Two-Stage Regression and Multilevel Modeling: A Commentary

Andrew Gelman

*Department of Statistics and Department of Political Science,
Columbia University, 1255 Amsterdam Ave (at 122 St.),
Room 1016, New York, NY 10027-5904
e-mail: gelman@stat.columbia.edu*

1 Introduction

These articles demonstrate, in several different examples, the effectiveness of two-level regression: the procedure of fitting several separate regression models, and then fitting a second, higher-level, regression to the estimated coefficients (for example, fitting a separate regression model to survey data from each of several countries, then regressing the coefficient estimates on country-level predictors). For simplicity, we will refer to the first- and second-level units as "persons" and "countries," respectively, but our points apply more generally.

2 Two-Level Regression and Multilevel Modeling

Two-level regression can be viewed as a special case of multilevel (hierarchical) modeling: we can obtain the fit-to-each-country-separately two-level results by fitting a multilevel model with the group-level variance parameter set to infinity (in which case the group-level model does no smoothing of the individual regression coefficients). It is generally preferable to estimate the group-level variance, and thus the appropriate amount of between-country smoothing, from the data. However, as pointed out by Jusko and Shively and by Duch and Stevenson, running separate regressions for each country is simpler to perform than multilevel modeling and gives essentially the same answers if the number of persons per country in the dataset is large. Some of the articles under discussion consider two-stage and multilevel regression to be competing methods, but any competition here arises largely from practical considerations of fitting. Such practical concerns are inevitable in statistical modeling. For example, in almost all uses of normal distribution, I suspect that the t distribution—which includes the normal as a special case—would be better (see, for example, Lange et al. 1989), but I rarely go to the trouble to fit this more complicated model.

As is usual in statistics, the model is more important than the particular method used to fix it. I applaud these articles in fitting models that allow parameters to vary by country—models that can be motivated empirically and also theoretically (as demonstrated by Kedar).

I am not so interested in arguments about estimation methods because I expect that when a method becomes less useful, practitioners will switch to improved estimators. Sometimes this process can be painful—for example, Lewis and Linzer and Huber et al. devote so much effort to constructing weights that I suspect it would be less effort for them to just bite the bullet and fit full multilevel models. But readers of this journal can make their own judgments about the most effective way to fit their varying-coefficient regressions.

3   Understanding the Fitted Model

Once a complex model has been fit, the crucial next step is to understand the estimated parameters, which in these examples include individual regression coefficients, level-2 (between-country) regression coefficients, and model estimates for the individual countries. With many countries being studied, it is helpful to display inferences graphically. Just as graphics are essential for revealing unanticipated patterns in raw data (Tukey 1977), they also give us a chance to learn in new ways about parameter estimates. I applaud the creative displays of inferences such as Fig. 2 of Jusko and Shively, Fig. 2 of Kedar, and Fig. 1 of Duch and Stevenson (although I am puzzled as to why some of the point estimates in that last figure fall outside the confidence intervals). The article by Bowers and Drake gives an excellent illustration of how graphical displays of data and inferences together can lead to better understanding; I particularly like their Fig. 2.

The graphs in these articles could be improved with a little effort. I discuss one example in each article, not to criticize these particular authors but rather to demonstrate the widespread potential for improvement. The points in Jusko and Shively's Fig. 3 could be connected by lines, and then each line could be labeled on the right of the graph. This would avoid the need for the reader to go back and forth several times between the symbols and the labels and would more directly reveal the patterns in the estimates. A similar comment applies to the figures of Huber et al., which in addition could be made much smaller: all of Figs. 2–5 could be placed on a single-page grid of figures, thus allowing the reader to parse all these results at once. Figure 2 of Bowers could be ordered from top to bottom, and the ordering could be explained in the caption. The countries in Fig. 1 of Duch and Stevenson could be ordered based on some meaningful country-level predictor rather than alphabetically—why waste this chance to reveal potential patterns? The country names in Kedar's Fig. 2 could be spelled out—there is lots of space to do so— the axes on the graphs clearly labeled, and the distracting dotted lines removed (see Tufte 1983). Figure 4 of Lewis and Linzer could be displayed as a scatterplot of standard errors versus some relevant country-level measure to better understand this variation.

More important, the information in the tables of all six of these articles could be displayed graphically (for general guidelines, see Gelman et al. 2002). I will not go through another long list, but, for example, Kedar's Table 4 could be put in a graph, taking up less space than the original table and giving the reader a sense of country-level variation in the predictors and outcomes of interest in this study. Given all the effort that has gone into collecting the data, formulating the model, and estimating parameters, why not take a better look at the data and the resulting inferences—especially since this actually takes up less space than displaying tables of numbers such as "0.507" that are essentially uninterpretable? I have made enough such mistakes in my own work to recognize them in others!

Such concerns may seem picky—but what is the point of fitting a model, if not to be able to understand it? The more effective our displays, the more models we can grasp, and the more effective can be our model fitting and data analysis.

   The articles under discussion also demonstrate in various ways the effectiveness of numerical summaries in understanding a fitted model. I particularly like Duch and Stevenson's discussion of computing predictive effects averaged over the sample, partly because this is an area of my own research (Gelman and Pardoe 2004).

   As always, analysts should be cautious about giving causal interpretations to inferences from observational data. For example, Huber et al. write, "An increase in the number of legislative parties . . . decreases the propensity of an individual to have party attachments." This claim comes from a between-country regression analysis controlling for several other factors including "number of elective parties," "vote for candidate," and "age of parties." It is a big leap to interpret differences between countries as a potential effect of a change within a country, especially given that increasing the number of legislative parties could change the values of the other predictors.

4   Looking toward the Future

Already, multilevel models are easier to write than two-level regressions (compare, for example, the simplicity of the models in Kreft and De Leeuw [1998] to the complicated matrix algebra in some of the articles here). As computer programs improve, we expect that soon it will be as easy to fit a multilevel model as it is currently to run a regression. At this point, political scientists and statisticians will no longer need to worry about technicalities such as weighted estimation. Even now, we can fit many multilevel models directly. For example, there is nothing wrong with the algebra in the article by Lewis and Linzer, but I would prefer to fit a model in Bugs (Spiegelhalter et al. 1994, 2002; see also, for example, Lancaster 2004) and put my effort into understanding the data and model, and forget about OLS, WLS, FGLS, EDV, and any other acronyms I might have missed.

   In the meantime, two-level analysis can be a useful technique in light of current computational limitations. The motivation for full multilevel modeling arises when the sample size in each group is less than overwhelming—from our own recent research, these examples include estimating state-level opinions from national polls (Park et al. 2004) and estimating an ideal point model of Supreme Court voting from at most nine votes per case (Bafumi et al. 2005). Such research endeavors are pushing us toward the development of software with which two-stage regressions—and more complicated models—can be fit routinely using multilevel modeling.

References

Bafumi, J., A. Gelman, and D. K. Park. 2005. "Issues in Ideal Point Estimation." *Political Analysis.*

Gelman, A., and I. Pardoe. 2004. "Average Predictive Effects for Models with Nonlinearity, Interactions, and Variance Components." Technical report, Department of Statistics, Columbia University.

Gelman, A., C. Pasarica, and R. Dodhia. 2002. "Let's Practice What We Preach: Using Graphs Instead of Tables." *American Statistician* 56:121–130.

Kreft, I., and J. De Leeuw. 1998. *Introducing Multilevel Modeling.* London: Sage.

Lancaster, T. 2004. *An Introduction to Modern Bayesian Econometrics.* Oxford: Blackwell.

Lange, K. L., R. J. A. Little, and J. M. G. Taylor. 1989. "Robust Statistical Modeling Using the t Distribution." *Journal of the American Statistical Association* 84:881–896.

Park, D. K., A. Gelman, and J. Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12:375–385.

Spiegelhalter, D., A. Thomas, N. Best, W. Gilks, and D. Lunn. 1994, 2002. BUGS: Bayesian Inference Using Gibbs Sampling. MRC Biostatistics Unit, Cambridge, England. (Available from www.mrc-bsu.cam.ac.uk/bugs/.)

Tufte, E. R. 1983. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

Tukey, J. W. 1977. *Exploratory Data Analysis.* Reading, MA: Addison-Wesley.