

# MLM NOTATION

Constantin Manuel Bosancianu<sup>1</sup>

WZB Berlin Social Science Center

*Institutions and Political Inequality*

January 14, 2021

The issue of notation for multilevel models is one of the more inconvenient issues related to the approach. At the same time, it is also an important one to grasp for anyone who is introduced in the theory and practice of such models. A major reason for this is that most software (R, SPSS, and Stata included, but not Mplus) will ask the user to specify the extended form of the multilevel model: what the fixed effects and random effects are for the model to be estimated. Going beyond this practical aspect to a more fundamental theoretical one, understanding the notation of a multilevel model allows the practitioner to comprehend the way coefficients are estimated in the model, how they are interpreted in the output, and why they change with the addition to or elimination from the model of a predictor.

## 1 Notation

Let's start with a simple model with one dependent variable,  $Y$ , and two independent variables,  $X_1$  and  $X_2$ . We could be talking about predicting satisfaction with democracy with age and income, or propensity to vote for a party with Left–Right position and political interest. Under OLS estimation, the regression equation is well-known:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i, \epsilon \sim N(0, \sigma^2) \quad (1)$$

Let's go beyond this, though, and assume that our sample consists of  $i$  individuals grouped in  $j$  countries—a fairly standard set-up for a multilevel model in political science. As you have seen from the readings so far, the model becomes:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \epsilon_{ij} \quad (2)$$

Equation 2 represents a minor extension of Equation 1. With this seemingly minor addition (every

---

<sup>1</sup>You can reach me at [manuel.bosancianu@gmail.com](mailto:manuel.bosancianu@gmail.com). If you spot any mistakes I'd be grateful if you sent me an email pointing it out; I'll update the document and credit the help offered.

coefficient has a subscript  $j$  now), a major implication has sneaked in: coefficients now vary from group to group. This is very much different from the typical setup of OLS regression with dummy variables, where we have to assume that the same slope governs the relationship between a  $X$  and the  $Y$  from group to group.<sup>2</sup>

This was only the level-1 model, though. We now continue by specifying a model for each of the coefficients in Equation 2.<sup>3</sup> This is essentially what makes them *random* coefficients: they are now assumed to vary from group to group, but collectively to have a normal distribution.<sup>4</sup> The mean and variance of this distribution can be explained with a statistical model, just as the mean and variance of the  $Y$ s could be explained with the model presented in Equation 1.

$$\begin{cases} \beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}, \text{ where } u_{0j}, u_{1j}, \dots, u_{kj} \sim N(0, \tau_k^2) \\ \beta_{2j} = \gamma_{20} + \gamma_{21}Z_j + u_{2j} \end{cases} \quad (3)$$

In Equation 3, I decided to make  $Z_j$  a predictor for both the intercept and the slopes at the level-1.<sup>5</sup> In simpler terms, we specified a statistical model for the distribution of each of the coefficients from the level-1 model (Equation 2). This model includes an intercept (e.g.,  $\gamma_{10}$ ), a slope for the effect of the level-2 variable on  $\beta_{1j}$  (which is itself an effect), and an error term (e.g.,  $u_{1j}$ ).

Bringing together the statistical models presented in equations 2 and 3, we obtain the extended form of the multilevel model:

$$\begin{aligned} Y_{ij} &= \gamma_{00} + \gamma_{01}Z_j + u_{0j} + (\gamma_{10} + \gamma_{11}Z_j + u_{1j})X_{1j} + (\gamma_{20} + \gamma_{21}Z_j + u_{2j})X_{2j} + r_{ij} \\ &= \underbrace{\gamma_{00} + \gamma_{10}X_{1j} + \gamma_{20}X_{2j} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{1j} + \gamma_{21}Z_jX_{2j}}_{\text{Fixed effects}} + \underbrace{u_{0j} + u_{1j}X_{1j} + u_{2j}X_{2j} + \epsilon_{ij}}_{\text{Random effects}} \end{aligned} \quad (4)$$

Reordering the terms presented in Equation 4, we distinguish between two types of coefficients. “Fixed”

---

<sup>2</sup>*Have to* is a bit too strong of a phrase here, which does not cover some *atypical* strategies. There are ways to go around it, such as interacting the group dummies with the substantive predictors in the model, although the structure gets very complex very quickly with a large number of groups in the data.

<sup>3</sup>Based on the theoretical framework we are operating with, we can decide on which coefficients we allow to vary.

<sup>4</sup>The original meaning, in fact, comes from experimental design, where this framework was used to measure the effect of factors for which we only took a random sample of levels. Think of testing the effect of varying levels of a new chemical compound on cancer—we couldn’t possibly test all the concentrations between, say, 0.1001% and 0.2000%, where we have reason to believe the optimum dosage is. This is why we would take a random sample of levels of concentration, test the effect of these on extent of cancer remission, and use statistical inference to get at the estimated effect of any level of concentration within these bounds.

<sup>5</sup>We could have equally plausibly specified a statistical model only for  $\beta_{1j}$ , and restrict  $\beta_{2j}$  to be fixed across groups. Keep in mind that this decision will be made most of the time (but not always!) on a theoretical basis, rather than a methodological one. Methodological restrictions might be required by estimation problems. Marco Steenbergen shows that in a fully specified 2-level model with  $P$  level-1 covariates and  $Q$  level-2 covariates, we will have  $(P+1)(Q+1)$  fixed effects, 1 level-1 random effect, and  $0.5(P+1)(P+2)$  level-2 variance and covariance components. With large  $P$ s and  $Q$ s, this becomes very quickly a large number of parameters to estimate with sometimes limited sample sizes at the level-2.

effects represent coefficients which the researcher can interpret in the same manner as in a traditional OLS regression. For example,  $\gamma_{10}$  is the change observed in  $Y_{ij}$  when  $X_{1j}$  increases by a unit. Other coefficients in this category can be interpreted in the same way as interactions in OLS regression:  $\gamma_{11}$  describes how the effect of  $X_{1j}$  on  $Y_{ij}$  changes when  $Z_j$  increases by a unit. In the context of MLM this is called a *cross-level* interaction, because it includes variables measured at different levels of the hierarchy. A second type of coefficients are so called “random” effects, which have no correspondent in OLS regression: these represent the residual (unexplained) variation in the slopes or intercepts which we allowed to vary. In the strict sense of the word these are not actually “effects”, but rather variances of effects; neither can they be interpreted in the same way as the “fixed” effects. Nevertheless, the terminology remains in use, and the combination between fixed and random effects in hierarchical models is the reason why they are known as “mixed effects” models.

The downside to such an extended form of a multilevel model is that it makes it more difficult to spot the hierarchical nature of the model. Unfortunately, this form is required by many statistical programs as input, and is how output is presented by many. One small benefit to this, according to [Luke \(2004, p. 11\)](#), is that it makes it obvious that  $\beta_{0j}$  and  $\beta_{1j}$  are not actually estimated directly, but rather by means of  $\gamma_{00}$  and  $\gamma_{10}$ .

## 2 3-level models

Although slightly more complex, the notation for 3-level models is a logical extension of the notation for 2-level ones. Let’s assume we have the situation of  $i$  voters clustered in  $j$  electoral districts, further on clustered in  $k$  countries (a set-up which could be investigated with CSES data, for example). We start with the level-1 model:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}X_{1jk} + \beta_{2jk}X_{2jk} + \epsilon_{ijk} \quad (5)$$

The level-2 model is:

$$\begin{cases} \beta_{0jk} = \gamma_{00k} + \gamma_{01k}Z_{1jk} + u_{0jk} \\ \beta_{1jk} = \gamma_{10k} + \gamma_{11k}Z_{1jk} + u_{1jk}, \\ \beta_{2jk} = \gamma_{20k} + \gamma_{21k}Z_{1jk} + u_{2jk} \end{cases} \quad (6)$$

Finally, the level-3 model<sup>6</sup>:

---

<sup>6</sup>Naturally, some of the  $\gamma$ s can be restricted not to vary across countries, so as to simplify the model a bit – here I’ve presented the more complex form, where both intercepts and slopes at the district-level are allowed to vary from country to country.

$$\left\{ \begin{array}{l} \gamma_{00k} = \varphi_{000} + \varphi_{001}W_{1k} + u_{00k} \\ \gamma_{01k} = \varphi_{010} + \varphi_{011}W_{1k} + u_{01k} \\ \gamma_{10k} = \varphi_{100} + \varphi_{101}W_{1k} + u_{10k} \\ \gamma_{11k} = \varphi_{110} + \varphi_{111}W_{1k} + u_{11k} \\ \gamma_{20k} = \varphi_{200} + \varphi_{201}W_{1k} + u_{20k} \\ \gamma_{21k} = \varphi_{210} + \varphi_{211}W_{1k} + u_{21k} \end{array} \right. \quad (7)$$

### 3 Alternatives to notation

The notation presented so far has been taken from [Raudenbush and Bryk \(2002\)](#), but is also employed by [Hox \(2011\)](#) and by [Steenbergen and Jones \(2002\)](#). Unfortunately, other authors rely on slightly different notation, which might pop up in published materials.

For example, [Snijders and Bosker \(1999, pp. 67–73\)](#) present the level-1 model in the following way:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + R_{ij} \quad (8)$$

At the level-2, the equations are:

$$\left\{ \begin{array}{l} \beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + U_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + U_{1j} \end{array} \right. \quad (9)$$

With the exception of the letters used to denote the errors at the level-1 and the level-2, [Snijders and Bosker \(1999\)](#) maintain the notation employed by [Raudenbush and Bryk \(2002\)](#) (this notation was established in the 1992 edition of their book). [Goldstein \(2011\)](#), on the other hand, deviates a bit more from this notation:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij} \quad (10)$$

$$\left\{ \begin{array}{l} \beta_{0j} = \beta_0 + u_{0j} \\ \beta_{1j} = \beta_1 + u_{1j} \end{array} \right. \quad (11)$$

[Goldstein \(2011\)](#) consistently used  $\beta$  to denote coefficients, although it's not entirely clear from the notation used in Equation 11 how we should designate the coefficient of a predictor  $Z_j$  for  $\beta_{0j}$ , since  $\beta_1$  is already taken. [Gelman and Hill \(2007\)](#), on the other hand, choose to rely on matrix notation:

$$Y_i = \alpha_{j[i]} + \beta_{j[i]}X_i + \epsilon_i \quad (12)$$

In Equation 12,  $X_i$  represents the matrix of predictors, while  $\beta$  designates the column vector of coefficients, and  $\alpha$  refers to the column vector of constants. In their notation,  $j[i]$  simply indexes the group  $j$  from where observation  $i$  originates. What makes the model a multilevel one is specifying a second model, for the intercepts and the slopes:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \gamma_0^\alpha + \gamma_1^\alpha v_j \\ \gamma_0^\beta + \gamma_1^\beta v_j \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right), \text{ for } j = 1, 2, \dots, J \quad (13)$$

In this second Equation  $v_j$  is a level-2 predictor for both the intercept and slope at level-1, while  $\sigma_\alpha^2$  and  $\sigma_\beta^2$  are the variances of the group-level residuals (p. 241).

It doesn't matter much which notation you use. The one used by Raudenbush and Bryk is probably the most common one, although it is also fairly lengthy, particularly in the case of a 3-level model; the one used by Gelman and Hill is the shortest one, although it is also technical and not particularly intuitive for someone without a solid mathematics background. The best strategy at this point is to choose one, learn it, and use it repeatedly in your work.

## References

- Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Goldstein, H. (2011). *Multilevel Statistical Models*. New York: Wiley.
- Hox, J. J. (2011). *Multilevel Analysis: Techniques and Applications*. New York: Routledge.
- Luke, D. A. (2004). *Multilevel Modeling*. Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Steenbergen, M. R., & Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science*, 46(1), 218–237.