Introduction and summary of the thesis

Joel G. R. Karlsson*

June 04, 2012

Introduction

Much applied economic research concerns complex data structures that need to be properly modelled. It is especially common with hierarchical structures such as children nested within families, pupils nested within schools, individuals nested within regional areas, or measurements nested within individuals. Failure to take into account such structures, in models, can lead to incorrect inferences. The failure to properly model also makes it impossible to capture the complexity that exists in the real world. What is known as multilevel analysis has been developed, and offer, when appropriate, model-based methods that deal with this issue. The four papers presented in Chapters 1-4 deal with complex hierarchical data structures that are common in the empirical economic literature and estimated with appropriate multilevel models.

Multilevel analysis at a glance

Generally[1], a multilevel model can be considered as a regression in which the parameters (the regression coefficients) are given by a probability model. This second model has parameters of its own, where these hyperparameters also are estimated from the data. This is the key part of multilevel modelling, varying coefficients and models for those varying coefficients (Gelman and Hill, 2007). In classical regression, it is sometimes possible to include varying coefficients by using indicator variables. The main distinguishing feature is hence the modelling of variation between groups.

*Centre for Labour Policy Research, Linnaeus University, Centre for Entrepreneurship and Spatial Economics (CEnSE), and the Swedish Board of Agriculture. Sweden. joel.karlsson@jordbruksverket.se. I'm grateful to Ghazi Shukur, Hkan Locking, Andreas Stephan, and seminar participants at Linnaeus University. for helpful comments and suggestions.

[1]The following section is inspired from examples given in Gelman and Hill (2007) Chapter 1, de Leew and Meijer (2008), Goldstein (2003) Chapters 1–2, Rabe-Hesketh and Skrondal (2008) and Hox and Roberts (2011).

Consider an educational study with data from students in many schools, predicting in each school the students' grades y on a standardised test given their scores on a pre-test x and other information. For simplicity, consider only two covariates, pre-test x and a school-level predictor z. Here the complex data structure is that pupils are nested within schools. Separate models can be fitted within each school, and the schools themselves can be modelled, depending on school characteristics (such as if the school is public or private and other information). The student-level regression and the school-level regression are here levels in a multilevel model, respectively level 1 and level 2. When there are two levels, as in this example, unit- and group-level are often used as synonyms to level 1 and level 2, in the multilevel literature.

A random-intercept model, where only the intercept varies between schools, for student i, $(i = 1, \ldots, n_j)$ within school j, $(j = 1, \ldots, N)$, can be written as:

$$\begin{aligned} & y_{ij} = \beta_{0j} + \beta_1 x_i + \epsilon_{ij} \\ & \beta_{0j} = \beta_0 + \delta z_j + \zeta_{0j} \\ & or \\ & y_{ij} = \beta_0 + \zeta_{0j} + \beta_1 x_i + \delta z_j + \epsilon_{ij}. \end{aligned} \tag{1}$$

The random intercept or level-2 residual ζ_{0j} is a school-specific component, which is constant for all students in the same school. The level-1 residual ϵ_{ij} is student-specific and varies between students i as well as schools j. Moreover, ζ_{0j} are independent over schools, ϵ_{ij} are independent over schools and students, and the two error components are independent and assumed to be normally distributed, i.e.

$$\epsilon_{ij} \sim N\left(0, \sigma_\epsilon^2\right), \ \zeta_{0j} \sim N\left(0, \sigma_{\zeta_0}^2\right)$$

$$and \text{ using } \perp for\ independence$$

$$(\epsilon_{ij}, \zeta_{0j}) \perp (\epsilon_{ij'}, \zeta_{0j'}) \quad \forall j \neq j'.$$

Note that the total residuals or error terms are homoskedastic, $Var(\zeta_{0j} + \epsilon_{ij}) = \sigma_{\zeta_0}^2 + \sigma_\epsilon^2$. Therefore, the intraclass correlation for responses y_{ij} and $y_{i'j}$ for school j, given the covariates is

$$\rho \equiv Cor(y_{ij}, y_{i'j} | x_i, z_j) = \frac{\sigma_{\zeta_0}^2}{\sigma_{\zeta_0}^2 + \sigma_\epsilon^2}. \tag{2}$$

A random-coefficient model, with both varying intercept and varying slope for the predictor pre-test x_i, can be written as:

$$y_{ij} = (\beta_0 + \zeta_{0j}) + (\beta_1 + \zeta_{1j})x_i + \delta z_j + \epsilon_i j. \tag{3}$$

The random slopes ζ_{1j} can be interpreted as interactions between schools and the covariate x_i. It is generally assumed that the three random terms are uncorrelated with x_i and z_j and that ϵ_{ij} are uncorrelated with both ζ_{0j} and ζ_{1j}. Both the random intercepts ζ_{0j} and random slopes ζ_{1j} are independend across schools and the level-1 residual is independent across schools, and students. Given the covariates, ζ_{0j} and ζ_{1j} are assumed to have a bivariate normal distribution with zero mean and covariance matrix:

$$\mathbf{\Omega} = \begin{bmatrix} \sigma^2_{\zeta_0} & \sigma_{\zeta_0\zeta_1} \\ \sigma_{\zeta_1\zeta_0} & \sigma^2_{\zeta_1} \end{bmatrix}, \ \sigma_{\zeta_0\zeta_1} = \sigma_{\zeta_1\zeta_0}.$$

The covariance matrix $\mathbf{\Omega}$ is not straightforward to interpret. First, the $\sigma^2_{\zeta_1}$ and $\sigma_{\zeta_0\zeta_1}$ do not only depend on the scale of the dependent variable but also the scale of the covariate. Second, the total variance is not constant, if we let $\xi_{ij} = \zeta_0 + \zeta_1 x_i + \epsilon_{ij}$, the total residual, $Var(\xi_{ij}|x_i, z_j) = \sigma^2_{\zeta_0} + 2\sigma_{\zeta_1\zeta_0} + \sigma^2_{\zeta_1} x_i^2 + \sigma^2_{\epsilon}$. Hence, the interclass correlation becomes an complicated function of x_i. The exception is the case for two students in a given school with $x_i = x_{i'} = 0$, when the expression for the interclass correlation is the same as for the random-intercept model. Finally, the values of $\sigma^2_{\zeta_0}$ and $\sigma_{\zeta_0\zeta_1}$ can be difficult to interpret since they depend on how much we add or subtract from the covariate. This is because the intercept variation is the variability at the vertical position of the school-specific regression lines where $x = 0$. Hence, it make sense to transform x so $x_i = 0$ is meaningful in some way. A useful way to interpret the magnitude of the estimated variances is to consider the intervals $\beta_0 \pm 1.96\sqrt{\sigma^2_{\zeta_0}}$ and $\beta_1 \pm 1.96\sqrt{\sigma^2_{\zeta_1}}$ that contains, respectively, about 95% of the intercepts and slopes in the population (Rabe-Hesketh and Skrondal, 2008).

A general mixed-effects model can be formulated as

$$y_{ij} = \sum_{h=1}^{r} \beta_h x_{hij} + \sum_{s=1}^{q} \zeta_{sj} z_{sij} + \epsilon_{ij}. \tag{4}$$

With r $(h = 1, \dots, r)$ fixed regressors x_{hij} that include an intercept and q $(s = 1, \dots, q)$ regressors z_{sij} with random regression coefficients ζ_{sj}. β_h are r fixed parameters and ζ_{sj} q zero mean random parameters. In many applications, z_{sij} are a subset of the regressors x_{hij}. The random-intercept model in Equation 1 and random-coefficient model in Equation 3 are all special cases of Equation 4.

The error components ζ_{sj} are group-specific and remain constant across observations, while the unit-specific component ε_{ij} varies between unit i as well as group j. The group-specific error components represent the combined effects of omitted group characteristics and unobserved group heterogeneity. Since ζ_{sj} is shared by all observations in same group, it induces within-group dependence among the total residual ξ_{ij}. Note that most applications of multilevel analysis assume for convenience that σ^2_ε are the same for all j and that the random coefficients ζ_{sj} are independent across all j.

Motivations for multilevel modelling

Multilevel analysis is not limited to the case of two levels and can readily be extended to the case of three or more levels. For simplicity this section, generally, focuses on the example of two levels but reasoning should be straightforward to extend to three or more levels.

Study effects that vary over groups. One of the basic uses of regression analysis is estimate the effect x on y, ceteris paribus. In many applications, it is not the overall effect of x that is of interest but how this effect varies in the population. Multilevel models allow the study of effects that vary by group—for example, schools. In classical regressions, this variation can be studied by using interactions, but estimates of varying effects can be noisy, especially for groups with few observations (Gelman and Hill, 2007).

Use all data to perform inference. Multilevel modelling creates a compromise between the noisy within-group estimate, especially for regressions for groups with small sample sizes, and the oversimplified regression estimate when the group indicators are ignored. The two alternatives to multilevel modelling are *complete pooling*, in which differences between groups are ignored, and *no pooling*, in which each group is analysed separately. Multilevel analysis can hence be defined as *partial pooling*. By partial pooling, multilevel analysis includes information that is ignored in no pooling and allows for variation that is retained in complete pooling. The result is *more efficient inference.*

Analysis of clustered data. When data are hierarchical, and hence clustered, or collected from clustered sampling, multilevel modelling provides a direct way to include indicators for clusters at all levels in complex hierarchical data. By using the cluster information, multilevel analysis provides correct standard errors, which generally are more conservative than the ones obtained by ignoring the presence of clustering (Goldstein, 2003).

Predictors at two, or more, different levels. In classic regression, it is not possible to include both group-level indicators and group-level covariates, as done in Equations 1 and 3, since the predictors would become collinear. One possible approach in classic regression is a *two-step analysis*, with first fitting the model with group indicators but without the group-level covariates, followed by a group-level regression of the estimated group indicators. But such approach is only possible with a large sample size in each group. Multilevel modelling is coherent and simultaneously incorporates both models at different levels (Gelman and Hill, 2007).

Predictions. If the sample size in each group is sufficiently large, it is possible to use *two-step analysis*, as mentioned above, for obtaining predictions for a new group. From multilevel models it is also possible to make such predictions for small sample sizes, and account for the uncertainty at both levels. Using the previous example of pupils nested in schools, compared to classical regression multilevel models can furthermore make predictions for a new student in a new school (Gelman and Hill, 2007).

Hence, it may be effective to apply multilevel modelling when it is of interest to:

- Account for more than one level of variation when estimating higher-level regression coefficients.
- Modelling variation of lower-level coefficients.
- Estimating coefficients for groups with small sample sizes.

Potential drawbacks of multilevel modelling

Additional modelling assumptions. Each level in a multilevel model can be seen as its own regression with its own set of assumptions such as additively, linearity, independence, equal variance, and normality (Gelman and Hill, 2007). It is, however, generally possible to check these assumptions, see for example Snijders and Johannes (2008), for an discussion about diagnostic checks in multilevel models, and Draper (2008), for a discussion about *MCMC* diagnostic in multilevel models. Furthermore, since multilevel modelling can be seen as partial pooling, classic regressions of no pooling and complete pooling are special cases of multilevel models. Returning to the example of pupils nested in schools, complete pooling and no pooling occur when school level variance equals zero or infinity.

Additional complexity. A main feature of multilevel analysis is its complexity. It is applied to complex data structures and includes coefficients varying by group. The complexity is often welcome. It does, however, add some new difficulties in understanding and summarising the model, as seen for the random-coefficient model in Equation 3.

Some words about estimation of multilevel models

Maximum likelihood (ML). A classical method for estimating the parameters of statistical models is determining the *ML*. The likelihood function is the joint probability density for all observed responses as a function of all model parameters. In the example of pupils nested in schools in Equation 1, the responses y_{ij} are a function of the parameters β_0, β_1, δ, $\sigma^2_{\zeta_0}$, and σ^2_ϵ. The underlying idea is to find parameter estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\delta}$, $\hat{\sigma}^2_{\zeta_0}$, and $\hat{\sigma}^2_\epsilon$ that maximise the likelihood function, hence making the responses as likely as possible. *ML* estimators has been shown to have good properties such as consistency and efficiency.

For linear multilevel models, the marginal likelihood (the joint probability of all responses given the covariates) can be evaluated and maximised. Linear multilevel models can therefore be estimated in most standard statistical software such as R, Stata, and SAS. However, for generalised linear multilevel models, the marginal likelihood does not have a closed form and must be evaluated by approximate methods. In most statistical software, these kinds of models are estimated with approximate quadrature methods. The main drawback with quadrature methods is that the quadrature points increase geometrically with the number of random coefficients (Goldstein, 2003), thus limiting the complexity of models that can be estimated. Estimation of generalised linear multilevel models and quadrature methods are discussed in e.g. Hedeker (2008), Goldstein (2003), and Snijders and Johannes (2008). In Chapter 3, generalised linear multilevel models and in Chapter 4, non-hierarchical multilevel models, are estimated by the adaptive quadrature[2].

Bayesian multilevel analysis and Markov Chain Monte Carlo (MCMC). For more complex models, *MCMC* methods are competitive estimation alternatives[3]. Because *MCMC* considers the whole distribution, rather than only the mode, it is generally slower. But as the complexity increases, the speed of *MCMC* becomes more competitive. *MCMC* also offers the opportunity to fit a varity of models and add complexity where needed. It is therefore possible to estimate multilevel models with *MCMC* that are not straightforward to estimate with other methods. In Chapter 1, multiple-membership linear multilevel models, and Chapter 2, multiple-membership generalised linear multilevel models and Chapter 4, non-hierarchical multilevel models are estimated by *MCMC* methods[4].

[2]The models were estimated in the GLLAMM (Generalized Linear Latent and Mixed Models) program for Stata, for an review see Rabe-Hesketh et al. (2002, 2005).

[3]Introductions to Bayesian multilevel analysis and *MCMC* can be found in e.g. Draper (2008) and Browne (1998).

[4]The multiple-membership models are estimated in Stata and MLwiN (Browne, 2009; Rasbash et al., 2009) using the *runmlwin* command (Leckie and Charlton, 2011). Estimation of non-hierarchical multilevel models are done in WinBugs 1.4, where Bugs is an acronym for Bayesian Inference using Gibbs Sampling.

Some clarifications on definitions

Multilevel or hierarchical. Multilevel models are also called hierarchical, for mainly two reasons: *first*, from the structure of the data (for example, pupils nested within schools); and *second*, from the model itself. With the parameters of the within-school regression at the bottom level, controlled by the hyperparameters of the second level model, multilevel models easily extend to the more complex structure three or more levels such as pupils within schools within school districts. Chapter 1 consider a model with farms within parishes within municipalities.

Multilevel models can also be non-nested—for example, individual observations that are nested within states and years. Neither state or year is, however, nested within the other. That is considered a model in which individuals, years and states are on different levels without the requirement of hierarchy. This type of non-nested multilevel model is often called *cross-classified* since the individual observations are cross-classified in years and states. Chapter 1 consider parishes cross-classified into regions and homogeneous areas of agricultural land; and Chapter 4 applies a cross-classified model to separate unit-specific variability from the variable-specific variability and statistical noise.

The data structure can also be non-nested in the sense that individuals can be members of more than one higher-level unit at the same time. For example, pupils may attend more than one schools. This type of non-nested multilevel model is often called *multiple-membership* since the individual observations are members of multiple higher level groups at the same time. Chapter 1 and Chapter 2 apply a multiple-membership model to capture the spatial spill-over effect of, respectively, neighbouring units and groups where the units are members of multiple neighbouring units or groups.

Random and fixed effects. Multilevel models, are known as random-effects or mixed-effects models. In multilevel models there are *random effects* in the sense that they are random outcomes of a process identified with the model that is predicting them. *Fixed effects* corresponds to either parameters that do not vary, for example fitting the same regression line for all schools or to parameters that vary but not modelled themselves (for example, a regression model that includes indicators for schools) Mixed-effects models, hence, include both fixed and random effects; for example, Equation 1 includes a (random) intercept that has a group-level model and fixed parameters β_0, β_1, and δ that are constant over groups. Moreover, as noted in Gelman and Hill (2007), fixed effects can be seen as a special case of random effects, where the higher level variance is equal to respective zero or ∞.

Summary of appended papers

The Capitalisation of Single Farm Payments on Farm Prices: An Analysis of Swedish Farm Prices Using Farm-Level Data

This first essay in the thesis estimates capitalisation effects of farm attributes, with a particular focus on decoupled support to farmers in the form of *single farm payments (SFP)*. The essay provides evidence from the Swedish market for farms by analysing a sample of 3 383 farm transactions sold all across the country with linked local *SFP*.

The implementation of the *Common Agricultural Policy reform (CAP)* 2003 and the subsequent decoupling of agricultural support payments has increased the attention towards capitalisation effects of such payments on the price of land. Signs of capitalisation would indicate that benefits are partly transferred towards landowners rather than to producers, and that payments intended to support farmers' incomes are transmitted to sectors other than agriculture. Moreover, capitalisation of support payments might reduce efficiency in the agricultural sector in that new farmers would face a higher entry cost, and existing farmers higher expansion costs (Ciaian et al., 2008).

In Sweden, the *SFP* have been implemented according to the historical model, that is the initial distribution of *SFP* entitlements was conditional upon historical yields.

Key empirical from a spatial multiple-membership model suggests that the local effect of *SFP* is negative while there is a positive between-region effect of *SFP* on farm prices. The quality of agricultural land was controlled for using the Swedish Board of Agriculture's homogeneous areas (SKO), with respect to mainly land fertility but soil quality, topography and climate have also been taken into consideration.

The regional capitalisation of *SFP*, after controlling for structural attributes, and regional characteristics, a one per cent increase in average *SPF* increase the average transaction price with 0.5 per cent. Furthermore, spatial heterogeneity was found for both regional and local levels, and a large spatial spill-over effect was found in-between neighbouring farms.

Getting a Full-Time Job as a Part-Time Unemployed—How Much Does Spatial Context Matter?

The second essay investigates the extent to which differences in the probability to exit from part-time unemployment to a full-time job are accountable for by spatial contextual factors and individual characteristics. The motivation for this study from a policy perspective is that the effect of contextual factors potentially can set limits for how successful the labour market policy can be. If contextual factors do affect the possibility to exit part-time unemployment for full-time employment, policy makers need to take into consideration the need of regionally adapted labour market policy measures. The influences on the probability of transition can be separated into three separate categories:

- Within-context effects, such as the population size of the context, or individual characteristics.
- Context effects or spatial heterogeneity, arising from individuals being grouped into geographical or administrative contexts, e.g. municipality or other classifications where the individual lives and works.
- Neighbourhood effects or spatial spill-over effects, as contexts which are close to each other in geographical space may share common environmental, social or demographic factors influencing the transition from part-time employment to full-time employment. Furthermore, contexts are formed by using geopolitical boundaries which may be unrelated to the transition of interest; we may hence use spatial smoothing of the distribution of probabilities of exit part-time unemployment to remove any artifactual variation due to the method of data aggregation.

In contrast to other recent studies on transition out of part-time unemployment (see Mnsson and Ottosson (2011) for an overview), the main focus in this study is on the effect of spatial contextual factors, for example how much of the probability to leave part-time unemployment depends on conditions regarding regional growth, local labour market competition, etc. However, instead of using a traditional econometric framework, we use a multilevel approach that enables simultaneous identification of how individual-level and contextual-level variables are related to individual-level outcomes. Furthermore, the multilevel analysis provides the framework to simultaneously investigate the unexplained contextual effect and contextual-level variables.

The results indicate that there is a contextual effect and that there are some spatial spill-over effects from neighbouring municipalities, and unemployment rate partly explains the context variability. Furthermore, the contextual effect is found to be especially large for individuals without a university degree.

Intergenerational Transmission of Education and Family Heterogeneity: A Study of Third-Generation Immigrants and Natives in Sweden

The third essay investigates the determinants of educational attainment for third-generation immigrants and natives in Sweden. Although much research has been undertaken to explain the determinants of educational attainment, there are still many questions unanswered. Some of those issues, like the impact of grandparents, the impact of different population groups over generations, as well as how to handle family heterogeneity, are investigated in this paper.

Using a mixed-effects model that includes unobserved family heterogeneity for linked register data, the main result is that the effect of parent's educational attainment is mainly due to the between-parental education effect of family income. The results suggest that the effect of parental education on the child's educational attainment are due to the family characteristics that influence both family income and parent's educational attainment, for example family culture and genetic endowments.

Moreover, immigrant paternal grandmothers from Europe were found to have a negative impact on educational attainment for third-generation immigrants. Immigrant paternal grandfathers from East Europe and maternal grandfathers from other European (not Nordic or East European countries) were found to have a positive impact.

A New Strategy for Performance Evaluation in the Case of Panel Data: Based on a Cross-Classified Multilevel Model

The fourth and last essay presents a new robust strategy for performance evaluation in the case of panel data that is based on routinely collected variables or indicators. The suggested strategy applies a cross-classified, mixed-effect model that can be implemented in most standard statistical packages. Moreover, it allows for correlated variables and missing observations, and it can be extended to include linear trends and heavy-tailed distributions. The strategy is implemented in two illustrative empirical examples, and the robustness is investigated in a Monte Carlo study. The simulation results indicate that the strategy performs well in cases of measurement error, and misspecification.

Finding the best and the worst performance is a common evaluation problem in several disciplines, including economics, finance, sociology, and education. Identification of the units (e.g. firms, organisations, individuals, projects) that have the best and worst performances provides benchmarking units. Such benchmarking units are central to any kind of performance evaluation since they supply both good and bad examples that can be evaluated and since they can also be used for comparative evaluation with other units. The paper offers a general empirical strategy for identifying these benchmarking units of ‘best performance’ and ‘worst performance’.

It is important to stress that each evaluated performance measure has it own characteristics regarding the nature and the definition of the problem. Even if the units and data differ, however, the empirical performance evaluation has possible complexities that are associated with empirical applications, including correlated variables, missing values, measurement errors, misspecification, and linear trends. Furthermore, when dealing with units observed over time (i.e. panel data), the dimensions of the data set must be correctly reduced if the goal is to obtain a performance ranking.

References

Browne, W. (1998). *Applying MCMC methods to multilevel models.* Ph. D. thesis, University of Bath, UK.

Browne, W. (2009). *MCMC estimation in MLwiN, v2.13.* University of Bristol, UK: Centre for Multilevel Modelling.

Ciaian, P., D. Kancs, and J. F. M. Swinnen (2008). Static and dynamic distributional effects of decoupled payments: Single farm payments in the European Union. *SSRN Working Paper Series*.

de Leew, J. and E. Meijer (2008). Introduction to multilevel analysis. In J. de Leew and E. Meijer (Eds.), *Handbook of Multilevel Analysis*, Chapter 1, pp. 1–75. New York, USA: Springer.

Draper, D. (2008). Baysian multilevel analysis and MCMC. In J. de Leew and E. Meijer (Eds.), *Handbook of Multilevel Analysis*, Chapter 2, pp. 77–139. New York, USA: Springer.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* New York, USA: Cambridge University Press.

Goldstein, H. (2003). *Multilevel Statistical Models* (3 ed.). London, UK: Kendall's library of statistics, Arnold.

Hedeker, D. (2008). Multilevel models for ordinal and nominal variables. In J. de Leew and E. Meijer (Eds.), *Handbook of Multilevel Analysis*, Chapter 6, pp. 237–274. New York, USA: Springer.

Hox, J. J. and J. K. Roberts (Eds.) (2011). *Handbook of Advanced Multilevel Analysis.* New York, USA: Routledge, Taylor & Francis Group.

Leckie, G. and C. Charlton (2011). *runmlwin: Stata module for fitting multilevel models in the MLwiN software package.* University of Bristol, UK: Centre for Multilevel Modelling.

Mnsson, J. and J. Ottosson (2011). Transitions from part-time unemployment: Is part-time work a dead end or a stepping stone to the labour market. *Economic and Industrial Democracy*.

Rabe-Hesketh, S. and A. Skrondal (2008). *Multilevel and Longitudinal Modeling Using Stata* (2 ed.). College Station, Texas, USA: Stata Press.

Rabe-Hesketh, S., A. Skrondal, and P. A (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics 128*(2), 301–223.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2002). Reliable estimation of generalized linear models using quadrature. *Stata Journal 2*(1), 1–21.

Rasbash, J., C. Charlton, W. Browne, M. Healy, and B. Cameron (2009). *MLwiN Version 2.1.* University of Bristol, UK: Centre for Multilevel Modelling.

Snijders, T. A. B. and B. Johannes (2008). Diagnostic checks for multilevel models. In J. de Leew and E. Meijer (Eds.), *Handbook of Multilevel Analysis*, Chapter 3, pp. 77–139. New York, USA: Springer.