

Fitting Multilevel Models When Predictors and Group Effects Correlate*

Joseph Bafumi and Andrew Gelman

*Prepared for the 2006 Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Random effects models (that is, regressions with varying intercepts that are modeled with error) are avoided by some social scientists because of potential issues with bias and uncertainty estimates. Particularly, when one or more predictors correlate with the group or unit effects, a key Gauss-Markov assumption is violated and estimates are compromised. However, this problem can easily be solved by including the average of each individual-level predictors in the group-level regression. We explain the solution, demonstrate its effectiveness using simulations, show how it can be applied in some commonly-used statistical software, and discuss its potential for substantive modeling.

1 Introduction

1.1 What's the Problem?

Researchers are leery of fitting random effects (better called modeled varying intercepts) in models where predictors and units may correlate. Such models have compromised estimates of uncertainty as well as possible bias (Hausman and Taylor 1981). This issue is most intuitively shown within the framework of a multilevel model. Equation (1) shows a individual-level equation where some outcome y_i is being predicted by modeled varying intercepts (or random effects) α_s and a predictor x_i .¹ The error in this regression is denoted ϵ_i . Equation (2) shows a group-level equation that estimates the mean of the varying intercepts α_0 's and the group-level error η_s .

$$y_i = \alpha_{s[i]} + \beta x_i + \epsilon_i, \text{ for } i = 1, \dots, n, \quad (1)$$

where $s[i]$ is the group s containing unit i .

$$\alpha_s = \alpha_0 + \eta_s, \text{ for } s = 1, \dots, S, \quad (2)$$

where η_s are group-level errors.

Assume that x_i and the varying intercepts $\alpha_{s[i]}$ correlate. If this correlation is not modeled, it will be absorbed into the error term η_s of (2), which results in the violation of a key Gauss-Markov assumption. To see why, substitute (2) into (1). The ϵ_i and η_s error terms combine to create a new error term, ϵ'_i :

$$y_i = \alpha_0 + \beta x_i + \epsilon'_i, \text{ for } i = 1, \dots, n. \quad (3)$$

The regression error ϵ'_i now correlates with the predictor in the model. This violation of a Gauss-Markov assumption may result in poor estimates of parameter uncertainty.

¹The subscript s is used because, later, we will look at an applied examples where the units are the American states.

This may make parameter estimates seem more precise than they really are. In turn, this will lead to inflated reports of statistical significance in regression analysis and a greater tendency to reject the null hypothesis than is warranted. Further, estimates may also be biased.

1.2 Why Do We Care?

Modeled varying parameters have been shown to have better statistical properties than their unmodeled or non-varying counterparts. This has been shown when analyzing data with low sample size per group (Park, Gelman and Bafumi 2004), when studying time series/cross sectional data (Western 1998; Beck and Katz 2006; Shor et al. 2005) and in a variety of contexts by Bartels (1996). It is the partial pooling that varying intercepts and varying coefficients undergo that provides the added benefit. With partial pooling, outlying groups provide some information toward parameter estimation but also are shrunk to the mean. The extent of information they provide and, inversely, the extent of their shrinkage, is determined by the amount of data in their (and in other) groups.

Such modeled varying parameters are popular in some areas of social science research but have been slower to gain popularity in others. This is unfortunate given the promise such a specification offers. Their relative obscurity can partly be blamed on the problematic correlation discussed above. To avoid the problem, econometricians have preferred to completely pool or not pool at all estimates that may vary across groups. In the next section, we reanalyze modeled varying intercepts in the context of these more popular competitors. Further, we propose a solution to the problematic correlation between predictors and group effects that allows researchers to more comfortably estimate modeled varying parameters.

2 Econometric Framework

One can estimate a model ignoring group effects. For example, with data that has individuals nested in states or regions, one could ignore the state or region effects and estimate a common (geographic) intercept for each individual in the model. This is shown in (4), where α_0 does not vary across groups.² This is a complete pooling model, since groups are completely pooled together as if they make no difference.

$$y_i = \alpha_0 + \beta x_i + \epsilon_i, \quad (4)$$

An improved model would allow for group effects. Equation (5) shows a model with varying intercepts denoted $\alpha_s^{\text{unmodeled}}$. The varying intercepts are superscripted unmodeled since an error term is not estimated. Rather, the variance of the parameters is set to infinity to allow for maximal variation given the data in the estimated group effects. For the unmodeled varying intercepts, this is equivalent to running separate regressions for each group. This specification is often referred to as the fixed effects or least squares dummy variable approach.³ Here, there is no shrinkage to the mean and chance outliers risk overinfluencing estimates.

$$y_i = \alpha_{s[i]}^{\text{unmodeled}} + \beta x_i + \epsilon_i, \quad (5)$$

$$\text{where } \alpha_s^{\text{unmodeled}} \sim N(\alpha_0, \infty).$$

The preferred specification for many data sets is varying intercepts that are modeled with error. This is shown in (6) and (7). An error, η_s , is estimated for the varying intercepts. A group effect is partially pooled contingent on how much data informs it and how much data informs the other effects. Each group borrows and offers information to

²The same can be said of β .

³Some distinguish the two in that fixed effects may refer to the procedure where group effects are subtracted out rather than being picked-up by a set of indicators. The two yield equivalent results but one must deal with the inflated sense of degrees of freedom that come about with fixed effects.

all other groups for optimal estimation.

$$y_i = \alpha_{s[i]}^{\text{modeled}} + \beta x_i + \epsilon_i, \quad (6)$$

$$\alpha_s^{\text{modeled}} = \alpha_0 + \eta_s, \quad (7)$$

However, it is well known that when the predictors and the groups in a model correlate, these models risk poor estimates of uncertainty. For this reason, econometricians rely most heavily on the specification where there is no pooling across groups, as in (5). This presents little costs when econometricians are really focused on β and do not care so much about the α_s . This is because (6), without the problematic correlation between the predictors and the groups, and (5) yield the same estimate for β . However, many of us are interested in the how the intercepts vary across groups or we may be interested in varying the coefficient β across groups. To satisfy this interest, it is important to deal with the problematic correlation.

Is there a solution? Can one fit a multilevel model with varying intercepts (or coefficients) when the units and predictors correlate? The answer is yes. And the solution is simple. The problem can usefully be viewed as an omitted variable bias. Once a model is as well specified as a researcher deems possible, and if the correlation between the units and the predictor still exists, one can remove the correlation with the predictor from the group-level error by calculating the mean of the predictor at each unit and including it as a group-level predictor. In equation (9), α_1 represents the coefficient for this new, second level regressor. By accounted for x_i 's correlation with the varying effects, the error term η_s is free of this pattern and a violation of the Gauss-Markov assumption does not occur. This leaves researchers free to estimate multilevel models with varying intercepts.⁴

$$y_i = \alpha_{s[i]}^{\text{modeled}} + \beta x_i + \epsilon_i, \quad (8)$$

⁴The same logic can be applied to modeled varying coefficients.

$$\alpha_s^{\text{modeled}} = \alpha_0 + \alpha_1 \bar{x}_s + \eta_s, \quad (9)$$

3 Simulations

Simulations can illustrate the problem and the solution further. First, we generate a random normal predictor of length 100 with a mean of 0 and a standard deviation of 2. Then, we generate an outcome (often called a dependent variable) that is equal to the predictor plus random normal noise with a mean of 0 and a standard deviation of 7. This ensures a strong, but not perfect, correlation between the two variables. Units effects are added to the outcome by adding a random normal component with a nonzero mean to each quarter of the data. So, for example, a set of random normal values with a mean of 1 and a tight standard deviation of .001 are added to the first 25 observations in the outcome.⁵

We start by predicting the outcome by the varying unit effects and the explanatory variable free of unit effects. These results should not be problematic since the units and the predictor do not correlate. Next, we estimate the same equation but with a predictor that, like the outcome, varies across the units. We estimate each equation 1000 times and record the coefficient and standard error of the key predictor. We plot a histogram of the t statistics (the coefficient divided by the standard error) calculated in each of the 1000 simulations.

Figure 1 shows that the t statistic for the model where the predictor does not correlate with the units tends to be smaller than the model where the correlation exists. The larger t statistic in the second plot results in an inflated sense of statistical significance in parameter analysis and a greater tendency to falsely reject the null hypothesis in research works, as mentioned earlier. This is the problem that researchers are cautioned against when estimating modeled varying intercepts. The problem is thought to be so severe, that this model is often cast aside as inviable.

⁵The means for the next three quarters are -1,-3 and 2. The standard deviations remain tight at .001

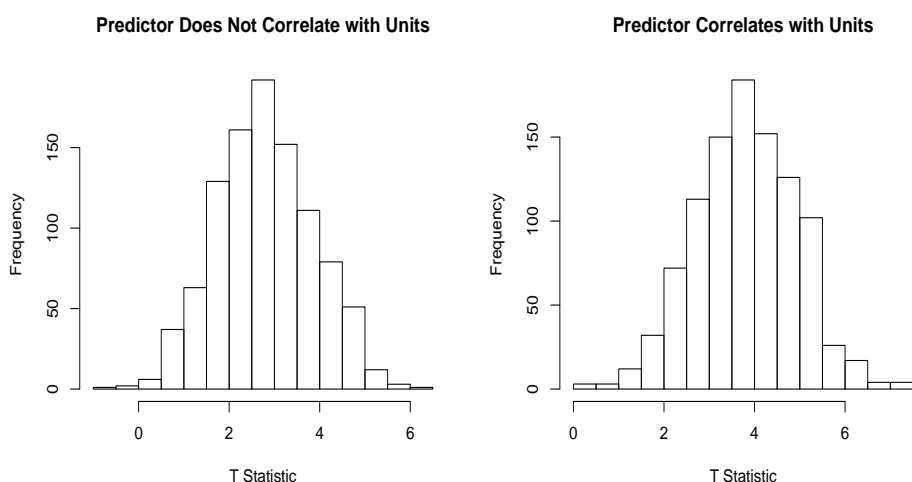


Figure 1: *Simulated t statistics for β in two varying intercept (modeled with error) models: one where the predictor does not correlate with the units and one where it does. 1000 regressions were fit on data sets of length 100. Four group effects (one for each quarter of the data) were created.*

To see if the solution highlighted above works as promised, let's run another simulation. We have seen the t statistics when the correlation between the units and the predictor does not exist and we have seen how the statistic has grown when the correlation is instituted. Now consider a model where the correlation exists but the mean of the predictor per unit is included as a group-level predictor. Figure 2 shows the result. The first two plots are as before. Now, we add a third plot, showing the t statistic when the group-level regressor is added to the model.

The t statistic for the key predictor looks virtually identical as in the model with no correlation between the group effects and the predictor. The additional group-level covariate successfully accounted for the correlation before it fell into the group-level error term and caused a problem. In the next section, we will discuss how to run multilevel models incorporating the above fix in commonly used software such as R and Stata.

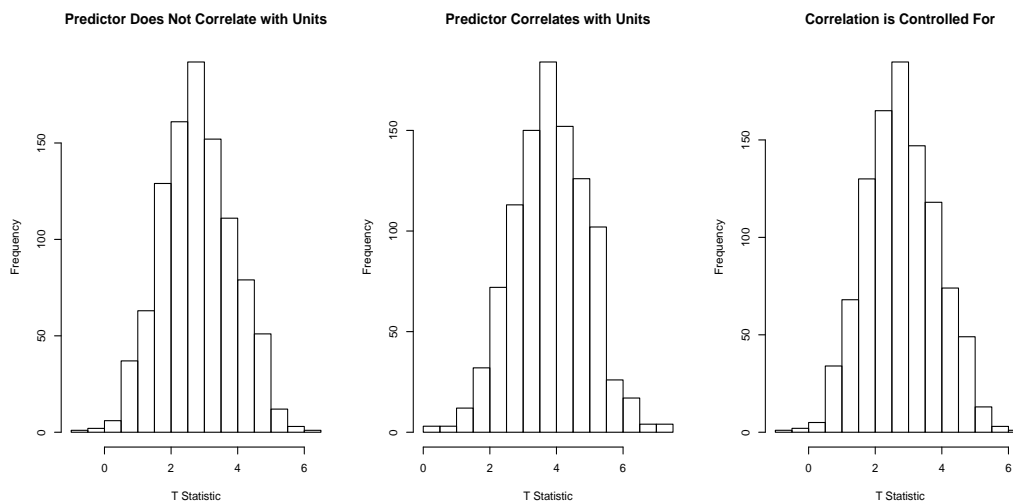


Figure 2: *Simulated t statistics for β in three models: one where the predictor does not correlate with the units, one where it does and one where it does and the correlation is controlled for with a second level regressor equal to the predictor at its mean per unit. 1000 regressions were fit on data sets of length 100. Four group effects (one for each quarter of the data) were created. The first two plots are the same as those shown in Figure 1. The key finding here is that the values of the t statistics are alike in a model without the problematic correlation and in a model with the correlation and the fix proposed in this paper.*

4 Practical Issues of Fitting

Multilevel models have begun to take hold in the political science literature as well as in other disciplines (for applied examples in political science, see Gelman and King (1990, 1993, 1994); Gelman and Little (1997); Reilly, Gelman and King (2001); Steenbergen and Jones (2002); Park, Gelman and Bafumi (2004); Bafumi (2004*a,b*); Gelman et al. (2005) and the collection of papers in Political Analysis, Volume 13). In response, major statistical software programs such as R and Stata now incorporate easy to use code to run these models. Most early multilevel modelers who did not write their own estimation code turned to Bugs (Bayesian inference using the Gibbs Sampler) to fit their models. This program proved to be very flexible although it often took a great deal of time to iterate, was prone to trap, required start-up costs and required an understanding of Bayesian updating. Not to be outdone, R and Stata programmers built in pre-packaged code so that practitioners could more easily and much more quickly fit multilevel models.

Bugs models are programmed just as one would write the mathematical notation for a multilevel model.⁶ The notation is repeated below.

$$y_i = \alpha_{s[i]} + \beta x_i + \epsilon_i, \quad (10)$$

$$\alpha_s = \alpha_0 + \alpha_1 \bar{x}_s + \eta_s, \quad (11)$$

In Bugs, a practitioner would specify priors and starting values and run the model for a pre-specified number of iterations in search of parameter convergence. Via the `lmer()` function in R and the `xtmixed` command in Stata, these programs can also run multilevel models. However, they do not easily allow for predictors at a secondary level such as \bar{x}_s . So, for example, if one has individuals nested in states with individual level and state level predictors (most applicably, an individual level variable calculated at it's mean per state), one needs to incorporate the group-level variables into the single equation.

This is accomplished in R and Stata by generating a variable equal in length to the individual-level predictors but varying only across the units or groups. So, this new group-level predictor will have the same value for each group of, say, states. This variable could be, as we discussed above, the mean per unit of an individual-level variable or some other substantive variable measured at the group level. The parameter estimate for the former will be equivalent to α_1 above.

We do not consider varying coefficients very deeply here but practitioners will often want to vary a predictor across the group effects and regress the varying coefficients by group-level variables. Again, this is intuitive in Bugs where two or more different equations with separate error terms are specified. In R or Stata, a practitioner would achieve the proper specification by interacting the expanded group-level covariate explained above with the variable that is modeled to vary across groups. The coefficient of this interaction will offer the estimate of the group-level covariate predicting the varying

⁶It is not our purpose to teach the programming language used in these software programs but to show that one can deal with group-level variables (such as the solution to dealing with a correlation between the units and predictors in a model) in each program.

coefficients. This coefficient and α_1 serve to wipe away a problematic correlation; but, as we shall see next, they also may offer an interesting substantive result.

5 Red/Blue Example

Gelman et al. (2005) wish to understand the paradox of how poor states vote Republican and rich states vote Democratic while individuals are known to vote in more traditional ways. We used a model where the Republican vote is predicted by varying state intercepts, varying coefficients for individual-level income, and average state income predicting both the varying intercepts and coefficients (at the group level). We included average state income in the model, partly to alleviate the possibility of a correlation between income and state effects.⁷ However, it also offers an interesting substantive result.

The main findings in Gelman et al. (2005) are revealed in Figure 3. It tells an individual-level and a state-level story. First, the probability of voting Republican increases for individuals as a whole as their income levels rise. However, income matters little or none at all for individuals in richer states (as shown by Connecticut) and a lot for individuals in poorer states (as Mississippi makes clear). Meanwhile, it also shows that as average state income (labeled with solid, black dots) goes up, states are less likely to be Republican. These findings are summarized by the group-level coefficients discussed above. The coefficient and standard error for average income predicting the varying intercepts are -1.7 and 0.2. It is negative and significant. Similarly, the coefficient and standard error for average income predicting the varying coefficients are -0.6 and 0.1. The parameter predicting the intercepts shows that as average state income increases, states are less likely to be Republican. The parameter predicting the varying income coefficients show that as average state income rises, income is a poorer predictor of the vote.

The inclusion of the correlating predictor measured at its mean per group is important

⁷Average state income is not calculated directly from individual level income but works in the same way. It is obtained from U.S. census data.

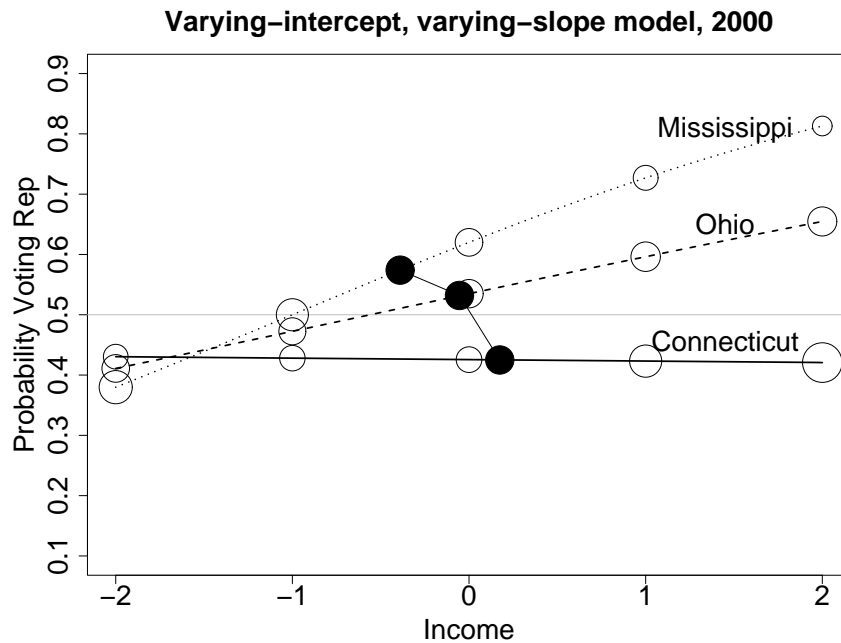


Figure 3: *Figure from Gelman et al. (2005). A varying intercepts and varying slopes model judging income's capacity to predict the vote across states and individuals. Open dots represent the relative portion in that income category in each state. Darkened dots represent the average income in the state.*

for alleviating methodological concerns in multilevel modeling. As above, it may also offer important substantive findings. This is because the new variable is an (cross-level) interaction and its inclusion resolves an omitted variable problem. It should be treated both as a methodological tool and, where useful, as a substantive covariate.

6 Conclusion

To date, many social scientists have been reluctant to fit regressions with modeled parameters that vary by group (or units). The reason is that uncertainty estimates can be highly problematic in these models when predictors and group effects correlate. Estimates may also be biased. We propose that this problem of modeling can be solved with more modeling. Practitioners can get around this problem by taking advantage of the multilevel structure of their regression equation.

Specifically, they can include the mean per group of the predictor in question as a

second level covariate predicting the varying intercepts. This will capture the problematic correlation before it falls into the group-level error term and creates a violation of an important Gauss-Markov regression assumption. This solution will leave researchers free to estimate modeled varying effects or, more generally, multilevel models when their data is best served with such a specification.

References

- Bafumi, Joseph. 2004*a*. “The Macro Micro Link in Vote Choice Models: A Bayesian Multilevel Approach.” Presented at the 2004 Annual Meeting of the Midwest Political Science Association, Chicago, IL.
- Bafumi, Joseph. 2004*b*. “The Stubborn American Voter.” Presented at the 2004 Annual Meeting of the American Political Science Association, Chicago, IL.
- Bartels, Larry. 1996. “Pooling Disparate Observations.” *American Journal of Political Science* 40:905–942.
- Beck, Nathaniel and Jonathan Katz. 2006. “Random Coefficient Models for Time Series/Cross-section Data: Monte Carlo Experiments of Finite Sample Properties.” Presented at the 2006 Annual Meeting of the Northeastern Political Methodology Group, New York University.
- Gelman, Andrew, Boris Shor, Joseph Bafumi and David K. Park. 2005. “Rich State, Poor State, Red State, Blue State: What’s the Matter with Connecticut?” Presented at the Annual Midwest Political Science Association Meeting, Chicago, IL.
- Gelman, Andrew and Gary King. 1990. “Estimating the Electoral Consequences of Legislative Redistricting.” *Journal of the American Statistical Association* 85:274–282.
- Gelman, Andrew and Gary King. 1993. “Why Are American Presidential Election Campaign Polls So Variable When Votes are So Predictable?” *British Journal of Political Science* pp. 409–451.
- Gelman, Andrew and Gary King. 1994. “A Unified Method of Evaluating Electoral Systems and Redistricting Plans.” *American Journal of Political Science* 38:514–554.
- Gelman, Andrew and Thomas C. Little. 1997. “Poststratification into Many Categories using Hierarchical Logistic Regression.” *Survey Methodology* 23:127–135.
- Hausman, Jerry A. and William E. Taylor. 1981. “Panel Data and Unobservable Individual Effects.” *Econometrica* 49:1377–1398.
- Park, David K., Andrew Gelman and Joseph Bafumi. 2004. “Bayesian Multilevel Estimation with Poststratification: State-level Estimates from National Polls.” *Political Analysis* 12:375–385.
- Reilly, Cavan, Andrew Gelman and Gary King. 2001. “Post-stratification without Population Level Information on the Post-stratifying Variable, with Application to Political Pooling.” *Journal of the American Statistical Association* 96:1–11.
- Shor, Boris, Joseph Bafumi, Luke Keele and David K. Park. 2005. “A Bayesian Multilevel Modeling Approach to Time Series Cross-Sectional Data.” Paper presented at the Annual meeting of the Midwest Political Science Association, Chicago, IL.
- Steenbergen, Marco R. and Bradford S. Jones. 2002. “Modeling Multilevel Data Structures.” *American Journal of Political Science* 46:218–237.

Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Modelling Approach." *American Journal of Political Science* 42:1233–1259.