Applied Multilevel Regression Modeling

Day 3: Random slopes

Constantin Manuel Bosancianu

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

July 31, 2019

Yesterday

Notation for MLMs is a simple extension of OLS notation, though it brings with it considerable implications.

Easy to write a model for how an intercept might systematically vary between groups.

To assess the need for a MLM, use the ICC. It tells you the share of *total* variance in outcome that is due to between-group variance $\Rightarrow$ higher is better.

Today

Centering in MLMs: why it's done, and how to do it.

How to specify *random slopes* in such models.

How to interpret and display effects from *cross-level* interactions.

Sample size considerations in MLM.

Centering predictors

Types

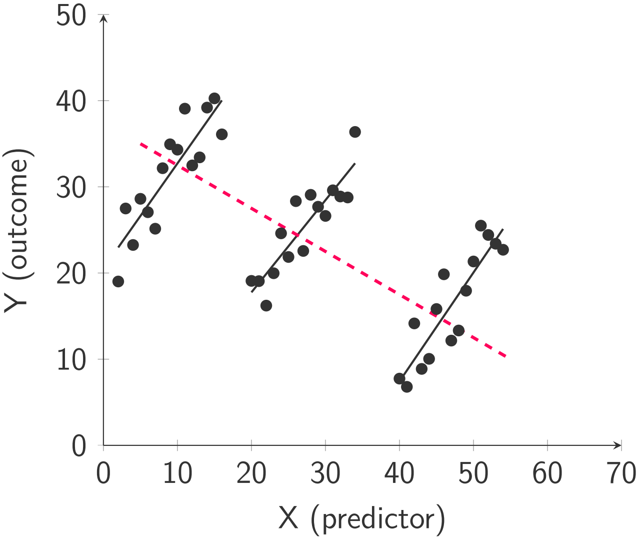
So far, you encountered the concept in regression, either to make intercept "more meaningful", or to be able to compare effect strength among predictors (*standardization*).

$$x - \bar{x} \tag{1}$$

With clustered data, centering represents a technical solution to a unique problem.

Without a correction, the coefficient for an individual-level variable is a mix of "within-group" relationship and "between-group" relationship.

Mix of patterns



Group-mean centering

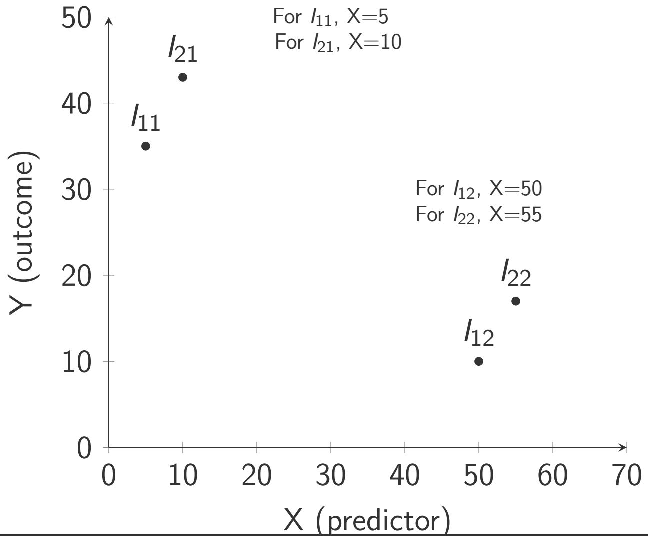
For this problem, centering offers a solution: “artificially” erase the between-group variation in the individual-level variables.

For each j group, one can center a variable X, with the formula

$$X_{centered} = X_{ij} - \overline{X_j} \tag{2}$$

$\bar{X}_j$ is the mean of variable X in the j group. This is called **group-mean centering**.

Centering variables



Centering variables

For group 1, $\overline{X}$=7.5. For group 2, $\overline{X}$=52.5.

	X raw	X centered
I_{11}	5	-2.5
I_{21}	10	2.5
I_{12}	50	-2.5
I_{22}	55	2.5

With group-mean centering, the only thing that is left over is the relative position of individuals within a group, e.g. the distance between I_{11} and I_{21} is still 5.

Grand-mean centering

There is a second type of centering, meant for level 2 variables: **grand-mean centering**.

$$Z_{centered} = Z_j - \overline{Z} \tag{3}$$

In this procedure, we subtract from each group's value on Z the mean of the Zs of all groups.

Practical advice

Recommendations made by Enders and Tofighi (2007):

- ✓ $X_{L1} \rightsquigarrow Y$: group-mean centering for X
- ✓ $X_{L2} \rightsquigarrow Y$: grand-mean centering for X
- ✓ $X_{L2} \times X_{L1}$: grand-mean for X_{L2}, group-mean for X_{L1}
- ✓ $X_{L2} \times X_{L2}$: grand-mean for both
- ✓ $X_{L1} \times X_{L1}$: group-mean for both

Unlike in standard regression, centering in MLMs can (and should!) change the magnitude of the estimate.

Random slopes

MLM specification (1)

We used a very simple model yesterday:

$$
\begin{cases}
EFF_{ij} = \beta_{0j} + \beta_{1j} * EDU_{ij} + \beta_{2j} * URB_{ij} + \epsilon_{ij} \\
\beta_{0j} = \gamma_{00} + \gamma_{01} * CORR_j + \upsilon_{0j} \\
\beta_{1j} = \gamma_{10} \\
\beta_{2j} = \gamma_{20}
\end{cases}
\tag{4}
$$

Our theory could extend to how the *effect* of a level-1 predictor varies based on a level-2 predictor.

Does the effect of income vary?

MLM specification (2)

$$
\begin{cases}
EFF_{ij} = \beta_{0j} + \beta_{1j} * EDU_{ij} + \beta_{2j} * URB_{ij} + \epsilon_{ij} \\
\beta_{0j} = \gamma_{00} + \gamma_{01} * CORR_j + \upsilon_{0j} \\
\beta_{1j} = \gamma_{10} + \gamma_{11} * CORR_j + \upsilon_{1j} \\
\beta_{2j} = \gamma_{20}
\end{cases}
\tag{5}
$$

If we choose to, we can also write up a model for the effect of urban residence on political efficacy.

MLM specification—extended form

The extended form:

$$
\begin{aligned}
EFF_{ij} &= \gamma_{00} + \gamma_{01} * CORR_j + \upsilon_{0j} + (\gamma_{10} + \gamma_{11} * CORR_j + \upsilon_{1j}) * EDU_{ij} + \\
&\quad + \beta_{2j} * URB_{ij} + \epsilon_{ij} = \\
&= \gamma_{00} + \gamma_{01} * CORR_j + \upsilon_{0j} + \gamma_{10} * EDU_{ij} + \gamma_{11} * EDU_{ij} * CORR_j + \\
&\quad + \upsilon_{1j} * EDU_{ij} + \gamma_{20} * URB_{ij} + \epsilon_{ij} = \\
&= \gamma_{00} + \gamma_{10} * EDU_{ij} + \gamma_{20} * URB_{ij} + \gamma_{01} * CORR_j + \gamma_{11} * EDU_{ij} * CORR_j + \\
&\quad + \upsilon_{1j} * EDU_{ij} + \upsilon_{0j} + \epsilon_{ij}
\end{aligned}
\tag{6}
$$

Extended form

The extended form is pretty useful as that's the way `Stata`, `R`, and even `SPSS` ask you to specify the model. `Mplus` is an exception in this case.

Fixed effects: γ_{00}, γ_{10}, γ_{20}, γ_{01}, and γ_{11}.

Random effects: υ_{0j}, υ_{1j}, and ϵ_{ij}.

$$\begin{cases} EFF_{ij} = \beta_{0j} + \beta_{1j} * EDU_{ij} + \beta_{2j} * URB_{ij} + \epsilon_{ij} \\ \beta_{0j} = \gamma_{00} + \gamma_{01} * CORR_j + \upsilon_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11} * CORR_j + \upsilon_{1j} \\ \beta_{2j} = \gamma_{20} \end{cases} \tag{7}$$

Only the highlighted quantities are being estimated in the model.

Our example—more predictors

	DV: political efficacy
(Intercept)	$2.960\ (0.041)^{***}$
Woman	$-0.130\ (0.008)^{***}$
Education (years)	$0.302\ (0.008)^{***}$
Income: 2nd quantile	$0.051\ (0.010)^{***}$
Income: 3rd quantile	$0.107\ (0.010)^{***}$
Income: 4th quantile	$0.217\ (0.010)^{***}$
Urban residence	$0.053\ (0.008)^{***}$
Corruption perceptions	$0.435\ (0.082)^{***}$
Num. obs.	32020
Num. groups: country	31
Var: country (Intercept)	0.051
Var: Residual	0.474

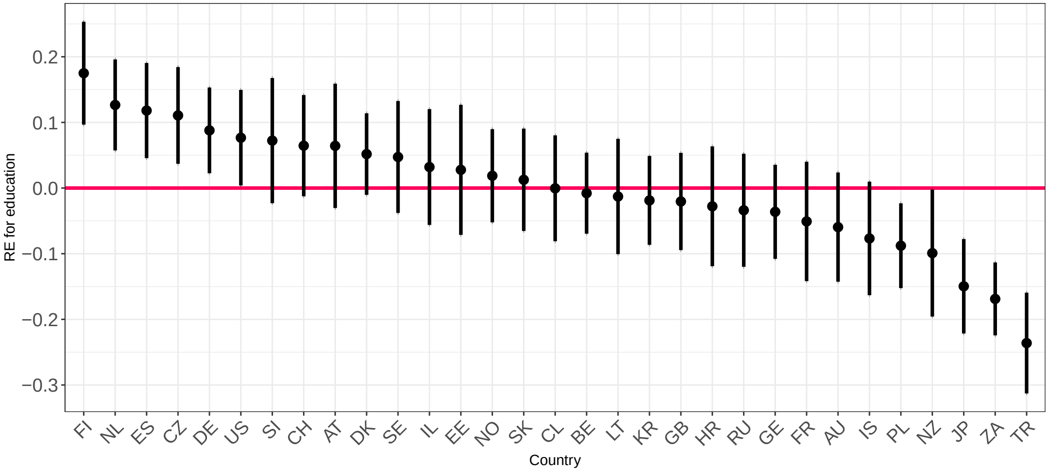
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Our example—a random slope

	DV: political efficacy
(Intercept)	2.960 (0.041)***
Woman	−0.131 (0.008)***
Education (years)	0.305 (0.020)***
Income: 2nd quantile	0.050 (0.010)***
Income: 3rd quantile	0.107 (0.010)***
Income: 4th quantile	0.217 (0.010)***
Urban residence	0.056 (0.008)***
Corruption perceptions	0.489 (0.077)***
Var: country (Intercept)	0.051
Var: country Education	0.010
Cov: country (Intercept) Education	−0.009
Var: Residual	0.471

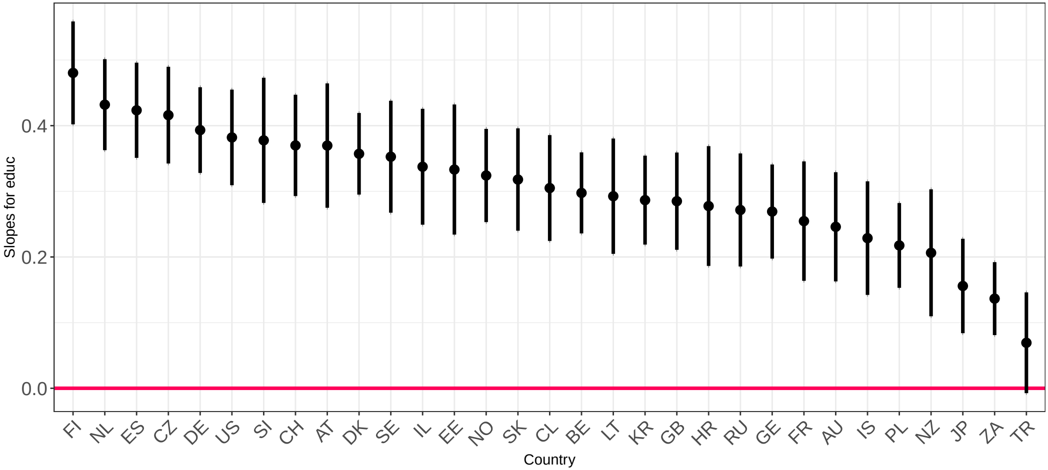
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Sample sizes excluded from table—identical to those in previous table.

Random effects

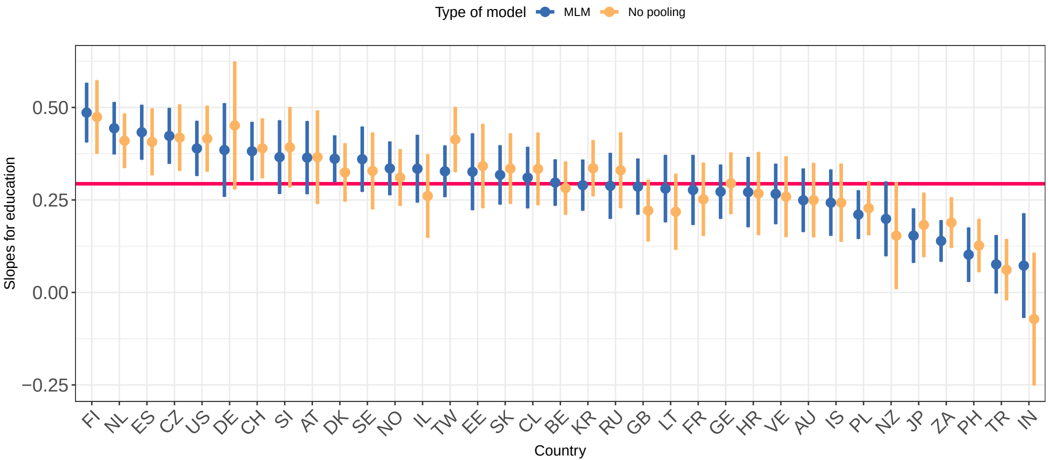


REs are deviations from the fixed-effect for education, not effects themselves!

Actual slopes



Comparison: no pooling & MLM with low N



20% of sample for IN and DE was used for demonstration

Adding predictors for random slopes

So far, we have only added a L2 predictor for the random intercept, but we can do the same for a random slope ⇒ we *systematically* explain how effects vary across contexts.

A comparison of 3 models:

- ✓ RI model with L2 predictor for intercept (Equation 4)
- ✓ RI+RS model with L2 predictor for intercept, but not for slope
- ✓ RI+RS model with L2 predictor for both intercept *and* slope(Equation 5)[1]

[1] The estimated models will differ a bit from the equations, because of more L1 predictors, but the fundamental features are there.

Model comparison

	RI	RI + RS with no pred.	RI + RS with pred.
(Intercept)	2.962^{***}	2.962^{***}	2.962^{***}
	(0.045)	(0.046)	(0.045)
Education	0.288^{***}	0.292^{***}	0.292^{***}
	(0.008)	(0.021)	(0.019)
Urban residence	0.044^{***}	0.045^{***}	0.045^{***}
	(0.007)	(0.007)	(0.007)
Perceptions of corruption	0.295^{**}	0.347^{***}	0.295^{**}
	(0.090)	(0.089)	(0.090)
Education * Perceptions			0.109^{**}
			(0.038)
Num. obs.	36486	36486	36486
Num. groups: country	35	35	35
Var: country (Intercept) ($\sigma_{v_{0j}}$)	0.072	0.072	0.072
Var: Residual ($\sigma_{\epsilon_{ij}}$)	0.476	0.473	0.473
Var: country Education ($\sigma_{v_{1j}}$)		0.014	0.011
Cov: country (Intercept) Education ($\rho\sigma_{v_{0j}}\sigma_{v_{1j}}$)		-0.008	-0.006

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Other predictors excluded, along with model fit statistics.

Small extensions—practice

Use the structure of the model presented in Equation 5 as a starting point, and write down 2 models:

- ✓ The specification in the Equation, *plus* `GINI` as predictor for the intercept (M1)
- ✓ The specification above, *plus* `GINI` as another predictor for the slope of education (M2)

How many parameters are estimated in each model?[2]

---

[2] If `GINI` would have been a predictor for the slope of `URB`, how many parameters would have been estimated?

Small extensions—results

	M1		M2	
	β	SE	β	SE
(Intercept)	2.959	$(0.047)^{***}$	2.963	$(0.046)^{***}$
Education	0.292	$(0.019)^{***}$	0.287	$(0.017)^{***}$
Urban residence	0.045	$(0.007)^{***}$	0.045	$(0.007)^{***}$
Perceptions of corruption	0.268	$(0.105)^{*}$	0.304	$(0.105)^{**}$
Gini (10-point)	−0.065	(0.120)	0.022	(0.123)
Education * Perceptions	0.109	$(0.038)^{**}$	0.049	(0.039)
Education * Gini			−0.142	$(0.045)^{**}$
Var: country (Intercept)	0.075		0.074	
Var: country Education	0.011		0.008	
Cov: country (Intercept) Education	−0.008		−0.006	
Var: Residual	0.473		0.473	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Sample sizes same as for previous specifications. Model fit measures and additional predictors excluded.

Cross-level interactions

Centering helps here, as it reduces the collinearity between main terms and interaction.[3]

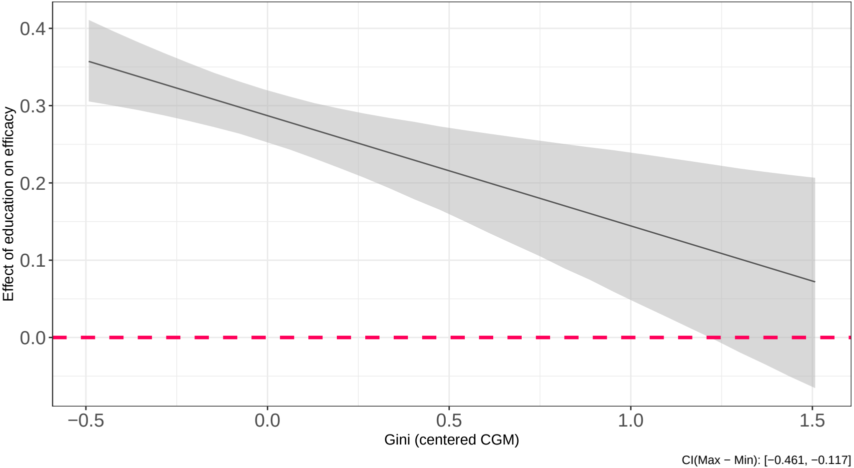
Education has been group-mean centered and standardized (2SD). Perceptions of corruption and Gini have been grand-mean centered and standardized (2SD).

Interpreted in the same way as interactions in regular multiple regression (Brambor, Clark, & Golder, 2005).

Always turn to graphs to present cross-level interactions.

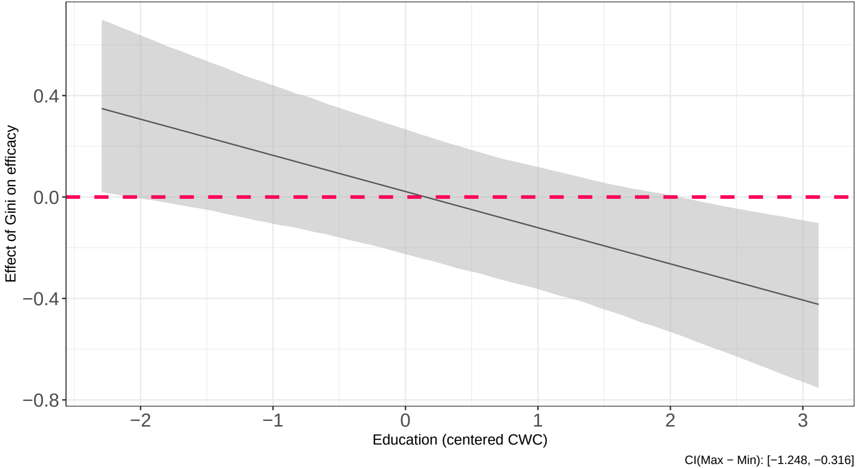
[3] Kam and Franzese Jr. (2007) point out that this happens because we are effectively changing what we are estimating, i.e. it's not so much a solution as a re-specification.

Education × Gini



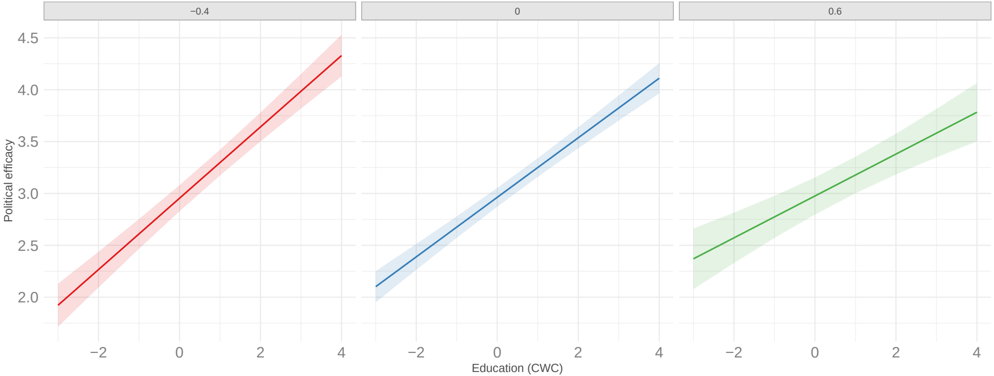
$\beta_{education}$ depicted at varying levels of `GINI`

Symmetry of interpretation



β_{Gini} depicted at varying levels of `EDU`

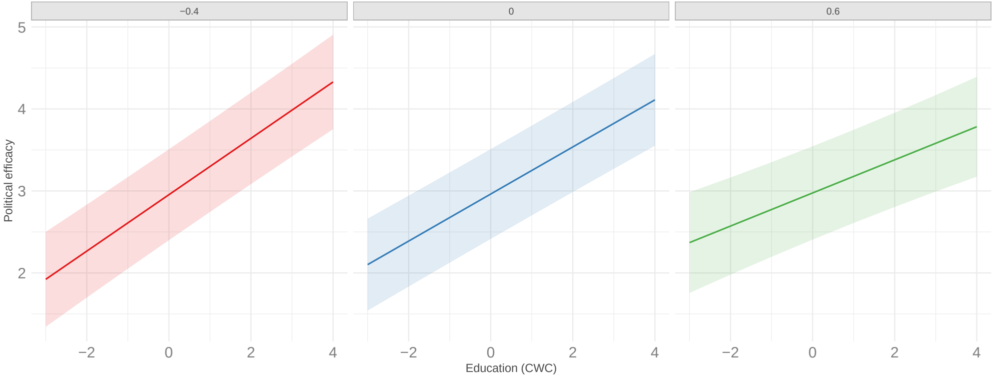
Use actual predictions of outcome



Panels vary based on different levels of `GINI`

However, this incorporates only fixed-effect uncertainty.

…with random-effect uncertainty



Panels vary based on different levels of `GINI`

Much wider CIs when incorporating both FE & RE uncertainty.

Sample size

Sample size

This is maximum likelihood estimation (more on this tomorrow!). Its desirable properties kick in only in large samples.

Usually, the concern is with the level 2 sample size. The minimum is something like 30 (Stegmüller, 2013), although a more desirable threshold is 50.

Even for low sample sizes ML estimates will be unbiased, but will likely be inefficient.[4]

---

[4] There is a bit of a difference here between different types of ML-based estimators. REML is generally better, and REML with a correction for degrees of freedom can compensate for these deficiencies (Elff, Heisig, Schaeffer, & Shikano, 2018). It can't work magic, though.

Sample size

It's fair, though, to take a more nuanced view of things.

Estimate	Continuous DV	Binary DV
L1 fixed-effects point estimates	5	10
L2 fixed-effects point estimates	15	30
Fixed-effects std. errors	30	50
L1 RE estimate	10	30
L2 RE estimate	10/30 (REML/FIML)	10/50 (REML/FIML)
L2 RE std. error	50	100

Level 2 minimum sample size requirements (McNeish & Stapleton, 2016)

L1 group sample size

In *vanilla* cross-national research it is typically not a concern, ranging from a few hundred to a few thousands.

Strength of MLM: can give an estimate for small groups, borrowing power from larger groups.

Interesting case: very many groups, but limited sample size at L1 in each group, e.g. a household survey.

You're facing a limit on the number of L1 parameters that you can allow to vary. For HH surveys, depending on the context, usually only 1 random slope.

Thank you for the kind attention!

References

Brambor, T., Clark, W. R., & Golder, M. (2005). Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, *14*(1), 63–82.

Elff, M., Heisig, J. P., Schaeffer, M., & Shikano, S. (2018). *No Need to Turn Bayesian in Multilevel Analysis with Few Clusters: How Frequentist Methods Provide Unbiased Estimates and Accurate Inference.* Retrieved from `https://doi.org/10.31235/osf.io/z65s4`

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138.

Kam, C. D., & Franzese Jr., R. J. (2007). *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor, MI: University of Michigan Press.

McNeish, D. M., & Stapleton, L. M. (2016). The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration. *Educational Psychology Review*, *28*(2), 295–314.

Stegmüller, D. (2013). How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science*, *57*(3), 748–761.