Applied Multilevel Regression Modeling

Day 7: Analyzing Change Over Time

Constantin Manuel Bosancianu

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

August 6, 2019

Over the past 6 days. . .

We have treated individuals as the L1 unit of analysis, and have offered frequent examples where such individuals are nested in higher-order units:

- ✓ voters nested in countries;
- ✓ students nested in classrooms, further nested in schools;
- ✓ MPs nested in committees.

The variation we were interested in explaining occurred between groups.

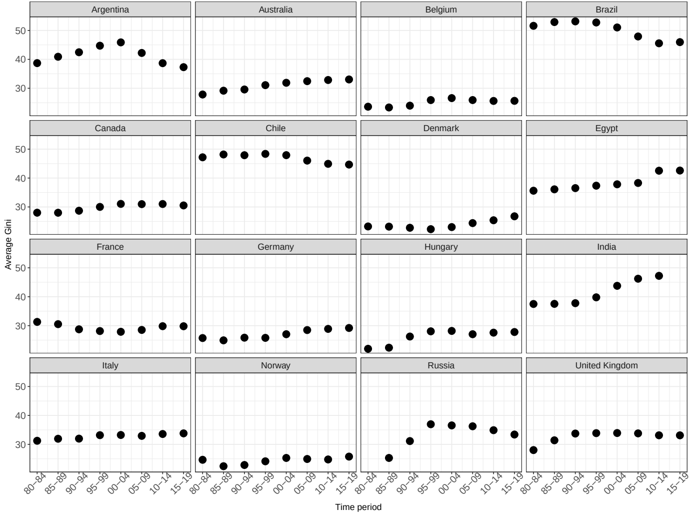
...but today...

We turn the logic around, by applying the MLM logic to an instance where the L2 “groups” are actually single units: individuals, companies, countries.

The L1 observations are measurements over time on these single units: performance on tests, share price, level of income inequality.

It will take some practice to adapt the interpretation of some of the measures learned last week, e.g. ICC, to this new data configuration, but the MLM framework applies very well.

Change over time—2 kinds of questions



How is the pattern of change over time—linear, non-linear? (L1 model)

Do different kinds of units experience different types of change? (L2 model)

Different perspective

	Cross-sectional	Longitudinal
ICC	High value denotes group differences	High value denotes individual differences
Focus	L2 & L1*L2 effects	L1 & L1*L2 effects
Cross-level interactions	Added bonus	Vital component

The approach should not be applied in every setting though:

- 3 or more measurement time points;[1]
- sufficient variance over time in outcome.

[1] 2 points make it difficult to distinguish pattern of change, and cannot disentangle change from measurement error.

Measuring time

You might opt for the natural metric (minutes, days, weeks, years), but you have more freedom than this: (1) number of sessions of treatment; (2) number of trips taken; (3) distance travelled; (4) school grade.

Another choice is the measurement interval—the MLM framework can accommodate:

- ✓ equally spaced intervals;
- ✓ varying spacing, depending on where the rate of change is higher.

Other choices: structured vs. unstructured schedule of measurement across individuals.

L1 trends

Describe change (1)

The first goal is to describe the change (*growth trajectory*) using empirical growth plots.

Use a nonparametric functional form, e.g. *lowess*, to allow the data to speak for itself, and then try to visually capture commonalities in trajectories.

Try to see how a parametric form works, e.g. OLS, in capturing the trends you have observed.

Finally, if still far from a proper description, alter the parametric form by allowing for non-constant change over time.

Describe change (2)

Use the automated tool we've relied on last week (`dplyr` + `broom` + `purrr`) to produce quick summaries for each individual:

- ✓ intercept and the slope (rate of change over time);
- ✓ correlation between intercept and slope

More informative: split these trends up based on theoretically-relevant time-invariant (i.e. individual-level) predictors.

Model specifications

Multilevel model for change (1)

- ✓ L1 model: describes change over time—*individual growth model*;
- ✓ L2 model: describes inter-individual variation in this rate of change

Assume a very straightforward linear form:[2]

$$WAGES_{ti} = \beta_{0i} + \beta_{1i}TIME_{ti} + \epsilon_{ti} \tag{1}$$

We're saying with that specification: each individual's wage has a linear growth pattern, described by individual growth parameters β_{0i} and β_{1i}.

[2] I find Singer and Willett's notation convention a bit imprecise, e.g. the way that the position of the indicator for child, i, changes from one subscript to another. This is why I use a slightly altered notation.

Multilevel model for change (2)

It's harder to claim that ϵ_{ti} are i.i.d, but we can accommodate a variety of error structures.

$$\begin{cases} \beta_{0i} = \gamma_{00} + \gamma_{01}AA_i + \upsilon_{0i} \\ \beta_{1i} = \gamma_{10} + \gamma_{11}AA_i + \upsilon_{1i} \end{cases} \tag{2}$$

The L2 (individual) model links the variation in individual growth trajectories to time-invariant characteristics of individuals.

Meaning of estimates

$$AA_i = 0 \Rightarrow \begin{cases} Initial\ status = \gamma_{00} \\ Rate\ of\ change = \gamma_{10} \end{cases} \tag{3}$$

$$AA_i = 1 \Rightarrow \begin{cases} Initial\ status = \gamma_{00} + \gamma_{01} \\ Rate\ of\ change = \gamma_{10} + \gamma_{11} \end{cases} \tag{4}$$

υ_{0i} and υ_{1i} are both normally distributed, with constant variance, and we also get $Cov(\upsilon_{0i}; \upsilon_{1i})$—the association between baseline value and rate of growth.

Composite model

$$
\begin{cases} WAGES_{ti} = \beta_{0i} + \beta_{1i} TIME_{ti} + \epsilon_{ti} \\ \beta_{0i} = \gamma_{00} + \gamma_{01} AA_i + \upsilon_{0i} \\ \beta_{1i} = \gamma_{10} + \gamma_{11} AA_i + \upsilon_{1i} \end{cases} \tag{5}
$$

$$
\begin{aligned} WAGES_{ti} = & (\gamma_{00} + \gamma_{01} * AA_i + \upsilon_{0i}) + \\ & + (\gamma_{10} + \gamma_{11} * AA_i + \upsilon_{1i}) * TIME_{ti} + \epsilon_{ti} = \\ = & \gamma_{00} + \gamma_{10} * TIME_{ti} + \gamma_{01} * AA_i + \gamma_{11} * TIME_{ti} * AA_i + \\ & + \upsilon_{1i} * TIME_{ti} + \upsilon_{0i} + \epsilon_{ti} \end{aligned} \tag{6}
$$

ϵ_{ti} likely have a more complex error structure than in previous examples.

Unconditional means & unconditional growth

Unconditional means is what we called in past days the "null model".

$$\begin{cases} WAGES_{ti} = \beta_{0i} + \epsilon_{ti} \\ \beta_{0i} = \gamma_{00} + \upsilon_{0i} \end{cases} \tag{7}$$

As before, we use it to determine the ICC: $\frac{\sigma_{\upsilon}^2}{\sigma_{\upsilon}^2 + \sigma_{\epsilon}^2}$ (it also quantifies the autocorrelation between ϵ_{ti}).

$$\begin{cases} WAGES_{ti} = \beta_{0i} + \beta_{1i} * TIME_{ti} + \epsilon_{ti} \\ \beta_{0i} = \gamma_{00} + \upsilon_{0i} \\ \beta_{1i} = \gamma_{10} + \upsilon_{1i} \end{cases} \tag{8}$$

Building up models

We follow-up on these models by adding predictors at the individual-level (L2) that explain both initial status and rate of change. So far I used racial background, but we can replace it with gender (FEM), as an example.

$$L2 = \begin{cases} \beta_{0i} = \gamma_{00} + \gamma_{01} * FEM_j + \upsilon_{0i} \\ \beta_{1i} = \gamma_{10} + \gamma_{11} * FEM_j + \upsilon_{1i} \end{cases} \tag{9}$$

Enough flexibility to be able to handle:

- ✓ differences between individuals in timing of measurement
- ✓ differences between individuals in number of time points measured

Interpretation of coefficients

- ✓ γ_{00}: average baseline value in the population for men ($FEM = 0$)
- ✓ γ_{10}: average rate of change in the population for men ($FEM = 0$)
- ✓ γ_{01}: average difference in baseline value between men and women
- ✓ γ_{11}: average difference in rate of change between men and women

Additional predictors for individual intercepts or slopes can be added, slightly altering the interpretations above.

Time-varying predictors

Adding L1 predictors

$$\begin{cases} WAGES_{ti} = \beta_{0i} + \beta_{1i} * TIME_{ti} + \beta_{2i} * COMP_{ti} + \epsilon_{ti} \\ \beta_{0i} = \gamma_{00} + \gamma_{01} * FEM_j + \upsilon_{0i} \\ \beta_{1i} = \gamma_{10} + \gamma_{11} * FEM_j + \upsilon_{1i} \beta_{2i} = \gamma_{20} \end{cases} \tag{10}$$

- ✓ γ_{10}: average rate of change for men, after controlling for $COMP_{ti}$
- ✓ γ_{20}: effect on $WAGES_{ti}$ of reaching a higher level of computer skills
- ✓ γ_{11}: the change in the gap in $WAGES_{ti}$ between men and women for every passing unit of $TIME$

Limitations

Severe imbalances in the number of time points of measurement for individuals will make estimating many random effects difficult.

A small number of measurement waves means you'll be constrained in how many random effects can be specified at L1.

Only linear and continuous change can be modeled with this specification.

Non-linear change and thresholds (1)

At particular turning points (PhD graduation, arrival of first child) both rates of growth and outcome level could change.

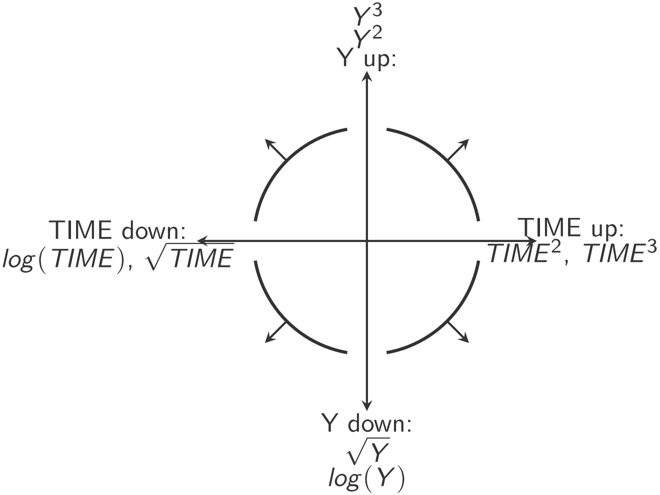
Adding such a predictor to the L1 specification would reveal any change in elevation.

Adding a recoded version of time ($POSTTIME$) to the model assesses changes in slope.

$$POSTTIME_{ti} = \begin{cases} 0, & \textit{all time points before threshold event} \\ TIME, & \textit{after threshold event} \end{cases} \tag{11}$$

Adding both models produces changes in both slope and elevation.

Non-linear change and thresholds (2)



Mosteller and Tukey's (1977, p. 84) set of rules for transformations.

Non-linear change and thresholds (3)

Model curvilinearity through the use of polynomials of *TIME* (the previous solution treated it as a nuisance to be corrected).

- ✓ *no change*: eliminating *TIME* as a predictor
- ✓ *linear change*: using *TIME*
- ✓ *quadratic change*: using *TIME* and $TIME^2$
- ✓ *cubic change*: using *TIME*, $TIME^2$ and $TIME^3$

Use GoF criteria to determine the moment at which you're overfitting the data.

Thank you for the kind attention!

References

Mosteller, F., & Tukey, J. W. (1977). *Data Analysis and Regression. A Second Course in Statistics*. Reading, MA: Addison-Wesley.