One Model, Many Guises: On Different Notation Conventions

Components of a mathematical model such as variables, constants, parameters and indices may in principle be represented by a symbol of your choice. The only constraints are of course the symbols universally used to describe mathematical operations, relations and definitions. Usually, naming conventions for different types of components will emerge within a scientific (sub) community (f.i. using ß for a parameter estimate). Usage of a particular notation convention will often reflect, “where you are coming from”, with respect to your educational and scientific background, quite similar to speaking with a particular regional accent. In multilevel analysis this typecasting by notation convention can even reveal which computer program you use to estimate your models, or which “introduction to multilevel modelling”-book you read!

To continue the analogy of notation conventions to different dialects of a language: Besides different conventions for the use of symbols (=accents), there are also conventions for model expression in terms of the arrangement, grouping and definition of model components. The same model can theoretically be expressed in many different, but equivalent ways so long as the mathematical constraints on the components are preserved (=same meaning, different grammar). The decision for a specific grammar to express your model in may be guided by the need for simplification, the highlighting of a feature or a component, but will often just be the one you were trained to use. There are two popular notation conventions to mathematically express a multilevel model: The *single equation expression* (also known as *the composite model*), and the fully equivalent *multiple equation expression*. Within the multiple equations expression there are different conventions for symbols used in the equations that depend on the type of data that is being analysed (f.i. different symbols for repeated measures data).

The *single equation expression convention* for all types of data:

the compound model

$$Y_{ij} = ß_{0ij}X_{0ij} + ß_{1j}X_{1ij} + ß_2X_{2j} + ß_3X_1X_{2ij}$$

$$ß_{0ij} = ß_0 + u_{0j} + e_{0ij}$$
$$ß_{1j} = ß_1 + u_{1j}$$

Notes: **i** indexes level-1 units; **j** level-2 units
X_0 is a constant; X_1 is a level-1 predictor; X_2 is a level-2 predictor
X_1X_2 is a cross-level interaction to explain slope variance by X_2

The *multiple equation expression convention* for data with no repeated measurements:

level 1: within group regression model

$$Y_{ij} = ß_{0j} + ß_{1j}X_{1ij} + e_{0ij}$$

level 2: between group differences in regression

$$ß_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$$
$$ß_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j}$$

Notes: **i** indexes level-1 units; **j** level-2 units
X_1 is a level-1 predictor; Z is a level-2 predictor
To see cross-level interaction insert the level-2 equations into the level-1 equation:

The most apparent difference is that the multiple equation expression emphasizes the idea that random intercepts and random slopes can be interpreted as latent variables that may each be explained separately by explanatory variables. The single equation expression shows us that the multilevel model is just a regression model with more than one kind of residual. It also highlights the concept of cross-level interaction. When the multilevel model is used to analyze repeated measures data (and is called: *the multilevel model for change*) different symbols are used for the variables, parameters and indices in the the *multiple equation expression*. The "convention" is now however completely defined by which textbook you studied as different authors are using different symbols. Those used to the single equation expression beware: the i –index is being used to identify individuals, which means it is now at level 2!!

Example: Three equivalent expressions of a linear growth model with one level-2 predictor

Singer & Willett:

level 1: individual change trajectory

$Y_{ij} = \pi_{0i} + \pi_{1i}T_{ij} + e_{ij}$!! **j** indexes level-1 units (time) and **i** level-2 (individuals)!!

level 2: inter-individual differences in change

$\pi_{0i} = \gamma_{00} + \gamma_{01}Z_i + \zeta_{0i}$

$\pi_{1i} = \gamma_{10} + \gamma_{11}Z_i + \zeta_{1i}$

Hox:

level 1: individual change trajectory

$Y_{ti} = \pi_{0i} + \pi_{1i}T_{ti} + e_{ti}$!! **t** indexes level-1 units (time) and **i** level-2 (individuals)!!

level 2: inter-individual differences in change

$\pi_{0i} = ß_{00} + ß_{01}Z_i + u_{0i}$

$\pi_{1i} = ß_{10} + ß_{11}Z_i + u_{1i}$

MLwiN:

the compound model (unchanged)

$Y_{ij} = ß_{0ij}X_{0ij} + ß_{1j}X_{1ij} + ß_2X_{2j} + ß_3X_1X_{2ij}$!! **i** indexes level-1 units (time) and **j** level-2 (individuals) !!

$ß_{0ij} = ß_0 + u_{0j} + e_{0ij}$

$ß_{1j} = ß_1 + u_{1j}$

To summarize:

If you use the *single equation expression* convention it is very likely that:
- you were trained in multilevel modelling in Europe
- you use MLwiN to model your data
- *you wish to emphasize the fact that the multilevel model is just a regression model with more than one kind of residual and highlight the concept of cross-level interaction*
- you are a very sensible person

If you use the *multiple equation expression*, convention it is very likely that:
- you were trained in multilevel modelling in the US
- you use HLM or Mplus to model your data
- *you wish to emphasize that random intercepts and random slopes can be interpreted as latent variables, each of which may be explained separately by explanatory variables*
- you are a very sensible person