questions about accuracy of the estimates. For the unbiased OLS estimate of θ_p, the reliability is denoted as $\rho(\hat{\theta})$.

c. The correlation between change θ_p and true initial status $\eta_p(t_I)$, $\rho_{\eta(t_I)\theta}$, where t_I indicates a designated time of initial status. The correlation is used to investigate whether those with lowest initial status make the most progress (negative value) or those with the highest initial status make the most progress (positive value). As discussed in Rogosa and Willett (1985), the choice of t_I is of critical importance because $\rho_{\eta(t)\theta}$ is functionally dependent on time (see App. A, Eq. A3).

d. The exogenous variable regression parameter γ in (2) and the standard error of its estimate: γ is taken to represent the "influence" of Z (on θ) and is often of primary interest in applications.

One important omission in this listing is the estimation of the individual θ_p ; methodology for improvements upon the unbiased estimate ($\hat{\theta}_p$) has a prominent history and central focus in empirical Bayes methodology (e.g., Morris, 1983). Because this estimation problem is not featured in the Bryk-Raudenbush methods, it is not part of the treatment here. Some general discussion of empirical Bayes estimates for measurement of change problems is given in Rogosa et. al. (1982).

4. Longitudinal Data Examples

We use three longitudinal data examples which are shown (in part) in Exhibit 1.

Example 1. Rat weight data, from HLM manual (Bryk, Raudenbush, Seltzer, Congdon, 1989). The rat data consist of 10 individuals, with weight measurements (Y) at 5 occasions (weeks 0,1,2,3,4) and a background measure (Z), the mother's weight, and are listed in Exhibit 1 .

Example 2. Artificial longitudinal data with known structure created by TPSIM (Rogosa & Ghandour, 1986). Artificial data allow comparisons among analysis procedures under a "controlled" setting with known parameter values. See Appendix A for technical details on the construction of these data. In keeping with the layout of the Rat weight data, these longitudinal data are five waves of observations on each of 200 individuals, with times of observation {0,1,2,3,4}, and with an exogenous measure Z for each individual. A scenario for these data might be longitudinal observations on academic achievement (Y) with a background measure of home environment (Z). The unobserved $\eta_p(t_i)$ follow the straight-line growth model in (1), with the observables Y including the measurement error ϵ . The data were constructed with Gaussian parameter distributions to fit the (untestable) assumptions of the HLM program-- θ~N(5, 5); $\eta(0)$~N(44, 52); $\eta(1)$~N(49,47); $\eta(2)$~N(54,52); $\eta(3)$~N(59,67); $\eta(4)$~N(64,92); Z~N(10,4); ϵ~N(0,12). After generating these artificial data, some observations (about

7%) were deleted at random to produce a moderate missing data situation. Missing observations are indicated by ***** in the first 9 cases shown in Exhibit 1.

Example 3. North Carolina Achievement Data (see Williamson, Applebaum, Epanchin, 1991). These education data are eight yearly observations on achievement test scores in math (Y), for 277 females each followed from grade 1 to grade 8, with a verbal ability background measure (Z). First 9 cases are shown in Exhibit 1.

Insert Exhibit 1 here

EXHIBIT 1
Longitudinal Data Examples

Rat Weight Data from HLM manual (Bryk, et. al., 1989)

rat	0	1	2	3	4	Z
	Observation time					
1	61	72	118	130	176	170
2	65	85	129	148	174	194
3	57	68	130	143	201	187
4	46	74	116	124	157	156
5	47	85	103	117	148	155
6	43	58	109	133	152	150
7	53	62	82	112	156	138
8	72	96	117	129	154	154
9	53	54	87	120	138	149
10	72	98	114	144	177	167

Artificial Data generated by TPSIM: first 9 cases

ID	0	1	2	3	4	Z
	Observation time					
1	51.94	50.05	*****	49.34	54.26	6.84
2	43.25	43.72	55.48	53.63	57.23	8.14
3	61.79	56.42	63.43	63.19	67.48	11.13
4	49.63	59.81	*****	59.55	65.36	9.73
5	38.88	48.95	51.30	51.71	64.20	8.18
6	28.18	40.34	41.55	48.88	55.14	8.65
7	47.57	58.84	68.12	73.50	82.65	13.65
8	48.05	53.06	60.85	62.10	70.38	9.10
9	45.18	*****	45.82	46.33	52.90	7.90

North Carolina Achievement Data (see Williamson, et. al., 1991): first 9 cases

ID	1	2	3	4	5	6	7	8	Z
	Observation time								
1	380	377	460	472	495	566	637	628	120
2	362	382	392	475	475	543	601	576	95
3	387	405	438	418	484	533	570	589	99
4	342	368	408	422	470	543	493	589	101
5	335	372	450	424	500	510	540	583	109
6	362	444	473	482	567	597	651	655	115
7	354	409	410	445	460	540	567	620	115
8	365	381	455	482	533	554	591	602	109
9	359	371	438	452	497	591	573	593	107

5. Analysis Approaches

The examples of this paper are used to illustrate and discuss the following longitudinal analysis approaches: (1) the HLM program of Bryk and Raudenbush, (2) computation of standard maximum likelihood estimates as in the TIMEPATH program, and (3) the data analysis ingenuity of a "Smart First Year statistics Student" (SFYS). A major shortcoming in our exposition is that no attention here is given to other, less-widely known hierarchical modeling estimation programs, either in use or under development, which should be considered as alternatives to the Bryk-Raudenbush program. One source of information about, and comparisons of, four hierarchical linear regression packages is Kreft, de Leeuw and Kim (1990).

5.1 Hierarchical Linear Models.

In this paper, that heading refers to the HLM procedures of Bryk and Raudenbush. The computing implementation is version 2.2 of the HLM program that was available fall 1993 (when these results were presented at the Rand Conference). Detailed technical descriptions can be found in the HLM manual (Bryk, Raudenbush, Seltzer, Congdon, 1989) and in the text of Bryk and Raudenbush (1992). The lead example in the HLM manual, the Rat data, is the template for these analysis examples.

No attempt is made here to describe their hierarchical models technology, but there are some details of the HLM implementation that are consequential for the specific examples and illustrations. In the running of the HLM program, a choice is available whether to "Center" in the individual unit model fit or not. Also there is a choice on using the exogenous variable in the between-unit model: i.e., in our examples whether or not to include the "Background" variable Z . This leads to a 2x2 structure for possible analyses-- no Center, no Background; Center, no Background; no Center, Background; Center, Background. In the Tables for the examples these are denoted by CNBN, CYBN, CNBY, CYBY (the first two the "unconditional" analyses and the last two "conditional"). In the HLM manual the Rat example is run with Centering and with the Background variable (CYBY). When there are missing data, there are some additional details, which are taken up in the discussion of the artificial data example in Section 7.

From a HLM run, estimates of the parameters under consideration here are obtained as follows. The estimate for σ_{θ}^{2} is the (2,2) entry of the "TAU" matrix. The estimate for $\rho(\theta)$ is listed as "Parameter Reliability Estimates" for the "time" variable (e.g., HLM manual p.26). An estimate of the correlation between θ and true initial status is obtained from the (1,2) entry of the "TAU (AS CORRELATIONS)" matrix. Discussion of how this quantity relates to $\rho_{\eta(t_1)\theta}$ is included in presentation of the examples and in Appendix C. The estimates of γ in (2) and its standard error are obtained from the "GAMMA-STANDARD ERROR-T STATISTIC TABLE" for C*BY runs.

5.2 Simple Maximum-likelihood estimates via TIMEPATH program.

In the case of complete data for Y and Z and the same observation times for all individuals, it is straightforward to implement computation of the closed-form maximum-likelihood estimates of parameters. Since 1981, we have used various versions of a computer program we call TIMEPATH (originally developed with the assistance of John Willett and Gary Williamson, current version written with Ghassan Ghandour; Rogosa and Ghandour, 1988). Examples of descriptive and inferential analyses can be found in Rogosa (in press) and, for earlier versions, in Williamson et al. (1991). The core of this program is the ordinary least-squares regression estimation of the growth curve model for each individual. As the empirical rate of change can be treated as an attribute of an individual (just like a measurement on Y or Z), the obtained slopes for each individual regression can be profitably used for various descriptive analyses, and such descriptive analyses may, in many situations, be more important and informative than the formal parameter estimation.

Estimation of the parameters is by the maximum likelihood estimates (mle) adapted from the results in Blomqvist (1977). Forms of these estimates are given in Appendix B. It is worth noting often and loudly that the mle does not necessarily have wonderful properties; estimates of ratios of variance components may have significant bias (e.g., $\rho(\theta)$ tends to be underestimated), and all parameter estimates may have poor precision.

These concerns are especially grave when the number of individual units is small. In the current TIMEPATH program (Rogosa & Ghandour, 1988), standard errors for these parameter estimates and confidence intervals for the parameters are obtained by bootstrap resampling methods in which rows (individual units) are resampled. In the Tables, the reported standard errors are just the standard deviation over 4000 bootstrap replications, and the endpoints of the reported 90% confidence intervals are just the 5% and 95% values of the empirical distributions from the resampling (i.e., the 200th values from the maximum and minimum values). More sophisticated and more accurate confidence intervals could be constructed using the methods in Efron and Tibshirani (1993).

When present, missing longitudinal observations are treated (deliberately) in a very simple manner; as discussed in Appendix B, the individual growth curves are fit to the data that are present. There are obvious improvements in which adjustments, such as weighting according to the observations present, would be implemented. Our purpose is not to strongly promote the use of this program for missing data situations, as serious multilevel estimation programs would be the technically correct approach, but this adaptation of the full data estimation may serve as a convenient benchmark for understanding the performance of the HLM program.

5.3 SFYS: "Smart First-Year Student".

To illustrate useful supplementary descriptive analyses and to describe methods that provide some useful checks and insights on the HLM results, we use the device of a "Smart First-Year-Student" (SFYS). Consider what a very smart first quarter undergraduate could do with the tools of a basic introductory statistics course-- descriptive statistics, plots, and basic OLS regression methods from a package such as MINITAB. (In addition, we assume that the SFYS has a considerable reserve of scientific common-sense about the longitudinal research problem and data analysis; in particular, the student would know to address questions about change directly, either by being unaware of or by knowing enough to avoid the standard psychometric measurement of change literature; see Rogosa, 1988). Main features of the SFYS data analysis are:

a. Y on t regression. For each individual, fit by OLS (e.g, the MINITAB REGRESS command) the Y on t regression. The estimated $\hat{\theta}_p$ values allow descriptions of the collection of individual rates of change (graphical and numerical) and comparisons across individuals . For example, the SFYS can directly estimate μ_θ and could reason that since the $\hat{\theta}_p$ values are "noisy" estimates of θ_p for each individual, the observed variance of the $\hat{\theta}_p$ provides an upper bound on σ^2_θ .

b. Describe relations with $\hat{\theta}$. Using the estimated $\hat{\theta}_p$ values, plots representing relations of change with initial status and relations with the exogenous measure Z are especially useful for diagnostic examination of the corresponding correlation or regression parameter estimates. SFYS would plot θ versus observed initial status and plot θ versus Z and examine the plots for anomalous data points, etc.

c. θ on Z regression. Fit θ on Z regression using OLS (e.g., MINITAB REGRESS) to estimate γ (from Eq. 2) and to obtain a standard error for the estimate.

6. Results for Rat Data

Table 1 presents parameter estimates obtained from the output of the set of HLM runs and from the TIMEPATH program. Below the point estimate in the TIMEPATH column are shown a standard error and the endpoints of a 90% confidence interval (both obtained from 4000 bootstrap replications). The HLM program provides standard errors only for the estimate of γ (more on that below).

Insert Table 1 here

6.1 Identities

a. Estimates for σ^2_θ, $\rho(\hat{\theta})$. The individual rates of change for the 10 rats are {28.8, 28.1, 36.3, 27.2, 23.4, 29.3, 25.6, 19.7, 23.6, 25.6} (as would be obtained by SFYS from OLS Y on t regressions for the data in Exhibit 1, also printed out by TIMEPATH and HLM). The sample variance of these 10 numbers is 19.714 . For estimating σ^2_θ, substitute into Equation (B1) the average mean-squared error 91.747 from the 10 individual fits and the SSt of 10 to obtain 19.714 − 91.747/10 = 10.539, the mle produced by TIMEPATH. Table 1 shows that 10.539 is precisely the estimate produced by the HLM runs in which no between-unit background variable (Z) was included (i.e., CNBN, CYBN). Also, the associated estimate of the reliability $\rho(\hat{\theta})$ is .535 (i.e., 10.539/19.714) both from TIMEPATH and from the CNBN, CYBN HLM runs. Note that the bootstrap standard errors provided by TIMEPATH are large, and confidence intervals wide, for both quantities; as would be

TABLE 1
HLM and Timepath Estimates for Rat Data

Quantities	σ^2_θ	$\rho(\hat{\theta})$	$\rho_{\eta(t)\theta}$	γ	$s.e.(\hat{\gamma})$
Estimates from:					
HLM					
CNBN: No Center, No Background	10.539	.535	-.123	--	--
CYBN: Center, No Background	10.539	.535	.50	--	--
CNBY: No Center, Background	5.518	.376	-.902	.147	.0727
CYBY: Center, Background	5.518	.376	-.666	.147	.0727
Timepath					
Estimate	10.539	.535	-.123	.147	.088
Standard Error	7.707	.176	.435		
90% CI	(1.335, 25.81)	(.151, .735)	(-.785, .704)	(.047, .319)	

expected, 10 observations do not appear sufficient to estimate between-unit parameters with any precision. Although HLM does present a chi-square significance test for $\sigma^2_\theta = 0$, the statistically significant test-statistic of 19.3 (p-value .022) really provides meager, if not misleading, information in this case on precision of estimation (i.e., significantly different from 0 does not imply good estimation).

b. Correlation between change and initial status. The $\rho_{\eta(t_I)\theta}$ column in Table 1 contains a variety of values, due to the effects of Centering and the effects of using the background variable. A natural definition of time of initial status for the Rat data is $t_I = 0$ (time of birth). The TIMEPATH point estimate of $\rho_{\eta(0)\theta}$ defined by Equations (B3, B4, A3) is −.123 which matches the HLM value with CNBN. But note the very wide confidence interval associated with this estimate, which shows the difficulty of estimating this quantity from only 10 observations (something you would not see from the HLM output). The value .500 obtained from the HLM CYBN illustrates the confusions that may result from Centering. With the times being {0,1,2,3,4} for all individuals, the (1,2) entry of the TAU (AS CORRELATIONS) matrix in the CYBN run is actually an estimate of $\rho_{\eta(2)\theta}$; that value of .500 could also be obtained using Equations (B3, B4, A3) with $t_I = 2$. Additional discussion on Centering is given in Appendix C. Appendix C also points out that HLM runs using Z (C*BY) contain no useful information on the relation between change and initial status.

c. Regression for exogenous variable, estimation of γ. The estimation of γ from (2) seems to be a focal point in many applications and expositions. The SFYS would examine a plot of $\hat{\theta}$ versus Z and then fit the OLS regression. The MINITAB output for this regression is (notation added)

```
The regression equation is   θ = 3.0 + 0.147 Z
Predictor       Coef       Stdev    t-ratio        p
Constant        2.97       11.85       0.25    0.809
Z            0.14687     0.07275       2.02    0.078
s = 3.833       R-sq = 33.7%     R-sq(adj) = 25.5%
```

The point estimate and the standard error from the OLS regression match exactly the CYBY output in the HLM manual (p.28) and Table 1 HLM results for CNBY and CYBY. So SFYS and HLM produce exactly the same results. That the HLM standard error is identical (to 6 decimal places) to that from the OLS $\hat{\theta}$ on Z fit is surprising, if not disconcerting. Also, see from Table 1 that the same point estimate is obtained from TIMEPATH but the bootstrap standard error is larger. (As a side note, even though the parameter t-ratios for both the constant and the slope are identical from MINITAB and HLM, see HLM manual p.28, the HLM p-values of .369 and .061 respectively differ from the OLS values above.)

6.2. Problems with the HLM Manual Presentation

The choice of presentation of only the CYBY HLM output in the HLM manual (pp. 22-29) has some unfortunate aspects. The only useful information from the CYBY output is the estimation of γ and its associated standard error (and we've seen above that SFYS can obtain exactly this from MINITAB). Moreover some serious confusions arise in the presentation, most notably with the quantities labeled as "reliabilities". As these problems may have misled users of HLM, some discussion here may be useful. The key issue here is the difference between a conditional and unconditional variance. As explained in the annotated output in p.26 of the HLM manual, in HLM runs that use the background variable Z (i.e., C*BY) the TAU matrix provides estimates not of the variance of θ, but of the conditional variance of $\theta \mid Z$. This conditional variance is the variance of θ with the part predictable by the background variable Z partialled out and is termed the "residual parameter variance" by Bryk et al. (1989, p.26). (The value of 5.52 for this conditional variance implies a value of .69 for $\rho_{\theta Z}$; the mle obtained by TIMEPATH is .79 with a standard error of .3). Although this conditional variance is of little substantive interest, it is correctly interpreted in the HLM manual. Such is not the case for what are labelled "RELIABILITY ESTIMATES". From the HLM manual page 26: "The reliability estimates are the proportion of the variance in the OLS within-unit estimates that are the parameter variance. Since they are ratios of 'true' to 'observed' variance, they can be interpreted as reliability coefficients....the reliability of the OLS growth rate estimates is .376." NOT! The reliability of the OLS estimates, which we have denoted as $\rho(\hat{\theta})$, has mle .535. The quantity that Bryk et al. (1989) label a reliability seems to be the ratio of the estimated residual variance of θ with Z partialled out to the estimated variance of $\hat{\theta}$ with Z partialled out (but also note that the prose description in Bryk &

Raudenbush, 1992, p. 137, contains a correct mention of reliability estimates obtained from the "unconditional model" i.e., C*BN). Furthermore, it seems clear from Appendix C that this quantity does not represent the reliability of alternative empirical Bayes (shrunken) estimates of the $\{\theta_p\}$, as that reliability would appear to be roughly comparable to $\rho(\hat{\theta})$.

7. Results for Artificial Data with Known Structure--Missing data examples

Technical details of the structure of these artificial data are given in Appendix A, and a brief depiction of these data is given in the discussion of Exhibit 1. From the artificial data generation, we start out with 5 longitudinal observations (Y) on each of 200 individuals (each individual having observation times {0, 1, 2, 3, 4}), along with an observation on Z for each individual. Call this data set the "full" data. As an advertised major strength of the HLM program is the treatment of the missing data situation, we created a data set with missing values, eliminating some Y-values by a random mechanism (remember, Z-values cannot be missing, HLM manual p.16). In the data set used for Table 2--which we term YMZF (Y Missing, Z Full)--about 7% of the Y-values from the full data were made to be missing. (The corresponding data set with some Z-values also missing is termed YMZM.)

Parameter estimates are shown in Table 2, which has the same structure and entries as Table 1, with the addition of the known parameter values shown below each of the symbols. For the HLM runs, time is coded {0,1,2,3,4}, the data set is YMZF, and the "listwise" deletion option was employed. (The alternative choice of "pairwise" deletion had the unattractive properties of producing individual unit fits that did not correspond to the OLS fits of the data present when Centering was used, and of producing estimates further away from the parameter values than did listwise). The TIMEPATH program was used on all three data set versions: full, YMZF, YMZM. Space limits the display to YMZF results, but all three runs produce nearly identical estimates (with slightly smaller standard errors for the full data). A general feature of Table 2 is that the TIMEPATH estimates are closer to the parameter values than are the corresponding HLM estimates. No assertion is being made that this will always or often be the case. On the other hand, this data example was not culled; this is the one example we did originally. Our main intent here is not a contest between HLM and simpler methods, but instead to better illuminate the output of HLM.

Insert Table 2 here

a. Estimates for σ^2_θ, $\rho(\theta)$. The TIMEPATH estimate 4.355 can be obtained from (B1), and is closer to the parameter value 5.0 than the HLM C*BN estimates of 4.1. (The conditional variance estimated by the HLM C*BY entries has parameter value 3.2). The TIMEPATH estimate for $\rho(\hat{\theta})$ (parameter value .806) of .781 (about one standard error from $\rho(\hat{\theta})$) is obtained from (B2); the corresponding HLM estimate is .74 .

b. Correlation between change and initial status. The value of $\rho_{\eta(0)\theta}$ is $-.310$ which TIMEPATH ($-.246$) comes closer to than the HLM CNBN entry ($-.213$). The CYBN entry is an estimate of $\rho_{\theta\eta(2)} = .310$. From Appendix C, the CNBY entry estimates $\rho_{\eta(0)\theta\cdot Z} = -.732$, and the CYBY entry estimates $\rho_{\eta(2)\theta\cdot Z} = -.274$.

TABLE 2
HLM and Timepath Estimates for Artificial Data, Y Missing

Quantities	σ^2_θ (=5.0)	$\rho(\hat{\theta})$ (=.806)	$\rho_{\eta(\tau),\theta}$ (=-.31)	γ (= .671)	$s.e.(\hat{\gamma})$
Estimates from:					
HLM					
CNBN: No Center, No Background	4.095	.740	-.213	---	---
CYBN: Center, No Background	4.108	.741	.350	---	---
CNBY: No Center, Background	2.514	.636	-.685	.6008	.0666
CYBY: Center, Background	2.526	.637	-.238	.6034	.0667
Timepath					
Estimate	4.355	.781	-.246	.613	.070
Standard Error	.576	.027	.080		
90% CI	(3.42, 5.29)	(.729, .818)	(-.370, -.106)	(.500, .732)	

c. Regression for exogenous variable, estimation of γ (= .671). The SFYS would obtain 200 θ_p values from successive OLS regressions of the existing Y-values on the corresponding t_i . After examining a plot of θ versus Z , OLS would be used to fit the θ on Z regression (as shown in the MINITAB output):

```
The regression equation is  θ = - 1.12 + 0.613Z
Predictor       Coef       Stdev    t-ratio        p
Constant     -1.1156      0.6778      -1.65    0.101
Z            0.61320     0.06729       9.11    0.000
s = 1.987       R-sq = 29.6%     R-sq(adj) = 29.2%
```

This estimate .613 (for γ = .671) is identical to the TIMEPATH entry in Table 2; the value from HLM is .60. The SFYS standard error of .0673 is very close to that from HLM. Furthermore, both the SFYS and TIMEPATH can handle the YMZM data, giving γ-estimates of .620.

Watch out for missing data on Z! There is a warning in the annotation of p.16 of the HLM manual not to have missing data on the between-unit variable, but some further demonstration may be sobering. For the YMZM data, the HLM CYBY run proceeds without complaint when told the missing data code is 999 for the within-unit model (missing Y-values), though this data set also contains 999 for missing Z-values (in the between-unit file). The result is a γ point estimate of .00015, with standard error .00066, which is very close to the results from a θ on Z OLS regression in which all missing Z are set to have the value 999.

8. Results for North Carolina Data

This third example is real education data previously analyzed using the maximum-likelihood estimates in TIMEPATH in an excellent expository paper by Williamson et al. (1991). Table 3 (which has the same structure and entries as Table 1) gives the results for parameter estimation. (Entries in our Table 3 correspond to point estimates reported in Table 3 of Williamson et al., 1991). On a basic descriptive level we note that these data conform well to the straight-line growth model; for example, the median value of R^2 for the 277 individual OLS fits is .963. The mean rate of change is 36.45, and the estimation from TIMEPATH shows that these data permit rather accurate assessment of rates of change; the reliability estimate is high, and the standard error of measurement of θ is 3.1.

Insert Table 3 here

8.1 Identities

a. Estimates for σ^2_θ, $\rho(\theta)$, μ_θ. The sample variance of the 277 θ values is 55.836, $\hat{\sigma}^2 = 403.486$, and SSt = 42; from (B1) we have 55.836 − 403.486/42 = 46.229. Table 3 shows this estimate matches the σ^2_θ estimate from HLM runs CNBN, CYBN. And the associated estimate of $\rho(\theta)$, .828, is 46.229/55.836. (Note also from the TIMEPATH results the more reasonable size of standard errors and confidence intervals for the estimation of the between-person moments with this larger sample of 277 cases compared with the n = 10 example in Table 1.) For estimating μ_θ, SFYS would

TABLE 3
HLM and Timepath Estimates for North Carolina Data

Quantities	σ^2_θ	$\rho(\hat{\theta})$	$\rho_{\eta(\tau)\theta}$	γ	$s.e.(\hat{\gamma})$
Estimates from:					
HLM					
CNBN: No Center, No Background	46.229	.828	.340	--	--
CYBN: Center, No Background	46.229	.828	.933	--	--
CNBY: No Center, Background	24.628	.719	.025	.336	.0254
CYBY: Center, Background	24.628	.719	.869	.336	.0254
Timepath					
Estimate	46.229	.828	.651	.336	.028
Standard Error	5.95	.019	.090		
90% CI	(36.67, 56.12)	(.792, .854)	(.513, .809)	(.291, .382)	

average the 277 θ values to obtain a point estimate 36.448, and use elementary theory and methods (e.g., MINITAB DESCRIBE command) to compute the standard error as .44897. Exactly the same results for estimating μ_θ are obtained from HLM; from HLM C*BN output the coefficient for the BASE variable for predicting θ in the GAMMA table is 36.448, with the same reported standard error .44897. (TIMEPATH gives a bootstrap s.e. of .4478).

b. Correlation between change and initial status. The initial descriptive information would be the correlation between θ and Y_1, which is .421, and the corresponding scatterplot. The mle for $\rho_{\eta(t_1)\theta}$ from TIMEPATH is .651. Because the $\{t_i\}$ were coded 1,...,8 the CNBN entry in the $\rho_{\eta(t_p)\theta}$ column in Table 3 does not give the mle for $\rho_{\eta(t_1)\theta}$. Appendix C shows how to obtain the .651 TIMEPATH estimate (for t_1 as initial status) from the CNBN and CYBN HLM runs.

c. Regression for exogenous variable, estimation of γ. The SFYS would first examine a plot of θ versus Z ; this scatterplot, shown in Figure 1, displays reasonable structure (and perhaps one anomalous point). Next, SFYS would fit the OLS regression for θ on Z. The MINITAB output for this regression is (notation added):

```
The regression equation is   θ = 0.84 + 0.336Z
Predictor        Coef        Stdev    t-ratio        p
Constant        0.844        2.713       0.31    0.756
Z             0.33569      0.02537      13.23    0.000
s = 5.851       R-sq = 38.9%     R-sq(adj) = 38.7%
```

The HLM results from C*BY match the values above for both the estimate of

γ and its associated standard error exactly (at least to the six decimal places accuracy provided by HLM). TIMEPATH gives the same point estimate.

Insert Figure 1 here

8.2 Problems with earlier HLM versions

The HLM version prior to version 2.2 which was used in the presentation to the October 1989 USC Conference ("Best Methods for Analyzing Change") produced wild results for these very well behaved data, even though the program happily converged after one iteration with no complaint or warning. A short discussion, which illustrates the value of simple data analysis checks, may be a useful caution for past users of the HLM program. For example, the estimates of σ_θ^2 produced by those HLM runs were 2,649 for CNBN and CYBN, 20,068 for CNBY, and 5,039 for CYBY! Remember that the observed variance of θ is 55.8, so that SFYS would know that the estimation was wildly amiss. Furthermore, the HLM estimate of γ is $-.122$ for CNBY, and 2.02 for CYBY; SFYS would have obtained from OLS of θ on Z a slope of .336 with a standard error of .025. In other settings the obvious checks or upper bounds may be less transparent, but the importance of such supplemental data analysis remains.

9. Discussion: Needs Assessment

1. The need to assess the performance of estimation procedures.

Multilevel statistical estimation and its computational implementations are complex endeavors, and the performance of these methods needs to be carefully checked at all possible opportunities. One approach for basic quality control is to conduct analyses of common canonical examples, especially examples with known structure (parameter values). And the longitudinal panel setting is one opportunity for such examples (see methods presented in Appendix A). Furthermore, such examples could be extended to three level settings by construction of a hierarchical grouping for the individual units (e.g., students' longitudinal progress in a collection of schools). An alternative strategy is for a number of investigators analyze and re-analyze common data sets (structure unknown); the main reason we gave attention to the limited Rat weight data is the opportunity to compare with the expository analysis in Bryk et al. (1989).[1]

2. The need for supplementary descriptive analyses and diagnostics.

Part of the "demystifying" mission of this paper and a main purpose for introducing the SFYS was to plead for more descriptive data analysis, especially as part of the use of the HLM program. Bryk and Raudenbush (1992, esp. Chap. 9) does contain some valuable illustrations. But examples in applications or other expositions have been lacking; formal parameter estimation seems to eclipse good description.

3. The possible need for re-examining past HLM analyses.

Many applications of HLM (whether these longitudinal analyses or other applications) do not give the necessary details for the reader to be sure that there are no serious problems in computation and interpretation of HLM output. This paper would be wildly successful if it motivates some users to re-examine past analyses (i.e., for proper performance of the HLM program or for appropriate interpretation of the output).

4. The need for standard errors.

Estimates of the precision of an estimate are critically important companions to parameter and variance component estimates. Especially in settings with small numbers of individual units, disregarding precision of estimation is dangerous.

5. The need for design guidance.

Basic design questions for these longitudinal studies remain rather neglected. Of the information we present, the bootstrap standard errors from TIMEPATH help a little in showing remarkably large standard errors for small studies (e.g., Rat data) and perhaps acceptable precision for the larger data sets ($n \geq 200$). But n is not the only important factor in these designs. On a similar note, one of the claims that seems to be made for the use of HLM is improved efficiency -- but questions like the following do not seem to have been addressed: How much precision is gained from using the full HLM machinery on these kinds of longitudinal designs? How can that be calculated?

6. The need to examine applicability to other settings.

More study would be needed before determining which of our specific results and concerns carry-over to other data analysis settings. The longitudinal setting is somewhat distinct in having a small number (say 4-6) of observations within each unit; some studies have few individuals, some have many (e.g., 10 in rat, 277 in North Carolina). In particular, the imbalance due to design or missing data is likely to be greater in other settings, such as school effects examples. Consequently, some numerical correspondences may differ in other applications. But it's clear that the simple MINITAB-style within-groups and aggregated descriptive analyses will provide a useful and sometimes critical supplement to the typical HLM analysis.

FOOTNOTE

Footnote 1.
Much earlier, we attempted to obtain the longitudinal Head Start data, extensively analyzed in Bryk and Raudenbush (1987), for comparative reanalysis and illustration of longitudinal data analysis methods. Regrettably, at that time we were informed that all copies of those data had been destroyed in a fire.

References

Blomqvist, N. (1977). On the relation between change and initial value. Journal of the American Statistical Association, 72, 746-749.

Bryk, A.S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101, 147-58

Bryk, A.S. & Raudenbush, S. W.(1992). Hierarchical linear models: Applications and data analysis methods. Sage Publications:CA:Lnd.

Bryk, A.S, Raudenbush, S.W, Seltzer,M. Congdon,R.T (1989) An Introduction to HLM: Computer Program and User's guide.

Efron, B.& D. Feldman (1991) Compliance as an Explanatory Variable in Clinical Trials. Journal of the American Statistical Association, 86, 9-17.

Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. New York: Chapman & Hall.

Fearn, T. (1975). A Bayesian approach to growth curves. Biometrika, 62, 89-100.

Holland, P. W. (1988). Causal inference, path analysis and recursive structural equation models. In C. Clogg (Ed.), Sociological Methodology 1988 Washington, D.C.: American Sociological Association. 449-484.

Hui, S. L., & Berger, J. O. (1983). Empirical Bayes estimation of rates in longitudinal studies. Journal of the American Statistical Association , 78 , 753-760.

Kreft, I.G., de Leeuw J., & Kim, K.S. (1990). Comparing Four Different Statistical Packages for Hierarchical Linear Regression: Genmod, HLM, ML2, and VARCL. CSE Technical Report 311, UCLA Center for Research on Evaluation, Standards, and Student Testing.

Morris, C.N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. Journal of the American Statistical Association, 78, 47-55.

Rogosa, D. R. (1987). Casual models do not support scientific conclusions: A comment in support of Freedman. Journal of Educational Statistics, 12, 185-195.

Rogosa, D. R. (1988). Myths about longitudinal research. In Methodological issues in aging research, K. W. Schaie, R. T. Campbell, W. M. Meredith, and S. C. Rawlings, Eds. New York: Springer Publishing Company, 171-209.

Rogosa, D. R. (1989). A growth curve approach to the analysis of quantitative change. Invited presentation at "Best Methods for Analyzing Change" Conference, Los Angeles, October 1989.

Rogosa, D. R. (1991). A longitudinal approach to ATI research: Models for individual growth and models for individual differences in response to intervention. In Improving inquiry in social science: A volume in honor of Lee J. Cronbach, R. E. Snow and D. E. Wiley, Eds. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 221-248.

Rogosa, D. R. (1993). Individual unit models versus structural equations: Growth curve examples. In Statistical modeling and latent variables, K. Haagen, D. Bartholomew, and M. Diestler, Eds. Amsterdam: Elsevier North Holland, 259-281.

Rogosa, D. R. (in press) . Myths and methods: "Myths about longitudinal research," plus supplemental questions. In The analysis of change, J. M. Gottman, Ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rogosa, D. R., and Ghandour, G. A. TPSIM: A program for generating longitudinal panel data with known structure. Stanford University, 1986.

Rogosa, D. R., and Ghandour, G. A. TIMEPATH: Statistical analysis of individual trajectories. Stanford University, 1988.

Rogosa, D. R. and Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50, 203-228

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychological Bulletin, 92, 726-748.

Rogosa, D. R., and Ghandour, G. A. (1991). Statistical models for behavioral observations (with discussion). Journal of Educational Statistics, 16, 157-252.

Williamson, G. L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analyses of academic achievement. Journal of Educational Measurement, Vol 28(1) 61-76.

APPENDIX A Construction of longitudinal data examples with known structure: Basic Relations for Straight-line Growth Models

To create examples of longitudinal panel data with known structure, we can use the basic relations and properties of collections of growth curves (developed in Rogosa et. al., 1982; Rogosa & Willett, 1985). The procedures discussed here are for data based on the individual straight-line growth curve model.

Simulation Procedure. Start by choosing the center for the time metric by specifying t^o (where $t^o = -\sigma_{\eta(0)\theta}/\sigma^2_\theta$). Then for the parameters of the straight-line growth model $\eta_p(t) = \eta_p(t^o) + \theta_p(t - t^o)$ specify the parameter distributions over individuals of the uncorrelated random variables $\eta(t^o)$ and θ (e.g., each distribution Gaussian, or each distribution Uniform) to generate the parameter values for each p. By specifying the variances for these distributions, the scale for the time metric $\kappa = \sigma_{\eta(t^o)}/\sigma_\theta$ is set. Then choose the discrete values of the times of observation $\{t_i\} = t_1 ,..., t_T$, which are substituted into (1′) to produce values for the $\eta_p(t_i)$ for p = 1, ..., n. The exogenous characteristic Z is generated with specified mean and variance, with the added specification of values for the two correlations of Z with $\eta(t^o)$ and with θ (under the constraint $(\rho_{Z\eta(t^o)})^2 + (\rho_{Z\theta})^2 \leq 1$). The final step is to create the fallible observables Y_{ip} as $\eta_p(t_i) + \epsilon$ for p = 1 , ..., n (the addition of measurement error according to the classical test theory model): e. g., drawing $\epsilon \sim N(0, \sigma^2)$ with a specified value for σ^2 .

Consequences for second-moments. The choices of the values above determine the population values of the familiar second-moments of η or Y for the artificial data. In practice, values of these quantities--variances, correlations, etc-- may be chosen first (say, to correspond to values familiar from empirical research or common-sense), and then solutions (explicitly or by trial-and-error) for the quantities in the simulation procedure above are obtained. The relations that provide values of these second moments for the $\eta_p(t_i)$ are:

variance

$$\sigma^2_{\eta(t)} = \sigma^2_{\eta(t^\circ)} + ((t - t^\circ)/K)^2 \sigma^2_{\eta(t^\circ)} \tag{A1}$$

covariance (also yields correlation, using Equation A1)

$$\sigma_{\eta(t_1)\eta(t_2)} = \sigma^2_{\eta(t^\circ)} + (t_1 - t^\circ)(t_2 - t^\circ)\sigma^2_\theta \tag{A2}$$

correlation between change and status

$$\rho_{\eta(t)\theta} = \frac{(t - t^{o})}{[\kappa^2 + (t - t^{o})^2]^{1/2}} \quad \text{(A3)}$$

correlation between exogenous variable, Z and status

$$\rho_{Z\eta(t)} = \frac{(t - t^{o})\rho_{Z\theta} + \kappa\rho_{Z\eta(t^{o})}}{[\kappa^2 + (t - t^{o})^2]^{1/2}} \quad \text{(A4)}$$

Technical specifications for Artificial Data Example. In terms of the model parameters, the values for the artificial data example are $t^{o} = 1$; $\sigma^2_{\theta} = 5.0$; $\sigma^2_{\eta(t^{o})} = 47$ (yielding $\kappa = 3.066$); $\rho_{Z\theta} = .60$; $\rho_{Z\eta(1)} = .60$; for $\theta \sim N(5, 5)$, $\eta(t^{o}) \sim N(49, 47)$, $Z \sim N(10, 4)$, $\epsilon \sim N(0,12)$. This configuration yields observables Y with population reliabilities {.813, .797, .813, .848, .885} at the observation times $t_i = \{0,1,2,3,4\}$. The upper triangle of the population correlation matrix for the $\eta(t_i)$ at the observation times $t_i = \{0,1,2,3,4\}$ is

0.951	0.808	0.627	0.463
	0.951	0.838	0.715
		0.966	0.896
			0.981

.

The upper triangle of the corresponding population correlation matrix for the Y-values at the five observation times is

0.765	0.656	0.52	0.392
	0.765	0.688	0.6
		0.802	0.76
			0.849

.

APPENDIX B

Forms for Estimates in TIMEPATH

$$\sum_{p=1}^{n} MSR_p / n \text{ , where for each } p\ MSR_p = \sum_{i=1}^{T} (Y_{ip} - \hat{Y}_{ip})^2 / (T-2)$$ Start with

an estimate for σ^2, the residual variance about the individual growth curve fits (OLS, Y on t), which has the form

as

MSR_p is the mean squared residual for the fit to individual p. Then the estimate for σ^2_θ can be written,

$$\hat{\sigma}^2_\theta = \hat{var}(\theta_p) - \hat{\sigma}^2/SSt \ , \qquad \text{(B1)}$$

where $\hat{var}(\bullet)$ indicates a sample variance of the quantity over p, and SSt is

the sum of squares for the time points $SSt = \sum_{i=1}^{T} (t_i - \bar{t})^2$. Then the estimate

of the reliability of θ_p is formed by

$$\hat{\rho}(\theta) = \hat{\sigma}^2_\theta / \hat{var}(\theta_p) \ . \qquad \text{(B2)}$$

The estimate for the correlation between change and initial status $\rho_{\eta(t_1)\theta}$, can be formed by substituting the following estimates into Equation (A3):

$$\hat{t}^{o} = \frac{-[\, c\hat{o}v(\hat{\eta}(0), \hat{\theta}) + (\hat{\sigma}^2 \sum_{i=1}^{T} t_i)/(T \times SSt)\,]}{\hat{\sigma}_{\theta}^{2}} \; ; \qquad \text{(B3)}$$

$$\hat{\kappa}^2 = -(\hat{t}^{o})^2 +$$

The estimate of $\beta_{\theta Z}$ is obtained from the ordinary least-squares fit of $\hat{\theta}_p$ on Z_p .

Missing data. For each individual p, an OLS fit of the observed Y_{ip} on the observation times t_i yields a $\hat{\theta}_p$ value (if there are at least two observations) and a MSR_p (if there are at least three observations). So taking the case of at least three observations present for each p, in the situation of missing Y_{ip} the $\hat{\theta}_p$ and MSR_p computed from the observed data values are simply substituted into the equations above. This treatment of missing data is deliberately kept primitive; slight (but not consequential for our example) embellishments would be to form a weighted estimate for σ^2 (weighting the by number of observations present), adjusting SSt to reflect differences in the $\{t_i\}$ over p, and so forth.

APPENDIX C

Additional Technical Notes

Correlation between Change and Initial Status

Three considerations--that $\rho_{\eta(t_I)\theta}$ depends on the $t = t_I$ as seen in (A3), that the HLM C*BN TAU (AS CORRELATIONS) (1,2) entry estimates $\rho_{\eta(0)\theta}$ (regardless of the values of the $\{t_i\}$), and that centering makes t = 0-- make for some complication in estimating $\rho_{\eta(t_I)\theta}$ from HLM (with no background variable). Let's take for illustration the complete data case. In TIMEPATH, estimates for $\rho_{\eta(t_i)\theta}$ are obtained (using Eq. A3) for all $\{t_i\}$, the value at t_1 usually being of primary interest. For HLM, the "easy" situation is a CNBN run with t_1 coded to be 0; then the mle for $\rho_{\eta(t_1)\theta}$ is obtained (as in the Rat data). From a CYBN HLM run, the mle for the correlation between $\eta(t)$ and θ is obtained, a quantity usually not of interest.

For an arbitrary set of $\{t_i\}$ with $t_1 \neq 0$, with the help of (A3) the mle for $\rho_{\eta(t_1)\theta}$ can still be obtained from HLM. These calculations may be of most interest to those (re-)examining prior HLM analyses. From (A3) create two equations in two unknowns (t^o, κ) using $t = 0$ and $t = \bar{t}$ and the corresponding HLM estimates TAU (AS CORRELATIONS) (1,2) entry from CNBN and CYBN. Solving for (t^o, κ) allows evaluation of (A3) for any t. Consider the North Carolina data with $\{t_i\} = \{1,...,8\}$. From Table 3 CNBN ($t_I = 0$) gave estimate .340 and CYBN ($t_I = 4.5$) gave .933. Solving the two (A3) equations (e.g., with Mathematica NSOLVE) yields $\{t^o, \kappa\}$ estimates {−0.729, 2.016}, which produces a $\rho_{\eta(t_1)\theta}$ estimate of .651 (matching the TIMEPATH entry in Table 3).

HLM runs that include the background variable do not appear useful for questions about change and initial status. From the CNBY run the parameter estimated appears to be the partial correlation $\rho_{\eta(0)\theta\cdot Z}$, and for CYBY the parameter is $\rho_{\bar{\eta}(t)\theta\cdot Z}$.

Reliability Calculation for OLS and Empirical Bayes Estimates

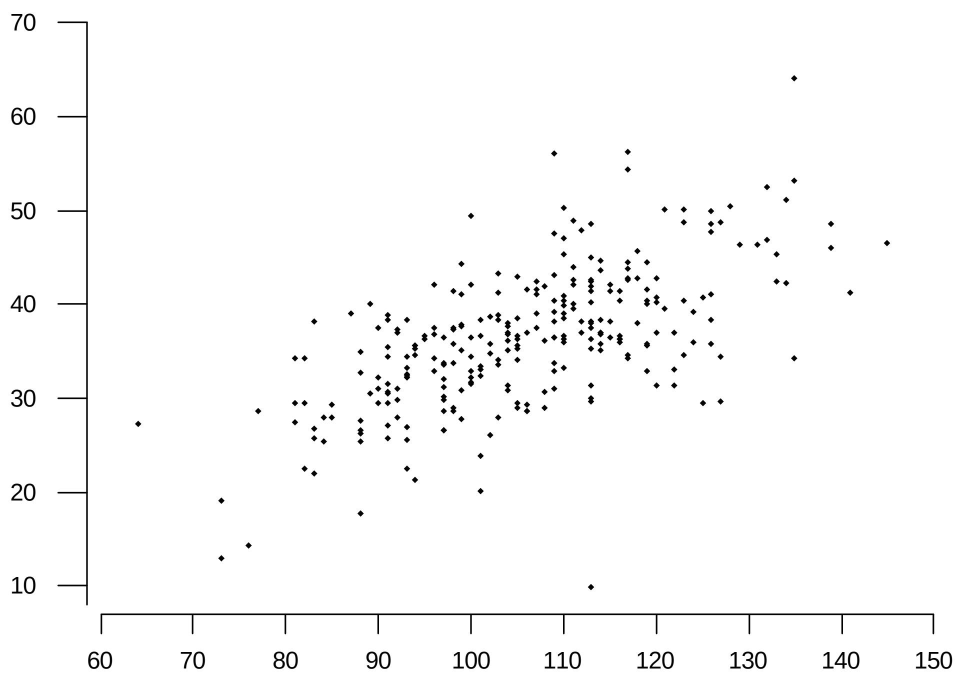
For better or worse, in behavioral science applications the properties of measures (estimates) are often judged by the obtained reliability coefficient. The reliability coefficient for the unbiased OLS estimate of the θ_p (written as $\hat{\theta}_p$) has the form from classical test theory $\rho(\hat{\theta}) = \sigma^2_\theta / (\sigma^2_\theta + V)$, where $V = \sigma^2/SSt$. One equivalence for this reliability is the square of the correlation between θ_p and $\hat{\theta}_p$. Empirical Bayes methods are often used in multilevel analyses to provide improved estimates of the set of the $\{\theta_p\}$. The familiar form of the estimate θ^{EB}, where the shrinking is toward the conditional mean $E(\theta | Z)$, is $(1 - B_p)\,\hat{\theta}_p + B_p E(\theta | Z)$, where the shrinkage coefficient B_p has the form $V_p/(V_p + \tau^2)$ (following approximately the standard notation as in Morris, 1983, Sec. 1). Setting τ^2 to be the conditional variance of $\theta | Z$, and taking the special case of $B_p = B$, $V_p = V$ (both assumed known), we can calculate the square of the correlation between θ^{EB} and θ as:

$$\rho(\theta^{EB}) = \frac{\sigma_\theta^2(\sigma_Z^2 - \sigma_{\theta Z}^2) + V\sigma_{\theta Z}^2}{\sigma_\theta^2[\sigma_Z^2(\sigma_\theta^2 + V) - \sigma_{\theta Z}^2]} \qquad (C1)$$

To illustrate (C1), consider the structure of the artificial data example from Appendix A; $\rho(\hat{\theta})$ has the value .806 and (C1) is .825. If the measurement error σ^2 were doubled, $\rho(\hat{\theta})$ becomes .676 and (C1) .726 . For $\rho_{\theta Z} = 0$, the two reliabilities are equal. In "real life" B and V must be estimated, so (C1) might be thought of as a rough upper bound on the reliability coefficient for the empirical Bayes estimates (under $V_p = V$).

Figure Captions

Figure 1. Scatterplot of $\hat{\theta}$ versus Z from the North Carolina data.



70
60
50
40
30
20
10
60
70
80
90
100
110
120
130
140
150