Applied Multilevel Regression Modeling

Day 1: The Basics

Constantin Manuel Bosancianu

WZB Berlin Social Science Center
Institutions and Political Inequality
bosancianu@icloud.com

July 29, 2019

Welcome! It's great to have you in the course!

Precursors and intellectual debts

I am following in the footsteps of a course taught for 4 years by Zoltán Fazekas (was a TA for 3 of those years). Learned a lot from him in that period.

Interspersed between slides and code files you find insights from work done by Michael Clark (University of Michigan), Anja Neundorf (University of Nottingham), Marco Steenbergen (University of Zurich), and others.

Course structure (1)

Week 1 showcases fundamental features of multilevel models (MLMs):

1. Ability to model (explain) variation in data at multiple levels;
2. Ability to explain variation between groups in *estimates*;
3. Ability to estimate effects for groups not encountered in our data.

We go through: (1) estimation; (2) interpretation; (3) model assessment and checking; and (4) graphical display of quantities of interest and predictions based on the models.

Course structure (1)

To explore these we make moderate use of notation for MLMs:

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i, \textit{ for } i = 1, 2, \dots, N \tag{1}$$

$$\alpha_j \sim \mathcal{N}(U_j \gamma, \sigma^2_\alpha), \textit{ for } j = 1, 2, \dots, J \tag{2}$$

$$\beta_j \sim \mathcal{N}(V_k \phi, \sigma^2_\beta), \textit{ for } k = 1, 2, \dots, K \tag{3}$$

Important to get used to the notation from the beginning. It lets you:

- ✓ Communicate your model specification easily to an audience;
- ✓ Advance on your own to more complex modeling strategies in MLM.

Course structure (2)

Week 2 extends the fundamentals to specific designs and data configurations:

- ✓ Limited dependent variables;
- ✓ Modeling change over time;
- ✓ Cross-classification;
- ✓ Deep interactions (post-stratification);
- ✓ Spatial modeling in the MLM framework.

To a limited extent, we can be flexible with the relative importance awarded to these.

Logistics

Typically, 90 min. lecture (.pdf), and 90 min. lab (.Rdata & .R).

Labs implement and extend topics covered in lecture. They also include small tasks (self-assessment).

Will also have consultations:

- ✓ Go over concepts discussed during lectures again;
- ✓ Discuss individual projects.

Scheduled *starting from Wednesday*, 1 hr. each day for the rest of the course.

ECTS credits

Attendance (minimum 80%) and readings ⇒ 4 credits. For more, add to these:

- ✓ two take-home assignments during the course ⇒ 6 credits
- ✓ two take-home assignments + final paper (after the course) ⇒ 8 credits

Take-home assignments are small tasks (6-8 hours each): assessing article using MLM, estimating and interpreting a few MLM specifications.

We can decide together on a convenient time for the final paper (has to have an original analysis, and not be a replication or extension).

Why MLM?

Value of MLM

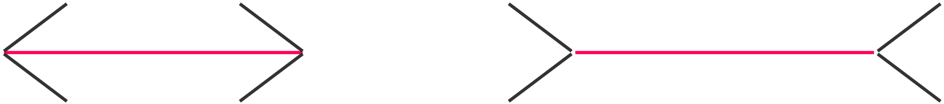
MLMs are uniquely suited to capturing one type of social complexity: the way individuals/firms/NGOs act or think may be context-dependent.

An example which I focus on are the cross-country differences in the likelihood that lower-income people participate in politics.

A more puzzling (and fascinating) one is the cross-country differences in how we *perceive* reality.

Müller-Lyer illusion

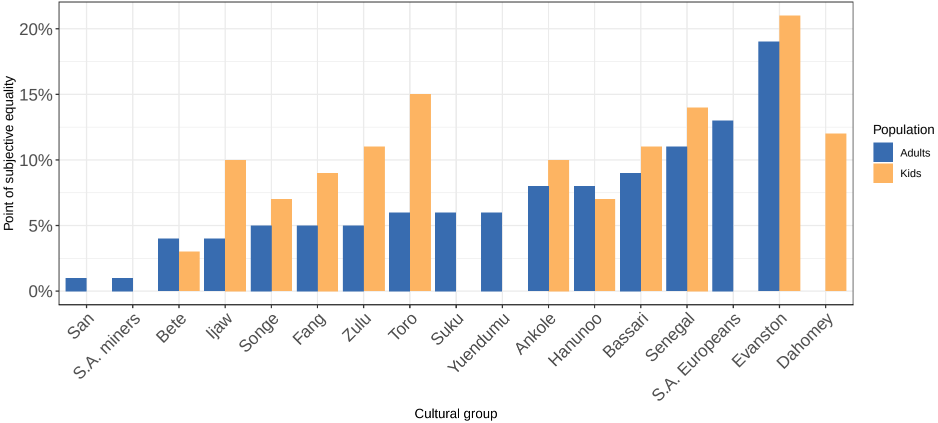
Which one is longer?



We've known about this illusion since 1889. However, since 1966 we also know that not all cultures experience this in the same way.

Cross-cultural variance

Segall, Campbell, and Herskovits (1966) find differences between cultures in how different people perceive these lines to be.



Adapted from McCauley and Henrich (2006)

Cross-cultural variance

The groups are from Uganda, Gabon, Republic of Congo, Nigeria, South Africa, Philippines, Central Australia, Guinea Coast, Illinois etc.

One hypothesis is that it has something to do with architecture: rectangular buildings vs. round huts.

Here we have but one example of how differences can arise both due to individual factors (age) and contextual factors (style of architecture).

Cross-national variance

Political science is replete with similar examples.

	Micro	Macro
Turnout	education	compulsory voting
	income	party polarization
Trust	education	post-communist country
Religiosity	age	income inequality
	gender	GDP

Even more examples in education research, this time focusing on differences between schools.

Social complexity

The point of this long introduction is that investigating social complexity often requires looking at how and based on what political behavior or attitudes vary.

Often, the factors which lead to these differences in behavior/attitudes are found at the contextual level.

Trying to see the world like this trains your mind: how individual actions shape context, and how context, in turn, shapes individual action.

Reasons for using MLMs

Substantive: systematically account for how outcomes (or *effects*) vary across groups, beyond what can be explained by unit-level factors.

Statistical:

- ✓ obtain accurate SEs for estimates in instances of clustered data;
- ✓ *model* the heteroskedasticity in the data.

There is the *nuisance* element to deal with, but it's the second statistical reason that's most important.

Quick OLS recap

OLS mechanics

$$Y_i = \beta_0 + \beta_1 * X1_i + \cdots + \beta_k * Xk_i + \epsilon_i,\ \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{4}$$

Here, Y is the dependent variable, $X1$ through Xk are the independent variables (IVs), and ϵ is the residual (error).

These ϵ_1, ϵ_2, $\ldots \epsilon_n$ have, collectively, a normal distribution with mean 0 and constant variance.

A quick example

	DV: Political efficacy
(Intercept)	2.155 (0.106)***
Age (decades)	0.019 (0.011)
Gender: woman	−0.189 (0.034)***
Education (no. of years)	0.046 (0.006)***
Income: 2nd quartile	0.021 (0.049)
Income: 3rd quartile	0.071 (0.049)
Income: 4th quartile	0.274 (0.052)***
Residence: urban	0.154 (0.035)***
R^2	0.105
Adj. R^2	0.101
Num. obs.	1683
RMSE	0.703

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

ISSP data, Citizenship module II, Belgium (2016).

Political efficacy is an index constructed by averaging 4 items, each measured on a 5-point scale.

OLS: matrix format

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{(n\times 1)} = \begin{bmatrix} 1 & X1_1 & \dots & Xk_1 \\ 1 & X1_2 & \dots & Xk_2 \\ \vdots & \vdots & \dots & \vdots \\ 1 & X1_n & \dots & Xk_n \end{bmatrix}_{(n\times(k+1))} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{((k+1)\times 1)} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{(n\times 1)} \tag{5}$$

In our case, $n = 1683$ and $k = 7$. The solution for b (population parameter) can be written as:

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{6}$$

$$V(\beta) = \sigma^2_{\epsilon}(X'X)^{-1} \tag{7}$$

OLS assumptions—structural part

Concerning the predictors, and the model specification used:

1. No specification error;
2. Predictors are measured without error;[1]
3. No perfect collinearity between predictors.

[1] If this is not the case, estimates are biased downward. If the error is random, estimates are unbiased but inefficient; if the error is systematic, all bets are off.

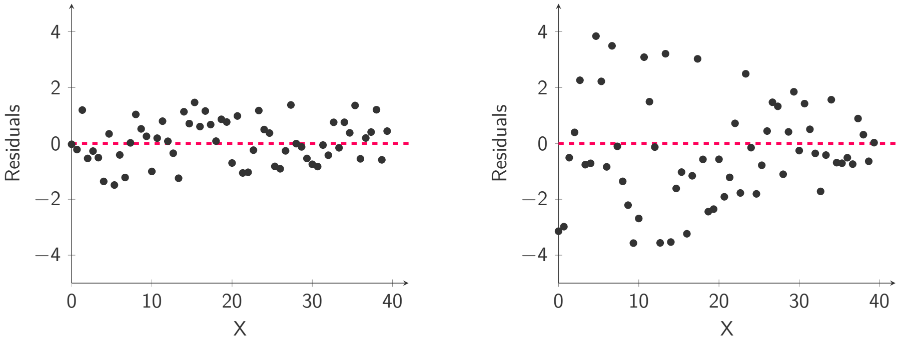
OLS assumptions—stochastic part

The residuals:

1. Average of the es is 0 along the length of Xs: $E(e|x_i) = 0$;
2. Variance is constant along the length of Xs: $V(e|x_i) = \sigma_e^2$. This is also called the assumption of "homoskedasticity"; when it does not hold, we are presented with "heteroskedasticity";
3. Errors are normally distributed: $e_i \sim \mathcal{N}(0, \sigma_e^2)$;
4. Errors are independent from each other: $cov(e_i, e_j) = 0$, for any $i \neq j$;
5. Predictors are measured without error, and are independent of the errors: $cov(X, e) = 0$.

One OLS assumption

Homoskedasticity: $\epsilon \sim \mathcal{N}(0, \sigma^2)$.[2] The residuals are a by-product of the estimation of $\hat{\beta}$.



(a) Homoskedasticity (b) Heteroskedasticity

[2]For an in-depth coverage, please consult the relevant sections in Berry (1993) or Fox (2016).

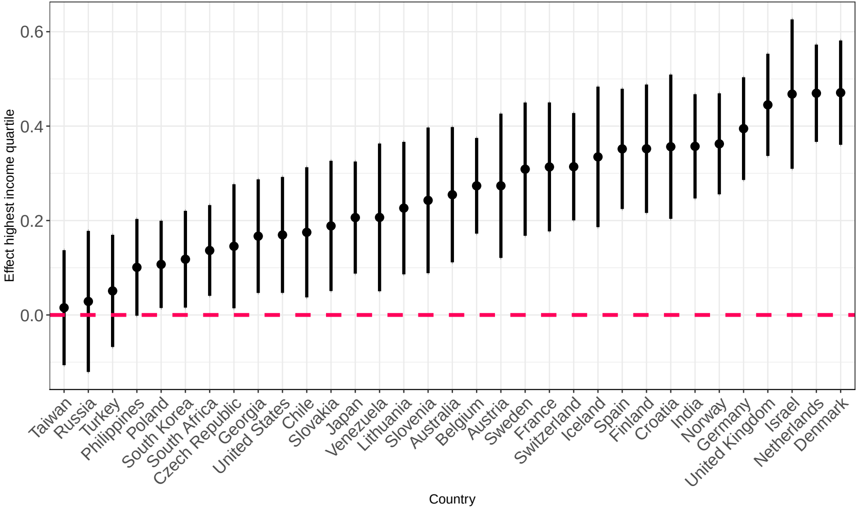
The case of clustered data

Level		
1	2	3
Voters	Districts	Countries
Students	Classrooms	Schools
Companies	Regions	Countries
Decision	Judge	District court

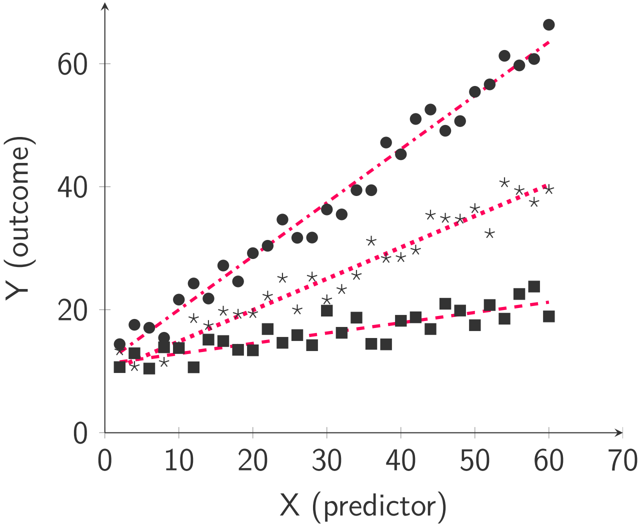
Two potential problems with applying OLS to these data configurations:

- ✓ heterogeneity of effects;
- ✓ sample size at L1.

Impact of urban residence



Consequences of heterogeneity



In this instance, applying an overall slope ("naive pooling") to the data will generate heteroskedasticity.

This can be addressed with country dummies (*fixed effects*), and a lot of interactions.

Fixed effects as solution

$J - 1$ dummy indicators for groups: LSDV approach. Computationally very fast, and conceptually accessible.

Problems with the strategy:

- ✓ Very cumbersome with large number of groups
- ✓ Cannot explain *why* slopes vary between groups

Sample size

Data displays *clustering*: two units drawn at random from same group are more similar to each other that two units drawn at random from different groups.

Effective sample size at L1 is $< n \Rightarrow \epsilon \nsim i.i.d.$

SEs are underestimated in this instance $\Rightarrow$ Type I error.

Cluster-corrected SEs as a solution

They implement a *post hoc* solution: adjust SEs (biased), while leaving $\hat{\beta}$ alone (*supposedly* unbiased).

If the heteroskedasticity is caused by effect heterogeneity, itself caused by a L2 dynamic at play $\Rightarrow$ model specification itself is incorrect (Freedman, 2006).

If this is the case $\hat{\beta}$s are incorrect (see also King & Roberts, 2015) and the Huber–White estimator is not very helpful.

MLM as a solution

Final points

Multiple benefits to using an MLM, instead of the previous two strategies (Steenbergen & Jones, 2002):

- ✓ Combine multiple levels of analysis in a *single* specification (addresses misspecification concerns);
- ✓ Explore and *model* causal heterogeneity, using substantive variables;
- ✓ Gain the ability to make predictions for new *contexts*.

Final points

These benefits naturally come with costs:

- ✓ Increased computational complexity (ML-based estimation);
- ✓ More stringent assumptions (operating at each level of the data);
- ✓ *Theoretically* more demanding, given potential linkages between predictors at different levels.

Thank you for the kind attention!

References

Berry, W. D. (1993). *Understanding Regression Assumptions*. Thousand Oaks, CA: Sage Publications.

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Freedman, D. A. (2006). On The So-Called "Huber Sandwich Estimator" and "Robust Standard Errors". *The American Statistician*, *60*(4), 299–302.

King, G., & Roberts, M. E. (2015). How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It. *Political Analysis*, *23*(2), 159–179.

McCauley, R. N., & Henrich, J. (2006). Susceptibility to the Müller-Lyer Illusion, Theory-Neutral Observation, and the Diachronic Penetrability of the Visual Input System. *Philosophical Psychology*, *19*(1), 1–23.

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The Influence of Culture on Visual Perception*. Indianapolis, IN: Bobbs-Merrill Company.

Steenbergen, M. R., & Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science*, *46*(1), 218–237.