Review of: Joop Hox
Multilevel Analysis, Erlbaum, 2002

Jan de Leeuw
UCLA Statistics

July 14, 2003

1 Target Audience

There are now quite a number of books on multilevel analysis targeted at applied researchers in the behavioral and social sciences. They are not intended primarily for methodologists, psychometricians, or statisticians. The idea is to provide information about this relatively new class of techniques to researchers who want to use them in their research.

The latest entry in the field is the book by Joop Hox, reviewed here. Earlier competitors are the classic work by Bryk and Raudenbush [1992], now Raudenbush and Bryk [2002], and the textbooks by Snijders and Bosker [1999] and Kreft and de Leeuw [1998]. The books by Longford [1993] and by Goldstein [1995], no matter how excellent, are written for an essentially different target group, with more statistical sophistication. And there are the edited books by Reise and Duan [2003] and by Leyland and Goldstein [2001]. Although these books cover the same material, they cannot really be compared, because they do not speak with a single voice, and their contributions vary wildly in terms of scope and level of difficulty.

2 Level and Content

So how does the book of Hox compare with the others ? Let's first look briefly at the technical level: how much statistics and mathematics is required to follow all the details ? The order of the existing books is fairly clear. Raudenbush and Bryk is the most demanding, followed by Snijders and Bosker, than by Hox, and finally by Kreft and De Leeuw. All an applied researcher needs to read Hox from cover to cover is some knowledge of and experience with regression and analysis of variance as it is practiced in the social and behavioral sciences.

More importantly, let's now look at the content of the book, chapter by chapter. What does Hox cover ? And how does this compare with the other multilevel textbooks ?

1: Introduction to multilevel analysis, 10 pages The motivational chapter, which is (surprisingly) very short. It emphasize some of the older sociological literature on ecological fallacies, cross-level inference, and the effect of group characteristics on individual behavior. Not very inspired and not very inspiring, I think.

2: The basic two-level regression model: Introduction, 25 pages This introduces the hierarchical linear model (HLM) in both single equation and multiple equation form. It discusses estimation (briefly), intra-class correlation, shrinkage and Empirical Bayes, residual plots, software and notation. A solid introduction, but the devil is in the details (cf. below).

3: Estimation and hypothesis testing in multilevel regression, 12 pages Maximum likelihood, Bayesian estimation, and the Bootstrap are explained, although not quite correctly in some cases (cf. below). Sometimes the discussion is quite surreal from the statistical point of view, because statistical questions are often decided on the basis of simulation studies. The AIC and BIC are introduced as yet two other pieces of statistical magic.

4: Some important methodological and statistical issues, 23 pages Here Hox discusses "analysis strategy", which is a stepwise approach of building up an appropriate model. He also discusses centering and interaction. This is a nice chapter, which stays close to the data. Unfortunately it derails a bit at the end, when the bugaboo of "explained variance" leads, as usual, to wishful juggling of variance ratio's.

5: Analyzing longitudinal data, 30 pages Repeated measures analysis fits nicely into the multilevel framework and this chapter explains competently why this is the case. It analyzes some examples in detail.

6: The logistic model for dichotomous data and proportions, 20 pages This chapter has a quick introduction to generalized linear models. It also discusses and compares the MQL, PQL, and numerical integration methods to fit generalized HLM's. It is informative, and has good examples. This is a very active research area in statistics at the moment, and it is impossible for a textbook to be completely up-to-date. Laplace expansion methods are not mentioned, for instance.

7: Cross-classified multilevel models, 16 pages If there are deviations from a strictly hierarchical data structure (students nested in primary schools and in secondary schools, but primary schools crossed with secondary schools) then design matrices become more complicated. This chapter has good explanations and examples.

8: The multilevel approach to meta-analysis, 17 pages Multilevel analysis, in particular as implemented in the program HLM, is one of the core techniques used in meta-analysis. The chapter describes the way meta-analysis fits into the multilevel framework, and gives examples. Reading about meta-analysis always makes me feel uncomfortable, but despite of that I think this chapter does a good job.

9: Multivariate multilevel models, 16 pages Multiple outcomes can be incorporated in multilevel analysis by adding an additional level, nested within the other levels. This is a clever trick which is explained and illustrated well in this chapter.

10: Sample sizes and power analysis in multilevel regression, 24 pages Sample size determination and power analysis always have certain ritualistic aspects, with researchers willing to make the wildest assumptions and guesses to satisfy their NIH or APA reviewers. Power analysis for multi-parameter models is notoriously problematic. In this chapter Hox reviews simulation studies and the few analytical results that are available.

11: Advanced methods of estimation and testing, 28 pages In this chapter, true to its daunting title, Hox give a brief introductions to profile likelihood, Bayesian statistics, MCMC methods, and the bootstrap. And he illustrates this with examples. I am not sure if it is wise to include a chapter like this, because it is unclear if the applied researchers get enough information to work with.

12: Multilevel factor models, 26 pages Here the book goes off the beaten path. It gives a short introduction to structural equation models (SEM), and then discusses the work of Goldstein, McDonald and Muthén, extending SEM to hierarchical data. This is no longer regression analysis, and new problems arise. Also, the reviewed developments are still incomplete and quite different from the random coefficient approaches to regression analysis. Again, one wonders if it was a good idea to include multilevel SEM in a book of this type.

13: Multilevel path models, pages, 12 pages More SEM. This chapter wanders off into the jungle of infinitely complicated multilevel models, with more and more elaborate black-box software to compute parameter estimates, standard errors and significance tests. It shows that the sins of SEM can be repeated at multiple levels, and that applied researchers need much more strong prior knowledge than they usually have to venture into these computer programs. One can argue that this chapter takes us too far astray, as before, but one can also argue that it promotes bad habits.

14: Latent growth curve model, 12 pages In this final SEM chapter specific latent variable models for growth curves (longitudinal data) are considered. More or less the same comments apply as in the previous chapters. Given the fact that Hox has decided to include SEM I guess this chapter had to be included as well.

We see that the number of topics covered in this book is huge. It easily beats out Raudenbush and Bryk, Snijders and Bosker, and Kreft and De Leeuw. As a necessary consequence, the book is "wide" rather than "deep". This is, of course, not always a problem, because Hox gives many references and those who want to go deeper can find details the relevant literature. On the other hand, discussing so many different topics at a rather superficial level can never make a really good book. Wide, rather than deep, tends to become shallow.

Not much of the contents is based on Hox own research, what we see is mostly a compendium of other people's results. As compendiums go, this one is done competently, although the book often fails if it tries to explain the technical aspects of the techniques. That creates a problem, especially in the later chapters, where more and more complicated models are explained less and less adequately. Fortunately in these later chapters Hox puts more and more emphasis on the examples.

One of the consequences of the "wide" approach taken by Hox is that some of the key concepts used in the text remain vague, presumably because they are too "technical". It is not explained why maximum likelihood is the ruling method of estimation in statistics, what the precise difference is between full and restricted maximum likelihood, and why we add two times the degrees of freedom to the log-likelihood to get the AIC. This has the unfortunate corollary that some of the concepts and techniques appear as black boxes, forcing the reader's local statistician in the role of magician. Of course many statisticians very much like to play that role, but the consensus these days is that too much of this role-playing is bad, both for the statistician and for the client.

3 Criticism

We now get to my main problem with the book. At too many places the author says things which are imprecise or even wrong. Let me give some examples from the first three chapters. I could make the list quite a bit longer, but the one below is probably long enough to make my point. The same sort of imprecisions and half-truths also occur in the discussions the more complicated models in the later chapters (although surprisingly enough somewhat less frequently).

- It is not true that data with a hierarchical structure necessarily come from a multistage sample (p. 5, also p. 81). There are many sampling designs that lead to hierarchical data. In fact Hox himself, on the same page, describes cluster sampling and stratified sampling. Multilevel techniques are even useful is we look at all pupils in all schools in the district, and we have no sample at all.
- Random coefficient (RC) models are more general than multilevel models, and mixed linear models are again more general. Variance components models are less general than mixed linear models. But certainly these terms do not describe the same class of models (p. 11), and in fact there is nothing inherently hierarchical about mixed models or RC models or variance component models.
- RC models do not get their name from the fact that intercepts and slopes vary (p. 12) over groups. Recent statistical literature discusses varying coefficient model in detail. in which regression coefficients are different but not necessarily random. In RC models, coefficients are modeled as random variables, and that explains their name.
- Across all classes, regression coefficients have a distribution with some mean and variance (p.12). This is fundamentally wrong. The regression coefficients in each class are random variables, they have a distribution, which is generally different for different classes. This basic error, which seems to confuse sample correlation and population correlation is repeated in other places (p. 31).
- The intraclass correlation is not the expected correlation between two randomly chosen units in the same group (p. 15). It is the population value of the correlation between any two units in the same group.
- The likelihood function is not the probability of observing the sample data (p. 16). In the case of continuous data, the probability of observing the sample data is always zero.
- The deviance does not have a chi square distribution (p. 16), but an approximate chi square distribution if the sample size is large (and in multilevel models it is not immediately obvious what sample size is).
- With large samples ML estimates are usually robust against misspecification (p. 37). Certainly not true in general. ML produces consistent estimates if the residuals are non-normal, but so do OLS estimates, which are even unbiased.
- It is never the case that the iterative ML algorithms sometimes never stop (p. 39). In the first place which algorithm is this ? EM or iterative reweighted least squares or Newton ? Second, with badly specified models and boundary solutions the convergence speed can become sublinear, but we will still have convergence (theoretically).
- Generalized least squares is not one-step of the maximum likelihood algorithm (p. 39). Which algorithm ? Which likelihood, for that matter ? And one step from where ?

- The distinction between GEE and ML is not “tricky” (p. 40). In fact, this distinction has been very clearly explained in very many places.

Hox is clearly most at ease when he is actually analyzing and presenting examples. One interpretation is that he himself is closest to the applied researchers, and the book summarizes what he has learned by listening to various methodologists and statisticians. Unfortunately, he has not always been paying attention, or, more likely, matters have not been explained correctly to him. The lack of understanding of the basics is solved, as in many similar social science methodology texts, by not explaining key concepts, by deferring to authorities, and by maximizing the number of references that supposedly give the necessary explanations. Thus the book becomes a uneasy hybrid of an encyclopedia of other people’s research and a manual for one or more computer programs. It is not clear to me why Hox wanted to write a book discussing so many elaborations of multilevel techniques, some of them quite esoteric and tentative, especially when his strength is clearly to perform and explain actual data analyses. I can recommend the book because of these recipes and examples, and as a possible gateway to the serious literature, but not as a textbook explaining the methodology.

A final minor point, but one that annoys me mightily, although perhaps irrationally so. The book looks needlessly amateurish because of the ugly typesetting, especially of in-line mathematical material. In many places, subscripts are too close or too high, symbols float above the line, and some parts of the book give the impression the type setter was drunk at the time (see, for example, the paragraph in the middle of page 34). Clearly the publishers’s software was inadequate for the job at hand. There is no excuse for such sloppiness, because anybody with a minimal TEX system can make something which looks a lot better.

References

A.S. Bryk and S.W. Raudenbush. *Hierarchical Linear Models. Applications and data analysis methods*. Number 1 in Advanced Quantitative Technqies in the Social Sciences. Sage Publications, Newbury Park, CA., first edition, 1992.

H. Goldstein. *Multilevel Statistical Models*. Number 3 in Kendall’s Library of Statistics. Arnold, London, second edition, 1995.

I. Kreft and J. de Leeuw. *Introducing Multilevel Modeling*. Sage Publications, London, 1998.

A.H. Leyland and H. Goldstein, editors. *Multilevel Modelling of Health Statistics*. Wiley Series in Probability and Statistics. Wiley, Chichester, 2001.

N.T. Longford. *Random Coefficient Models*. Number 11 in Oxford Statistical Science Series. Clarendon Press, Oxford, 1993.

S.W. Raudenbush and A.S. Bryk. *Hierarchical Linear Models. Applications and data analysis methods*. Number 1 in Advanced Quantitative Technqies in the Social Sciences. Sage Publications, Thousand Oaks, CA., second edition, 2002.

S.P. Reise and N. Duan, editors. *Multilevel Modeling. Methodological advances, issues, and application.* Multivariate Applications. Erlbaum, 2003.

T.A.B Snijders and R.J. Bosker. *Multilevel Analysis. An introdution to basic and advanced multilevel modeling*. Sage Publications, London, 1999.