

## Reply to Discussants: Longitudinal Data Analysis Examples with Random Coefficient Models

David Rogosa    Stanford University  
Hilary Saner    RAND Corporation

We start with a series of short items, and one longer example follows.

1. Properties of maximum likelihood estimates. Morris provides a very useful discussion of some of the deficiencies of maximum likelihood in multilevel settings. Results from a small simulation illustrate the extent of bias for the m.l.e. in our longitudinal setting (for number of individuals, denoted by  $n$  or  $k$ , small). For  $k = \{10, 15\}$ , with the Table 2 value  $\rho(\theta) = .806$  we obtain average  $\hat{\rho}(\theta)$  values  $\{.71, .77\}$  respectively, and for  $\rho(\theta) = .50$  we obtain average  $\hat{\rho}(\theta)$  values  $\{.37, .41\}$  (bias increasing for smaller  $\rho(\theta)$ ). It was an oversight in our paper not to make clear that both HLM (see restricted maximum likelihood in Bryk & Raudenbush 1992, pp. 45-46) and the analyses we report in Tables 1-3 use estimates based on  $(k - 1) \times V/S$  (in Morris' notation), whereas the m.l.e. uses  $k \times V/S$ . In order for our estimation to obtain the equivalences with HLM results, we used, for example,  $k - 1$  in the denominator of  $\hat{\text{var}}(\theta_p)$  in Eqs. B1, B2, whereas the m.l.e. from Blomqvist (1977, p.748) uses  $k$ . These estimates will tend to be less biased than the m.l.e.; in the simulation above we obtain values  $\{.75, .78\}$  and  $\{.42, .43\}$ .

2. Extensions to non-normality. Morris and Mason give important pointers to settings (and computational programs) outside the basic Gaussian distributions for continuous random effects. One slight extension of our examples is to have Uniform (rather than Normal) distributions for  $\eta(t^o)$  and  $\theta$ , for which the same equivalences of HLM with SFYS and TIMEPATH are found.

3. HLM version 3 update. Subsequent to the RAND conference, in 1994 a new version of the HLM program became available. For the examples in our Tables 1 and 3, exactly the same numerical results are obtained with the new version 3.01. For the example in Table 2 (and also the "continuous time" example below) noticeable, but not consequential, numerical differences (1 part in 500 to 1 part in 10000) are obtained between versions 2.2 and 3.01. The "group mean" Centering option in version 3 is equivalent to the CYB\* results we report, as this centering option gives the same "OLS Level-1 regressions". As Raudenbush notes, in 1994 a new HLM manual appeared, providing improved exposition along with the 1992 text.

4. Basic equivalences with HLM. We regret that Raudenbush misrepresents the

tone and intent of our exposition; he begins his comments by quoting the phrase "surprising, if not disconcerting" badly out of context. That phrase (in the Rat data example) was used only in reference to the s.e.( $\hat{\gamma}$ ) equivalence in the sentence: "That the HLM standard error is identical to that from the OLS  $\theta$  on Z fit is surprising, if not disconcerting." This equivalence may well be "surprising" to users in light of years of HLM literature (at least back to Bryk & Raudenbush, 1987) claiming increased precision or accuracy. Moreover, the s.e.( $\hat{\gamma}$ ) equivalence is intuitively "disconcerting" because advanced multilevel estimation of the  $\theta$  on Z regression, especially in the case of  $\sigma_{\theta}^2$ ,  $\sigma_{\theta \cdot Z}^2 > 0$ , seemingly should be different than simple OLS for a "noisy"  $\theta$  on Z.

Furthermore, Raudenbush's extended discussion of his recent chapter (Raudenbush, 1993) in this context is disingenuous. His chapter does demonstrate that HLM can be used to reproduce the standard results for textbook analysis of variance designs (t-tests up through standard anova), and in his Table 13.13 he does use HLM to produce a version of the individual growth analysis. But nowhere does he present any alternative analysis for the individual growth setting, nor do his examples include an exogenous variable (for equivalences involving estimation of  $\gamma$ ). Thus, there is no real overlap with the results of our paper. In fact, a theme of his chapter is the absence of a classical or simpler equivalent to the HLM analysis of the individual growth problem, which is what our paper provides for our examples and also can provide for his Table 13.13 results. (To bring this full circle, the examples that Raudenbush treats in that chapter would be presented using MINITAB, not HLM, in advanced texts, such as Neter, Wasserman and Kutner, 1990.)

5. Bootstrap methods. Raudenbush's generally positive comments on our bootstrap procedures have one aspect needing clarification. As we state in our paper, this bootstrap procedure simply resamples individuals (rows in the data structures in Exhibit 1), which would hold with no computational difficulties even for the extended ("unbalanced") case of different observation times for each individual and missing observations. His recommendation for multilevel resampling may well be important in other (more complex) multilevel estimation settings, but do not appear relevant for the two-level individual growth problem. One message from our bootstrap calculations is that point estimates from HLM applications having small numbers of level-2 units are taken far more seriously than is merited.

6. Reliability (unconditional and conditional) versus precision. To restate what

should be clear and uncontroversial, the estimate of  $\rho(\theta)$  provides traditional information about the measurement properties of  $\theta$ , although the standard error ( $\text{Sqrt}[\hat{\sigma}^2/\text{SS}_t]$ ) for the OLS estimate may well be preferable. Also, (as discussed in our paper) the older HLM manual (Bryk et al. 1989) is incorrect to label the reliability (.376) obtained from the CYBY Rat Data run as providing "reliability of OLS growth rate estimates." Now, Raudenbush, in his discussion, defends a different interpretation of the CYBY reliability as a "conditional reliability". That quantity (from C\*BY) is equivalent to a reliability for the estimate of the partial variate  $\theta \cdot Z$  (part of  $\theta$  not predictable by  $Z$ ); it is simply sample estimates in the ratio,  $\text{var}(\theta | Z)/\text{var}(\theta)$ . Properties and interpretations for this measure hark back to the discredited residual change score (see Rogosa et al., 1982), and interpretations and properties become even more tenuous with a fallible  $Z$ .

7. "Let's be careful out there". One theme of our efforts is the need for caution in both applying and proselytizing for multilevel methods (see also Mason, Morris). Our work on a simple two-level problem<sup>1</sup> indicates that guides, cross-checks, equivalences, etc. will prove even more useful and important for the complex multilevel settings in which HLM and other multilevel programs are now being applied.

#### Examples with "Continuous" observation time

In the spirit of give 'em what they ask for, we follow up on comments by Raudenbush and by Goldstein with additional examples. The issue raised was that in our longitudinal examples, the same (integer) times of observation for all individuals (e.g., {0,1,2,3,4}) gave examples with extreme balance, and that balance is seen to lead to the equivalences we illustrate. Now the longitudinal research example, by its design, will typically have more "balance" than say a school effects example (e.g., closer to equal numbers of observations per unit) but we can expand the examples in our paper to include times of observations that differ over individuals (within reasonable constraints) for both complete data and missing data situations.

The examples are constructed with the same underlying population of individual growth curves (and thus the same values of the parameters of interest) as in the example in Table 2. Starting with a longitudinal design with integer times of observation {0, 1, 2, 3, 4} (call these the "anchor" times), actual times of observation for each individual are created by adding successively to each of the anchor times realizations of a Uniform random variable with range [-.5, .5]. Examples of resulting observation times for individuals are: {-.23, .52, 1.79, 3.12, 4.30} and {-.22, 1.29, 1.99, 3.22, 3.58}. The Y-

observations are constructed based on these times of observation in the same manner as the example in Table 2. (This structure updates our examples to bring them closer to the longitudinal example in Bryk & Raudenbush, 1992, Chap. 6).

Analysis of a sample of 100 individuals, each with 5 longitudinal observations, produces the same kind of correspondences seen in the three examples in our paper, even for the SFYS. Start with  $\gamma$  ( $=.671$ ). HLM (version 3.01) C\*BY gives estimates .590 with standard error .108, and SFYS (just computing the 100  $\theta$ -values and doing an OLS fit of  $\theta$  on  $Z$ ) gets estimate .581, and s.e. .108. For  $\mu_\theta$  ( $=5.0$ ), HLM C\*BN gives estimate 5.25, s.e. = .250, and SFYS gets estimate 5.24 with s.e. = .249. Crude method of moments estimates were adapted from App. B by averaging over individuals quantities depending on the  $t_i$  (e.g. SSt). For estimating  $\sigma_\theta^2$  ( $= 5.0$ ), HLM CNBN (and also grand-mean centering) gives 5.09, CYBN 5.08, and our estimate is 5.04. For  $\rho(\theta)$  ( $= .806$ ), HLM CNBN gives .812, CYBN gives .811 and our estimate is .813. Finally, for estimating  $\rho_{\eta(0)\theta}$  ( $= -.31$ ), HLM CNBN gives  $-.399$ , and our estimate is  $-.402$ .

Constructing a further example with approximately 16% of the 500 Y-observations above made missing (a moderate missing data situation) produces the same story. For  $\gamma$  ( $=.671$ ), HLM CNBY gives estimate .6010 with standard error .1141, CYBY gives estimate .6018 with s.e. .1139, and SFYS gets estimate .5945 with s.e. .1125. For  $\mu_\theta$  ( $= 5.0$ ), HLM C\*BN gives estimate 5.17 with s.e. = .262, and SFYS gets estimate 5.13 with s.e. = .258. For  $\sigma_\theta^2$  ( $= 5.0$ ), HLM CNBN (and grand-mean centering) give 5.199, CYBN 5.186, and our estimate is 5.115. For  $\rho(\theta)$  ( $= .806$ ) HLM CNBN gives .757, CYBN gives .756 and our estimate is .770.

## References

[author note: only the Neter, et al (1990) and Raudenbush, S.W. (1993) references are not given in main paper; is it necessary to list the others a second time?]

- 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*.
- Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models*, Third edition. Homewood IL: Irwin.
- Raudenbush, S.W. (1993). Hierarchical linear models and experimental design. In Edwards, L. K. (Ed.), *Applied analysis of variance in behavioral science*. New York: Marcell Decker, 459-496.
- Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 92, 726-748.