NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Aleksey Oshchepkov, Anna Shirokanova

MULTILEVEL MODELING FOR ECONOMISTS: WHY, WHEN AND HOW

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 233/EC/2020

Aleksey Oshchepkov[1], *Anna Shirokanova*[2]

MULTILEVEL MODELING FOR ECONOMISTS: WHY, WHEN AND HOW[3]

Multilevel modeling (MLM, also known as hierarchical linear modeling, HLM) is a methodological framework widely used in the social sciences to analyze data with a hierarchical structure, where lower units of aggregation are 'nested' in higher units, including longitudinal data. In economics, however, MLM is used very rarely. Instead, economists use separate econometric techniques including cluster-robust standard errors and fixed effects models. In this paper, we review the methodological literature and contrast the econometric techniques typically used in economics with the analysis of hierarchical data using MLM. Our review suggests that economic techniques are generally less convenient, flexible, and efficient compared to MLM. The important limitation of MLM, however, is its inability to deal with the omitted variable problem at the lowest level of data, while standard economic techniques may be complemented by quasi-experimental methods mitigating this problem. It is unlikely, though, that this limitation can explain and justify the rare use of MLM in economics. Overall, we conclude that MLM has been unreasonably ignored in economics, and we encourage economists to apply this framework by providing 'when and how' guidelines.

JEL codes: C18, C50, C33, A12

Key words: multilevel modeling, hierarchical linear modeling, mixed effects, random effects, fixed effects, random coefficients, clusterization of errors.

[1] National Research University Higher School of Economics, Moscow, Centre for Labour Market Studies, Senior Research Fellow, E-mail: aoshchepkov@hse.ru. The corresponding author.

[2] National Research University Higher School of Economics, Saint-Petersburg, Laboratory for Comparative Social Research, Senior Research Fellow, E-mail: a.shirokanova@hse.ru

[3] The study was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project '5-100'. We are thankful to Malcolm Fairbrother, Rostislav Kapelyushnikov, Hartmut Lehmann, Arthur Lewbel, Anna Lukyanova, Eduard Ponarin, Francesco Sarracino, Boris Sokolov, and Yukai Yang for their remarks and comments on the earlier drafts. The paper has benefited much from discussions at iCare-5 conference on Applied Economics in Perm in 2017 and 1st Italian Workshop of Econometrics and Empirical Economics in Milan in 2018.

1. Introduction

Many datasets used in the social sciences, including economics, have a hierarchical structure, where lower units of aggregation are 'nested' in higher units called clusters or groups. Examples include data on individuals nested in households, employees nested in firms, regions in countries. Likewise, panel data represent unit-time observations nested within units (see Fig.1 and Fig. 2 for visualized examples). A hierarchical structure is innate to datasets used in meta-regression analyses (Stanley & Jarrell, 1989).

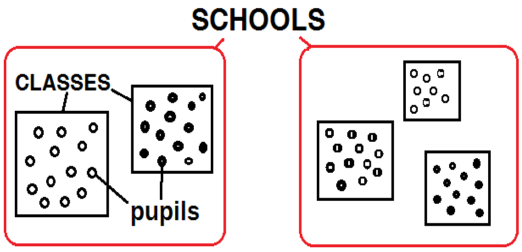


Fig.1. Cross-sectional 'nested' data: pupils (Level 1) are nested in classes (Level 2), while classes are nested in schools (Level 3).

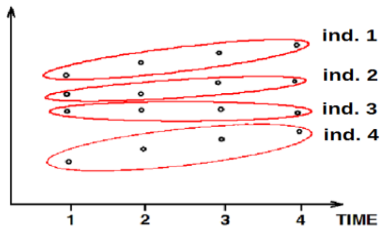


Fig.2. Panel data: individual-time observations (Level 1) are nested in individuals (Level 2).

In many disciplines, such data are analyzed using *multilevel modeling* (MLM), also known as *hierarchical linear modeling* (HLM) (Bryk & Raudenbush, 1992; Gelman & Hill, 2007; Goldstein, 1995; Hox, Moerbeek & Van de Schoot, 2010; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Rabe-Hesketh & Skrondal, 2012; Snijders & Bosker, 2012).[4] Over the last quarter of a century, the framework has become commonplace in education (Raudenbush & Bryk, 2002), public health (Diez-Roux, 2000), sociology (Schmidt-Catran & Fairbrother, 2014), management, marketing, and international business studies (Carter, Meschnig & Kaufmann, 2015; Hoffman, 1997; Ozkaya et al., 2013); it has 'blossomed in political science' (Kedar & Shively, 2005) and even been hailed 'unnecessarily ubiquitous' in psychology (McNeish, Stapleton & Silverman, 2016).

The puzzle is that MLM has been relatively unknown and rarely used in economics. Authors outside of economics have mentioned this fact (e.g., Bell & Jones, 2015; Dieleman & Templin, 2014; McNeish, Stapleton & Silverman, 2016), but the quantitative evidence is much more impressive. Our analysis of the Web of Science Core Collection shows that the relative number of journal articles containing MLM-related terms in economics is several times smaller than that in other social sciences.[5] Moreover, this gap has been widening over time (see Fig.3).

[4] The adjective 'linear' should not cast confusion that MLM/HLM deals exclusively with linear models. The modern MLM framework covers a wide range of non-linear models for binary, ordinal, multinomial, and count outcomes (see e.g., Hedeker & Gibbons, 2006; Snijders & Bosker, 2012). This review focuses on linear models for continuous outcomes to keep things simple and easily comparable with the OLS-based techniques.

[5] We searched for articles containing MLM-related terms in titles, abstracts, or keywords. Two alternative sets of search terms were used. The narrow set includes 'hierarchical linear', 'multilevel model*' and 'multilevel analysis', while the wide set

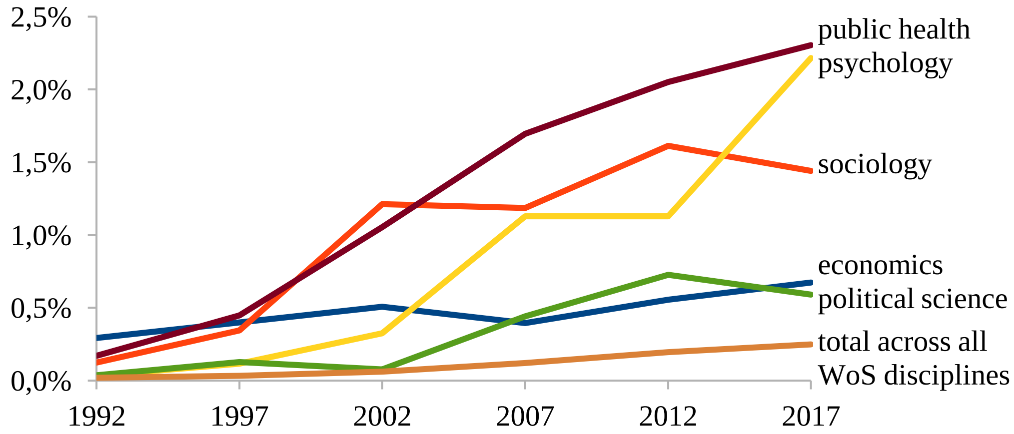


Fig. 3. Relative number of articles with MLM-related terms across disciplines, 1992-2017.

Notes. Articles containing any of the terms ‘hierarchical linear’, ‘multilevel model*’, ‘multilevel analysis’, ‘random effect*’, ‘random coefficient*’ and ‘mixed effect*’ in their titles, key words, or abstracts from Web of Science Core Collection, filtered by discipline according to the Web of Science classification. The numbers are provided for every 5th year (1992, 1997, etc.). The starting point is year 1992, when the first edition of the popular textbook on hierarchical linear modeling by Bryk and Raudenbush (1992) was published.

The evidence of this gap in use is supported by our review of articles published in top journals. For instance, the number of MLM-related articles published in the top five economics journals in 2002-2017 was over ten times smaller than that in the top sociology journals (see Table 1).

Table 1. Number of publications containing MLM-related terms in top economics and sociology journals, 2002-2017.

	Economics Journals	Items	Sociology Journals	Items
1	*Quarterly Journal of Economics* (#1)*	0	*Annual Review of Sociology* (#1)	4
2	*Review of Economic Studies* (#9)	4	*American Sociological Review* (#2)	32
3	*American Economic Review* (#10)	2	*Sociological Methods & Research* (#3)	18
4	*Journal of Political Economy* (#11)	1	*American Journal of Sociology* (#6)	17
5	*Econometrica* (#22)	6	*European Sociological Review* (#20)	79

*The number in parentheses (#) is the journal’s rank by journal impact factor in Web of Science for 2017. Top journals of general scope or methods journals were chosen for comparison. The set of search terms in titles, keywords, and abstracts (one entry counted if matching several search terms): *hierarchical linear; multilevel analysis; multilevel model; random effect; random coefficient; mixed effects*. Authors’ calculations.

The existence of the gap is further corroborated by a review of econometric textbooks. The latest editions by Green (Green, 2017) and Wooldridge (Wooldridge, 2010a) mention MLM in the chapters on panel data analysis as an unconventional way either to adjust standard

additionally includes ‘random effect*’, ‘random coefficient*’ and ‘mixed effect*’ (controlling for cases when ‘mixed effects’ reflect results’ uncertainty). The former helps to find MLM-using papers but can underestimate their number because of differences in terminology, especially within economics due to the ongoing ‘fixed effects’ vs. ‘random effects’ debates. The latter, by contrast, can overestimate the number of published MLM-using articles. To construct Fig.3, we used the wide set of search terms and treated it as an optimistic estimate for the number of MLM-using articles in economics. Fig. A1 (in Appendix A) shows the difference when the narrow set of search terms is applied.

errors for clusterization or to model differences in the intercepts and slopes across panels. Notably, very few references to core MLM papers are provided. Our review of a few popular textbooks on panel data analysis yields similar evidence of the neglect of MLM (see Appendix B).

The question is why economists are ignoring MLM, a well-developed methodological framework widely used in other disciplines. The understanding of this issue could benefit both economists and non-economists. If MLM is better than methods used in economics, then economists should use this framework in their research. On the other hand, if MLM has serious flaws – at least in economists' eyes – then highlighting them for practitioners would help to avoid dubious conclusions and would, potentially, define ways for improving MLM, while attracting substantial numbers of empirical social scientists towards economic methods of data analysis.

To shed some light on this issue, we reviewed a large body of interdisciplinary methodological literature and contrasted MLM with techniques typically used by economists when analyzing data with a hierarchical structure, including panel data. Our review suggests that MLM may be more appropriate than the cluster-robust estimation technique when dealing with the clusterization of errors, be more flexible than the fixed effects model when dealing with unobserved heterogeneity, and, finally, be more parsimonious and efficient than the models with cross-level interactions and the estimated dependent variable models used to model coefficient heterogeneity across higher level units. We also show that these advantages of MLM tend to be more pronounced in a panel-data setting. Therefore, MLM can be successfully employed in the analysis of hierarchical data instead of using separate econometric techniques. Acknowledging this fact, some economic authors do apply MLM, but they do not generalize the (dis)advantages of this framework as compared to the econometric techniques popular in economics.[6] In this regard, our study provides an important theoretical rationale for the decision to use MLM in economic studies.

Our review reveals a weak spot in the core MLM framework, although this is unrelated to the hierarchical data setting. MLM ignores the possible omitted variable problem at the lowest level of data by assuming that explanatory variables at that level are not correlated with

[6] We may refer to at least two characteristic examples. One is a study by Oort et al. (2012), where the authors argue that MLM can be extremely useful in the analysis of why and how agglomeration economies affect firms' productivity. According to the authors, the benefits of MLM in this case are that: 1) it allows taking into account the similarity of firms within clusters (regions or industries) and 2) it allows modeling the heterogeneity of the impact of agglomeration on productivity across clusters. The other example comes from Bell, Johnston & Jones (2015). The authors re-analyze the study by Reinhart & Rogoff (2010) and show that taking into account the heterogeneity of slopes across countries using MLM triples the size of standard errors and makes the effect of government debt on economic growth statistically insignificant. Our review generalizes and explains the advantages of MLM in both applications.

errors (similar to the Gauss-Markov's assumption of exogeneity in the OLS regression). In this regard, economic techniques are preferable as they can be complemented by quasi-experimental methods (e.g, IV, see Angrist & Pishke, 2009, 2010), yet these methods are poorly integrated into the MLM framework.

Overall, MLM's neglect of the omitted variable problem at the lowest level of the data hierarchy is its major limitation compared to methods popular in empirical economic research. Although this problem is often central in empirical economic studies, it is unclear to what extent this limitation can explain infrequent use of MLM by economists. The fact is that a variable omitted at the lowest level of data is often not the key issue in the analysis of hierarchical data. Moreover, the application of quasi-experimental methods often presents a challenge in practice (French & Popovici, 2011; Young, 2019). Most likely, many economists seem to be unaware of the MLM framework in general, without acknowledging its possible advantages and shortcomings. As we find, even those studies that apply MLM-related methods are vaguely connected to the core MLM literature (see Appendix C). This corroborates the evidence that economics as a discipline is more insular and hierarchical than other social sciences (Fourcade et al., 2015; Pieters & Baumgarrner, 2002), which, among other things, means a stronger imprint on current economic research of traditional mainstream approaches (Kapeller et al., 2017). Therefore, our most general conclusion is that MLM has been unreasonably ignored and seriously underrated in economics, and thus we encourage economists to apply this framework in their analyses of data with a hierarchical structure.

We also have a message to practitioners that tend to apply MLM 'by default': become more aware of the endogeneity issue when using MLM. Its innate vulnerability to the omitted variable problem at the lowest level should be acknowledged when interpreting any estimation results. In cases when a researcher is interested in a single-level model only, without analyzing the contextual effects and coefficient heterogeneity across higher-level units, the 'good old' OLS regression with cluster-robust standard errors and fixed effects at higher levels should be retained as an valid alternative to MLM.[7]

The rest of the paper is organized as follows. Section 2 describes the general MLM framework as oriented towards applied economists. Section 3 provides a detailed comparison of MLM with techniques used in economics to analyze data with a hierarchical structure. Section 4 provides a summary and practical recommendations on the use of MLM by economists.

[7] For an earlier discussion, see Gorard (2007) and McNeish (2014).

2. A (very) brief introduction to MLM

In the most general sense, MLM is a regression framework designed to analyze data with a multilevel structure. Introductory sections of textbooks on MLM usually juxtapose it with two straightforward approaches to handling such data: aggregating individual-level variables to the group level and disaggregating group-level variables to the individual level. MLM is presented as an intermediate approach that allows combining individual- and group-level predictors in single regression framework, avoiding the drawbacks of those two approaches. Below we provide a brief introduction to the MLM framework, necessary for understanding MLM and for further comparisons between MLM and 'economic methods'. A much more extensive treatment can be found in Gelman & Hill (2007), Raudenbush & Bryk (2002), Rabe-Hesketh & Skrondal (2012), Snijders & Bosker (2012) and other textbooks.

2.1 The multilevel model

A two-level model can be represented as a system of three simultaneous equations[8]:

$$Y_{ij} = \beta_{0j} + \beta_{1j} * X_{ij} + r_{ij} (1)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * W_j + u_{0j} (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * W_j + u_{1j} (3)$$

Equation 1 refers to the lowest level of data, e.g., individual level (i, level 1), while Equations 2 and 3 refer to a higher level, i.e., group level (j, level 2). In Equation 1, Y_{ij} is the dependent variable; X_{ij} denotes the individual-level explanatory variables; β_{0j} and β_{1j} are the intercept and slope coefficients which are allowed to vary across groups, and r_{ij} is the individual-level error term.

Equation 2 serves to model the variation in intercepts β_{0j} across groups, assuming that these intercepts depend on some group-level explanatory variable W_j with the coefficient γ_{01}. The constant term γ_{00} is the expected value of β_0 when $W_j = 0$, while u_{0j} is the error term that represents the remaining variability in the intercepts after controlling for W_j. In a similar way, Equation 3 models variation in slopes β_{1_j}.[9] They are assumed to depend on W_j with the coefficient γ_{11}, where γ_{10} is the constant term and u_{0j} is the error term.

[8] Three- and higher level models are constructed following the same logic. In all equations, we use the notations proposed in Raudenbush & Bryk (2002).

[9] Coefficient β_{1j} at each individual-level variable X is modeled through its own Equation 3 and, thus, the number of Equations 3 corresponds to the number of Xs.

All error terms in the model – r_{ij}, u_{0j}, and u_{1j} – are assumed to be identically (normally) and independently distributed and averaged at zero, given the values of *X* and *W*: $r_{ij} \sim N(0,\ \sigma^2)$, $u_{0j} \sim N(0, \tau_{00})$, and $u_{1j} \sim N(0, \tau_{11})$. Terms u_{0j} and u_{1j} are called 'random effects', thus assuming that intercepts β_{0j} and slopes β_{1j} contain random group-level components. [10]

Equations 1, 2, and 3 can be combined into one equation, so that the whole model is presented as:

$$Y_{ij} = \gamma_{00} + \gamma_{01} * W_j + \ \gamma_{10} * X_{ij} + \gamma_{11} * X_{ij} * W_j + u_{0j} + u_{1j} * X_{ij} + r_{ij} (4).$$

Therefore, a typical multilevel model consists of a fixed part ($\gamma_{00} + \gamma_{01} * W_j + \gamma_{10} * X_{ij} + \gamma_{11} * X_{ij} * W_j$) and a random part ($u_{0j} + u_{1j} * X_{ij} + r_{ij}$), which explains another popular name for this model, the *mixed-effects model* (Raudenbush & Bryk, 2002; Luke, 2004).

A typical multilevel model, introduced above, is a 'complete' multilevel model which allows both the intercepts and slopes for each *X* to vary across groups and includes explanatory variables at the individual and group levels. Not all multilevel models have to be complete. The general logic of the multilevel model-building routine, as reflected in many MLM textbooks, is to move from the simplest to more complex models, and this process terminates when the model fits a research issue and its theoretical considerations. The modeling usually begins with the null model (aka the 'empty', 'ANOVA' model) that contains no predictors:

$$Y_{ij} = \beta_{0j} + r_{ij} (5)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} (6)$$

In the combined form this model reads as follows:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} (7)$$

In the null model, the total variance of Y_{ij} may be decomposed into group variance τ_{00} and individual variance σ^2. Individual values of Y_{ij} are represented as deviations from the group mean β_{0j}, which, in turn, is treated as a deviation from the grand mean, γ_{00}. As a result, each individual Y_{ij} is modelled as a deviation around the grand mean. The proportion of variance in *Y* accounted for by the group level, $\frac{\tau_0^2}{\sigma^2 + \tau_0^2}$ (where τ_0 is the group-level variance of *Y*), is called the intra-class correlation coefficient (ICC). The higher the ICC, the stronger the similarity between individual *Y*s within groups. A high ICC is often viewed as a justification to explore the data with MLM techniques (e.g., Luke, 2004, pp. 18-21). Additionally, the null model

[10] While individual-level errors r_{ij} and random effects u_{0j} and u_{1j} are assumed to be not correlated, covariance of u_{0j} with u_{1j} is allowed to be non-zero, i.e., $\mathrm{Cov}(u_{0j}, u_{1j}) = \tau_{10}$.

shows the initial magnitude of the random variance which MLM is to explain, and it serves as a benchmark for more complex models.

As a next step, the empty model is extended to include individual-level predictors into Equation 5 and group-level predictors into Equation 6, if necessary. Finally, the slopes are allowed to vary across groups, which adds Equation 3 to the model. At each step, it is advised to analyze whether a further model extension is justified.

There are two distinct approaches as to which coefficients should be modeled as fixed and which as random. In the first approach researchers randomize only the coefficients of interest (Barr et al., 2013; Brauer & Curtin, 2017). The second approach is more data-driven and consists in first randomizing all the coefficients and then fixing those whose variance is not different from zero (Gelman & Hill, 2007; Matushek et al., 2017). [11]

2.2 Estimation

A linear multilevel model (Equation 4) typically includes two categories of parameters: the fixed effect estimates (coefficients) including the intercept (γ_{00}) and slopes (γ_{10} and γ_{01}) for the individual and group-level predictors, cross-level interactions (γ_{11}); and the variance components – the individual-level variance (σ^2), the group-level intercept and slope variances (τ_{00} and τ_{11}), and their covariance (τ_{10}).[12] The methods commonly used in the MLM literature to estimate all these parameters are the full information maximum likelihood (FIML) and the restricted maximum likelihood (REML).[13] Iterative procedures are used to obtain the coefficients as long as closed-form solutions are not available for the multilevel models with several random effects (e.g., see Hox, 2010; McCoach et al, 2018).

2.3 Parallels in empirical economics

[11] Nested linear models are compared with chi-square-based measures, such as deviance and -2 log likelihood (e.g., Snijders & Bosker, 2012: 98-99). The significance of random effects is usually evaluated using mixture chi-square tests (e.g., Morrell, 1998; Snijders & Bosker, 2012, pp. 99-100).

[12] Random effects, u_{0j} and u_{1j}, are characterized by variance components and not estimated directly, although they may be derived after estimations, see Section 3.

[13] Both FIML and REML assume the normal distribution of level-one and level-two variances but differ in how parameters are estimated. While FIML yields simultaneous estimation of both the coefficients and variance components, REML involves a few steps. Firstly, an OLS regression is fit. Secondly, variances and the covariance are estimated by maximizing the likelihood of the OLS residuals. Finally, coefficients are estimated using GLS. The general advantage of REML is that it tends to provide less biased estimates for variance components than FIML, but the differences between the two methods in practice are usually small (e.g., see Hox, 2010). The two advantages of FIML over REML is that FIML is easier computationally and allows the comparison of nested models that differ not only in the random part but also in the fixed part, i.e., it allows the researcher to run a statistical test on whether to include another explanatory variable to the multilevel model or not.

Although the MLM framework as a whole may be unfamiliar to economists, especially when presented in hierarchical form, as in the system of Equations 1–3, many of its elements have clear analogues in the standard regression analysis.

First of all, Equation 1 basically is a regression of *Y* on *X*, assuming that the intercepts and slopes differ across the groups as if Equation 1 was estimated for each group separately. Equations 2 and 3 are the group-level regressions of those intercepts and slopes, respectively, on some group-level variable *W*. A complete multilevel model can be estimated in two steps. In the first step, Equation 1 is estimated separately for each group. In the second step, Equations 2 and 3 are estimated as regressions with dependent variables that were estimated at the first step. This two-step empirical strategy is well known in applied economics (e.g., Saxonhouse, 1976).

Equation 4 can be viewed as a regression of *Y* on the individual-level variable *X* and the group-level variable *W*, including their interaction, and it can potentially be estimated using an OLS-based technique that could take into account heteroscedasticity and the clusterization of errors.[14]

A two-level model without Equation 3 can be viewed as a random-effects model used in panel data analysis (e.g., Green, 2003, p. 293; Wooldridge, 2002, p. 257), while a multilevel model with more than two levels can be treated as a nested error components model (e.g., Hsiao, 2014, p. 453).

Lastly, a complete multilevel model, net of group-level regressors *W*s, is actually the same as the random-coefficient model (Swamy, 2012[1971]).

All this suggests that MLM is able to serve many goals of empirical economists, such as to estimate the impact of individual- and/or group-level variables on *Y*, modeling their possible interactions; to account for the clusterization of errors; to model differences in the intercepts and slopes across groups; to measure the extent to which selected regressors may explain the overall initial variation in *Y* as well as cross-group variation in parameters (the explanatory power of *Wj* can be treated as a percentage decrease in random variance, τ_{00}, after *W* has been included).

The principal estimation method employed within the MLM framework – maximum likelihood – is routinely used in economics to estimate models with limited dependent variables

[14] Note that simple OLS is not efficient here as the composite error term $\varepsilon_{ij} = u_{0j} + u_{1j} * X_{ij} + r_{ij}$ violates the usual assumptions of homoscedasticity and the absence of serial correlation. Errors are dependent (clustered) within each group as they include u_{0j} that is common to any individual within a group, and heteroscedastic as their variance depends on u_{0j} and u_{1j} which vary across groups and also depend on X_{ij}.

(e.g, probit or logit models). This makes the neglect of MLM in empirical economics even more puzzling and calls for a detailed comparison of MLM with standard economic methods.

3. MLM vs. Economic methods

This section provides a more detailed comparison of MLM with the econometric techniques typically used in economic studies when analyzing nested data. To keep things simple, we assume that the data have only two levels, individual level (level 1) and group level (level 2), but all the conclusions may be generalized to data with more than two levels.

We distinguish three methodological challenges inherent in multilevel data – clustering, omitted variables, and heterogeneous coefficients across groups. Then we compare solutions to each problem in empirical economics (hereafter, 'economic methods') with those existing within the MLM framework, to investigate which approach is preferable and why. We start with formulating a problem, then review a popular economic technique used to tackle it, discuss how the problem is addressed within MLM, and, finally, compare the two. There are also separate sections on panel data and on technical and computational issues.

3.1. Clustering

Problem statement

Let us assume that an economist has to estimate the impact of an individual-level *X* on some *Y* using data with a hierarchical structure. The usual initial step is to estimate the following regression by OLS:

$$Y_{ij} = \beta_0 + \beta_1 * X_{ij} + \varepsilon_{ij} \quad (8),$$

where *i* represents individuals; *j* denotes groups; β_0 and β_1 are parameters of interest; and ε is the individual-level error term.

If the set of the Gauss-Markov assumptions holds true, then the OLS estimates of βs will be best, linear, and unbiased (aka BLUE, see, e.g., Wooldridge, 2002). However, these assumptions may not hold in practice. One assumption that is often violated in a hierarchical data setting is the absence of a serial correlation between ε_{ij}s.[15] Individuals belonging to the same group may be more similar to each other than individuals randomly chosen from the

[15] See Section 3.2 for the violation of the assumption of zero correlation between X and ε.

population.[16] This similarity results in an intra-group correlation of ε_i (and also in X_i), which violates the Gauss-Markov assumption.

If this correlation is overlooked, OLS underestimates the standard errors of βs, increasing the probability of a Type I error, i.e., detecting an effect that is not present (e.g., Bertrand, Duflo & Mullainathan, 2004; Moulton, 1990; Wooldridge, 2003). Due to clustering, the default OLS estimate of the variance of β_1 is inflated by a factor of $[1 + \rho_\varepsilon\, \rho_x\ (\frac{Var(\mathrm{N_g})}{\bar{N}_\mathrm{g}} + \overline{\mathrm{N_g}} - 1)\ ]$ – sometimes called the 'Moulton factor' or the 'design effect' – where ρ_ε is the intra-class correlation of residuals, ρ_x is the intra-class correlation of regressor X, $\overline{N_g}$ is the average cluster size, $Var(\mathrm{N_g})$ is the variance of cluster sizes. Even when ρ_ε is small, the magnitude of this inflating factor may be large enough to produce inflated standard errors (see Moulton, 1990).

Another common research task in a multilevel data setting is to estimate the impact of some group-level W on Y. The standard approach is to estimate the following equation by OLS:

$$Y_{ij} = \beta_0 + \beta_1 * X_{ij} + \beta_2 * W_j + \varepsilon_{ij} \text{ (9),}$$

where coefficient β_2 is of interest, and X is a control. In this case, the problem of error clusterization becomes more severe as ρ_x for W is equal to 1 and, thus, the design effect for the variance of β_2 is greater than that for the variance of β_1.

An economic solution

There are a few possible approaches to dealing with the clusterization of errors used in economics. The most popular one is to estimate the cluster-robust standard errors (CRSE, see, e.g., review by Cameron & Miller, 2015) that adjust the OLS standard errors for correlation within clusters. This technique may be considered as an extension of the Eicker-Huber-White 'sandwich' robust estimator for variance (Eicker, 1967; Huber, 1967; White, 1980) to the case of clustered errors (Liang & Zeger, 1986). It provides a consistent estimate of the variance if the number of clusters (*Ncl*) goes to infinity.[17]

[16] For instance, if we consider countries as groups, this similarity may exist because individuals living in the same country share common cultural norms and are affected by the same economic and political institutions. Similar substantive sources for the similarity – sharing some common group properties – may exist between any units belonging to the same group.

[17] In the general case, the variance of the OLS estimate is $Var(\widehat{\beta_1}) = \frac{Var(\sum_k x_k\varepsilon_k)}{(\sum_k x_k^2)^2}$. When errors are serially uncorrelated (and homoscedastic), this variance is reduced to $\frac{\sigma_\varepsilon^2}{\sum_k x_k^2}$, but when errors are serially correlated, $Var(\widehat{\beta_1}) = \frac{\sum_k \sum_s x_k x_s E[\varepsilon_k \varepsilon_s]}{(\sum_k x_k^2)^2}$. The CRSE technique uses OLS residuals for the k^{th} and s^{th} observations, $\widehat{\varepsilon_k}$ and $\widehat{\varepsilon_s}$, to get an estimate $\widehat{\varepsilon_k}\widehat{\varepsilon_s}$ for $E[\varepsilon_k \varepsilon_s]$. Although $\widehat{\varepsilon_k}\widehat{\varepsilon_s}$ is a poor estimate, $E[\varepsilon_k \varepsilon_s]$, averaging these estimates over all the clusters gives a consistent estimate of the variance when *Ncl* goes to infinity.

MLM solution

A multilevel model that corresponds to Equation 8 is:

$$Y_{ij} = \beta_{0j} + \beta_1 * X_{ij} + r_{ij} \quad (10)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (11),$$

where all notation is the same as in Equations 1–3. This is a 'random-intercept model' (RIM). In the mixed/combined form, this model reads as follows:

$$Y_{ij} = \gamma_{00} + \beta_1 * X_{ij} + u_{0j} + r_{ij} \quad (12)$$

A multilevel model estimating the impact of a group-level variable (corresponding to Equation 9) is:

$$Y_{ij} = \beta_{0j} + \beta_1 * X_{ij} + r_{ij} \quad (13)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * W_j + u_{0j} \quad (14)$$

In a single-equation form, this model is:

$$Y_{ij} = \gamma_{00} + \beta_1 * X_{ij} + \gamma_{01} * W_j + u_{0j} + r_{ij} \quad (15)$$

Compared to Equations 8 and 9 of the economic approach, the error terms in Equation 12 and 15 is composite and consists of two components: u_{0j} and r_{ij}. RIM explicitly distinguishes a group-specific component u_{0j} responsible for the correlation of errors within groups, while the strength of this correlation is measured using the ICC coefficient (see 2.1). As the maximum likelihood (either FIML or REML) used to estimate Equations 12 and 15 delivers not only the coefficients and their variances, but also the variance (τ_{00}) of u_{0j}, the clusterization of errors is explicitly taken into account during the estimations.

The issue of a small *Ncl* is important for MLM as for CRSE (e.g., see Maas & Hox, 2005; Stegmuller, 2013), even though it is not evident how *Ncl* affects the estimated variance of the coefficients and other parameters of a multilevel model. Below we compare OLS-CRSE with MLM regarding the minimum *Ncl* and other related issues.

Comparison

The point estimates obtained via OLS (Equations 8 and 9) and MLM (Equations 12 and 15) do not differ significantly, as documented both in Monte Carlo simulations (e.g., see Clarke, 2008; Huang, 2017; McNeish et al., 2016) and in applied examples (e.g., Arcenaux & Nickerson, 2009).[18] Therefore, the key difference between the methods concerns the estimation

[18] This is true for linear models but not true for the non-linear ones. While a coefficient in the standard regression reflects the impact of one unit change in a regressor X on Y, holding all other regressors constant, a coefficient in a multilevel model reflects the impact of one unit change in a regressor X on Y, holding all other regressors constant *and having the same random*

of standard errors. When errors are clustered, OLS-CRSE and MLM may provide quite different results, and it is not clear *a priori* which approach is more appropriate as both have limitations.

A disadvantage of the MLM solution is that it relies on the distributional assumption of normality for the random effects, while the CRSE technique is 'model-free'. In practice, however, this disadvantage seems to be crucial only in extreme scenarios, as maximum likelihood estimates are generally robust to mild violations from normality (e.g., Bell et al., 2019; Hox, 2010).

Both OLS-CRSE and MLM become problematic when *Ncl* is small. In OLS-CRSE, *Ncl* is important for estimating the robust variances of coefficients, while the estimation of the coefficients is not conditional on *Ncl*. The common lower bound for *Ncl* to have a consistent estimate for the variances is about 50 (e.g., see Angrist & Pishke, 2009; Cameron & Miller, 2015; Wooldridge, 2010a, pp. 884-894).

In MLM, the estimation of all parameters is dependent on *Ncl*. In general, MLM allows a lower *Ncl* than OLS-CRSE, but an exact lower bound differs for different types of estimated parameters. Monte Carlo simulations show that individual-level coefficients and their standard errors are estimated with a negligible bias even when *Ncl* = 5 (see Stegmuller, 2013), which is in line with results for *Ncl* = 10 (Maas & Hox, 2005). Estimates for group-level coefficients exhibit a 'reasonable' bias when *Ncl* = 30 (Maas & Hox, 2005), but may become substantially biased when *Ncl* < 20 even at ICC = 0.1 (Stegmuller, 2013). Concerning the variance components, the effect of *Ncl* on their standard errors is 'definitely larger than on standard errors of coefficients', and at *Ncl* = 10 the estimated standard errors become 'clearly unacceptable' (see Maas & Hox, 2005).

While a sufficient number of clusters is an issue that has received attention both in the CRSE and MLM literature, the issue of cluster size has been much less examined. In the MLM literature, the general consensus is that the cluster size has a smaller impact than the number of clusters (e.g., Max & Hox, 2005). A smaller cluster size implies a smaller design effect, which suggests that clustering is less pronounced in the data and, thus, needs correcting to a lesser extent. However, when clusters are sparse, adjustment methods may underperform. The literature suggests that MLM is more vulnerable to small cluster size than CRSE, as it needs to estimate variance components. As shown by McNeish (2014), when the number of observations within clusters is less than five, variance components tend to be overestimated

effect. This difference does not matter in a linear model, but it does in a non-linear setting where random effects are not added (e.g., see McNeish, 2014).

(see also Clarke, 2008; Stegmuller, 2013). This implies the overestimation of within-cluster correlations (which may lead to the decision to adjust for clustering when it is not needed) and less reliable estimates for random effects (see Section 3.3).[19] CRSE, however, may face problems in the opposite case, when the cluster sizes are too big, compared to their number, even if *Ncl* is at 'acceptable' levels (see Hansen, 2007).

Does variability in cluster sizes matter? Theoretically, the design effect should increase with the variability in cluster sizes, but the literature, both economic and MLM, is inconclusive. As Max & Hox (2005, p.88) note:

> *We carried out some preliminary simulations to assess whether having balanced or unbalanced groups has any influence on multilevel ML [maximum likelihood] estimates, with a view to including this in the simulation design. However, despite extreme unbalance, there was no discernible effect of unbalance on the multilevel estimates or their standard errors*.

Some Monte Carlo studies that report results separately for balanced and unbalanced clusters show that estimated parameters tend to have a large relative bias in the latter case, but the size of that bias remains acceptable when *Ncl* is larger than 50 (e.g., Clarke, 2008; McNeish, 2014). However, little is known about cases when a highly unbalanced structure is combined with a relatively low number or clusters and/or sparse clusters.

Economic studies rarely focus on the issue of different cluster sizes, although the basic CRSE assumption that clusters are of equal size is rarely met in practice (except in the case of balanced panel data). Simulation results presented in Cameron et al. (2008) indicate that differing cluster sizes can increase the value of a cluster-robust t-statistic, but the results are reported only for *Ncl* = 10. Carter et al. (2017) consider the variability in cluster sizes as a part of a more general 'cluster heterogeneity' issue and suggest reporting and using an 'effective' number of clusters, which is the actual number of clusters adjusted for heterogeneity. No 'rule of thumb', however, is suggested.

Finally, general arguments can be found against using robust standard errors as a solution to the error clusterization (or heteroscedasticity) problem. As argued by King & Roberts (2015), a substantial difference between classical and robust standard errors merely signals model misspecification(s), which should motivate a researcher to improve the original model.

[19] Five or fewer observations per cluster seem unrealistic for cross-sectional surveys, but are commonplace for longitudinal data.

In this regard, MLM looks preferable to OLS-CRSE as the former assumes a 'less misspecified' model that includes the random effects u_{0j} omitted from Equations 8 and 9 used in economics. However, the recommendation may be exactly the opposite in cases when a researcher is just interested in the 'correct' standard errors of the coefficients and not interested in the variance components or cluster-specific inference. When *Ncl* is large enough, CRSE is very attractive as it delivers consistent estimates for standard errors without any parametric assumptions and without modeling random effects, or estimating variance components, which may save time and effort. Moreover, improving the original model may be difficult, as is detecting model misspecifications (e.g., determining whether the covariance structures are properly specified).

3.2. Omitted variable

Problem statement

Another assumption frequently violated in a simple OLS regression (Equation 8) is the zero correlation between X and ε (the exogeneity assumption). If this assumption does not hold, then the OLS estimate of β_1 is biased and inconsistent. A common reason is an omitted variable (OMV) that affects both X and Y. In this case, any statistically significant relationship between X and Y may be driven by OMV and thus may be spurious. In a multilevel data setting, OMV may belong either to the individual (OMV-1) or to the group level (OMV-2).[20]

In a similar vein, the OMV problem may exist in a regression where one is interested in estimating the impact of a group-level variable (W) on Y (Equation 9). However, as it is unlikely that an individual-level OMV could affect group-level factors, the whole OMV problem is reduced to the OMV-2 case.[21]

An economic solution

In economics, the most popular approach to dealing with OMV-2 in Equation 8 is to introduce N-1 group dummies and then apply OLS[22] (aka the 'least squares dummy variables estimator', LSDV, see Baltagi, 2008, p. 12; Wooldridge, 2002, pp. 485-486):

[20] OMV-2 can be viewed as another consequence of clusterization within groups. If individuals cluster, the average levels of Y or/and X differ systematically across groups. Three cases are possible. First, Ys differ, Xs do not. Second, Xs differ, but Ys do not. Third, both Ys and Xs differ, i.e., higher or lower Ys are associated with higher Xs *at the group level*. While cases one and two do not imply any serious problems, the third case implies dependence between Y and X at the group level. If the researcher is interested in the individual-level dependence, then this group-level dependence should be taken into account, else one faces the OMV-2 problem (e.g., see discussion in Kennedy, 2008, pp. 282-285).

[21] Although the impact of group-level factors is usually conditional on the composition of groups, any omitted compositional characteristic of a group is not an individual-level but rather a group-level variable.

[22] Additionally, all the procedures that solve the OMV-1 problem can also potentially solve the OMV-2 problem.

$$Y_i = \beta_0 + \beta_1 * X_i + \sum_{j=1}^{N-1} \theta_j d_j + \varepsilon_i \text{ (16)},$$

where d_j is a group-specific dummy variable and θ_j is its coefficient. By including group dummies, the researcher eliminates any unobservable group-specific characteristics from the errors, thus ensuring that $Corr(\varepsilon_i, X_{ij}) = 0$.

In practice, especially in a panel data setting, an LSDV model is estimated in two steps as a fixed effects (FE) model. In the first step, the 'within-transformation' of all variables is applied:

$$Y_{ij} - \bar{Y}_{.j} = \beta_1 * (X_{ij} - \bar{X}_{.j}) + \varphi_{ij} \text{ (17)},$$

where $\bar{Y}_{.j}$ and $\bar{X}_{.j}$ are group means of Y and X, respectively, and φ is a new error term resulting from the transformation. As a result, the unobservable group effects (assumed to be fixed) are completely removed from the equation (as $d_{ij} = d_{.j}$). In the second step, Equation 17 is estimated by OLS. This approach gives exactly the same results as the estimation of Equation 16 by OLS but is preferable as it avoids the estimation of additional *N*-1 coefficients, θ_j.[23]

The OMV-1 problem can be treated with quasi-experimental techniques, e.g., the method of instrumental variables (IV), difference-in-difference (DiD), matching, regression discontinuity (RD), (see Angrist & Pishke, 2009, 2010). This general methodological direction is well developed and widely applied in empirical economics (all the techniques, however, have limitations).

However, in case of the OMV-2 problem in Equation 9 the FE model is not possible as the within transformation 'throws out the baby with the bathwater', i.e., it removes all group-level variables of interest. The only possibility, again, is to use quasi-experimental methods (e.g., IV), but caution is required as these methods work properly only with large samples, while the number of observations at a higher levels is limited by the number of groups, which is usually substantially fewer than the total number of observations.

MLM solution

Within the MLM framework, solving the OMV problem is much less straightforward than solving the problem of error clusterization. MLM assumes 1) a zero correlation between X and r, i.e., $Cov(X_{ij}, r_{ij}) = 0$ and 2) a zero correlation between X and u, i.e., $Cov(X_{ij}, u_{0j}) = 0$ (e.g., Raudenbush & Bryk, 2002, p. 255). In other words, both error components are assumed

[23] For applications when these coefficients may present substantive interest, see Section 3.3.

to be uncorrelated with the explanatory variables, and the MLM framework does not consider possible violations of these assumptions. This fact may partly explain the absence of a discussion of causal inference in early MLM textbooks. As noted by Oakes (2004, p. 1934):

> *…of the five major texts focused on multilevel modeling, per se (Goldstein, 1995; Kreft & De Leeuw, 1998; Snijders & Bosker, 1999; Heck & Thomas, 2000; Raudenbush & Bryk, 2002), only Goldstein (1995) has an index citation for 'cause'. Goldstein devotes very little attention to the issue and essentially notes multilevel models are not adequate for causal analysis*.[24]

In its more recent development, however, the MLM framework has started to overcome this limitation. The basic MLM framework can be modified to allow a possible violation of the assumption of $Cov(X_{ij}, u_{0j}) = 0$. In case of OMV-2 in Equation 8, the conventional modification is to introduce the group average of *X* which will absorb all possible correlations between *X* and group-level random effects:

$$Y_{ij} = \gamma_{00} + \beta_1 * X_{ij} + \beta_2 * \overline{X_{.j}} + u_{0j} + r_{ij} \quad (18)$$

Studies show that this modification provides the estimates of β_1 and its standard errors equivalent to the FE model (e.g., Bell & Jones, 2015; Fairbrother, 2014; Goetgeluk & Vansteelandt, 2008; Huang, 2017). Acknowledging this solution, more recent MLM textbooks have included special sections on causal inference (e.g., Gelman & Hill, 2007; Kim & Swoboda, 2011). The same modification can be applied in Equation 9 (see Huang, 2017).

The possible violation of the other exogeneity assumption of MLM, $Cov(X_{ij}, r_{ij}) = 0$, has not yet found any conventional solutions in the literature. However, there is a growing body of techniques integrating MLM with quasi-experimental methods, e.g., IV (Kim & Frees, 2006; Kim & Swoboda, 2011; Spencer & Fielding, 2000), matching (Arpino & Mealli, 2011; Zubizarreta & Keele, 2017), the general Rubin causal model framework (Hill, 2013; Feller & Gelman, 2015).

Comparison

The assumption of MLM of a zero correlation between the explanatory variables and the error components seems to be an important flaw in economists' eyes. As Bell & Jones (2015,

[24] See also Bell & Jones (2015, p. 134) for a similar observation.

p. 136) note, *'…one must consider why RE is not employed more widely, and remains rarely used in disciplines such as economics and political science. The answer lies in the exogeneity assumption of RE models: that the residuals are independent of the covariates […]'.* In a similar vein, Huang (2017, p. 4) notes, *'The issue related to the correlation with the error terms is not new and is one reason certain disciplines shy away from using MLM because of the potential for biased estimates*' (see also Kim & Frees, 2006). Economists generally find this assumption 'unreasonably strong' (Cameron & Miller, 2015, p. 332; Bell & Jones, 2015). Consequently, they usually favor FE models over RE models, as the former provide consistent estimates even if this assumption does not hold.

However, this limitation of MLM is important only in case of OMV-1, since MLM does not provide any conventional solutions in this case. In the case of OMV-2, the modified multilevel model (Equation 18) not only provides the estimates for β_1 equivalent to the FE model, but also retains all the flexibility of the RE model by estimating the impact of group-level explanatory variables[25] and extending the model for random slopes (see Section 3.3). In this regard, as pointed by Bell & Jones (2015, p. 135), 'a well-specified RE model can be used to achieve everything that FE models achieve, and much more besides' (for supporting simulations, see Dieleman & Templin, 2014).[26]

In practice, MLM usually tends to center *X*s at the group level:

$$Y_{ij} = \gamma_{00} + \beta_1 * (X_{ij} - \overline{X_{.j}}) + \beta_3 * \overline{X_{.j}} + u_{0j} + r_{ij} \text{ (19)},$$

where $\beta_3 = \beta_1 + \beta_2$. This model is known as the 'hybrid' model (Allison, 2009; Neuhaus & Kalbfleisch, 1998), the 'within-between' estimator (Dieleman & Templin, 2014), the 'within-between RE model' (Bell & Jones, 2015), or the 'including-the-group-mean model' (Huang, 2017). According to Bell & Jones (2015, p.142), the specification in Equation 19 has three advantages over the specification in Equation 18. First, Equation 19 clearly separates the 'within' and 'between' effects which may be useful for research purposes. Second, while Equation 18 involves a correlation between X_{ij} and $X_{.j}$, the group mean centering X_{ij} in Equation 19 gets rid of this correlation, leading to more stable and precise estimates. Finally, if multicollinearity exists between multiple $X_{.j}$s and other time-invariant variables, $X_{.j}$s can be simply removed from the equation without causing a heterogeneity bias.

[25] There are methods for fitting group-level predictors within the FE framework (Hausman & Taylor, 1981; Pluemper & Troeger, 2007; 2011). Their detailed treatment is out of the scope of our paper.

[26] Interestingly, specification in Equation 18 has been known in economics at least since Mundlak (1978) and is sometimes called the 'correlated random effects' model (Schunck, 2013; Wooldridge, 2010b), or simply 'Mundlak's approach'. So, why would economists not use model in Equation 18 as an alternative to the FE model, since it is at least as good as the latter?

The simultaneous testing of the within-group and between-group effects of the variable of interest presents a strong advantage over the standard FE model for several reasons. For one thing, individual- and group-level dependencies between the same variables can be quite different (Robinson, 1950). A combination of a statistically non-significant β_1 and a significant β_3 is a signal to refocus the analysis of the dependencies between *X* and *Y* from the individual to the aggregated level and, possibly, to revise the theoretical considerations.

Secondly, the 'within transformation' in the FE model that removes the between-group variation often results in insignificant coefficients for variables that have low within-group variance. Usually, it is unclear whether this happens because those variables bear no effect on *Y* after taking into account the unobserved group-level heterogeneity, or because of a reduced variation in *X*. By contrast, if β_3 in Equation 19 remains statistically significant, it means that *X* is still related to *Y*, but this relationship is more pronounced at the group level.

Finally, both Equation 18 and Equation 19 may provide a convenient alternative to conducting the Hausman test, often used in econometrics to decide which model, FE or RE, to choose (especially in a panel data analysis).[27]

To sum up, MLM outperforms the FE model in dealing with unobserved heterogeneity at the group level. In a more general sense, the current view in empirical economics (especially in the analysis of panel data) saying that the FE approach to the OMV problem is generally more preferable than the RE approach should be abandoned, as 'a well-specified RE model can be used to achieve everything that FE models achieve, and much more besides'.[28]

[27] Traditionally, this test is conducted in two steps. At the first step, FE and RE models are estimated. At the second step, the FE and RE estimates are compared. If the null hypothesis, that differences between FE and RE estimates are not systematic, is rejected, then RE estimates should be preferred as more efficient. If the null hypothesis cannot be rejected, FE estimates are preferred as consistent. Instead of this routine, $\beta_1 = \beta_3$ can be tested in Equation 19, or $\beta_2 = 0$ in Equation 18. If these hypotheses are rejected, the FE model is preferable to the RE model (e.g., Green, 2007; Wooldridge, 2010b). At the same time, estimating the 'hybrid'/'within-between' model that encompasses both the FE and RE makes redundant the choice between these models and the Hausman test itself (Bell & Jones, 2015).

[28] At the very least, a correlation between group-level unobservable effects and individual-level Xs should not be considered a critical issue in choosing between the FE and RE models. As put by Snijders & Bosker (2012, p. 48): "*An often mentioned condition for the use of random coefficients models is the restriction that the random coefficients should be independent of the explanatory variables. However, if there is a possible correlation between group-dependent coefficients and explanatory variables, this residual correlation can be removed, while continuing to use a random coefficient model, by also including effects of the group means of the explanatory variables. Therefore, such a correlation does not imply the necessity of using a fixed effects model instead of a random coefficient model*."

3.3. Heterogeneity of coefficients

Problem statement

A simple OLS regression (Equation 8) assumes that the intercept β_0 and slope β_1 do not vary across groups. As this assumption may be too restrictive in a multilevel data setting, it may be useful to model or test for the differences in coefficients across groups.

The heterogeneity of the slopes also comes up as a problem if it is not taken into account in the pooled regression (Equation 8). Consider Equation 8 where β_1 has a group-specific random component u_{1j}:

$$Y_{ij} = \beta_0 + (\beta_1 + u_{1j}) * X_{ij} + \varepsilon_{ij} = \beta_0 + \beta_1 * X_{ij} + u_{1j} * X_{ij} + \varepsilon_{ij} \quad (20)$$

In this case, the composite error term $(u_{1j} * X_{ij} + \varepsilon_{ij})$ is heteroscedastic, hence a simple OLS regression provides inefficient, although consistent, estimates. Therefore, taking into account the variability of the slopes across groups may produce an insignificant slope coefficient, a fact that was discussed as applied to economics (Bell, Fairbrother & Jones, 2019) and other social sciences (e.g., see Heisig & Schaeffer, 2019; Schielzeth & Forstmeier, 2008).

An economic solution

When economists want to model differences in parameters in the intercepts and/or in slopes across groups, there are two major options – to model parameters as either ‘fixed and different’, or as ‘random and different’ (Hsiao, 2008, p. 178). In the ‘fixed and different’ case, it is assumed that the parameters are fixed for each group and, thus, do not have any random components. There are two ways to model this. First, by introducing group dummies and their interactions with individual-level regressors:

$$Y_i = \beta_0 + \beta_1 * X_i + \sum_{j=1}^{N-1} \theta_j d_j + \sum_{j=1}^{N-1} \vartheta_j d_j * X_i + \varepsilon_i \quad (21),$$

where $d_j * X_i$ is the interaction of *X* with dummy for group *j*. In this case, β_0 and β_1 are the intercept and slope coefficients for the reference group, while the intercept and slope coefficients for any other group k will be $\beta_{0k} = \beta_0 + \theta_k$ and $\beta_{k1} = \beta_1 + \vartheta_k$, respectively. If θ and/or ϑ are insignificant for some group, then this group has the same intercept and/or slope as the reference group. If all ϑ are equal to zero, then Equation 21 turns into Equation 16, which models intercept heterogeneity only.

The other way to model differences in the intercepts or/and slopes across groups is to estimate Equation 8 separately for each group.[29] The Chow test (Chow, 1960) is used to gauge the extent to which this approach is more appropriate than estimating a single equation for all the groups. Although it may show that the estimation of separate equations has more predictive power than one pooled equation, it is not able to indicate which groups have similar parameters (and thus can be pooled). In this regard, a model with multiple interactions, Equation 21, is more informative as it allows testing the differences in coefficients between specific groups.

In the 'random and different' case, economists tend to estimate a 'random-coefficient regression model' (Hsiao & Pesaran, 2008; Swamy, 1970, 2012 [1971]; Swamy & Tavlas, 1995), identical to Equation 20. A limitation of this model is that it does not allow any group-level explanatory variables. Therefore, if it is part of the research goals to understand the sources of parameter heterogeneity and/or estimate the impact of certain group-level factors (W) on intercepts and/or slopes, the choice is restricted to the 'fixed-coefficients' approach. There are two ways to proceed in this case as well. First, by estimating a single equation with group-level dummies and multiple interaction terms:

$$Y_{ij} = \beta_0 + \beta_1 * X_{ij} + \sum_{j=1}^{N-1} \theta_j d_j + \sum_{j=1}^{N-1} \vartheta_j d_j * X_i + \beta_2 * W_j + \beta_3 * X_{ij} * W_j + \varepsilon_{ij} \quad (22),$$

where $X_{ij} * W_j$ is the interaction between an individual-level variable X and a group-level variable W. In this case, the impact of X on Y is modeled as conditional on W. For example, if β_3 is statistically significant and greater (less) than 0, then the impact of X is equal to $(\beta_1 + \beta_3 * W_j)$ which increases (decreases) with W.

Secondly, by following a two-step procedure: 1) estimate group-specific Equations 8, and 2) regress the obtained β_0s and β_1s on W. This approach is also known as the 'estimated dependent variable model' (EDV model, e.g., Hanushek, 1974; Pagan, 1984; Saxonhouse, 1976). One issue here is how to correct for the heteroscedasticity of errors in the second-step regression, which exists because the coefficients obtained in the first step are estimated with different degrees of precision. Several alternatives are available: 1) the weighted least squares estimator (WLS) with weights inversed to the estimated standard errors of the dependent variables; 2) OLS with a robust estimation of the standard errors (White, 1980); or 3) the feasible generalized least square estimator (FGLS). As Monte Carlo simulations suggest, the second method is more efficient than WLS, while FGLS may be preferable in cases when

[29] If the errors of group-specific equations are correlated, then their joint estimation is more efficient, and the model belongs to the class of 'seemingly unrelated regressions' (Zellner, 1962). These models, however, are out of the scope of this paper.

reliable information is available about the sampling variances of the estimated dependent variable (Lewis & Linzer, 2005).

Both approaches – the model with interactions and the EDV model – provide consistent estimates of the intercepts and slopes. Lewis and Linzer, however, advise against the former as '*it assumes, almost certainly incorrectly, that there would be no residual in a regression of individual-level coefficients on the country-level variables'* (2005, p. 347). Bryan and Jenkins favor the two-step EDV model as well, mentioning its three advantages: 1) it 'shows why a small number of countries can affect the reliability of estimate'; 2) its 'estimates are unbiased (with correct SEs) and so can be used to benchmark the other methods'; and 3) it 'leads naturally to a graphical summary of country-level variations in outcomes' (2016, p. 6) in which one plots the country intercepts fitted at step 1 against group-level explanatory variables *W*. None of these advantages of EDV, though, seems to be crucial and thus both approaches can be used in practice.

MLM solution

To model parameter heterogeneity across groups, MLM users typically estimate the following model:

$$Y_{ij} = \beta_{0j} + \beta_{1j} * X_{ij} + r_{ij} \quad (23)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (24)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (25)$$

In MLM parlance, this is called the 'random-intercept random-slopes model' (RIRSM, Raudenbush & Bryk, 2002, p. 151). It differs from the RIM model presented in Equations 10–11 by the addition of Equation 25, which models cross-group variation in the slope coefficients. Thus, this model allows both the intercepts and the slopes to vary across groups. In the mixed form, this model is:

$$Y_{ij} = \gamma_{00} + \gamma_{10} * X_{ij} + u_{0j} + u_{1j} * X_{ij} + r_{ij} \quad (26)$$

This model can be easily extended to include group-level explanatory variables *W*, which results in the complete multilevel model presented in Section 2.1.

Comparison

To begin with, consider a case where the goal is to model only heterogeneity in the intercepts. If the *group size* is small, the 'fixed-coefficients' approach (Equation 21) provides unreliable estimates of group-specific intercepts. This may be important for studies where these intercepts should be interpreted, e.g., as group differences in the mean levels of the dependent variable. MLM is better equipped in this case as the estimates of random effects for groups with a relatively small number of observations receive lower weights and converge to the mean (zero) level. MLM relies on an empirical Bayesian estimation and uses ‘shrunken’ group-level residuals that are estimated as:

$$\hat{u}_{0j} = c_j\,(\overline{Y_{.j}} - \hat{\beta}_{0j} - \hat{\beta}_{1j} * \overline{X_{.j}}) \quad (27)$$

where $\overline{Y_{.j}}$ and $X_{.j}$ are group means of Y and X, respectively (therefore, the term in parentheses is the mean residual for group j); $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ are the RE estimators from Equation 12; c_j is a shrinkage factor, $c_j = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \left(\frac{\hat{\sigma}_u^2}{n_j}\right)}$ (Raudenbush and Bryk, 2002, pp. 86-92). This factor is far smaller than the one for groups with a low number of observations (n_j) or comparatively large within-group variance. Thus, the whole estimate $\hat{u}_{0j}$ will converge to zero for such groups.

However, if the *number of groups* is low, the ‘fixed-coefficients’ approach scores better than MLM, as the latter needs to estimate variance components. According to Maas and Hox (2005), at least 30 groups are needed for reliable variance estimates in MLM. Stegmueller (2013) recommends at least 15 groups though the number may be higher in some contexts (see Bryan & Jenkins, 2016).

An advantage of MLM over the ‘fixed-coefficients’ approach is that all the estimated intercepts are assumed to be random drawings from the same normal distribution and thus can be easily extended to out-of-sample groups. The group effects in this case are ‘exchangeable’ between groups in the sample and thus may be generalized to groups out of the sample. In contrast, the estimated intercepts in the FE model are not ‘exchangeable’ between groups. Taking the example from Bryan & Jenkins (2016, Supplement, p. 5): ‘*estimates from an FE approach (intercepts and coefficients) relate specifically to the set of countries included in the sample and cannot be generalized out of sample. As an example, FE estimates from a dataset including respondents from the original 15 European Union member states could not be applied to describe outcomes for the 12 new member states with their very different institutions and history.*’ Such an advantage of MLM, however, is not relevant in cases when the ‘exchangeability’ assumption is unrealistic (Bryan & Jenkins, 2016), or when all the groups of interest are already in the sample.

Next, consider a case where the goal is to model heterogeneity not only in the intercepts but also in the slopes. Here, two sub-cases should be distinguished: without and with group-level explanatory variables.

When there are no group-level variables, the standard random-coefficient model and the MLM random slope model are equivalent. Compared to the EDV model, MLM produces more efficient estimates as it is based on an empirical Bayesian estimation (e.g., Raudenbush and Bryk, 2002, pp. 86-92). Moreover, MLM is more parsimonious as it needs to estimate only two parameters (average slope γ_{10} and its random component, u_{1j}), compared to estimating separate regressions across each group resulting in dozens of parameters. Finally, MLM is more convenient for interpretation than the model with interactions (Equation 21), which becomes cumbersome when the number of individual-level regressors *X* and groups is large.

Alternatively, when group-level explanatory variables are involved, economists choose between a model with interactions and an EDV model, both of which provide estimates that are less efficient than MLM, as supported by Monte Carlo simulations (Heisig et al., 2017). For a model with interactions, it is difficult to combine fixed group effects with *W***X* interactions in one equation, as fixed effects displace the main effects of *W* from the regression, undermining the interpretation of the whole model (Brambor et al., 2006; Bell & Jones, 2015).

In practice, the MLM's advantage in efficiency, however, is contingent on the issue under study and on the nature of the data used. In cases when the number of observations within groups is small relative to the number of groups, MLM produces more efficient estimates as it uses both within- and between-group information. In cases where considerable information is available within the groups relative to the number of groups, 'the added effort of fitting a more complex single-stage linear hierarchical model would provide little advantage relative to the simple two-stage EDV method' (Lewis & Linzer, 2005, p. 348).

3.4. Panel data analysis

Panel or longitudinal data are a form of data with a hierarchical structure where unit-time measurements are nested in units. In this case, individuals or other units of analysis (e.g., firms or countries) are taken as higher-level units, while unit-time observations are taken as lower-level units. The comparison of MLM with economic methods in a panel data setting deserves special attention due to the enormous popularity of these data in economic studies. Both MLM and all the economic methods discussed above have important time-related specifics.

In a panel data setting, the basic equation of interest typically looks as follows:

$$Y_{it} = \beta_0 + \beta_1 * X_{it} + \varepsilon_{it} \quad (28),$$

where *i* represents units of analysis (e.g., individuals or countries); *t* is time period; β_0 and β_1 are the parameters of interest; and ε is the individual-level error term. Unlike Equation 8, the grouping variable here is the unit of analysis itself. However, the nature of methodological challenges – clustering, omitted variables, heterogeneous coefficients – remains the same as in the cross-sectional case.

Clustering

The assumption of a null serial correlation between $\varepsilon_{ij}s$ in Equation 28 is usually implausible in a panel data setting as different measurements of the same unit of analysis, by definition, tend to be similar to each other. The positive autocorrelation of $\varepsilon_{it}s$ is expected, which under OLS leads to the underestimation of standard errors. Economic and MLM solutions to this problem are similar to the solutions existing in the cross-sectional case: economists apply CRSE (see also Arellano, 1987), while the MLM solution is to estimate a RIM model (Equation 12). The comparative (dis)advantages of these two approaches remain the same as in the cross-sectional case: OLS-CRSE needs a larger number of clusters, while MLM is more vulnerable to small cluster sizes. Therefore, in a 'large *N*, small *T*' panel data setting, CRSE is more appropriate than MLM, while for 'small *N*, large *T*' panels, MLM looks more appropriate than CRSE in dealing with clustered error terms.

Another expectation, non-existent in a cross-sectional case but natural in a panel data setting, is that observations within one unit that are closer in time to each other should be more strongly correlated. This demands modelling varying correlations between observations within a cluster. In contrast to CRSE, MLM can model this correlation explicitly, by introducing random components into the coefficients (see below).

When panel data are unbalanced, the issue of unequal cluster sizes arises. Unbalanced panel data exhibit a higher 'Moulton factor' than balanced data and may complicate the estimation of standard errors in both strategies (similar to the cross-sectional case). In non-balanced designs, MLM can provide more precise and reliable estimates than the FE model for a particular group with few observations – provided that the missing values are random (Diez-Roux, 2000; Rice & Jones, 1997).

Omitted variable

Similar to the cross-sectional case discussed above, to deal with OMV-1 in a panel data setting economic techniques like CRSE or FE models may be complemented by quasi-experimental methods, while MLM generally ignores the problem, assuming exogeneity.

Regarding OMV-2, the 'gold standard' in economics in a panel data setting is to estimate the FE model (Equation 17), while MLM leverages the 'within-between RE model' (Equation 19). MLM is more flexible in dealing with OMV-2 than the FE model as it allows the inclusion of the time-invariant characteristics of units. Moreover, the 'within-between RE model' estimates the within-unit and between-unit effects, while the FE model considers only the within-unit effects. Thus, the FE model can be misleading when variance exists mostly between units.[30] Finally, removing all the between-unit variation, the FE model magnifies the relative importance of possible data errors, which are common in repeated surveys (e.g., see Brown & Light, 1992) and, thus, increases the attenuation bias (see Angrist & Pishke, 2009, p.168).

Heterogeneity of coefficients

Both intercepts and slopes in Equation 28 may differ across the units of analysis. Economists tend to model these differences using either the model with interactions (Equation 22) or the two-step EDV model, while the MLM solution is to estimate the RIRSM model (Equations 23–25 or Equation 26). In the cross-sectional case, as discussed above, the MLM solution is generally more convenient, parsimonious, and efficient, and these advantages become even more evident in a panel data setting with large *N* and relatively small T.[31]

Moreover, by allowing random intercepts and slopes, MLM helps to model the abating serial correlation between ε_{it}. With random intercepts and slopes, Equation 28 becomes:

$$Y_{it} = \gamma_{00} + \gamma_{10} * X_{it} + u_{0i} + u_{1i} * X_{it} + r_{it} \text{ (29)}$$

Therefore, $Cov(\varepsilon_{it}, \varepsilon_{i,t-1}) = Cov(u_{0i} + u_{1j} * X_{it} + r_{it}, u_{0i} + u_{1i} * X_{it-1} + r_{it-1}) = \sigma_{u0}^2 + X_{it} \cdot \sigma_{u1} + X_{it-1} \cdot \sigma_{u1} + X_{it} \cdot X_{it-1} \cdot \sigma_{u1}^2$. At the same time, $Var(\varepsilon_{it}) = \sigma_{u0}^2 + 2 \cdot X_{it} \cdot \sigma_{u1} + X_{it}^2 \cdot \sigma_{u1}^2 + \sigma_r^2$. As a result, $Corr(\varepsilon_{it}, \varepsilon_{i,t-1})$ is no longer constant within the units of analysis, decreasing in time (e.g., see Singer & Willet, 2003).

Missing values and panel attrition

Many datasets contain missing values at least for some variables. In longitudinal surveys, it is always difficult to follow up all the respondents. The original sample tends to shrink for various reasons (refusal to answer, moving, death), which is called 'panel attrition'. The big

[30] The classic example is the effect of changes in household incomes on their consumption. Consumption of some item should be affected not by transitory fluctuations of the income but by changes in the permanent income, which are reflected in between-household differences in incomes.

[31] It is not by chance that the MLM framework has underpinned a large class of 'growth curve models' widespread in individual psychology, which 'allow for the estimation of inter-individual variability in intra-individual patterns of change over time' (e.g., see Curran et al., 2010; Hedecker & Giddons, 2006).

question is whether panel attrition leads to problems with statistical inference. Three patterns of missing data are usually distinguished, depending on whether and how the missingness is related to the dependent variable (Rubin, 1976).

The first case is when missing values are completely at random (MCAR) and not related to the dependent variable, which does not create any problems for inferences for economic methods or MLM. An important sub-case here is when missing values are related to some fully observable regressor(s) (Little, 1995; Moffit, Fitzgerald & Gottshalk, 1999). The whole issue of missing data is then reduced to tackling the unbalanced panels.

The second case is missing at random (MAR) when missing values can depend on the observed (past) values of the dependent variable. As many authors note, likelihood-based inference is valid under MAR, whereas other methods of inference require MCAR for their validity (e.g., Hedecker & Gibbons, 2006; McNeish, 2014). Therefore, MLM is preferable to economic methods in this case.

The third case is when missing values are not random (MNAR), which is the most unpleasant for inference using MLM or economic methods. MNAR means the missing values are related to the *unobserved* values of the dependent variable. The key problem is that there is no direct approach to test MNAR (e.g., versus MAR) as it depends on the unobservable. There are two general approaches to providing valid inferences in this case – selection models and pattern-mixture models.

Selection models were proposed as a two-step procedure by Heckman (1976). The first step models the probability of dropping out for each subject, while the second stage uses the predicted propensities as a covariate in the main model.

The idea behind panel-mixture models is to divide the sample into groups by their missing-data pattern. The grouping categorical variable is then used in the main analysis. It can be used to show how much those groups differ in terms of the outcome, or whether controlling for that variable changes the estimated parameters of interest (see Hedecker & Gibbons, 2006).

Selection models have become an extremely popular tool to deal with panel data attrition in economics (since Hausman & Wise, 1979), while pattern-mixture models are much less utilized. Within the MLM framework, both approaches are used more evenly (Hedecker & Gibbons, 2006). However, the comparative (dis)advantages of selection models and pattern-mixture models are not well discussed in the literature. Therefore, it is difficult to say which approach, MLM or economic models, should be used in the case of panel data attrition under MNAR.

3.5. Software and computation

The review of an unfamiliar framework like MLM would be incomplete without a discussion of its estimation and computation. The need to estimate variance components clearly distinguishes MLM from linear regressions and makes estimation more difficult and computationally expensive. To do that, MLM relies on the maximum likelihood estimation without closed-form solutions, a method which can fail to converge. This is especially relevant as one of the two modeling strategies in MLM is to 'keep it maximal', randomizing all the parameters first and then fixing those whose variance is not different from zero. In this regard, depending on the algorithm under the hood of particular multilevel modeling software, MLM models can take considerably more time.

How do statistical packages compare by performance for MLM? While the general advice is to 'be cautious when fitting complex models with a large number of macro-micro interactions' (Stegmueller, 2013, p. 759), specialized statistical packages, HLM (Raudenbush et al, 2019) or Mplus (Muthén & Muthén, 1998-2017), seem to have serious computational advantages over the generalized ones such as Stata, SAS, or R. A performance test of five packages estimating MLM shows that Stata is 'by far the slowest of the software programs, and the difference was not trivial' running up to five times more slowly than SAS and up to 50 times more slowly than Mplus (McCoach et al., 2018). To date, effective MLM requires additional investment in learning and / or acquiring specialized software which may present obstacles for some economists.

4. Summary and Conclusions

Multilevel modeling is a full-fledged methodological framework developed to analyze data with a hierarchical structure. It is applied in many social sciences including psychology, sociology, management, public health, and political science. By contrast, economists rarely use it in their analyses of hierarchical data, relying on other econometric techniques. Which approach is preferable? Can economists benefit from the use of a method common in other social sciences? To understand these issues, we reviewed a large amount of methodological literature and contrasted techniques popular in empirical economics with MLM.

The main conclusions of this review are summarized in Table 2. When a researcher wants to estimate the impact of an individual- or group-level variable on some Y and is exclusively concerned with the clusterization of errors, then it is not *a priori* clear which approach, OLS-CRSE or MLM, is preferable. While distributional assumptions are rarely considered as a flaw of the MLM framework, the choice should depend on the characteristics of the dataset used, namely on the number of clusters (groups) and their sizes. Both approaches suffer with a low number of clusters, but MLM generally allows a lower number (at least 20) than OLS-CRSE (30–50). On the other hand, MLM is more vulnerable to the small cluster sizes, demanding more than 5 observations per cluster.

When a researcher wants to estimate the impact of an individual- or group-level variable on some *Y* and is concerned with the omitted variable problem at the lowest level of data (OMV-1), then MLM is clearly a bad choice since this framework offers little to solve this problem. Simple OLS is a bad choice as well, but it may be potentially upgraded using some quasi-experimental techniques, compatible with CRSE. However, if a variable is believed to be omitted at the group level (OMV-2), then MLM (in guise of the 'hybrid' model) is preferred over the standard FE model. MLM becomes the best reasonable choice when the goal is to estimate the impact of a group-level variable and there is no opportunity to apply a quasi-experimental technique.

In modeling parameter heterogeneity across groups, MLM is generally more convenient and provides more efficient estimates than either the EDV 2-step procedure or models with cross-level interactions. A potential disadvantage of MLM, again, is that its framework, unlike EDV, cannot be as easily combined with quasi-experimental methods.

In sum, our review shows that MLM performs at least as well as the economic methods typically applied in a hierarchical data setting. The main limitation is its neglect of the omitted variable problem at the lowest level of the data hierarchy. This limitation should not be exaggerated, though, as the quasi-experimental techniques helping to mitigate this problem often present a challenge even for economists. There is also a growing literature that attempts to integrate quasi-experimental methods popular in economics with the MLM framework (Feller & Gelman, 2015; Kim & Swoboda, 2011; Zubizarreta & Keele, 2017).

Finally, estimation issues are a complementary concern when comparing MLM with economic methods. Flexibility brings complexity, and in practice MLM usually requires more sophisticated and computationally intensive estimation methods than any OLS-based technique. This practical burden may be one of the less obvious hurdles for economists on their way to exploring and using MLM. In general, however, computational time is not a specific MLM problem as it is conditional on the complexity of the model and the chosen estimation strategy, including the choice of computational algorithm and software. Specialized software, such as HLM or Mplus, is able to make the estimation process more efficient and less time-consuming, compared to generalized statistical packages like Stata.

Our review leads to the conclusion that there are few modeling-related reasons for not applying MLM in economic research. MLM covers many typical economic research questions for a hierarchical data structure, including panel data. Rather, there seems to be other reasons hampering the adoption of MLM in economics, related to the disciplinary isolation and intellectual tightness of economics which tend to systematically discount, and raise additional barriers to, ideas and findings outside the discipline (Fourcade et al., 2015; Kapeller et al., 2017). Against this backdrop, we find it particularly important to compare MLM with the standard econometric methods, deciding in a balanced manner, which approach fits which empirical task better. The results of this review suggest that MLM should gain popularity in economics, similar to other social sciences.

Table 2. MLM vs. economic methods in solving problems inherent to multilevel data.

Problem		Formulation	Economic solution	MLM solution	Which is better and when?
Clusterization of errors		In regression $Y_{ij} = \beta_0 + \beta_1 * X_{ij} + \varepsilon_{ij}$ simple OLS underestimates the standard error (SE) of coefficient β_1	OLS-CRSE	Random intercept model (RIM): $Y_{ij} = \beta_{0j} + \beta_1 * X_{ij} + r_{ij}$ $\beta_{0j} = \gamma_{00} + u_{0j}$	It depends. 1) CRSE is ‘model-free’, but MLM is generally robust against mild violations from normality; 2) both OLS-CRSE and MLM are problematic when *Ncl* is too small but MLM allows a lower *Ncl*; 3) MLM is more vulnerable to the small *cluster size*.
		In regression $Y_{ij} = \gamma_0 + \beta_1 * X_{ij} + \beta_2 * W_j + \varepsilon_{ij}$ simple OLS underestimates SE of coefficient β_2	OLS-CRSE	RIM with group-level explanatory variables: $Y_{ij} = \beta_{0j} + \beta_1 * X_{ij} + r_{ij}$ $\beta_{0j} = \gamma_{00} + \gamma_{01} * W_j + u_{0j}$	
Omitted variable	Lowest level	In regression $Y_{ij} = \beta_0 + \beta_1 * X_{ij} + \varepsilon_{ij}$ some individual-level variable may be omitted. OLS estimate of β_1 is biased.	Quasi-experimental methods (e.g., IV)/	No conventional solution	MLM is generally worse as it ignores OMV-1 the problem itself, quasi-experimental methods are not yet incorporated to the MLM framework.
	Higher level	In regression $Y_{ij} = \beta_0 + \beta_1 * X_{ij} + \varepsilon_{ij}$ some group-level variable may be omitted. OLS estimate of β_1 is biased.	Option1: FE model Option 2: IV or other quasi-experimental methods.	‘Hybrid’ model: $Y_{ij} = \gamma_{00} + \beta_1 * \left(X_{ij} - \overline{X_{.j}}\right) + \beta_3 X_{.j} + r_{ij}$ $\beta_{0j} = \gamma_{00} + u_{0j}$	MLM is better than the FE model as it provides the equivalent estimates but more general and flexible (allows group-level explanatory variables and may be extended to the random coefficients model.)
		In regression $Y_{ij} = \gamma_0 + \beta_1 * X_{ij} + \beta_2 * W_j + \varepsilon_{ij}$ some group-level variable may be omitted. OLS estimate of β_2 is biased.	FE model is not applicable. Use IV for W_j or other quasi-experimental methods.	‘Hybrid’ model: $Y_{ij} = \gamma_{00} + \beta_1 * \left(X_{ij} - \overline{X_{.j}}\right) + \beta_3 X_{.j} + r_{ij}$ $\beta_{0j} = \gamma_{00} + \gamma_{01} * W_j + u_{0j}$ It absorbs bias in β_1 and reduces bias in γ_{01}.	MLM is the only choice, unless quasi-experimental methods are applicable.
Heterogeneity of coefficients across groups		Modeling and/or testing heterogeneity of coefficients in regression: $Y_{ij} = \beta_{0j} + \beta_{1j} * X_{ij} + \varepsilon_{ij}$	Option 1: random coefficients model (RCM). Option 2: regression with (Dummy$_j$ * X_{ij}) interactions. Option 3: separate group-specific regressions.	Random intercept-random slope model (RIRSM): $Y_{ij} = \beta_{0j} + \beta_{1j} * X_{ij} + r_{ij}$ $\beta_{0j} = \gamma_{00} + u_{0j}$ $\beta_{1j} = \gamma_{10} + u_{1j}$	MLM is equivalent to Option 1, more parsimonious than both Options 2 and 3 and more efficient than Option 3.
		Explaining heterogeneity of coefficients	Option 1 (RCM) is not applicable. Option 2: regression with X_{ij}* Wj interactions. Option 3: 2-step EDV model	RIRSM with group-level explanatory variables: $Y_{ij} = \beta_0 + \beta_{1j} * X_{ij} + r_{ij}$ $\beta_{1j} = \gamma_{10} + \gamma_{11} * W_j + u_{1j}$	MLM is better; it is more parsimonious than both Options 2 and 3 and more efficient than Option 3. However, MLM, unlike Options 2 or 3, cannot be combined, if necessary, with quasi-experimental methods.

Notes. The first column lists problems typically emerging in economics when analyzing data with a hierarchical structure, while the second column briefly describes them. The third and fourth columns show standard solutions to these problems in empirical economics and MLM, respectively. The last column contains a conclusion on which approach is more advantageous (unbiased, efficient, convenient, or parsimonious) and briefly explains why and when.

References

1. Agresti, A. and Franklin C. (2012) *Statistics: The Art and Science of Learning from Data*. 3rd ed. NY: Pearson Prentice Hall.
2. Allison, P. (2009) *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage.
3. Allison, P (2014) Problems with the Hybrid Method. *Statistical Horizons*. Available at http://www.statisticalhorizons.com/problems-with-the-hybrid-method (accessed July 28, 2017).
4. Angrist, J. and Pishke, J.-S. (2009) *Mostly Harmless Econometrics*. Princeton: Princeton University Press.
5. Angrist, J. and Pishke, J.-S. (2010) The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *American Economic Review* 24(2): 3-30.
6. Arellano, M., (1987) Computing Robust Standart Errors for Within Group Estimators, *Oxford Bulletin of Economics and Statistics* 49(4): 431-434.
7. Arpino, B. and Mealli, F. (2011) The specification of the propensity score in multilevel observational studies. *Computational Statistics and Data Analysis* 55: 1770-1780.
8. Baltagi, B.H., Song, S.H. and Jung, B.C. (2001) The unbalanced nested error component regression model. *Journal of Econometrics* 101(2): 357-381.
9. Baltagi, B. (2008) *Econometric analysis of panel data, 3rd edition*. Chichester: John Wiley & Sons.
10. Barr, D.J., Levy, R., Scheepers, C. and Tily, H.J. (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* 68(3): 255-278.
11. Barrow, L., and Rouse, C.E., (2005) Causality, Causality, Causality: The View of Education Inputs and Outputs from Economics. *FRB of Chicago Working Paper* No. 2005-15. Available at http://dx.doi.org/10.2139/ssrn.869440 (accessed September 4, 2018).
12. Bell A., R. Johnston and K. Jones (2015) Stylized fact or situated messiness? The diverse effects of increasing debt on national economic growth, Journal of Economic Geography, 15(2), 449–472.
13. Bell, A. and Jones, K. (2015) Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods* 3(1): 133-153.
14. Bell, A., Fairbrother, M. and Jones, K. (2018) Understanding and misunderstanding group mean centering: a commentary on Kelley et al.'s dangerous practice. *Quality & quantity* 52.5: 2031-2036.
15. Bell, A., Fairbrother, M. and Jones, K. (2019) Fixed and random effects Models: Making an Informed Choice. *Quality & quantity* 53(2): 1051-1074.
16. Beron, K.J., Murdoch, J.C. and Thayer, M.A. (1999) Hierarchical linear models with application to air pollution in the South Coast Air Basin. *American Journal of Agricultural Economics* 81(5): 1123-1127.
17. Bertrand, M., Duflo, E. and Mullainathan, S. (2004) How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics* 119(1): 249-275.
18. Brambor, T., Clark, W.-R. and Golder, V. (2006) Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis* 14(1): 63-82.
19. Brauer, M. and Curtin, J.J. (2017) Linear Mixed-Effects Models and the Analysis of Nonindependent Data: A Unified Framework to Analyze Categorical and Continuous Independent Variables that Vary Within-Subjects and/or Within-Items. *Psychological methods*. Epub ahead of print, 27 November. DOI: 10.1037/met0000159.
20. Brown J., and Light A. (1992) Interpreting Panel Data on Job Tenure. *Journal of Labor Economics* 10(3), 219–257.
21. Bryan, M. and Jenkins, S. (2016) Multilevel Modeling of Country Effects: A Cautionary Tale. *European Sociological Review* 32(1): 3-22.
22. Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical linear models for social and behavioral research: Applications and data analysis methods*. Chicago: Chicago University Press.

23. Cameron, A. J., Gelbach, and D. Miller (2008) Bootstrap-based Improvments for Inference with Clustered Errors. *Review of Economics and Statistics* 90(3): 414-427.
24. Cameron, A.C. and Miller, D.L. (2015) A practitioner's guide to cluster-robust inference. *Journal of Human Resources* 50(2): 317-372.
25. Carter, C.R., Meschnig, G. and Kaufmann, L. (2015) Moving to the next level: Why our discipline needs more multilevel theorization. *Journal of supply chain management* 51(4): 94-102.
26. Carter A., Kevin Schnepel, and D. Steigerwald (2017) Asymptotic Behavior of a t-Test Roust to Cluster Heterogeneity, *Review of Economics and Statistics* 99(4): 698–709.
27. Chow, G. (1960) Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica* 28(3): 591-605.
28. Clark, T. and Linzer, D. (2015) Should I Use Fixed or Random Effects? *Political Science Research and Methods* 3(02): 399-408.
29. Clarke P. (2008) Theory and Methods: When Can Group-level Clustering Be Ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology and Community Health* 62(8): 752-758.
30. Curran, P.J., Obeidat, K. and Losardo, D. (2010) Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development* 11(2):121-136.
31. Dieleman, J.L. and Templin, T. (2014) Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: A simulation study. *PloS one* 9(10): e110257.
32. Dielman, T.E. (1989) *Pooled cross-sectional and time series data analysis*. NY: M. Dekker.
33. Diez-Roux, A.V. (2000) Multilevel analysis in public health research. *Annual review of public health* 21(1): 171-192.
34. Eicker, F. (1967) Limit theorems for regressions with unequal and dependent errors. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1: 59-82.
35. Fairbrother, M. (2014) Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets. *Political Science Research and Methods* 2(1): 119-140.
36. Feller, A. and Gelman, A. (2015) Hierarchical models for causal effects. In R. Scott and S. Kosslyn (eds.) *Emerging Trends in the Social and Behavioral Sciences: An interdisciplinary, searchable, and linkable resource*. John Wiley & Sons. DOI: 10.1002/9781118900772
37. Field, A. (2013) *Discovering statistics using IBM SPSS Statistics*, 4th ed. London: Sage.
38. Fielding, A. (2004) The role of the Hausman test and whether higher level effects should be treated as random or fixed. *Multilevel modelling newsletter* 16.2: 3-9.
39. Fourcade, M., Ollion, E., and Algan, Y. (2015) The superiority of economists. *Journal of economic perspectives* 29(1): 89-114.
40. French M., and I. Popovichi (2011) That Instrument Is Lousy! In Search of Agreement When Using Instrumental Variables Estimation in Substance Use Research. *Health Economics* 20: 127-146.
41. Gelman, A. (2006) Multilevel (Hierarchical) Modeling: What It Can Do and Cannot Do. *Technometrics* 48(3): 432-435.
42. Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. NY: Cambridge University Press.
43. Gill, J. and Womack, A. (2013) The Multilevel Model Framework. In J.S. Simonoff, M.A. Scott, and B.D. Marx (eds.) *The SAGE Handbook of Multilevel Modeling* (pp. 3-20). London: SAGE Publications.
44. Goetgeluk, S., Vansteelandt, S. and Goetghebeur, E. (2008) Estimation of controlled direct effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.5: 1049-1066.
45. Goldstein, H. (1995) *Multilevel Statistical Models*, 2nd ed. London: Edward Arnold; New York: Halsted Press.
46. Gorard, S. (2007) The dubious benefits of multi-level modeling. *International Journal of Research & Method in Education* 30(2): 221-236.

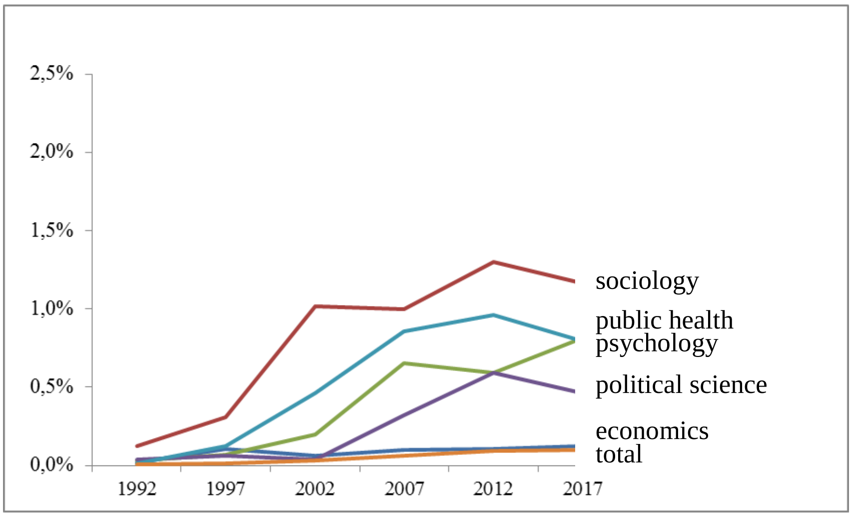
47. Greene, W.H. (2003) *Econometric Analysis,* 5th ed.. NJ: Pearson.
48. Greene, W.H. (2007) *Econometric Analysis,* 6th ed. NJ: Pearson.
49. Greene, W.H. (2011) *Econometric Analysis,* 7th ed. NJ: Pearson.
50. Greene, W.H. (2017) *Econometric Analysis,* 8th ed. NJ: Pearson.
51. Hansen C. B. (2007) Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data When T Is Large. *Journal of Econometrics* 141: 597–620.
52. Hanushek, E.A. (1974) Efficient estimators for regressing regression coefficients. *American Statistician* 28: 66-67.
53. Hausman, J.A. (1978) Specification tests in econometrics. *Econometrica: Journal of the econometric society* 46(6): 1251-1271.
54. Hausman, J.A., and Taylor, W.E. (1981) Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398.
55. Hausman J. A. & D. Wise (1979). Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment, *Econometrica* 47(2): 455-473.
56. Heck, R.H. and Thomas, S.L. (2000) *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum.
57. Heckman, J. (1979) Sample selection bias of the ordinary least squares estimator of the tobit model. *Econometrica* 47: 153-161.
58. Hedeker D., and R. Gibbons (2006) *Longitudinal Data Analysis*, New Jersey: Wiley
59. Heisig, J.P., Schaeffer, M. and Giesecke, J. (2017) The Costs of Simplicity: Why Multilevel Models May Benefit from Accounting for Cross-Cluster Differences in the Effects of Controls. *American Sociological Review* 82(4): 796-827.
60. Heisig, J.P., and Schaeffer, M. (2019) Why You Should Always Include a Random Slope for The Low-Level Variable Involved in a Cross-Level Interaction. *European Sociological Review* 35(2): 258-279.
61. Hill, J. (2013). Multilevel Models and Causal Inference. In J.S. Simonoff, M.A. Scott, and B.D. Marx (eds.) *The SAGE Handbook of Multilevel Modeling* (pp. 201-220). London: SAGE Publications.
62. Hoffman, D. (1997) An Overview of the Logic and Rationale of Hierarchical Linear Models. *Journal of Management* 23(6): 723-744.
63. Hoover, K. (2016) *Causality in Economics and Econometrics* In: Palgrave Macmillan (eds.) The New Palgrave Dictionary of Economics, Live Edition. Palgrave Macmillan, London. DOI: 10.1057/978-1-349-95121-5_2227-1.
64. Hox, J.J., Moerbeek, M. and Van de Schoot, R. (2010) *Multilevel analysis: Techniques and applications*. London: Routledge.
65. Hsiao, C. (1986) *Analysis of Panel Data, Vol. 11 of Econometric Society Monographs*. NY: Cambridge University Press.
66. Hsiao, C. (2007) Panel data analysis — advantages and challenges. *Test* 16(1): 1-22.
67. Hsiao, C. (2014) *Analysis of Panel Data,* 3rd ed. NY: Cambridge University Press.
68. Hsiao C. and H. Pesaran (2008) *Random Coefficient Models*, in L. Matyas and P. Sevestre (eds.) The Econometrics of Panel Data, pp.185-214.
69. Huang, F.L. (2018) Multilevel Modeling and Ordinary Least Squares Regression: How Comparable Are They? *The Journal of Experimental Education* 86(2): 265-281.
70. Huber, P.J. (1967) The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1: 221-233.
71. Imbens, G.W. and Rubin, D.B. (2008) Rubin Causal Model. In: Palgrave Macmillan (eds.) *The New Palgrave Dictionary of Economics* (pp. 229-241). Palgrave Macmillan, London.
72. Imbens, G.W. and Wooldridge, J.M. (2009) Recent developments in the econometrics of program evaluation. *Journal of economic literature* 47(1): 5-86.
73. Kapeller, J., Aistleitner, M., and Steinerberger, S. (2017). Citation patterns in economics and beyond: Assessing the peculiarities of economics from two scientometric perspectives. No. 60. *ICAE Working Paper Series*. Available at http://hdl.handle.net/10419/171435 (accessed September 25, 2019).

74. Kedar, O. and Shively, P. (2005) Introduction to the Special issue on multilevel models. *Political Analysis* 13 (4): 297-300.
75. Kennedy, P. (2008) *A guide to econometrics*. Cambridge, MA: MIT press.
76. Kim, J.-S. and Frees, E.W. (2006) Omitted variables in multilevel models. *Psychometrika* 71: 659-690.
77. Kim, J.-S. and Frees, E.W. (2007) Multilevel modeling with correlated effects. *Psychometrika* 72(4): 505-533.
78. Kim, J.-S. and Swoboda, C. (2011) Handling Omitted Variable Bias in Multilevel Models. In J. Hox and J.K. Roberts (eds.) *Handbook of Advanced Multilevel Analysis* (pp. 197-217). London: Routledge.
79. King, G., and 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.
80. Koop, G. and Tobias, J.L. (2004) Learning about heterogeneity in returns to schooling. *Journal of Applied Econometrics* 19(7): 827-849.
81. Kreft, I.G.G. and De Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.
82. Laird N., & J. Ware (1982) Random Effects Models for Longitudinal Data. *Biometrika* 38(4): 963-974.
83. Lewis, J. and Linzer, D. (2005) Estimating Regression Models in Which the Dependent Variable Is Based on Estimates. *Political Analysis* 13: 345-364.
84. Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73(1): 13–22.
85. Little A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American StatisticalAssociation*, 90:1112-1121, 1995.
86. Luke, D.A. (2004) *Multilevel modeling. Series: Quantitative Applications in the Social Sciences, Vol. 143*. Thousand Oaks, CA: Sage.
87. Maas, C.J. and Hox, J.J. (2004) The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational statistics & data analysis* 46(3): 427-440.
88. Maas, C.J. and Hox, J.J. (2005) Sufficient Sample Sizes for Multilevel Modeling. *Methodology* 1(3): 86-92.
89. Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. and Bates, D. (2017) Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94: 305-315.
90. McCoah B., G. Rifenbark, S, Newton, X. Li, J. Kooken, D. Yomtov, A. Gambino, and A. Bellara (2018). Does the Package Matter? A Comparison of Five Common Multilevel Modeling Software Packages, *Journal of Educational and Behavioral Statistics*, 43(5): 594-627.
91. McNeish, D. (2014). Modelling Sparsely Clustered Data: Design-based, Model-based, and Single-level Methods. *Psychological Methods*, 19(4), 552-563.
92. McNeish, D. and Stapleton, L.M. (2016) Clustered Data with Very Few Clusters. *Multivariate Behavioral Research* 51(4): 495-518.
93. McNeish, D., Stapleton, L.M. and Silverman, R.D. (2016) On the Unnecessary Ubiquity of Hierarchical Linear Modeling. *Psychological methods* 22(1): 114-140.
94. Moffit, Fitzgerald and Gottshalk (1999) Sample Attrition in the Panel Data; the Role of Selection on Observables, *Annales d'Économie et de Statistique*, 55/56: 129-152.
95. Morrell, C. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics*, 54, 1560–1568.
96. Moulton, B. (1990) An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units. *The Review of Economics and Statistics* 72(2): 334-338.
97. Mundlak, Y. (1978) Pooling of Time-series and Cross-section Data. *Econometrica* 46(1): 69-85.
98. Munnell A (1990) Why has productivity growth declined? Productivity and public investment. New England Economic Review 30: 3-22.
99. Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

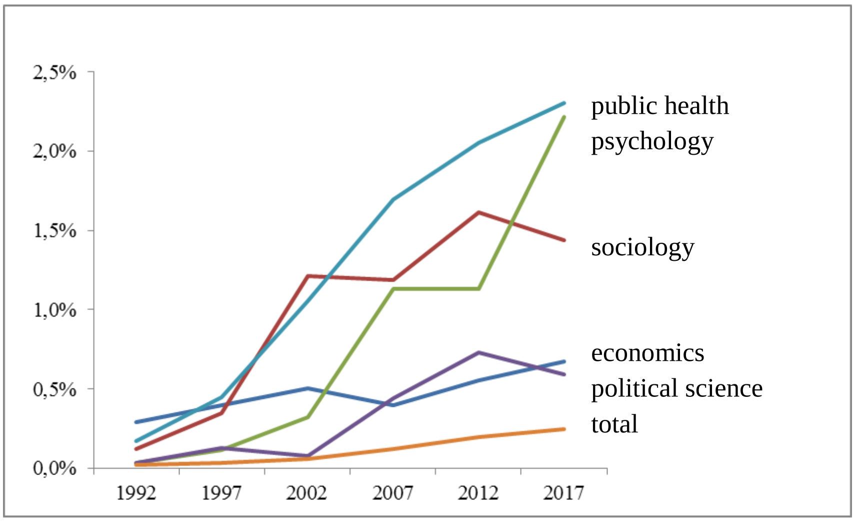
100. Neuhaus, J.M. and Kalbfleisch, J.D. (1998) Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 54(2): 638-645.
101. Oakes, M. (2004) The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social Science & Medicine* 58(10): 1921-1952.
102. Oort F, M. Burger, J. Knoben, O.Raspe (2012) Multielvel Approaches and the Firm-Agglomeration Ambiguity in Economic Growth Studies, *Journal of Economic Surveys*, 26(3), 468-491.
103. Ozkaya, H.E., Dabas, S., Kolev, K., Hult, T., Dahlquist S. and Manjeshwar, S.A. (2013) An assessment of hierarchical linear modeling in international business, management, and marketing. *International Business Review* 22(4): 663-677.
104. Pagan, A. (1984) Econometric Issues in the Analysis of Regressions with Generated Regressions. *International Economic Review* 25(1): 221-247.
105. Pieters R., and H., Baumgarrner (2002) Who Talks to Whom? Intra- and Interdisciplinary Communication of Economics Journals. *Journal of Economic Literature*, 40(2): 483-509.
106. Pluemper, T. and Troeger, V. (2007) Efficient Estimation of Time-invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. *Political Analysis* 15(2): 124-139.
107. Pluemper, T. and Troeger, V. (2011) Fixed-Effects Vector Decomposition: Properties, Reliability, and Instruments. *Political Analysis* 19(2): 147-164.
108. Rabe-Hesketh S. & A. Skrondal (2012). *Multilevel and Longitudinal Modeling Using Stata*, 3rd Edition, Stata Press.
109. Raudenbush, S.W. (1989) The analysis of longitudinal, multilevel data. *International journal of educational research* 13(7): 721-740.
110. Raudenbush, S. W. and Bryk, A.S. (2002) *Hierarchical linear models: Applications and data analysis methods, Vol. 1*. Thousand Oaks, CA: Sage.
111. Raudenbush, S.W., Bryk, A.S, Cheong, Y.F. & Congdon, R. (2019). *HLM 8 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.
112. Reinhart, C., and K. Rogoff. (2010). Growth in a Time of Debt. *American Economic Review*, 100 (2): 573-78.
113. Rice, N. and Jones, A. (1997) Multilevel models and health economics. *Health economics* 6(6): 561-575.
114. Robinson, W. (1950). Ecological Correlations and the Behavior of Individuals, *American Sociological Review*, 18(3), 351-357.
115. Rubin, D. (1976). Inference and missing data. *Biometrika*, 63: 581-592.
116. Saxonhouse, G. (1976) Estimated Parameters as Dependent Variables. *American Economic Review* 66(1): 178-183.
117. Schielzeth, H., and W. Forstmeier. (2008). Conclusions Beyond Support: Overconfident Estiamtes in Mixed Models. *Beahavioural Ecology* 20(2): 416-420.
118. Schmidt-Catran, A.W. and Fairbrother, M. (2015) The random effects in multilevel models: Getting them wrong and getting them right. *European Sociological Review* 32(1): 23-38.
119. Schunck, R. (2013) Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *The Stata Journal* 13(1): 65-76.
120. Singer J., and J. Willet (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* NY: Oxford University Press.
121. Snijders, T. (1996). Analysis of longitudinal data using the hierarchical linear model. *Quality and Quantity* 30(4): 405-426.
122. Snijders, T.A.B. and Bosker, R.J. (2012) *Multilevel Analysis*, 2nd ed. London: Sage.
123. Snijders, T.A.B. and Bosker, R.J. (1999) *Multilevel Analysis: An Introduction to basic and advanced multilevel modeling*. London: Sage.
124. Spencer, N. and Fielding, A. (2000) An Instrumental variable consistent estimation procedure to overcome the problem of endogenous variables in multilevel models. *Multilevel Modeling Newsletter* 12(1): 4-7.
125. Stanley, T.D. and Jarrell, S. (1989) Meta-Regression Analysis: A Quantitative Method of Litareture Surveys. *Journal of Economic Surveys* 3(2): 161-170.

126. Steenbergen, M.R. and Jones, B.S. (2002) Modeling multilevel data structures. *American Journal of Political Science* 46(1): 218-237.
127. Stegmueller, D. (2013) How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science* 57(3): 748-761.
128. Swamy, P.A.V.B. (1970) Efficient Inference in a Random Coefficient Regression Model. *Econometrica* 38(2): 311-323.
129. Swamy, P.A.V.B. (2012) [1971]. *Statistical inference in random coefficient regression models*. Berlin: Springer Science & Business Media.
130. Swamy, P.A. and Tavlas, G.S. (1995) Random coefficient models: theory and applications. *Journal of Economic Surveys* 9(2): 165-196.
131. Van Eck, N.J. and Waltman, L. (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84(2): 523-538.
132. Van der Meer, T., Grotenhuis, M. and Pelzer, B. (2006) Influential Cases in Multilevel Modeling: A Methodological Comment. *American Sociological Review* 75(1): 173-178.
133. Wakefield, J. (2009) Multi-level modeling, the ecologic fallacy, and hybrid study designs. *International Journal of Epidemiology* 38(2): 330-336.
134. White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica:* 48(4): 817-838.
135. Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*. Cambridge MA: MIT Press.
136. Wooldridge, J. (2003). Cluster-Sample Methods in Applied Econometrics, *American Economic Review* 93(2): 133-138.
137. Wooldridge, J. (2005) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics* 20(1): 39-54.
138. Wooldridge, J. (2010a). Econometric Analysis of Cross Section and Panel Data, 2nd ed. Cambridge, MA: MIT Press.
139. Wooldridge, J. (2010b). Correlated Random Effects Models with Unbalanced Panels, http://econ.msu.edu/faculty/wooldridge/docs/cre1_r4.pdf. Accessed 19 May, 2017.
140. Young, A. (2019). Consistency without Inference: Instrumental Variables in Practical Application, Mimeo (Draft version: August 2019).
141. Zellner, A. (1962) An Efficient Method of Estimating Seemingly Regression Equations and Tests for Aggregation Bias. Journal of the American Statistical Association 57: 348-368.
142. Zubizarreta, J. and Keele, L. (2017) Optimal Multilevel Matching in Clustered Observational Studies: A Case Study of the Effectiveness of Private Schools Under a Large-Scale Voucher System. The Journal of American Statistical Association 112(518): 547-560.

Appendix A.
Relative number of articles with MLM-related terms across disciplines.



(a)



(b)

Figure A1. Share of Web of Science Core Collection articles having MLM-related terms in title, abstract, or keywords, a narrow **(a)** and wide **(b)** set of search terms, every five years, 1992-2017.

Appendix B.
MLM in Econometrics Textbooks

Table B1. Popular *econometrics* textbooks analyzed for MLM coverage

Author	Title	Year (Edition)	Total number of citations*
William H. Greene	Econometric Analysis	2003(5), 2007(6), 2012(7), 2017(8)	64,068
Jeffrey M. Wooldridge	Econometric analysis of cross-section and panel data	2002(1), 2010(2)	30,544
Badi H. Baltagi	Econometric Analysis of Panel Data	2008(3)	15,671
Cheng Hsiao	Analysis of Panel Data	2014(3)	11,077

*Total number of citations in Google Scholar (as of 10 August, 2017).

In Greene's 5th edition (2003), a multilevel model was referred to as a 'random parameters model' or 'hierarchical model' and presented in the chapter on panel data in the form of a random-slopes random-intercept model (see Equations M9-M11). In that section, the sole author's comment was that such a model is 'extremely versatile' (Greene, 2003, p. 286). Further in the book, this model was called a 'multilevel' model and 'hierarchical regression model' and appeared in sub-section 13.8 'Random coefficients model' as an extension of a random coefficients model which allows 'the means of the coefficients to vary with measured covariates' (Greene, 2003, p. 319). The author notes that this model can still be fit by least squares.

In Greene's 6th edition (2007), the MLM framework received more attention. This edition contains a separate sub-section 9.8.4 'Hierarchical linear models' within section 9.8 'Parameter heterogeneity' in the chapter on panel data analysis (Greene, 2007, pp. 233-237). Here the author presents a two-level model as a system of two equations similar to Equations (1) and (2). Like in the 5th edition, Greene notes that this model is the same as the random-coefficients model (reviewed in subsections 9.8.1 and 9.8.2) with the addition of the interaction between individual- and group-level explanatory variables. He also presents two applications of this model in economics (Beron et al., 1999; Koop & Tobias, 2004). Interestingly, in subsection 9.8.3 Greene refers to the 'hierarchical linear model' as well and notes that if unobserved group effects are correlated with individual-level explanatory variables, then a two-step estimation procedure based on fixed-effect estimator should be applied. (Section 3 in our paper discusses this approach). Additionally, we encountered a note related to MLM in subsection 9.7.1 'Nested random effects' (Greene, 2007, p. 214). The author formulates a 4-level (!) model (students nested in classes nested in schools nested in school districts) as an extension of the standard random-effects model. (In terms of the MLM framework, it is a random-intercept model that distinguishes variance components at four levels of aggregation.) Greene discusses possible estimation approaches for such a model in balanced and unbalanced samples with reference to Baltagi et al. (2001) and provides the example of Munnell (1990) with reference to the econometrics textbook by Baltagi (2008).

In Greene's 7th edition (2012), hierarchical linear models appear twice. First, like in the 6th edition, they appear in a sub-section of the section 'Parameter heterogeneity' of chapter 11 dedicated to panel data models (Greene, 2012, pp. 420-421). Compared to the previous edition, the length of that sub-section shortened and the only proposed approach to estimating the two-level model was a two-step approach based on the fixed-effects model. The second mentioning of MLM was in section 15.8, chapter 'Simulation-based estimation and inference and random parameters model' (Greene, 2012, pp. 639-641) which follows a section on the random-parameters model. The key practical example on MLM remains the same (Beron et al., 1999). The **latest Greene's edition** to date has the same references and contents on MLM (Greene, 2017).

In **Wooldridge's first edition**, we found only one reference to MLM in section 11.5 'Applying Panel Data Methods to Matched Pairs and Cluster Samples' with an explicit reservation about this model. As the author notes, *'in some fields, an unobserved effects model for a cluster sample is called a hierarchical model. In the hierarchical models literature, c_i [which is the group-level random effect] is often allowed to depend on cluster-level covariates[...] But this is equivalent to simply adding cluster-level observables to the original model and relabeling the unobserved cluster effect*' (Wooldridge, 2002, p. 329).

In **Wooldridge's second edition**, we found an extended reference to MLM in sub-section 20.3.2 'Cluster Samples with Unit-Specific Panel Data', section 20.3 'Cluster sampling', chapter 'Stratified Sampling and Cluster Sampling' (see Wooldridge, 2010a, pp. 876-877):

> *Here we consider the case where each unit belongs to a cluster and the cluster identification does not change over time. In other words, we have panel data on each individual or unit, and each unit belongs to a cluster. For example, we might have annual panel data at the firm level where each firm belongs to the same industry (cluster) for all years. Or, we might have panel data for schools that each belong to a district. This is a special case of a hierarchical linear model (HLM) setup. Models for data structures involving panel data and clustering are also called mixed models (although this latter name typically refers to the situation, which we treat later, in which some slope parameters are constant and others are unobserved heterogeneity). In the HLM/mixed*

models literature, more levels of nesting are allowed, but we will not consider more general structures; see, for example, Raudenbush and Bryk (2002).

Further, Wooldridge formulates a 3-level model in the combined form and notes that this model can also include explanatory variables at any level. He emphasizes that in this case, if one wants to estimate their impact, one should assume that there are no group-level unobservables affecting the dependent variable.

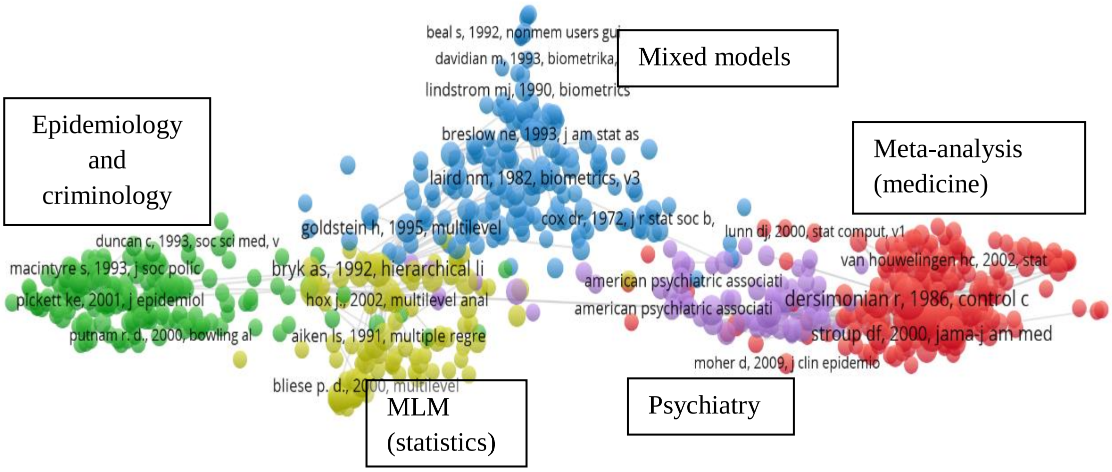
Under this assumption, the OLS estimator is consistent but the feasible generalized least squares (FGLS) estimator is more appropriate as it allows the researcher to take clusterization into account. Wooldridge also notes that a simpler strategy is available, which is to apply usual random-effects estimator at the individual level, effectively ignoring the clusters in estimation, but computing a fully robust variance matrix estimator.

The two selected textbooks focusing on panel data econometrics, Baltagi (2008) and Hsiao (2014), slightly differ in their coverage of MLM. In Baltagi's 'Econometric analysis of panel data' we did not manage to find any direct references to 'multilevel' or 'hierarchical' models. However, one may encounter a 4-level random-intercept model in section 9.6 'The unbalanced nested error component model' (Baltagi, 2008, p. 180), which resembles Greene's section 'Nested random effects' (Greene, 2007, p. 214) and contains the same applied example (Munnell, 1990).

Likewise, Hsiao's 'Analysis of panel data' does not contain any of the terms 'hierarchical linear model' or 'multilevel model'. However, a 2-level model is presented in section 6.5 'Coefficients that are functions of exogenous variables' (Hsiao, 2014, p. 193). The following section, 6.6, titled 'Mixed fixed- and random-coefficients model', considers a case when some coefficients of a 2-level model may be fixed while others are random, which is the core of the MLM framework (Gelman & Hill, 2007). Finally, in the section 'Data with multilevel structure' Hsiao, like Baltagi and Greene, formulates a 'multiway error component model' and discusses possible methods of its estimation (Hsiao, 2014, p. 453). But none of them goes further to acknowledge (or formulate) these techniques as a methodological framework of its own.

Appendix C.
Disciplinary fields within the MLM literature

For our citation analysis we selected all the Web of Science Core Collection papers (articles, chapters, reviews, etc.) published between 1992 and 2017 that contain any of the following terms in their titles, abstracts, or keywords – ‘hierarchical linear’, ‘multilevel model’, ‘multilevel analysis’, ‘random effect’, ‘random coefficient’, or ‘mixed-effect’ (the ‘wide’ set of search terms). Next, we collected the reference lists of those documents and constructed a co-citation network of documents that shows ‘communities of documents’ cited together in the MLM literature. To make the clusters less cumbersome, we limited the sources to the 10,000 most cited papers (having at least 10 citations). The resulting map is presented in Figure C1.



.

Fig. C1. Co-citation groups in references lists from the 10,000 most cited MLM-related works.
Notes. Web of Science Core Collection; 1992-2017; 716 documents with 10 or more co-citations; search terms in title, keywords, or abstract: ‘hierarchical linear’, ‘multilevel model’, ‘multilevel analysis’, ‘random effect’, ‘random coefficient’, or ‘mixed-effect’. Made by the authors with VOSviewer (Van Eck & Waltman, 2010). Each circle is a document. The larger the circle, the more co-citations it has. The choice of colors is arbitrary; positions of circles in the picture are relative. Papers which are closer in this graph are cited together more often. Lines are co-citations from other publications in Web of Science Core Collection.

Each circle in this figure represents a separate published document. The larger the circle, the more co-citations this document has. All documents may be classified into five large clusters – ‘communities of references’ – which are referred to together in the literature. These clusters roughly correspond to disciplinary and methodological boundaries in epidemiology and criminology, multilevel modeling as a statistical framework, early statistical texts on mixed models, psychiatry, and meta-analysis in medicine (from left to right). Econometric or economics papers or books do not constitute a separate methodological tradition among the highly cited MLM-related papers.

Figure C2 represents the same analysis but focuses on journals where MLM-related articles are published. The only economics journal visible in this figure, *Econometrica*, is located far away from other journals publishing MLM-related articles, i.e., it is rarely cited along with any other journals applying similar models.

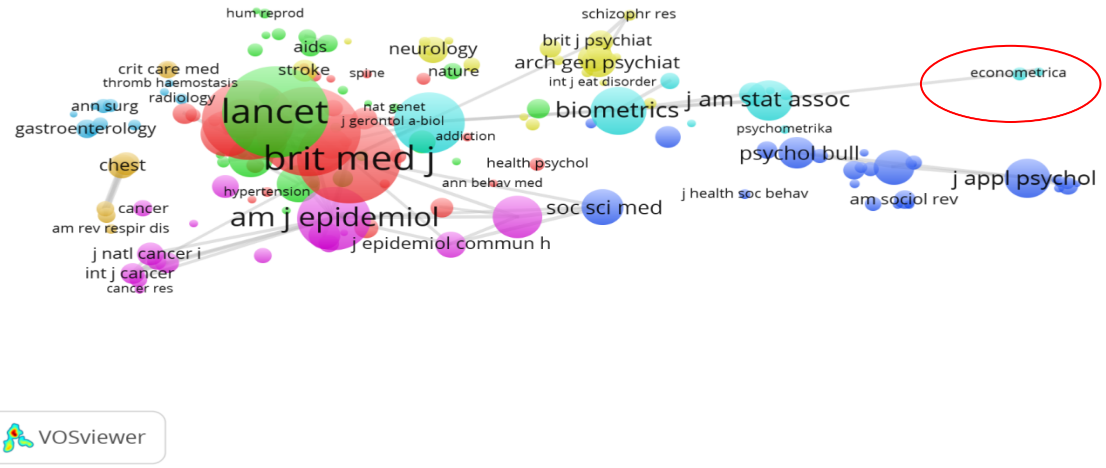


Fig. C2. Co-citation network of journals with articles related to MLM.
Notes. Web of Science Core Collection; 1992-2017; 160 journals (5,000 articles) with 100 or more co-citations; search terms in title, keywords, or abstract: ‘hierarchical linear’, ‘multilevel model’, ‘multilevel analysis’, ‘random effect’, ‘random coefficient’, or ‘mixed-effect’. The graph built with VOSviewer (Van Eck & Waltman, 2010). Each circle is a journal cited within the MLM literature along with another journal. The larger the circle, the more co-citations the journal has. Color clusters show the journals most commonly cited together (the choice of colors is arbitrary). Positions of circles in the picture are relative. Journals that are closer on this graph are cited together more often in publications from Web of Science Core Collection.

We also analyzed the position of economists within the MLM-related literature from the side of economics. Relying on the disciplinary classification of sources in the Web of Science, we explicitly distinguished the segment of economics papers. Of all the MLM-related documents from 1992 to 2017, we selected those identified with economics (n=1,661) and built a co-citation network of references for this segment only. To avoid cluttering, we used the works co-cited at least 10 times and obtained a segment of 36 papers. The results are presented in Figure C3.

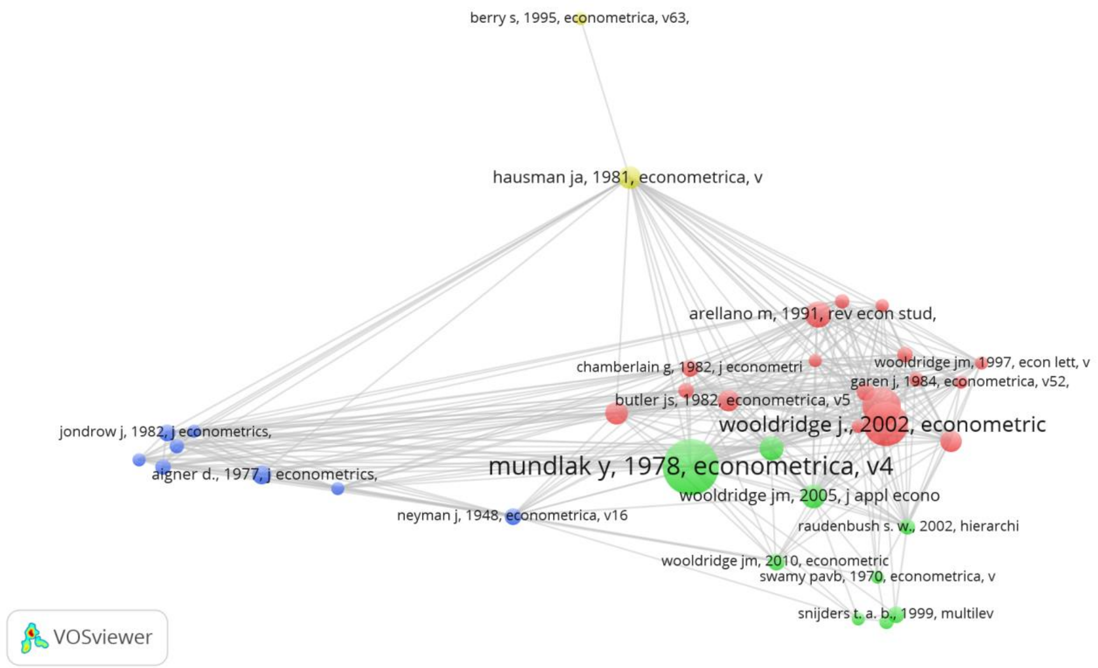


Fig. C3. Co-citation network of 36 economic papers containing MLM-related terms and co-cited 10 or more times, 1992-2017.

Notes. Web of Science Core Collection; 1,661 papers, search terms in title, keywords, or abstract: ‘hierarchical linear’, ‘multilevel model’, ‘multilevel analysis’, ‘random effect’, ‘random coefficient’, or ‘mixed-effect’. The graph built with VOSviewer (Van Eck & Waltman, 2010). Each circle is a paper. The larger the circle, the more co-citations it has with other papers from Web of Science Core Collection containing the same search terms. The choice of colors is arbitrary; positions of circles in the picture are relative. Papers which are closer in this graph are cited together more often. Lines are co-citations.

Figure C3 shows four clusters of papers. The largest of them (central right in the graph) contains 17 works by M. Arellano, J. Butler, G. Chamberlain, J. Garen, J. Hausman, C. Hsiao, and J. Wooldridge. The second largest cluster (bottom) is most often co-citing C. Hsiao, Y. Mundlak, and J. Wooldridge. This cluster also features classic works on MLM by H. Goldstein, S. Raudenbush, and T. Snijders, and which are, however, located at the very edge of the graph. Their marginal position in the economic literature resembles the marginal position of economic authors and journals within the MLM literature mentioned above. This suggests that economists have their own citing (and thinking) tradition that goes somewhat apart from other social sciences. Although multilevel classics (Bryk, Goldstein, and Raudenbush) are not unknown, they are marginal to the overall discussion on MLM-related models in economics.

Aleksey Oshchepkov
National Research University Higher School of Economics, Moscow, Centre for Labour Market Studies, Senior Research Fellow, E-mail: aoshchepkov@hse.ru. The corresponding author