PENNSTATE
COLLEGE of EDUCATION
CENTER FOR THE STUDY OF HIGHER EDUCATION

Multilevel Analysis in Higher Education Research: A Multidisciplinary Approach

John J. Cheslock
Center for the Study of Higher Education
The Pennsylvania State University
jjc36@psu.edu

&

Cecilia Rios-Aguilar
Center for the Study of Higher Education
University of Arizona
ceciliar@email.arizona.edu

August 2010
Working Paper No. 4

This paper was prepared for publication in *Higher Education: Handbook of Theory and Research*, Volume 26.

Multilevel Analysis in Higher Education Research: A Multidisciplinary Approach

Abstract:

Within higher education research, both hierarchical linear models (HLM) and econometric panel models are commonly employed in studies examining multilevel data. These two statistical traditions are interesting to compare, because despite a number of underlying similarities, they differ in complementary but often confusing ways. The confusion arises from varying terminology and model presentation, which makes almost identical models appear different. Econometrics textbooks focus on how the multilevel structure can be exploited to advance overall causal inference, while HLM texts primarily highlight opportunities to examine heterogeneity across groups. This chapter highlights the core similarities between these two traditions so HLM-trained researchers can use their existing knowledge base to read econometric-based articles and vice versa. By contrasting these approaches, this chapter helps applied higher education researchers learn the full range of benefits allowed by the advanced analysis of multilevel data.

Introduction

Most quantitative studies in higher education utilize cluster data sets where students are nested within schools, faculty are nested within higher education institutions, or higher education institutions are nested within states. Higher education researchers also employ panel data sets where individual entities (e.g., higher education institutions) have multiple observations, with the observations representing different periods of time. Because individual observations are essentially nested within entities in panel data, its structure is very similar to a cluster data set. This data structure, where individual observations are nested within groups, is present in other types of data sets as well. The term multilevel data is typically used to describe this structure. Multilevel data often contains independent variables representing the individual observations (which we will call level-1 variables) and independent variables representing the groups in which the observations are nested (which we will call level-2 variables).

A researcher restricted to using a standard OLS regression could only produce limited insight into the variation contained within these sorts of multilevel data sets. Tests of statistical significance would be inaccurate because the correlation between error terms within groups would lead to incorrect standard error estimates. Bias in level-1 coefficients could not be reduced by controlling for all observed and unobserved level-2 variables that are consistent across the group. The researcher could not effectively measure variation across groups in their level-1 intercepts and coefficients and could not examine how those group-specific intercepts and coefficients covary with level-2 variables. And individual estimates of these level-1 intercepts and coefficients could not be produced for specific groups with smaller sample sizes.

Multilevel models allow researchers to produce accurate standard errors, reduce level-1 coefficient bias, measure variation across groups in their level-1 coefficients, and produce level-1 estimates for specific groups with smaller sample sizes. Higher education researchers are often aware that some of these potential benefits exist, but few are conscious of the wide range of remedies just mentioned. Which remedies are known by an individual researcher? The answer often depends upon the statistical tradition within which the researcher was trained.

Within higher education journals, researchers almost exclusively draw from two traditions when performing advanced regression analysis of multilevel data (Cheslock & Rios-Aguilar, 2008). The first class, hierarchical linear models (HLM) were brought into prominence by Bryk & Raudenbush (1992) and are clearly the technique of choice within higher education research. The second class of models, which are growing in use in higher education, are commonly referred to as econometric panel models, but they have received less attention from higher education researchers.

The econometric approach is often very simple: the basic econometric models focus on exploiting the multilevel structure of the data to reduce bias created by the lack of controls for unobserved group-level characteristics. These models seek to control for these characteristics by exclusively using variation within groups when estimating the results. Econometric textbooks discuss this approach in-depth and the applied researcher, after reading these texts, will likely seek research contexts where within-group variation is of special interest.

The applied researcher would, however, search for very different contexts after reading an HLM text. While the option to estimate results solely using variation within groups is briefly noted within these texts, it is not featured as a major opportunity created by multilevel data. Attention is placed instead on variation across groups in terms of outcomes or in terms of the relationship between outcomes and individual-level explanatory variables. Applied researchers will consequently be drawn to studies for which heterogeneity across groups is especially relevant and of interest.

Higher education researchers would benefit from the perspectives emphasized in both traditions. When researchers are not able to experimentally assign the policy or practice of interest, the econometric focus on reduction in coefficient bias for the overall coefficients will be of special interest. But educational researchers are often not interested in just the overall relationship, and the HLM approach emphasizes opportunities to thoroughly examine how these relationships vary across groups or group-level characteristics.

Most higher education researchers typically view multilevel data solely through the HLM or econometric lens. Clearly, higher education research would be enhanced if researchers knew and employed the primary insights from both traditions. But to do so presents a number of challenges. Applied educational researchers are usually trained within just one tradition, and because econometrics and HLM often differ in very significant ways, researchers may not be able to easily translate their mastery of one tradition into an understanding of the other. For example, key terms, such as random-effects and fixed-effects, mean very different things across these two traditions. Also, models that produce almost identical overall results vary substantially in presentation. HLM models are presented using multiple equations while econometric models are typically based on a single equation. As a result, many applied researchers do not even realize the underlying similarities across these models, thereby complicating efforts to critically examine the work of others.

Even more troubling is the evidence that many education researchers are not even properly employing the statistical tradition in which they were trained (Cheslock & Rios-Aguilar, 2008; Dedrick et al., 2009). Researchers are often not fully exploiting the opportunities presented by multilevel data to gain deeper insights into the phenomenon under study. Instead, they are simply employing the most basic HLM or econometric model without deep thought about how to craft and interpret the model to maximize insight.

The goal of this chapter is to improve future analysis of multilevel data. To guide this work we provide a framework containing a simple yet insightful presentation of the core benefits of multilevel data. The econometric and HLM approaches to multilevel data are presented and then contrasted in terms of numerous elements, including the benefits they feature.[1] We also examine a number of studies from the literature that demonstrate how researchers can reap the potential benefits of multilevel data.

This chapter seeks to make a targeted contribution to the literature. Except for several attempts to generalize the issues, we focus solely on issues pertaining to multilevel data. Even more specifically, we only discuss multilevel analysis that occurs within a regression framework. This restriction reflects the regression-based focus of econometrics and the knowledge base of the authors. Readers who primarily employ other statistical approaches, such as structural equation modeling (SEM), should gain some relevant insights, but researchers using regression-based approaches should benefit the most from this chapter.

Our targeted audience is applied higher education researchers, not methodologists. Advanced methodologists who specialize in econometrics or HLM will typically not be confused or influenced by differences in model presentation and terminology or by the emphasis assigned to the various benefits of multilevel models. In contrast, many applied researchers will be, and consequently, we have designed this chapter with them in mind.

[1] To our knowledge, only one previous paper has spent considerable space directly comparing statistical models from these two specific traditions (HLM and econometrics), and this paper was prepared for a conference and not disseminated into the education research community (Chaplin, 2003).

Generalizing the Challenge

Before delving into specifics, some generalization of the phenomenon under study would be helpful. The challenges faced by higher education researchers when employing multilevel data are not unique. In almost all methodological areas, we must ask: From which areas of the academy should higher education researchers borrow methods, and how do we effectively communicate with each other if we choose methods from multiple areas?

No large body of methods has been intentionally designed for the particular context and questions faced in higher education research. Consequently, we must borrow from elsewhere. Such borrowing would be simplest if we restricted ourselves to methods from one other academic field. Higher education programs could easily funnel all of their students into the same methodological courses from that field which would simplify advising. Communication at higher education conferences or within higher education journals would be seamless as all researchers would employ similar terminology and methods. But this simplicity would come at a substantial cost. Higher education researchers perform studies that are similar in nature to those conducted in sociology, psychology, economics, political science, anthropology, geography, educational psychology, and other areas of study. Each of these fields has developed methods that are appropriate for the context and research questions of interest in their disciplines. Whereas the methods of any one field would be appropriate for some higher education researchers, they would be inappropriate and of little help to others.

Thus, it appears we will produce greater insights if we borrow our methods from multiple fields. However, the nature of academe presents challenges to doing so. Academic disciplines have become silos across which fairly little communication occurs (Becher & Trowler, 2001). Each field invests heavily in the development of their own methods, but little effort is expended on contrasting and reconciling the methods developed within these fields. Consequently, higher education researchers are provided with little guidance when seeking to identify the methodological approach that would best complement their research agenda.[2] Efforts to compare studies that use methods from different disciplines are also challenging because few authors have sought to reconcile the varying terminology and model presentations that make similar statistical approaches across fields appear quite different. Also, past work has typically not provided guidance for researchers seeking to simultaneously utilize the insights of multiple statistical traditions.

This chapter seeks to address these challenges for a small part of the methodological world: statistical models appropriate for multilevel data. To help researchers make informed choices when selecting among different statistical traditions, we will clearly explain the benefits of each approach. We will also reconcile varying terminology and model presentation so researchers trained in one tradition can utilize their knowledge when examining research using alternative statistical traditions. In addition, we seek to aid those who wish to conduct research using a methodological approach that is not located solely within one methodological silo. Higher education researchers have the ability to utilize the primary insights of multiple traditions, and we will provide advice on how to do so.

[2] Of course, knowledge is only one of the challenges faced by higher education researchers when choosing the methodological approach in which to specialize. Students within higher education programs often face difficulty gaining access to graduate methodological classes in particular fields, and some programs, such as economics, teach in a manner that requires high levels of prior mathematical training. For these and several other reasons, higher education programs have historically funneled their students into methodology courses taught within educational psychology.

The Problems and Opportunities Presented by Multilevel Data

Before we compare particular traditions, we need to provide a framework within which comparisons can be made. The most helpful framework would clearly articulate the primary benefits associated with the advanced analysis of multilevel data. Most textbooks outline numerous benefits, but these lists vary substantially across books and often include a number of abstract benefits. This may be one of the reasons that many higher education researchers have not clearly articulated their motivation for using a multilevel model (Cheslock & Rios-Aguilar, 2008). Furthermore, most researchers are not strategically utilizing multilevel data to advance a core objective of their study. Multilevel models are employed, but they are not employed strategically. Perhaps this phenomenon explains the following statement by Smart: "the results obtained thus far from the use of HLM have not suggested any dramatically different conclusions from those based on the use of more conventional analytical procedures" (2005, p. 466).

We believe a conceptual framework that promotes a deep understanding of the potential benefits of multilevel analysis will encourage more strategic utilization of multilevel methods. We present here our list of the most important benefits. This list was developed with the applied use of multilevel models in mind, so the benefits connect with the potential objectives of a research study. Our list and our subsequent discussion are focused on simplicity. The core benefits of multilevel analysis are simple, but they are often presented alongside numerous technical complexities. This presentation style leads to partial comprehension of the core benefits, and until those benefits are deeply understood, the researcher runs the risk of employing advanced models for trivial ends. We hope our simpler approach helps promote that deep understanding.

Our list is designed to answer the basic question: Why should researchers use advanced models when using multilevel data in a regression framework? We believe there are four primary advantages to a regression-based multilevel model. The importance of each individual benefit depends upon the particulars of the study in question, which means the main justification for using a multilevel model will vary across studies. The four primary benefits of such a model are:

1. Improved estimation of standard errors.
2. The opportunity to estimate level-1 coefficients solely using within-group variation, solely using between-group variation, or using a combination of the two.
3. Examination of whether and to what extent key level-1 coefficients vary across groups and group characteristics.
4. The estimation of level-1 coefficients for a specific group even if the number of observations for that group is relatively small.

Improved estimation of standard errors

When higher education researchers who use HLM justify their methodological choices they typically note the first benefit (Cheslock & Rios-Aguilar, 2008). This benefit occurs because the standard error estimates from an ordinary least squares (OLS) regression are only correct when the error terms across observations are uncorrelated. But such an assumption is unlikely to hold in studies using multilevel data. Two individuals in the same group are likely to share traits that lead to membership in that group, and they will share numerous experiences as participants of that group. If these traits and experiences are important determinants of the outcome under study, and if one's regression model does not fully control for them, a portion of these traits and experiences will be included in the error term. As a result, the error terms of individuals in the same group will be correlated.

The most common multilevel structure within higher education research is students nested within higher education institutions. The aforementioned benefits of employing a multilevel approach are very relevant in this case. Past research has clearly demonstrated that attendance patterns across institutions are stratified in terms of academic preparation, socioeconomic status, and other traits that influence most outcomes. Furthermore, higher education institutions vary in culture, educational opportunities, and other aspects that would also impact most educational outcomes. However, researchers will never have detailed enough information to sufficiently control for these student and institutional traits, so error terms across students at the same institution will tend to be correlated.

Multilevel models do not require an assumption of uncorrelated error terms, so this issue is remedied. But one could also relax this assumption using a much simpler approach. Robust standard errors that allow for clustering at the group level can be estimated for an OLS regression, and the resulting standard errors will be correct even in the presence of correlated error terms (Huber, 1967; Rogers, 1993; White, 1980). While a multilevel model may still be more efficient in that it minimizes variance, this benefit may be relatively minor compared to the other three benefits mentioned above. In other words, when students are nested within institutions benefits two through 4 will often provide much more compelling justifications for employing a multilevel model.

The opportunity to estimate level-1 coefficients solely using within-group variation, solely using between-group variation, or using a combination of the two.

When estimating coefficients for explanatory variables measuring characteristics of the individual observations that are nested within groups (i.e., level-1 explanatory variables), a pooled OLS regression uses a very different combination of within-group and between-group variation than does a multilevel model.[3] To understand how pooled OLS and multilevel models differ, one must first be able to answer the following question: What does it mean to estimate a regression solely using within-group or between-group variation? Consider a study that examines how the parental income of a student predicts the amount of grants that student receives from the higher education institution she or he attends. Such a study uses data containing students from a number of different higher education institutions.

Regressions using only within-group variation would only compare institutional grants across students who attend the same institution but who have different parental income levels. If higher education institutions primarily distribute grants based on financial need, we would expect lower-income students to receive more institutional grant dollars than upper-income students who attend the same institution. If true, analysis of within-group variation would reveal a strong negative relationship between parental income and institutional grants.

Regressions that solely rely upon between-group variation would ignore differences among students at the same school and focus instead on variations among students attending different schools. This regression would focus on how the average parental income of an institution's students covaries with the average level of institutional grants received by these students. In other words, it would utilize individual-level data aggregated at the group level. Such analysis may not reveal a strong negative relationship because higher-income students attend very different schools than lower-income students, and these differences across schools

[3] Pooled OLS simply means an OLS regression that examines a data set that combines data from a number of different groups.

may influence grant levels. For example, schools that disproportionately enroll higher-income students charge higher tuition, which will increase the need for institutional grants. Furthermore, these schools are often relatively wealthy, which will make it easier for them to forego the revenue required to offer more generous financial aid packages. Consequently, the average grant offered by schools enrolling wealthier students may be similar or even higher than that of schools enrolling lower-income students, even if wealthy students at each school receive less institutional grant dollars than their lower-income colleagues at the same institution.

Figure 1 illustrates how between-institution variation can lead to different results than reliance upon within-institution variation. Regression lines HT, MT, and LT represent separate regressions for a high-tuition, medium-tuition, and low-tuition institutions. Each of these regressions has the same negative slope, but the level of each regression line rises with the tuition level of the institution. These regression lines demonstrate a strong negative relationship between parental income and institutional grants when within-institution variation is isolated. Regression line BI represents a regression based on the average parental income and institutional grants for each institution. Line BI is positive because the high-tuition school provides more grant aid and enrolls students with higher parental incomes than the low-tuition school. So, while the regression lines examining within-institution variation reveals a strong negative relationship, the regression line examining between-institution variation indicates a positive one.

A pooled OLS regression will produce neither the within-group nor the between-group estimate but will instead produce a weighted combination of the two. Raudenbush and Bryk (2002) present the following equation, which details the specific weights assigned to each:

(1)

where β_t represents the pooled OLS coefficient, β_b represents the between-group estimate, β_w represents the within-group estimate, and η represents the ratio of the between group sum of squares on the independent variable (X) to the total sum of squares on X (p. 137). This equation reveals that the pooled OLS regression estimate will fall between the between-group and within-group estimates and will be closer to the between-group estimate when most of the variation in X occurs between groups and closer to the within-group estimate when most of the variation in X occurs within groups. For the institution grant/parental income example, the pooled OLS estimate will depend upon the variation in parental income. If parental income levels vary substantially across students at the same institution, then β_t will be relatively close to β_w. If parental income levels are similar across students at the same institution but vary substantially across institutions then β_t will move towards β_b.

In contrast, a multilevel model will use a very different combination of between-group and within-group variation with the relative contributions depending upon the particular model estimated. As we shall elaborate upon later, two basic multilevel model types are typically estimated. The first type incorporates the fixed-effects model within econometrics and the group-mean centered model within HLM, while the second type incorporates the random-effects model within econometrics and HLM models that include grand-mean or no centering. The first model type simply produces the within-group estimate, while the second produces another weighted combination of the within- and between-group estimates. The weighted combination used in a random-effects/grand-mean centered model will always produce a coefficient that falls somewhere between the pooled OLS and the within-group estimates. This coefficient will be similar to the pooled OLS estimate when the unexplained variation within groups is relatively large and will be similar to the within-group estimate when the unexplained variation across groups is relatively large. We will explain this point more thoroughly later in the chapter.

To summarize, the relationship between the dependent variable and a level-1 independent variable can be described in a number of ways when one uses a multilevel data set. The most commonly used descriptions include the:

- Between-group estimator
- Pooled OLS estimator
- Random effects/Grand-mean centered HLM
- Within-group estimator (Fixed-effects/Group-mean centered HLM)

This list reflects the ordering of the magnitude produced by each estimator. The between-group and within-group estimators will produce the most extreme values, while the pooled OLS and random-effects/grand-mean centered estimators will produce results within those two extremes. In some studies, the between-groups estimator will possess the most positive value and the within-group estimator will possess the most negative value. In other studies, the opposite will occur.

Which estimator should be employed? When you are primarily interested in the individual-level (level-1) independent variables rather than the group-level (level-2) variables, the within-group estimator is often of special interest. Why? Consider a research study that is seeking to estimate the causal effect of a level-1 independent variable. If the regression analysis does not properly control for group-level determinants of the outcome that are correlated with the individual-level variable of interest, then the estimated coefficient for that individual-level variable will be biased. The estimated coefficient will contain the true causal effect plus a portion of the effect of the group-level determinant.

The within-group estimator addresses this problem by implicitly controlling for all group-level determinants so the effects of these determinants are not mistakenly assigned to individual-level explanatory variables. It is easy to understand intuitively why implicit controls are present: The within-group estimator solely compares individuals who have the same values for all group-level variables. Such comparisons are of great interest when we cannot effectively measure many important group-level determinants, which prevent direct statistical controls for these determinants. Of course, the within-estimator is not a panacea. As our section describing examples from the literature will reveal, one still needs to employ substantial assumptions to interpret the within-group estimator as producing a causal result. Furthermore, "going within" can create new problems, discussed in more detail below. The within-group estimator, however, can be a helpful part of an array of tools used to measure causal relationships. This estimator will usually get us closer to the true causal effect, and when the primary issue is missing controls for group-level determinants, this estimator can be extremely valuable.

The range of estimators that can be employed alongside multilevel data is also helpful for studies that do not seek to describe a causal relationship. Consider again an examination of how institutional grants vary by a student's parental income. We may wish to simply understand how these grants are allocated across students with various income levels and not really care if parental income levels are (or are not) the actual cause of the observed differences in grant receipt. We may be quite interested in both the within-institution and between-institution patterns because this information would provide unique insights into how financial aid is distributed across students and institutions.

Another example would be a study of racial/ethnic gaps in graduation rates. If we find lower *overall* graduation rates for underrepresented minorities, we could increase our understanding of this relationship by examining both the between-institution and within-institution effects in isolation. Some questions that could be answered by this inquiry include: Do under-represented minorities disproportionately enroll at institutions with lower graduation rates for all students? Do they have lower graduation rates than fellow students who are enrolled at the same institutions? Additional interesting comparisons can be made by including statistical controls for other individual-level traits. These examples demonstrate how multilevel data can help a researcher provide a much richer description of the relationship(s) under study.

Cronbach (1976) provides further illumination of this issue by noting that different types of analyses often relate to fundamentally different research questions that often produce very different results. As demonstrated by the previous paragraph, the within-group and between-group estimators typically relate to clearly defined questions regarding level-1 variables. The pooled OLS and the random effects/grand-mean centered HLM models utilize combinations of within-group and between-group variation and are consequently much harder to connect to research questions. Consequently, Cronbach (1976) dissuades researchers from utilizing such composite variables because they are "rarely of substantial interest" (p. 228).

Examination of whether and to what extent key level-1 coefficients vary across groups and group characteristics.

The previous section focused on estimating an overall relationship between variables that is consistent in magnitude across groups. But some authors claim that such an overall relationship is of limited use in educational research. Labaree (2003) notes that educational "researchers are unlikely to establish valid and reliable causal claims that can be extended beyond the particulars of time, place, and person" (p. 14). Berliner (2002) highlights the "power of contexts" within education, which results in relationships varying substantially across settings (p. 19).

Multilevel data allow a researcher to investigate across settings, especially for individual-level explanatory variables. The variation across groups in level-1 relationships can be estimated so we can understand the degree to which findings vary across contexts. Consider again research that examines how institutional grants vary by the financial resources of a student/their family. Would we expect this relationship to differ across higher education institutions? Probably, because institutional financial aid is offered for a number of reasons only one of which is income related. In some instances, schools provide aid to increase access for low-income students, which results in a strong negative relationship between institutional grants and income. In other cases, schools use aid to attract students with particular traits, such as strong academic preparation. In this case if upper-income students applying to a particular school are disproportionately well prepared the result would be a positive aid/income relationship. If higher education institutions differ substantially in their motivation for offering institutional aid, the relationship between aid receipt and parental income will vary substantially by school. Multilevel data and methods will allow a researcher to measure the magnitude of that variation.

In most studies, researchers will want to better understand the variation across groups in level-1 coefficients. While it is helpful to know that this variation exists, knowledge about the relationship between such variation and group-level characteristics typically provides even greater insights. Which institutions concentrate their institutional aid on lower-income students? Which focus more of their aid on upper-income students? A researcher can answer these questions by using multilevel data to identify institutional characteristics that are correlated with patterns of aid distribution.

Often, one's conceptual framework can identify group-level characteristics that predict level-1 relationships. This point is illuminated by a third reason for a higher education institution to offer aid. Colleges and universities that are not at enrollment capacity can theoretically improve their economic situation by offering institutional aid to students who are on the margin of enrolling (Breneman, 1994; McPherson & Shapiro, 1998). These students are more likely to attend if they are provided institutional aid, so the aid offer increases net tuition revenue by the sticker price minus institutional aid. The increase in cost is relatively small as the student will be filling an empty seat because the institution is not at enrollment capacity. This logic closely follows microeconomic theory and is often used as a framework for explaining the allocation of institutional aid. This framework has one clear prediction: Variables that measure whether an institution is below enrollment capacity will have a strong influence on the relationship between institutional aid and parental income.[4]

The estimation of level-1 coefficients for a specific group even if the number of observations for that group is relatively small.

If a researcher has a large number of observations for each group, the estimation of level-1 coefficients for each group is fairly easy. But in most multilevel data sets researchers have a relatively small number of observations for some groups.[5] How do we produce valid estimates for each group in this scenario? Gelman and Hill (2007) helpfully explain that a multilevel model can produce estimates for each group that are a weighted average of the "no pooling" and "complete pooling" estimates. The "no pooling" estimate for a particular group is produced by

---

[4] This prediction clearly assumes that parental income is likely to be correlated with the propensity of students to be on the margin of attendance. Because price responsiveness and the number of submitted applications vary by a student's parental income, such an assumption is realistic.

[5] To understand what we mean by a "small number," consider the sample size requirements for statistical analysis in general. Now apply those requirements to each group in the analysis.

analyzing only those observations in the group. The “complete pooling” estimate is produced by simultaneous analysis of all observations in the data set.

For groups with small sample sizes, the multilevel estimate will be closer to the “complete pooling” estimate. We have little confidence in the “no pooling” results for these groups because the results are based on a small amount of information. When the number of observations grows, however, a group’s estimate can move closer to the “no pooling” estimate.

Producing results for each group allows us to reap some of the benefits mentioned earlier. Doing so helps us examine the variation across groups as well as how this variation relates to group characteristics. But in some cases researchers will be interested primarily in estimates for specific groups. For example, a researcher conducting a mixed-methods study may seek to identify specific higher education institutions that concentrate their aid on low-income students as well as specific institutions that concentrate their aid on upper-income students. The identification of such institutions can be achieved through the analysis of a multilevel data set that produces estimates of level-1 coefficients for each institution.

Generalizing the Potential Benefits of Multilevel Data

Shadish, Cook, and Campbell (2002) present a helpful validity typology that sheds insight into the potential benefits of multilevel data. They outline four types of validity. Statistical conclusion validity occurs when statistics are appropriately used to infer whether the independent and dependent variables covary. Internal validity refers to whether the covariation reflects a causal relationship. Construct validity involves making inferences from the sampling particulars of a study to the higher-order constructs they represent. External validity refers to whether the cause-effect relationship holds over different persons, settings, treatment variables, and measurement variables.

Each of the four benefits discussed earlier are related to this validity typology. [6] The first benefit of multilevel data, improved estimation of standard errors, improves statistical conclusion validity. The opportunity to estimate level-1 coefficients solely using within variation, which is the second benefit, can improve internal validity in many contexts. The third benefit examines whether results consistently hold across groups so this benefit closely relates to external validity. If we know that the results vary little across individual settings, then we can be more confident when using overall results to help shape policy for a particular setting.

Later in this chapter, we will highlight how the econometric approach to multilevel data focuses on the second benefit, while the HLM approach concentrates on the third and fourth benefits. In other words, econometrics fixates on using multilevel data to improve internal validity while HLM is heavily concerned with issues relating to external validity. The trade-off between internal and external validity is the most major and most discussed tension within the research design literature (Shadish, Cook, & Campbell, 2002, p. 96). Researchers regularly face choices where one option improves internal validity at the expense of external validity and vice versa. A theme of this chapter is that almost all previous research solely utilized either the econometric or the HLM tradition when examining multilevel data. Consequently, previous work has chosen to use multilevel data to improve internal validity or external validity but rarely both. Such tradeoffs are not required for the analysis of multilevel data, however, and we will promote an approach to research that simultaneously advances both forms of validity.

---

[6] The fourth benefit connects to multiple elements of Shadish, Cook, and Campbell's validity typology. Because the upcoming discussion focuses on the tradeoffs between validity types, we will focus solely on the first three benefits.

This discussion has emphasized studies that examine research questions of a causal nature. Internal validity is only relevant to such questions. But multilevel data can also advance non-causal studies in which external and statistical conclusion validity are still of concern. And non-causal studies can often be advanced by estimating relationships separately for within-group variation, between-group variation, and total variation. Each type of variation produces answers to distinct questions. So, the general points made within this section apply to non-causal studies as well.

Overview of Econometric & HLM Approach to Multilevel Data

One cannot contrast the econometric and HLM-based approaches to multilevel data until one develops a basic understanding of the key elements of both traditions. Therefore, both are introduced here. We do not cover the full range of issues for each approach; we focus on basic topics that are relevant for the purposes discussed in our introduction. We place a heavy priority on simplicity and accessibility, so that a strong core understanding of each tradition can be easily attained. Previous handbook chapters discuss each tradition in more depth, and we hope our work will help readers develop a deeper understanding of the more advanced points discussed there (Ethington, 1997; Zhang, 2010).

This chapter divides multilevel methods into two categories: HLM-based multilevel models and econometric-based multilevel models. Some justification for that decision is required. In our review of the higher education literature we were able to easily assign every reviewed article that used advanced multilevel models to one of these two groups (Cheslock & Rios-Aguilar, 2008). For most journal articles, one can quickly identify the tradition within which the author has been trained because of differences in terminology, model presentation, and motivation. While more general approaches to multilevel data exist (e.g., Gelman & Hill, 2007), the econometric and HLM approaches dominate current practice.

The names used for each category also deserve some mention. We use HLM for the first category to reflect the often-used terminology within educational research, which demonstrates the influence of Bryk and Raudenbush (1992) and Raudenbush and Bryk (2002). These textbooks, both entitled "Hierarchical Linear Models," have dominated the literature and the acronym HLM has come to represent multilevel models among educational researchers. Of course, many textbook authors who present models similar to those of Bryk and Raudenbush do not use the term hierarchical linear model or HLM (e.g., Heck & Thomas, 2009). Often, they simply use the term multilevel model, and sometimes define this term in a very specific way that does not include the basic econometric models. In this chapter, we use the term multilevel model in a more general way: to describe any model that was designed for use on multilevel data. We experimented with a more general name that is consistent with prior usage, but "models designed for multilevel data" proved clumsy.

When describing econometrics and HLM, we focus on the issues emphasized within each one. Overall, both traditions cover similar ground, but the emphasis placed on particular elements differs dramatically. For example, most econometrics textbooks spend little time discussing how researchers can examine heterogeneity across groups using multilevel data, but a lack of emphasis does not mean that this approach is completely absent. Some prominent journal articles in economics, such as Rivkin et al. (2005), examines how level-1 coefficients vary across groups. Econometrics also contains a class of models, called random coefficient models that are designed to examine heterogeneity across groups. These models, however, are rarely covered in econometrics textbooks, and most economic journal articles use multilevel structures to reduce bias rather than to examine heterogeneity. Researchers who are not methodologists will primarily grasp the benefits that are featured within a tradition, so our overview will discuss what is emphasized in each.

Introduction to HLM

When higher education researchers justify their decision to employ HLM they cite Raudenbush and Bryk (2002) more than any other text (Cheslock & Rios-Aguilar, 2008). Given the influence of this book, we will focus heavily on its presentation of HLM models. We will pay special attention to terminology, model presentation, featured data set types, and the benefits by which the models are motivated.

Raudenbush & Bryk (2002) feature examples based on cluster data sets where students are nested within schools. These data sets are combined with models that are presented as a series of three equations. The first equation represents a student-level regression that produces estimates for separate intercepts (β_{0j}) and slope coefficients (β_{1j}) for each school. In the next two equations, the school-level intercepts and slope coefficients are then regressed on school-level variables (W_j).

Level 1 model (Students) $$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}, \quad (2)$$

Level 2 model (Schools) $$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}, \quad (3)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j} \quad (4)$$

Students are denoted by the subscript *i* while schools are denoted by *j*. Raudenbush & Bryk (2002) call this "full" model the "intercepts- and slopes-as-outcomes model" and also present a number of models which restrict some parameters to be constant across groups.

The benefits featured within this HLM-tradition lie in the estimation of β_{0j} and β_{1j}. These parameters describe the student-level results for school *j*, and the estimation of these parameters can provide insights into how student-level results vary across schools or school-level characteristics. (In other words, these parameters can help establish the third benefit of multilevel data presented in our framework above.) The variation in β_{0j} and β_{1j} describes the differences across schools, and the cross-level effects (γ_{01} and γ_{11}) describe the variation across school-level characteristics. Raudenbush and Bryk (2002) describe how the covariation of β_{0j} and β_{1j} can also be helpful in certain settings.[7]

The primary examples used in Raudenbush and Bryk (2002) come from a study where the dependent variable is mathematics achievement and the independent variable is socio-economic status. Such a study is greatly augmented by analysis of how student-level relationships vary across schools. If high-SES students perform better than low-SES students then a school has a strong positive relationship ($\beta_{1j} > 0$). If no relationship exists ($\beta_{1j} = 0$), and students of different SES-levels achieve similar levels of academic success. The variation in β_{1j} across schools describes how the SES-achievement relationship fluctuates across schools. If W_j represents a school's financial resources, then γ_{11} permits analysis of whether additional funding is associated with a school's SES-achievement relationship. Thus, the covariation of β_{0j} and β_{1j} indicates whether a school's average level of achievement is related to its SES-achievement relationship.

[7] For these points, our focus on simplicity is obscuring some important technical details. The HLM framework does not include direct analysis of the variation in β_{0j} and β_{1j} using equations (1)-(3). Instead, this variation is measured by estimating a different set of equations which do not contain W_j in equations (2) and (3). Raudenbush and Bryk (2002, p. 77-80) call this restricted version the "random coefficients" model. Using this model, one can examine the variation in β_{0j} by estimating the variance of u_{0j}. The variation in β_{1j} can be examined by estimating the variance of u_{1j}.

Raudenbush & Bryk (2002) clearly emphasize how multilevel data allow a researcher to examine whether group-specific level-1 parameters vary across groups or group characteristics. Their presentation, however, also notes the second benefit of multilevel data featured in our framework. They also discuss how multilevel data allows one to estimate overall level-1 (i.e., student level) parameters solely using within-group variation, solely using between-group variation, or using a combination of the two. These considerations primarily arise during discussions of centering.

Centering involves the choosing of alternative locations of the independent variable. Researchers typically center at the grand-mean $\left(\overline{X}\right)$ or at the group-mean $\left(\overline{X}_j\right)$. To center at the grand mean, one simply subtracts the overall mean from each of the predictors.

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}) + e_{ij} \qquad (5)$$

To center at the group mean, one subtracts the group mean from each of the predictors.

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_j) + e_{ij} \quad (6)$$

Centering at the grand mean has minor implications for researchers because it just changes the interpretation of the intercept. However, group-mean centering fundamentally changes the analysis, as this form of centering causes β_{1j} to be estimated solely using within-group variation.[8] Without group-mean centering, β_{1j} is computed using a weighted combination of within-group and between-group variation.

Raudenbush and Bryk (2002) explain the implications of centering and clearly recommend the use of group-mean centering for studies focused on overall level-1 coefficients (pp. 135-141 & 261-263). Their recommendation is accompanied by an explanation of positive

[8] Within HLM, the within-group estimator can also be obtained by including group means of each level-1 variable in each level-2 equation. In other words, we would add $\overline{X}_j$ as an independent variable to equations (3) and (4). Raudenbush and Bryk (2002) also discuss this approach.

attributes of the within-group estimator. These discussions, however, are brief and occur in less prominent parts of the text. For example, the major explanation of these issues (on pages 135-141) occurs within the "special topics" portion of chapter 5. Raudenbush and Bryk (2002) reserve the prominent parts of their text for analysis of how results vary across groups or group-level characteristics. In other words, they feature the third benefit of multilevel data documented in our conceptual framework rather than the second.

While Raudenbush and Bryk (2002) primarily consider data sets where individuals are nested within organizations, they devote chapter 6 to data sets that contain multiple observations for the same unit. Their primary example features a dataset containing multiple observations over time for each student. These data are multilevel because the individual observations are nested within the student.

Raudenbush & Bryk (2002) recommend a very different model for these data structures. This model's level-1 equation estimates the growth rate for the outcome of interest.

$$Y_{ti} = \pi_{0i} + \pi_{1i} a_{ti} + \pi_{2i} a_{ti}^2 + \cdots + \pi_{Pi} a_{ti}^P + e_{ti} \qquad (7)$$

The level-2 equations then examine these growth parameters.

$$\pi_{pi} = \beta_{p0} + \beta_{p1} X_{1i} + r_{pi} \qquad (8)$$

In these equations, a_{ti} represents the age of student *i* at time *t*.

This growth model fundamentally differs from the earlier version presented in equations (2)-(4). The inclusion of numerous age variables in the growth model allows the researcher to examine how student-level variables (X_{1i}) alter the growth trajectory of the outcome. In contrast, the previous model examined how explanatory variables influenced the outcome at a point in time. The growth model also differs in that it does not include any individual-level variables (X_{ti}). For the example of individual observations nested within students, X_{ti} would contain a variable measuring some aspect of the student at time *t*. The above equations only incorporate X_{1i}, which measures some aspect of the student that is consistent throughout the period under study.

Raudenbush and Bryk (2002) briefly note that time-varying student variables (X_{ti}) could be incorporated into analysis of time-related multilevel data (p. 183). They discuss how the inclusion of these covariates allows the researchers to estimate the results solely using within-group variation by employing group-mean centering. For a multilevel data set containing multiple observations over time for each student, within-group analysis would only examine variation over time for the same student.

Introduction to Econometrics

Econometricians refer to their multilevel models as panel models, probably because they focus primarily on panel data sets that contain multiple observations over time for the same unit. When educational examples are provided in textbooks, schools or states are typically the unit of analysis, and the explanatory variables of interest are time-varying. The structure of the examples reflects a heavy emphasis on the second benefit of multilevel data featured in our conceptual framework. In other words, econometrics stresses how multilevel data provide an opportunity to estimate level-1 coefficients solely using within-group variation, solely using between-group variation, or using a combination of the two. More specifically, it concentrates on how the within-group estimator can advance causal inference in many studies.

Unlike HLM models, econometric multilevel models are usually presented within a single equation.

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 W_j + \alpha_j + \mu_{ij} \qquad (9)$$

As in the first HLM models presented, *i* represents the individual observations and *j* represents the groups in which these observations are nested. This equation allows the intercept to vary across groups by including α_j, but does not allow the slope coefficient to vary across groups. Econometricians have developed a model that allows the slope coefficients to vary, but this model, which is called a random coefficients model, receives very little attention in most econometric textbooks and the literature.

Although a large number of econometric models can be applied to multilevel data, two models are typically featured. The fixed-effects model solely uses within-group variation to estimate results, while the random-effects model uses a combination of within-group and between-group variation. Like Raudenbush and Bryk (2002), econometrics books generally recommend the use of the within-group estimator (i.e., the fixed-effects model) when examining explanatory variables at level one. Formally, they state that one should only use the random-effects model whenever $\mathrm{Cov}(X_{ij}, \alpha_j) = 0$. Wooldridge (2009) notes that this scenario "should be considered the exception rather than the rule" (p. 493). As Zhang (2010) explains, a formal test exists for choosing between the fixed-effects and random-effects model.

To illuminate the points just made, we will use a research example featured in Zhang (2010), a recent Handbook chapter that examined econometric models in greater detail. This example considers the impact of non-resident tuition on non-resident enrollment using institution-level data that contain observations for multiple years for each institution. In this case, X_{ij} represents non-resident tuition and α_j represents other factors that help explain an institution's non-resident enrollment. Zhang (2010) notes that institutional prestige, geographical location, academic programs, and other factors would be included in α_j. One could theoretically remove these factors from α_j by including them as control variables in the regression, but the researcher is unlikely to have sound measures of these constructs that fully capture their influence on non-resident enrollment. As a result, at least part of their influence will remain in α_j. Within the econometric framework the key question is whether $Cov(X_{ij}, \alpha_j) = 0$. If the answer is "yes", the random effects estimator is preferred, because it is more efficient than the fixed-effects estimator. As Zhang (2010) notes, however, the answer is likely to be "no". Institutions that are prestigious and located in attractive geographical locations are likely to charge different non-resident tuition levels than institutions without these traits. Consequently, the fixed-effects estimator is recommended because the use of within-group variation will implicitly control for those elements of prestige and geography that are constant over the period of study.

The fixed effects estimator isolates within-group variation by group-mean centering each element of equation (9).[9] This transformation eliminates α_j because the group-mean of α_j is α_j. The same process eliminates any group-level explanatory variables (W_j). What remains results in the following equation:

$$(Y_{ij} - \overline{Y}_j) = \beta_1(X_{ij} - \overline{X}_j) + (\mu_{ij} - \overline{\mu}_j) \qquad (10)$$

This model no longer requires X_{ij} and α_j to be uncorrelated because α_j has been removed as a source of variation.

The random-effects model is typically represented by a similar yet more complex model:

$$(Y_{ij} - \lambda\overline{Y}_j) = \beta_0(1-\lambda) + \beta_1(X_{ij} - \lambda\overline{X}_j) + \beta_2(1-\lambda)W_j + (v_{ij} - \lambda\overline{v}_j) \quad (11)$$

[9] One can also think of the fixed-effects model as adding a dummy variable for each group. From this perspective, we are employing a fixed-effects model whenever we add dummy variables for any classification in which each observation is in one, but no more than one, category. Older versions of the Carnegie classification would be a good example from higher education research. For a multilevel data set with a large number of groups, the addition of dummy variables for each group creates computational challenges, which is why equation (10) is used instead.

In this equation, $v_{ij} = (\alpha_j + u_{ij})$ and λ represents a complicated formula. When the unexplained variation within groups (u_{ij}) is relatively large, then λ will be close to 0. When the unexplained variation across groups (α_j) is relatively large, then λ will be close to 1. This technical detail helps one interpret the results of the random-effects model. As we noted in our conceptual framework, the coefficient from the random effects model will fall somewhere between the coefficients from the pooled OLS and fixed-effects models. This relationship becomes clearer when one realizes that equation (11) becomes identical to the pooled OLS model when λ equals 0 and identical to the fixed-effects model (equation 10) when λ equals 1.

Although the fixed-effects model is recommended in most circumstances, econometric texts also highlight several limitations of this model. One cannot examine group-level explanatory variables (W_j) using the fixed effects model because only within-group variation is used. Examination of individual-level explanatory variables (X_{it}) can sometimes be complicated by measurement error (Ashenfelter & Kreuger, 1994; Griliches, 1979). In some settings, measurement error can comprise a major share of the within-group variation in X_{it}, which can lead to substantial coefficient bias.

Although most econometric texts focus on panel data sets that contain multiple observations over time for the same unit of analysis, the fixed-effects and random-effects models can also be applied to other types of multilevel data. Econometric textbooks recognize these other data types, but pay them relatively little attention. For example, Wooldridge (2009) only devotes two pages to the subject.[10] Matched pair samples are typically featured, and special attention is placed on data sets where siblings are nested within families. Within-group analysis

[10] The coverage of other multilevel data structures is slightly larger in more advanced econometric textbooks, such as Wooldridge (2002). These books, however, still place a much larger emphasis on panel data. Individual journal articles, such as Moulton (1990) and Wooldridge (2003), focus on cluster samples in much more depth. These articles as well as advanced econometrics textbooks assume a strong existing knowledge base in econometrics and mathematics, so their usefulness will vary considerably across researchers.

of these data allows the researcher to control for unobserved “family effects”. Cluster samples, where each observation belongs to a well-defined group, do receive some attention for their ability to control for unobserved “cluster effects.” Most multilevel studies within higher education use cluster samples where students or faculty are nested within higher education institutions or units (e.g., students) within an institution.

When discussing multilevel data, econometrics textbooks clearly concentrate on the second benefit featured in our conceptual framework. Some books do not even recognize the third and fourth benefit while other books only briefly note them. Wooldridge (2009) devotes two paragraphs to these two benefits, noting that we could study the distribution of α_j’s or that we could study estimates for particular groups, but he essentially concludes that “the sense in which the α_j can be estimated is generally weak” (p. 486). His pessimism may partially reflect his focus on data structures that typically contain a small number of observations per group. As noted earlier, econometrics does contain a random coefficients model that allows β_1 to vary across groups, but this model rarely receives substantial coverage in texts (see Zhang, 2010, for an accessible introduction to this model).

Zhang (2010) covers several other relevant econometric models. The most prominent is the “difference-in-differences” (DD) model, which for panel data examines how changes in the outcome for the treatment group differs from the changes in the outcome for the control group during a period in which the treatment was instituted (Meyer, 1995). This estimator essentially examines how differences within groups over time differ across groups, so its core structure could be described as multilevel. In terms of causal inference, the DD estimator is similar to the within-group estimator in that it assumes that within-group differences in the explanatory variable are unrelated to within-group differences in the error term. In many ways, however, the multilevel structure of the DD model is fundamentally different than the other multilevel structures we discuss. For example, the DD model typically contains a very small number of groups. In some applications, just one treatment group and one control group are examined. This chapter primarily focuses on multilevel data structures containing large number of groups, so even though the DD model is an important part of econometrics, we do not cover it in depth here.

Comparing Econometric and HLM Models

The previous sections demonstrate core similarities between econometrics and HLM when employing multilevel data. Both traditions contain models that allow a researcher to reap the multiple benefits of multilevel data featured in our conceptual framework. Furthermore, the most often employed models within each tradition are very similar. Consider the random-intercepts model within HLM, which fixes the slope coefficient as follows:

Level 1 model $$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}, \qquad (12)$$

Level 2 model $$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}, \qquad (13)$$

$$\beta_{1j} = \gamma_{10} \qquad (14)$$

With group-mean centering, this model is extremely similar to the econometric fixed-effects model. With grand-mean or no centering, the econometric random-effects model is its counterpart. Cheslock and Rios-Aguilar (2008) found that most HLM papers in higher education journals employed some version of this random-intercept model and all econometric papers employed the fixed-effects or random-effects model.

HLM and econometric texts often provide similar advice. When seeking to describe the overall relationship between an explanatory variable(s) (X_{ij}) and an outcome (Y_{ij}), both traditions recommend the within-group estimator for most studies. They both suggest examining the variation in β_{1j} when one believes the level-1 relationships vary across groups, and each tradition recommends the use of cluster data, panel data, and other forms of multilevel data. The difference between HLM and econometrics lies in the emphasis placed upon particular pieces of advice. Most HLM texts pay relatively little attention to the virtues of the within-group estimator while most econometric texts devote even less attention to the benefits of studying heterogeneity across groups. Cluster data examples are featured within HLM while panel data examples are primarily used within econometrics.

The advanced methodologist may not be influenced by these differences in emphasis, but most applied researchers will gain a deeper appreciation of these benefits when illustrated by examples and insightful extended discussions. Once such an understanding is developed by researchers it may well alter the topics, designs, and models they employ. HLM-trained researchers will disproportionately employ cluster data in studies that seek to study heterogeneity across groups, while econometric-trained researchers will disproportionately use panel data to answer causal questions about the overall relationship between X_{ij} and Y_{ij}. When HLM-trained researchers employ panel data, they will disproportionately examine the specific model featured within their tradition. In other words, they will study the growth trajectory of Y_{ij} and only examine the influence of explanatory variables that do not vary over time (X_i). In contrast, econometric-trained scholars will tend to utilize panel data to study simple changes in Y_{ij} and the influence of time-varying explanatory variables (X_{ij}).

Mainstream higher education researchers are not doomed to this fate. A deep understanding of the benefits emphasized in both traditions would allow them to focus on the data structure and benefit most helpful for the particular study at hand. Whether HLM models or econometric models are employed is often not the major issue. These benefits can be realized within both traditions.

That said, some real differences do exist across traditions. Some differences relate to minor technical issues that will rarely change, in a significant way, the results of a study. For example, the random effects model within econometrics is usually estimated using generalized least squares, while the basic random-intercept model within HLM is usually estimated using both maximum likelihood [ML] and generalized least squares [GLS]. By default, the HLM software (version 6) estimates the variance-covariance components via ML, and the parameters are estimated using GLS (Raudenbush, Bryk, Cheong, Congdon & du Toit, 2004).

More substantial differences exist between the advanced models within HLM and econometrics, and these differences often reflect the emphasis within each tradition. One area of advanced work within econometrics focuses on how to continue to promote causal inference when the assumptions of the fixed-effects and random-effects models no longer hold. We present the basic structure of these models again, with the new version including a coefficient (δ) in front of α_j to highlight the effect of unobserved group-level factors on the outcome of interest.

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 W_j + \delta\alpha_j + \mu_{ij} \qquad (15)$$

The basic fixed-effects model assumes that β_1 does not vary across *i*, δ does not vary across *i*, and α_j is not a function of . These assumptions often do not hold, especially in studies employing panel data, and advanced econometric models add extensions that require fewer assumptions but increase the complexity of the estimation procedure (Chamberlain,1982, 1984; Mundlack, 1978a, 1978b).

Advanced HLM-based models are more likely than advanced econometrics to focus on estimating valid results for individual groups or estimating how group-specific results vary across groups or group characteristics. This goal is often complicated by data sets that contain a relatively small number of groups, a relatively small number of observations per group, relatively little variation within groups for some explanatory variables, or other limitations. Advanced techniques can help researchers produce more valid analysis of heterogeneity across groups when facing these sorts of data challenges (Gelman & Little, 1997; Kenny et al., 2002; Snijders, 2005).

These advanced models, however, are not the focus of this chapter. Instead, our interest lies in the most basic models, which are quite similar across traditions. These similarities are difficult to perceive due to differences in notation, model presentation, and terminology. The notation differences are the least substantial, but they likely cause initial confusion. The primary examples in an HLM book use cluster data where subscript *i* denotes the individual observations nested within group *j*. Econometric books use panel data examples where subscript *i* represents the unit within which individual observations over time are nested. To demonstrate the similarities across traditions, we presented the econometric models using the notation employed in HLM for cluster data.

These two traditions also differ substantially in the equations used to represent similar models. HLM models are presented in multiple equations, while the econometric version presents the model within one equation. Once one combines the multiple HLM equations into one equation, the similarities between HLM and econometrics become more transparent, but some researchers may not be able to undertake these transformations. The equations representing the within-group estimator differ in even more confusing ways. The econometric version is produced by subtracting group means from all variables in the model, while the HLM version group-mean centers the explanatory variables but not the dependent variable(s). The resulting equations appear to be fundamentally different, and very few applied researchers would realize that despite the differences in centering, they both only use within-group variation to estimate results.

Terminology differences are probably the most annoying, especially in the case of random-effects and fixed-effects. In econometrics, these terms originally referred to whether α_j was a random variable or a parameter to be estimated, but they are now primarily used to describe the basic models discussed in this chapter. Within HLM, the terms do not describe overall models, but they instead represent specific elements within models. Random effects are the error terms associated with the coefficient estimates. In contrast, fixed effects represent the non-random parts of these coefficients. When a researcher "fixes" level-1 coefficients, they are not allowing level-1 coefficients to vary across level-2 groups. The confusion created by the varying terminology is perhaps best represented by the following relationship: When a researcher "fixes" all level-1 slope coefficients in HLM without group-mean centering, they create a HLM model that is essentially identical to the random effects model in econometrics.

Examples from the Literature

We turn now to past articles that employed multilevel data in a manner that helped the authors achieve the core objectives of their study. We will organize our review of these articles around the four benefits of multilevel analysis discussed earlier. As before, we spend little time on the first benefit, improved standard error estimates, as this benefit is ubiquitous and easy to obtain. We focus instead on the latter three benefits and seek papers whose primary contribution to knowledge would not have occurred without advanced multilevel analysis. The first subsection concentrates on papers that reaped the second benefit featured earlier in this chapter, and the second subsection focuses on the third benefit. We close this section by discussing a paper that realizes the fourth benefit as well as the other three.

For each paper, we will discuss the research questions, the limited insights that standard analysis can provide into the answers to these questions, the deeper insights that multilevel analysis can provide, and how the authors employed multilevel models to reveal these deeper insights. Numerous details about each paper are omitted so that our discussion focuses on the basic intuition that explains how deeper insights are produced by multilevel analysis. For each paper, we try to discuss how these insights are vital given the goals of the study.

Papers realizing benefit #2: The opportunity to estimate level-1 coefficients solely using within-group variation, solely using between-group variation, or using a combination of the two.

Unsurprisingly, most econometric-based higher education papers followed the lead of econometric texts and focused on the second benefit of multilevel analysis.[11] These papers employed the within-group estimator to implicitly control for unobserved group-level variables. Panel data were used primarily and causal questions were mostly investigated. Of course, the within group estimator also aids analysis of other data structures and helps answer research questions that investigate descriptive or associational relationships. To demonstrate this point, we will examine one panel data example (Archibald & Feldman, 2006), one cluster data example (Goldhaber & Brewer, 1997), and one matched-group data example (Dale & Kreuger, 2002). While all three examples consider causal questions, our description of Gelman et al. (2007) in a later section will demonstrate how one can skillfully use between-group and within-group variation to describe non-causal relationships.

[11] There are some prominent econometric papers, however, that focus on our third benefit of multilevel data. For example, Rivkin et al. (2005) used a data set containing students nested within teachers to estimate the variation in student test scores across teachers.

Archibald and Feldman (2006) use panel data to examine how state laws and constitutional provisions that alter state taxing and spending policies influence state appropriations. These laws and provisions were motivated by a desire to restrain state spending, and such restraints could limit the amount of state dollars that flow to higher education institutions. Archibald and Feldman (2006) focus on two particular types of restraints: Tax and Expenditure Limitations (TELs) and Supermajority Requirements (SMRs). Since the late 1970s, 23 states adopted a TEL and 13 states added an SMR.

One could examine the impact of TELs and SMRs on state appropriation for higher education by using state-level data for a recent year. Such cross-sectional analysis would solely use between-state variation in the given year to identify the results. In other words, the average state appropriation effort for states with a TEL would be compared with the average effort for states without a TEL. Would such an analysis be valid? Archibald and Feldman (2006) clearly want to estimate a relationship that can be interpreted as a causal relationship between appropriations and TELs/SMRs. The implicit goal of their study is to understand the distinct role played by these laws and to predict the future impact on higher education if more states add these types of spending restraints. Can cross-sectional analysis of one year of data produce a causal result?

The answer will likely be “no”. States that instituted TELs and SMRs may disproportionately possess citizens who prefer a smaller state government, and these states would have spent less on higher education even if formal spending restraints were never instituted. In this scenario, the cross-sectional results will be biased downwards and could reveal a negative relationship even if there was no causal relationship. Alternatively, states with TELs and SMRs may have a history of extensive government spending, and these spending restraints may have been instituted to moderate this tendency. In this case, the cross-sectional results will be biased upwards and could reveal a positive relationship even if TELs and SMRs cause state appropriation effort to decline.

To address this problem, Archibald and Feldman (2006) use state-level panel data that span a forty-year period (1961-2001) and employ an econometric fixed-effects regression so that only within-group variation identifies the results. For this panel sample, the state is the group and the specific measurements for a year are the observations nested within the group. To use only within-group variation is to use only variation over time within states. So, the regression results are essentially produced by comparing higher education funding levels in states after they added a TEL or SMR with the funding levels in those states before the spending restraint was instituted. This comparison becomes more complex when you add controls for other determinants of appropriation effort and you adjust for timing differences across states in the adoption of TELs/SMRs. The core intuition, however, remains even after these complexities are added.

Statistical controls and timing considerations can be easily incorporated into a multilevel models framework. Within econometrics, one can add these considerations to the basic fixed effects model by estimating either of the following equations.

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 M_{ij} + \beta_2 W_j + \alpha_j + \mu_{ij} \qquad (16)$$

$$(Y_{ij} - \overline{Y_j}) = \beta_1 (X_{ij} - \overline{X_j}) + \beta_2 (M_{ij} - \overline{M_j}) + (\mu_{ij} - \overline{\mu_j}) \qquad (17)$$

For this model, subscript *j* represents the state and subscript *i* represents the particular year in which variables are measured for that state. This model assumes one primary explanatory variable (X_{ij}), one time-variant control variable (M_{ij}), and one time-invariant control variable (W_j).

In practice, Archibald and Feldman (2006) estimate an even more elaborate model because they also include time fixed-effects.

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 M_{ij} + \beta_2 W_j + \alpha_j + \gamma_i + \mu_{ij} \qquad (18)$$

The inclusion of time fixed-effects (γ_i) means that controls will be added for factors common to all states that impact the state appropriation effort in a given year. These controls cause the effort for a particular state in a particular year to be effectively measured relative to other states for that year. So, the regression results will be produced essentially by comparing *relative* state appropriation effort in states after they added a TEL or SMR with the *relative* efforts in those states before the spending restraint was instituted. Time fixed-effects are vital when the explanatory variable (X_{ij}) of interest is correlated with key unobserved factors that vary over time and are common to all states. Archibald and Feldman (2006) faced this scenario because, as states added TELs and SMRs over time, other forces that shape state higher education funding were also changing.

The estimation of equation (18) could, however, still produce a biased result. For example, states that added a TEL/SMR could experience other state-specific changes that impact state higher education funding and occur around the time that TELs/SMRs were instituted. If sufficient controls are not included to account for these state-specific changes, the estimated effect of TELs/SMRs will not represent the causal effect. These scenarios for bias, however, are much less plausible than the concerns that arise when employing cross-sectional analysis. All techniques for estimating a causal effect require some assumptions, and the goal of researchers is to make the assumptions as reasonable as possible and to replicate results across a number of different assumptions and data sets to check the robustness of one's findings.

In estimating equation (18), Archibald and Feldman (2006) found that both TELs and SMRs lead to substantial declines in state appropriation effort for higher education. The authors could have estimated similar results using the HLM framework. If they had employed a group-mean centered random-intercept model, the within-group estimator would still have been produced.

Goldhaber and Brewer (1997) use a cluster data set rather than a panel, but they face similar estimation challenges. They use a dataset containing students nested within schools to examine the impact of numerous student, teacher, and school variables on tenth-grade mathematics achievement test scores. Results from eighth-grade mathematics exams are included as a control variable, so the estimation method seeks to measure the "value-added" between the eighth- and tenth-grades. While the authors examine a wide range of explanatory variables, measures of the experience, certification, and education of mathematics teachers receive primary attention.

The authors could simply estimate the impact of these teacher characteristics by using an OLS regression, but the OLS coefficient, which is a weighted combination of the between-group and within-group relationships, typically produces little insight by itself. The covariation between a school's average test scores and the average characteristics of the school's mathematics teachers represents the between-group relationship. The within-group relationship is estimated by examining how mathematics achievement varies across students at the same school who have teachers with different levels of experience, certification, or education.

Do Goldhaber and Brewer (1997) want to measure the between-group relationship, the within-group relationship, or a weighted combination of the two? To answer this question, one needs to examine the purpose of their paper. The authors are primarily interested in the causal effect of teacher characteristics, as they wish to understand whether future increases in the level of teacher certification and education would produce learning gains. In their methodology section, they discuss threats caused by omitted variables that are determinants of mathematical achievement and correlated with teacher characteristics. Many of these omitted variables are at the school level. Schools differ in resources, average unobserved student traits, and average unobserved teacher traits. The standard notion of a wealthy suburban school that enrolls well-prepared students, attracts motivated and skilled teachers, and has ample resources to offer numerous programs illustrates this point. If teachers with strong credentials and extensive education disproportionately work within these schools, we would find a positive relationship between these teacher characteristics and test scores even if no true causal relationship existed.

The within-school estimator would implicitly control for these school-level differences by only examining the variation in mathematics achievement across students who attend the same school. Of course, this estimator could still be subject to bias. What if schools disproportionately assigned teachers with greater credentials, experience, and education to those classes within the school that contain students who would perform well under any teacher? This staffing pattern could again produce an upward bias and possibly produce a positive relationship even if no true causal relationship existed.

This discussion reveals how different estimators face very different internal validity threats. The within-group estimator requires assumptions regarding the distribution of teachers across classes within a school while the between-group estimator requires assumptions regarding the distributions of teachers across schools. Because an OLS regression uses both types of variation, that approach requires both sets of assumptions.

Goldhaber and Brewer (1997) estimate OLS, econometric random-effects, and econometric fixed-effects models, and their results do not vary substantially across these approaches. They could have produced similar results using related HLM models. Interestingly, their interpretation of the results as well as the motivation for their methodological choices would likely have differed if they worked within a different tradition. For example, organizational sociology highlights other differences between the within-group and between-group estimator as they emphasize Cronbach's (1976) point that "the aggregate variable represents a very different construct than the individual-level variable" (p. 20). At the individual-level, the education of an individual teacher could impact the students within his/her classroom due to changes instituted by that teacher. At the aggregate-level, the average education of a school's teachers could impact the culture and norms within the entire school, which could then affect all students within that school. By going within, aggregate-level considerations are eliminated, and the results do not capture the indirect effects of teacher's education on student's academic performance through altered culture and norms within a school. This point is rarely emphasized within economic frameworks but receives prominent attention within sociology.

Archibald and Feldman (2006) and Goldhaber and Brewer (1997) represent fairly conventional approaches to the within-group estimator and demonstrate the utilization of that estimator for panel and cluster samples. In both cases, the groups were formed naturally by linking individual observations to states or linking students to schools. In contrast, Dale and Kreuger (2002) formed a data set where students were grouped by the set of colleges that accepted and/or rejected them. A student who was rejected by school A and accepted by school B would be joined by all other students who were both rejected by schools with a selectivity level similar to school A and accepted by schools whose selectivity was similar to school B. For this matched-applicant sample, the multilevel data structure was created artificially by the researchers to control for unobserved student traits.

Dale and Kreuger (2002) employ this grouping pattern to study how attendance at a selective higher education institution impacts future earnings. Students with strong pre-college academic preparation and high socioeconomic status (SES) disproportionately attend more selective institutions. Because academic preparation and SES also impact future salaries, conventional analysis will produce biased estimates if one insufficiently controls for these student traits. Unfortunately, observable measures would only partially capture these traits. Dale and Kreuger's insight is straightforward: Because admissions committees are able to observe more extensive information about each student than researchers can, the decisions of these committees can provide additional information regarding a student's academic preparation. By grouping students based on the decisions of admissions committees, one can partially control for the additional information available to committees through the use of the within-group estimator. When producing results, this estimator will only compare students who were accepted and rejected by a comparable set of colleges.

For standard multilevel structures such as the case of students nested within schools, the researcher can easily assign individual observations into groups. For studies such as Dale and Kreuger (2002), assignment is more complicated. Which colleges are comparable in terms of selectivity? The authors decided to match applicants by the average SAT score (within 25 point intervals) of each school at which they were accepted or rejected, but this choice was somewhat arbitrary, so they also ran analysis using several other definitions of comparable selectivity to ensure that their findings were robust. These complications highlight the creative manner in which Dale and Kreuger created their multilevel data structure. They formed groups so that membership contained key information that observable measures of the student did not contain. In other words, Dale and Kreuger did not realize they had a multilevel data structure and then try to exploit it to advance their study. Instead, they determined what multilevel data structure would best advance their study, they developed a way to create that data structure, and then they exploited that structure.

As in the earlier papers, Dale and Kreuger (2002) still need to make substantial assumptions to interpret the within-group estimator in a causal manner. They must assume that, for students who were accepted and rejected at comparable colleges, the variation in enrollment decisions is uncorrelated with other factors that determine future earnings. In other words, students in the same matched applicant groups who attended institutions with different levels of selectivity cannot systematically differ in unobserved traits that determine future earnings.

Unlike Brewer and Goldhaber (1997), Dale and Kreuger found that the within-group estimator produced drastically different results than that produced by the OLS estimator. The OLS model indicated that students who attended a school with a 100 point higher average SAT score earned about 8 percent boost in future earnings. In contrast, the within-group estimator produced results that essentially were equal to zero. Because the OLS estimator is a blend of the between-group and the within-group estimator, the difference in results suggests that the between-group variation was driving the OLS result. In this case, the between-group estimator compares students who were accepted and/or rejected at very different sets of institutions. For example, it would compare students accepted at Harvard with students who were rejected at or who never even applied to Harvard.

The three journal articles reviewed to this point clearly demonstrate how advanced analysis of multilevel data can provide more robust estimates of the relationship between individual-level explanatory variables and outcomes. OLS results are typically not informative because they contain a weighted combination of the between-group and within-group relationships. A separate description of each type of variation usually produces more insight. In studies seeking a causal estimate, the within-group estimator is of special interest because it allows the researcher to control for unobserved group-level variables. The value of this estimator depends upon the particulars of the study. As we have demonstrated, the plausibility of assumptions required to rule out internal validity threats varies across studies for each estimator. Multilevel analysis advances causal inference most when the within-group estimator requires more reasonable assumptions than do conventional approaches.

Our examples demonstrated that the assumptions required to support the within-group estimator were still fairly strong, which demonstrates that multilevel analysis is often not a panacea. Alternative methods, such as experimental design, regression discontinuity, or instrumental variables, often rely upon assumptions that are more realistic.[12] These alternative approaches, however, are not available for many research questions within higher education, and, even when they are, they can seriously compromise external validity in pursuit of internal validity.

These considerations are especially relevant for institutional researchers. These researchers typically cannot randomly assign the treatment or exposure of interest, and they rarely get to observe natural experiments at their institution where part of the assignment was

[12] Some papers that use alternative multilevel structures may also possess more realistic assumptions than the three papers we reviewed. As noted earlier in the chapter, difference-in-differences models can produce compelling results in certain contexts (Dynarski, 2000; Cornwell et al, 2006). Analysis of within-family differences can also be convincing, because siblings possess a number of shared traits and experiences (Ashenfelter & Rouse, 1998). Both difference-in-differences models and sibling studies are very common in economics.

random and could be analyzed using the regression discontinuity or instrumental variables framework. Furthermore, institutional researchers are often constrained by time considerations, and multilevel analysis, which can be instituted relatively easily, is often the most promising option for advanced inquiry.

Papers realizing benefit #3: Examination of whether and to what extent key level-1 coefficients vary across groups and group characteristics.

While examining educational phenomena, researchers must recognize the "power of contexts" and the "ubiquity of interactions" that are present (Berliner, 2002, p. 19). Multilevel data allow researchers to model variation across groups through the use of cross-level interactions. Such interactions can test whether group characteristics moderate—increase or diminish—the strength of the individual-level relationships. In this section, we present three illustrative examples of studies that take advantage of this feature of multilevel analysis to provide more insightful and relevant findings. Our collection of papers include a classic example (Raudenbush & Bryk, 1986), a recent study of K-12 education (Langenkamp, 2010), and an example from higher education (Umbach, 2007).

Raudenbush and Bryk's (1986) study of the relationship between a student's SES and her mathematics achievement is well-known, because they use this example in their subsequent books to demonstrate many of the key insights they reveal (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002). Furthermore, they investigate important topics: How do schools differ in average mathematics achievement? Does the SES-achievement relationship vary across schools? Do average achievement and the SES-achievement relationship vary across specific school characteristics? These questions seek to understand how the relationship between SES and mathematics achievement intersects with the multilevel structure of education where students are nested within schools.

Multilevel analysis is vital, because the overall SES-achievement relationship may contain little information. If the SES-achievement relationship varies substantially across schools, then we cannot use overall results to describe the situation at individual schools. Furthermore, we cannot identify the specific features of educational institutions that contribute to a stronger or weaker SES-achievement relationship. To properly craft policy, we need to understand which schools or school characteristics have more equitable levels of student achievement.

Although numerous features of educational institutions (e.g., composition, size, and location) could alter the distribution of achievement among their students, Raudenbush and Bryk (1986) concentrate on two school characteristics: sector (public vs. Catholic) and average SES level. Before examining the impact of these characteristics, they first test whether the individual-level coefficients vary across groups. In other words, they estimate the following model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}) + e_{ij} \qquad (19)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (20)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \qquad (21)$$

Their results clearly indicate that $\text{Var}(\beta_{0j}) \neq 0$ and $\text{Var}(\beta_{1j}) \neq 0$, which implies that average achievement (represented by β_{0j}) and the SES-achievement relationships (represented by β_{1j}) varies across schools.

They then seek to explain this variation. They add a variable measuring the amount of time spent on homework to equation (19) and they add a sector variable, an average SES level variable, and the interaction between the two to equations (20) and (21). Their results for β_{1j}

indicate that the SES-achievement relationship is weaker at Catholic schools relative to public schools and is stronger at high-SES schools than at lower-SES schools. In other words, the gap in achievement between high-SES and low-SES students is smaller at private Catholic schools and at schools whose students have lower average SES levels.

Langenkamp (2010) examines the academic performance of students as they transition from middle school to high school. This stage of the educational process often contains risk, because students must renegotiate their relationships with both teachers and peers. Previous research showed that transitions can lead to declines in academic performance, especially for those at risk academically prior to the transition.

Langenkamp's primary contribution is to examine whether the transition from middle school to high school depends upon the context of the school district. She views the feeder pattern used to assign middle school students to high schools as the most important contextual element. In a "uniform" district context, one middle school cohort feeds into one high school. In a "mixed" district context, many middle schools feed into a high school. The degree to which students must renegotiate relationships varies drastically across these two contexts. In mixed school districts, a student's existing social ties are reconfigured and the opportunities for new social ties are created.

Students at the same school may experience the transition to high school differently, and the extent of these differences may vary by school district context. Previous research established that academically vulnerable students will find it more difficult to navigate the transition, but we don't know whether this problem is more or less severe in mixed school districts. Academically vulnerable students may find it even more difficult to transition in a mixed context, because the need to renegotiate social relationships may distract them from their studies and cause them to fall further behind. Alternatively, the changing environment may disrupt problematic social relationships that were limiting past academic performance.

Multilevel analysis allows one to elegantly test these competing hypotheses. Langenkamp (2010) uses cross-level interactions to examine whether the feeder pattern of a school district alters the relationship between low middle school achievement and early high school academic performance. She also includes interactions that allow the impact of middle school popularity and teacher bonding to vary across feeder systems. Two measures of early high school academic performance, low math course placement and first year course failure, are used as outcomes that capture academic struggles that occur during the transition to high school.

While many of her cross-level interactions were not statistically significant, Langenkamp (2010) did find that students with low achievement in middle school were less likely to fail initial high school coursework when residing in a mixed school district. Specifically, she found "that among low-achieving middle school students, the probability of course failure is almost 0.50 for students in a uniform context, whereas those in a mixed context have a predicted failure rate of only 0.20" (p. 12). One potential explanation for these results is that low-achieving students benefit greatly from the opportunity for new relationships among incoming student cohorts. These results are useful for policy considerations, but they also help us understand more deeply the role of social relationships in determining academic performance, which was the core phenomenon under study.

Umbach (2007) examines how faculty engagement in good practices varies by the appointment status of the faculty member. Part-time and full-time non-tenure-track faculty are growing in importance within higher education, and we consequently need to understand how these contingent faculty differ from tenured/tenure-track faculty in their teaching practices.

Umbach studies good practices related to faculty-student interactions, course preparation, and teaching techniques that past research has linked to increases in student learning.

The performance of contingent faculty relative to their tenured/tenure-track counterparts may vary substantially across higher education institutions. In same contexts, contingent faculty face a welcoming culture and are provided with key resources, such as office space and training that can lead to better instruction. In other settings, however, important resources and support are not provided. Because Umbach (2007) is interested in the *relative* performance of contingent faculty, the performance of tenured/track-track faculty, which also varies substantially across institutions, is another important consideration. In combination, these points suggest we may be less interested in estimating an overall contingent effect than in identifying a number of context-specific contingent effects.

Umbach considers a wide range of institutional characteristics when seeking to explain variation in the contingent effect. That list includes an institution's contingent faculty share, location, Carnegie classification, control, selectivity, and size. Each of these variables could capture important differences across institutions that impact the relative performance of contingent faculty. For example, as the proportion of contingent faculty grows on a campus, elements of culture and policy that impact contingent faculty performance could be altered. The importance of research to the institution's mission could also have an effect. Tenure-track faculty could be more consumed by research, so contingent faculty, who can focus more heavily on teaching, may perform relatively well at research universities. Alternatively, contingent performance could suffer if the culture of research universities is less welcoming and supportive of non-tenure-track faculty.

To examine these possibilities, Umbach (2007) employed a HLM-based model that allowed the slope coefficient for part-time faculty to vary across institutions and institutional characteristics. Surprisingly, this coefficient did not vary across institutions for most outcomes; the only exception was non-class-related interactions. For this practice, Umbach found substantial differences across Carnegie classifications. Relative to their tenure-track/tenured counterparts, part-time faculty were least likely to engage in non-class-related interactions when teaching at Doctoral and Master's universities.

In each of the three articles reviewed, the level-1 relationship under study is only partially illuminated by an overall result. Much deeper insights can be obtained through multilevel analysis of how this relationship varies across groups and the characteristics of those groups. Theoretical propositions can be refined, because individual theories predict that level-1 relationships and group characteristics covary in specific ways. Multilevel analysis can also aid policy development, as findings can inform efforts to optimally match policies with the specific context of individual schools.

Papers realizing all four benefits.

Our final example is notable for three reasons. First, Gelman et al. (2007) presented numerous level-1 coefficients for specific groups in their sample. In other words, they reaped the fourth benefit of multilevel analysis. We rarely came across papers, especially in higher education research, that used multilevel data in this manner. Second, Gelman et al. (2007) also realized the other three benefits of multilevel analysis. We did not find another paper that examined multilevel data in such a comprehensive fashion. Finally, Gelman et al. (2007) used both the between-group estimator and the within-group estimator to describe overall relationships, but they were not seeking to describe a purely causal relationship. Most analysis of within-group variation fixates on causal questions, but use of this variation, especially alongside analysis of between-group variation, can also advance associational or descriptive studies.

Gelman et al. (2007) examined the relationship between voting patterns and income. Since the 2000 U.S. presidential election, we have used the terms red state and blue state, where red states vote Republican and typically have lower average incomes while blue states vote Democratic and typically have higher average incomes. After recent presidential elections, county-level analyses were also conducted, and the results identified many wealthy counties that voted Democratic and many lower-income counties that voted Republican. This national-level and county-level analysis contradicts long-held beliefs that portrayed the Democrats as the party of the poor and Republicans as the party of the rich.

Gelman et al. (2007) used the within-group and between-group estimators to explain how this "apparent paradox is no paradox at all" (p. 365). By examining relationships at multiple levels, they demonstrated that voters are more likely to support the Democrats when they live in richer states, but within any given state, richer voters are more likely to vote Republican. The relationship between voting patterns and income depends upon whether you compare states or compare voters within states.

Multilevel analysis also allows one to examine whether the within-state relationship varies across states. Gelman et al. (2007) find substantial variation. A citizen's likelihood of voting Republican substantially increases with income in poor, rural, Republican-leaning states, but only slightly increases with income in rich, urban, Democratic-leaning states. The authors describe succinctly their finding: “In poor states, rich people are very different from poor people in their political preferences. But in rich states, they are not.” (p. 365)

Most readers would be primarily interested in the overall relationship between income and voting patterns or how this relationship varies across state characteristics, but some may wish to observe estimates for a particular state of interest. The fourth benefit of multilevel analysis, producing results for a particular group even if the number of observations for that group is relatively small, allowed Gelman et al. (2007) to report estimates for all 50 states. The authors also used state-specific findings to help demonstrate national voting patterns. By graphing state-specific results for a low-income red state (Mississippi), a middle-income purple state (Ohio), and a high-income blue state (Connecticut), the authors were able to communicate, in just one figure, their three major findings: (a) the between-state relationship is negative (b) the within-state relationship is positive, and (c) the within-state relationship diminishes with the income of the state.

This paper does not seek to test whether income is *causing* support for Republicans, but the authors did investigate whether demographic measures of voters helped explain the correlation between income and voting patterns. When they dropped all African-American respondents from the sample, the impact of income fell by about half. But the basic pattern remained, even when controls for other variables, such as gender, age, and education, were incorporated.

The multilevel analysis was the heart and soul of this paper, as conventional analysis would not have produced any of the major insights that were revealed. Gelman et al. (2007) was unusually comprehensive when exploiting the multilevel structure of voting patterns. They used a variety of multilevel analysis, from relatively simple models to models that allowed both intercepts and slope coefficients to vary across groups. They used these models to study the correlation between voting patterns and income from a variety of levels. The above text described a rich set of findings, and we still ignored several additional levels of analysis. For example, county-level estimates revealed that high-income counties were the most Republican in southern states and low-income counties were the most Republican in western states. Examination of separate time periods found that most of the main findings of the paper strengthened considerable over the last 15 years.

Our understanding of higher education would be deepened drastically if we also conducted analysis at all the levels that are available in the data used in a study. Students are nested within teachers who are nested within departments which are nested within colleges which are nested within institutions. These institutions are then housed within sectors and states. When one also considers that multiple time periods can be examined at each of these levels, one realizes the extreme complexity that can be contained within any particular relationship in higher education, but also the detailed understanding that may be obtained when we exploit these sources of variation.

Conclusion

Past chapters in this Handbook have discussed hierarchical linear models and econometric panel models separately (Ethington, 1997; Zhang, 2010). This chapter presented both traditions simultaneously and demonstrated a number of core similarities between them. At the same time, we explained why these core similarities are hard to detect. Varying notation, model presentation, and terminology create artificial differences that confuse readers. HLM and econometrics often utilize different data structures and realize different benefits of advanced multilevel analysis. Many researchers will focus on the aspects highlighted within each tradition and not realize that the basic econometric and HLM models can handle similar data structures and reap similar benefits.

The differences in these methodological traditions create substantial challenges for the field of higher education. If a researcher is trained solely in one tradition, he/she will often be at a disadvantage when trying to understand research using the other tradition. Scholarly communication is inhibited as a result. Scholars trained in econometrics may disproportionately employ panel data and seek to advance causal inference, whereas individuals trained in HLM may disproportionately employ cluster data and examine heterogeneity across groups.

Higher education researchers typically have more flexibility in their choice of methods than other academics because a variety of methodological approaches are accepted within the field. Consequently, one can borrow from the methodological tradition which can best advance the study at hand. The benefits of multidisciplinarity are not easily obtained, however. Producing deep insights into the world through the analysis of data is an arduous task, and working within an individual tradition simplifies the process. Each methodological tradition has a distinct approach ranging from terminology to the basic logic and goals underlying the analysis. A researcher seeking to engage multiple traditions may waste considerable time trying to learn and trying to decipher these differences, and this chapter is designed to minimize the time lost in that process.

As researchers develop a deep understanding of multilevel models, we hope they will begin to realize that for most mainstream studies the primary issue is not the choice of tradition. Instead, the challenge is to skillfully employ multilevel models to help one realize the core objectives of the study at hand and to answer questions such as: Which benefits allowed by the advanced analysis of multilevel data would best advance these objectives? Which data sets and models would best allow one to reap these selected benefits?

The choice of methodological tradition within which to work is often a much less important issue. For studies employing basic multilevel models the choice will often depend on practical considerations. For example, the focus within HLM on the study of heterogeneity across groups has resulted in a wide array of accessible computing options for HLM models that allow slope coefficients to vary across groups.

In this chapter, we examined only the most basic models and topics pertaining to the analysis of multilevel data. Our approach promotes a deep understanding of core issues, but it does not introduce the reader to a number of important technical and advanced considerations. Ethington (1997) and Zhang (2010) discuss many of these topics, and available textbooks cover even more ground. In addition to Raudenbush and Bryk (2002), Snijders and Boskers (1999) and Heck and Thomas (2009) provide strong coverage of HLM-related issues. Wooldridge (2009) and Stock and Watson (2007) provide a good introduction to econometric panel models, while Baltagi (2008), Cameron and Trivedi (2005), and Wooldridge (2002) expose readers to advanced considerations.

Although higher education researchers almost always work within the HLM or econometric framework, more general approaches to multilevel data exist. Gelman and Hill (2007) incorporate features of both the HLM and econometric traditions, and a forthcoming book by Singer and Willet is also expected to provide broad coverage. The different perspective of these books often allows for new insights. For example, Gelman and Hill (2007) provide numerous examples of how advanced analysis of multilevel data allows one to estimate level-1 coefficients for a specific group even if the number of observations for that group is relatively small. Their focus leads to accessible and insightful explanations of how these level-1 coefficients are estimated in practice.

The benefits of multilevel models, while important, are limited. Quantitative higher education researchers consequently also need to master additional methodologies, and an understanding of the topics covered in this chapter can aid that process. Consider the goal of causal inference. Multilevel analysis is of limited help when the primary explanatory variable of interest is at the group level, when the variation within groups is inadequate or swamped by measurement error, or when the primary omitted variables of concern are at the individual level. In many settings, alternative techniques, such as the difference-in-differences, instrumental variables, and regression discontinuity models, may provide more convincing results.
These models are similar to the within-group estimator in that they use only a portion of the variation in the treatment or exposure, and the goal of the researcher is to find settings where that portion of the variation promotes internal validity. Once a researcher understands how to effectively select studies to maximize the benefits of the within-group estimator, he/she can easily learn how to employ these alternative techniques as well. As noted earlier, the difference-in-difference model essentially examines how differences within groups vary across groups, so an understanding of multilevel frameworks is especially helpful for this model.

Multiple traditions exist in other parts of the methodological world, including the qualitative domain. We hope that our chapter will inspire other researchers to integrate multiple approaches, and this handbook is a natural home for such work. Such efforts drastically improve scholarly communication and help researchers develop a broad methodological approach that incorporates the strengths of multiple traditions.

We also hope that readers will use this chapter as motivation to fully reap the benefits of multilevel data. Higher education researchers regularly utilize multilevel models, but they often employ them without seeking to fundamentally advance their study. Instead, the models appear to be used simply to fulfill expectations of future readers who want to see advanced techniques. For support of our claim, we return to the following quote from Smart (2005): “the results obtained thus far from the use of HLM have not suggested any dramatically different conclusions from those based on the use of more conventional analytical procedures” (p. 466). Advanced techniques are not intrinsically valuable; their worth depends upon the extent to which they advance the study in question. This chapter has provided a simple and accessible introduction to the benefits of advanced analysis of multilevel data, and we hope that future research will more fully reap these benefits.

References

Archibald, R., & Feldman, D. (2006). State higher education spending and the tax revolt. *Journal of Higher Education, 77*, 618-644.

Ashenfelter, O. & Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review,* 84(5), 1157-1173.

Ashenfelter, O. & Rouse, C. (1998). Income, schooling, and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics*, 113(1), 253-284.

Baltagi, B. H. (2008) *Econometric analysis of panel data. 4th edition.* Chichester: John Wiley & Sons.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines.* Buckingham: Society for Research in Higher Education/Open University Press

Berliner, C. (2002). Comment: Educational research: The hardest science of all. Educational *Researcher, 31*(8), 18-20.

Breneman, D. (1994). *Liberal arts colleges: Thriving, surviving, or endangered?* Washington, D.C.: Brookings Institution.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Cameron, A. & Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Chamberlain, G. (1982), Multivariate regression models for panel data, *Journal of Econometrics*, 18, 5-46.

Chamberlain, G. (1984). Panel data. In Z. Griliches & M. Intriligator (ed.), *Handbook of Econometrics* (Vol. 2, pp. 1247-1318). Amsterdam: North-Holland.

Chaplin, D. (2003). *Hierarchical linear models: Strengths and weaknesses*. Paper prepared for the November 2003 meetings of the Association for Public Policy and Management.

Cheslock, J., & Rios-Aguilar, C. (2008, November). Reaping (or not) the benefits of hierarchical data. Paper presented at the Association for the Study of Higher Education Meeting. Jacksonville, FL.

Chronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis,* Occasional paper, Stanford: Stanford Evaluation Consortium, Stanford University.

Cornwell, C., Mustard, D., & Sridhar, D. (2006). The enrollment effects of merit-based financial aid: Evidence from Georgia's hope program. *Journal of Labor Economics*, 24(4), 761-786.

Dale, S., & Krueger, A. (2002). Estimating the payoff to attending a more selective college: An application o selection on observables and unobservables. *Quarterly Journal of Economics, 107*(4), 1491-1527.

Dedrick, R., et al. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research 79*(1), 69-102.

Dynarski, S. (2000). Hope for whom? Financial aid for the middle class and its impact on college attendance. *National Tax Journal*, 53(3), 629-662.

Ethington, C. (1997). A hierarchical linear modeling approach to studying college effects. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 12, pp. 165–194). New York: Agathon Press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gelman, A. & Little, T. (1997) Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, 23, 127–135.

Gelman, A., Shor, B., Bafumi, J., & Park, D. (2007). Rich state, poor state, red state, blue state: What's the matter with Connecticut. *Quarterly Journal of Political Science, 2*, 345-367.

Goldhaber, D., & Brewer, D. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources 32*(3), 505-523.

Griliches, Z. (1979). Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy*, 87(5), S37-S64.

Heck, R., & Thomas, S. (2009). *An introduction to multilevel modeling techniques*, 2nd edition. New York: Routledge/Taylor & Francis.

Huber, P. (1967). *The behavior of maximum-likelihood estimates under non-standard conditions.* Paper presented at the Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA.

Kenny, D., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, 83, 126-137.

Laboree, D. (2003). The peculiar problems of preparing educational researchers. *Educational Researcher, 32*(4), 13-22.

Langenkamp, A. (2010). Academic vulnerability and resilience during transition to high school: The role of social relationships and district context. *Sociology of Education, 83*(1), 1-19.

McPherson, M., and Schapiro, M. (1998). *The student aid game: Meeting need and rewarding talent in American higher education*. Princeton: Princeton University Press, 1998.

Meyer, Bruce D. (1995) "Natural and quasi-experiments in economics" *Journal of Business and Economic Statistics*. 13(2). pp. 151-161.

Moulton, B. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics*, 72(2), 334-338.

Mundlack, Y. (1978a). On the pooling of time series and cross section data. *Econometrica, 46*, 69-85.

Mundlack, Y. (1978b). Models with variable coefficients: Integration and extension. *Annales de l'INNSEE, 30-31*, 483-509.

Raudenbush, S., & Bryk, A. (1986). A hierarchical model for examining school effects. *Sociology of Education, 59*(1), 1-17.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, 2nd edition. Newbury Park, CA: Sage.

Raudenbush, S., Bryk, A., Cheong, Y., Congdon, R., du Toit, M. (2004). *HLM 6: Linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International Inc.

Rivkin, S., Hanushek, E., and Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.

Rogers, W. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin, 13,* 19-23.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental & quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Smart, J. (2005). Attributes of exemplary research manuscripts employing quantitative analyses. *Research in Higher Education*. 46(4), 461-477.

Snijders, T. (2005). Power and sample size in multilevel linear models. In B Everitt & D. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science*, (Vol. 3, pp. 1570–1573). Chicester: Wiley, 2005.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Stock, J., & Watson, M. (2007). *Introduction to econometrics*, 2nd edition. Addison-Wesley Series in Economics. Boston, MA: Pearson

Umbach, P. (2007). How effective are they? Exploring the impact of contingent faculty on undergraduate education. *The Review of Higher Education, 30*(2), 91-123.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*, 817-830.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, 93(2), 133-138

Wooldridge, J. (2009). *Introductory econometrics: A modern approach,* 4th edition. South-Western Cengage Learning.

Zhang, L. (2010). The use of panel data models in higher education policy studies. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 25, pp. 307–349). New York: Agathon Press.

Figure 1: Within-Group and Between-Group Regression Lines