Clustering Standard Errors or Modeling Multilevel Data?

Ban Chuan Cheah

This version: May 2009

Abstract

Multilevel models are used to revisit Moulton's (1990) work on clustering. Moulton showed that when aggregate level data is combined with micro level data, the estimated standard errors from OLS estimates on the aggregate data are too small leading the analyst to reject the null hypothesis of no effect. Simulations using similar data suggest that even when corrected for clustering, the null hypothesis is over-rejected compared to the estimates obtained from multilevel models. The relationship between survey sampling and Moulton's correction is also explored. The parallel between these two areas is extended into multiway clustering. Simulations using a data set with students clustered within classrooms and classrooms within schools suggest that the over-rejection rate from multilevel models is smaller than those corrected for clustering. This is particularly true when the number of clusters (classrooms) is small. The results suggest that modeling the clustering of the data using a multilevel methods is a better approach than fixing the standard errors of the OLS estimate.

1 Introduction

This note revisits the results of Moulton's (1990) study on how using grouped aggregate level data such as state level unemployment rates with individual or micro level data such as individual earnings within a state can result in a downward bias in the estimated standard errors of OLS estimates. The analyst can conclude the state level effects are statistically significant when they are in fact not because the standard errors are *too small*.

The insight provided by Moulton's work was that individuals within the aggregated level such as state are clustered so that they are in fact more similar to one another than individuals from another state. Thus, the assumption that observations are independent and identically distributed is violated. This finding has led many statistical software packages to implement a correction for clustering.[1]

[1] For instance in STATA the correction is implemented using the "cluster" option for the standard error of a regression command.

Bertrand, Duflo and Mullainathan (2004) study the effects of over-rejection in difference-in-difference estimates and find substantial over-rejection even after correcting for clustering[2]. They note that in order to correct for the standard error due to clustering one has to account for the presence of a common random effect at the group level. Thus, they identify the effect of clustering as this common effect rather than an assumption that individuals within a cluster are more alike. The presence of a common group level effect is the essence of multilevel models.[3] Multilevel models have the benefit of allowing for partial pooling of coefficients toward the completely pooled OLS estimate which according to Gelman (2006) can be a more effective estimation strategy. Bafumi and Gelman (2006) observe that while multilevel models have become accepted in some social sciences they have been slow to gain popularity in others such as economics.

This note finds that even after correcting for clustering, there is a tendency to over-reject the null hypothesis of no effect. The hypothesis that there is no effect is implemented by introducing a cluster level random number which is estimated over many repetitions. The proportion of over-rejection is reduced when the multilevel structure of the model is explicitly modeled. The effect of the degree of clustering measured as the average number of units within a cluster is also investigated. When the number of individuals within a cluster is reduced, multilevel models still outperform the standard clustering correction.

The next section revisits Moulton's study and the data set used. The multilevel model to be estimated and the results of the simulations are then presented. An investigation into multiway clustering is considered by linking Moulton's correction to the literature on survey sampling and then applying the analytic approach from used in analysis of survey data. Over-rejection of the null hypothesis in multiway clustering is explored using a similar strategy on a data set of students clustered within classrooms and classrooms within schools. This data set allows an exploration into the robustness of the results when the number of individuals within a cluster (classroom) is small. Random numbers at the school level are over-rejected by all types of standard error correction except for multilevel models. Clustering is examined at the classroom level only and when classrooms are *nested* within schools. The simulations suggest that modeling data using its multilevel structure is a better approach than fixing the standard errors of an OLS estimate.

2 Revisiting Moulton (1990)

Failure to account for the grouped nature of the aggregate level data can lead to erroneous conclusions. Moulton (1990) demonstrated this by replicating Topel's

[2]Row 2 of their Table II, pg. 257.

[3]Multilevel models are sometimes referred to as random coefficients models or varying intercepts models. They are also referred to as hierarchical linear models (HLM) in psychology and education or mixed effects model in health and biostatistics because they allow for both random and fixed effects.

Variable	Source
x_1Estimated Rate of state employment growth	Computed from BEA
x_2Current state relative employment disturbance	Computed from BEA
x_3Predicted state disturbances	Computed from BEA
x_4Live Birth Rate per 1,000 population (2001)	Table 87
x_5Legal abortions per 1,000 women (2000)	Table 104
x_6Death rate from heart disease per 100,000 population (2001)	Table 119
x_7Death rate from suicide per 100,000 population (2001)	Table 119
x_8Death rate from cancer per 100,000 population (2001)	Table 119*
x_9Marriage rate per 1,000 population (2000)	Table 126*
x_{10}Percentage of persons age 5−17 enrolled in public elementary and secondary schools (2000)	Table 244
x_{11}Total Land Area in Sq Km	Table 359
x_{12}Total Water Area in Sq Km	Table 359
x_{13}Elevation of Highest Point in Meters	Table 363
x_{15}Total Black Elected Officials (2001)	Table 417
x_{16}Daily newspaper circulation per capita (2002)	Table 1132

Table 1: Variables used in model

(1986) study on the impact of local labor market conditions on individual level wage rates. Individual level data from the March Current Population survey is merged onto state level data on state employment growth rate, relative and predicted state disturbances. Topel (1986) finds that these state level variables are significant in their impact on individual level wages. Moulton (1990) concluded that the standard errors of these variables were significantly biased downward because they did not account for the grouped level nature of the aggregate data.

He further demonstrated this problem by including state level variables on abortion rates, death rates, newspaper circulation and other state level variables that he considered to be irrelevant to individual level wages. In addition, he also constructed random numbers for each state. These state level variables were then merged onto the individual level data. Surprisingly, some of these 'irrelevant' variables (including the random number) were significant. These variables are listed as x_4 through x_{16} in Table 1.

This note replicates Moulton's results using more recent data and extends the analysis to ask whether the downward bias in the standard error can be further improved by taking into account the multilevel nature of the data. The individual level data is the March 2003 Current Population survey. The data contains 81,588 individuals who are civilian employees currently employed in the labor force. Following Moulton, only individuals older than 20 years and whose computed weekly earnings exceed $40 are included.[4]

State level wage and salary employment numbers were obtained from the Bu-

[4]This data is available for download at www.bls.census.gov/cps/cpsmain.htm. Accessed May 2, 2005.

reau of Economic Analysis Annual State Personal Income estimates.[5] This data allows the creation of disturbance terms detailed in Topel (1986). A quadratic trend was fitted to the log value of employment to compute employment growth for each state.[6] The current state relative employment disturbance was computed as the deviation of the estimated residual for each state from the residual estimated at the national level.[7] The predicted state disturbance was computed as the next period forecast using an ARMA model with exponential smoothing. This forecast was computed as a deviation from the national forecast.[8]

Additional state variables that Moulton considered 'irrelevant' were also included in the data. These were obtained from the 2003 Statistical Abstract of the United States. The full list of state level variables are listed in Table 1. I have followed Moulton's naming convention for the variables using $x_4$to x_{16}. If the variable used in Moulton's original paper was not available from the 2003 Statistical Abstract, I substituted it with another variable and have indicated this with an asterisk in the *Source* column. Cancer deaths (x_8) was substituted for perinatal deaths and divorce rates was substituted with marriage rates (x_9). Only the original variable x_{14}which was per capita state legislative appropriations for arts agencies was not substituted and is left out.

Moulton (1990) provided the results of an OLS regression of wage rates on the individual level covariates[9] and the all of the variables listed in Table 1. Included in his specification is also a random variable. This study takes Moulton's approach one step further. Focusing solely on the regression of wages on all of the above individual and state level variables *including* a state level random variable it asks: If a random variable is drawn 1,000 times, what is the number of times that the t-test will show that the random variable is significant when it is in fact not. Three types of t-tests are examined: OLS, Moulton's clustering, and multilevel models. In other words, what is proportion of over-rejection when the null hypothesis of no relationship between the random number and the dependent variable is true?

3 Multilevel models and the relationship between survey sampling and clustered standard errors

3.1 Multilevel models

Multilevel models have the following structure:

Consider the following simplified model:

$$\text{Wages}_{ij} = \beta_{0j} + r_{ij}$$

[5]This data is series SA04 and is available for download at `www.bea.govbea/regional/spi/default.cfm`. Accessed June 3, 2005.

[6]This corresponds to Moulton's x_1

[7]This corresponds to Moulton's x_2

[8]This corresponds to Moulton's x_3

[9]Individual level variables included are age, gender, education level, race, marital status, whether the person lived in a central city and the census division of the individual.

where i denotes individuals and j denotes states and $r_{ij} \sim N(0, \sigma^2)$. Individuals are clustered within states. The intercept β_{0j} is modeled as:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} S_j + u_{0j}$$

where S_j are state level characteristics or covariates such as unemployment rates and $u_{0j} \sim N(0, \tau_{00})$. Thus, u_{0j} is the common group level random effect noted by Bertrand, Duflo and Mullainathan (2004). Therefore, the full equation estimated is

$$\text{Wages}_{ij} = \gamma_{00} + \gamma_{01} S_j + u_{0j} + r_{ij} \tag{1}$$

There are two variance components, u_{0j} which represents variation between state means (τ_{00}) and r_{ij} which represents variation among individuals within states (σ^2) and also one covariate S_j.

When person level covariates are included, the equations estimated are:

$$\text{Wages}_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$$

where

$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad u_{1j} \sim N(0, \tau_{10})$$

and

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]$$

Combining equations give:

$$\text{Wages}_{ij} = \gamma_{00} + \gamma_{01} S_j + u_{0j} + (\gamma_{10} + u_{1j}) X_{ij} + r_{ij}$$

and rearranging terms,

$$\text{Wages}_{ij} = \gamma_{00} + \gamma_{01} S_j + \gamma_{10} X_{ij} + u_{1j} X_{ij} + u_{0j} + r_{ij} \tag{2}$$

The two combined equations, (1) and (2) are sometimes referred to as a varying intercepts (or random intercepts) and varying intercepts and slopes (or random coefficients) model respectively. In both equations, the intercept term is $\gamma_{00} + u_{0j}$ while $\gamma_{10} + u_{1j}$ is the slope in equation (2). It is not necessary to assume random slopes with person level covariates. In this case, the equation to be estimated is simply

$$\text{Wages}_{ij} = \beta_{0j} + \beta_1 X_{ij} + r_{ij}$$

where the dependence of $\beta_1$on group j is removed.

The coefficients and variance components are estimated using maximum likelihood methods. While the coefficients in OLS and clustered standard errors are identical (only the standard errors are different), typically the coefficients in multilevel models will not be the same as the OLS coefficients. This is because multilevel models attempt to estimate the variance components. As such, multilevel models are more computationally intensive and may affect the estimates of the coefficients if the estimation of the variance components are inaccurate.[10]

[10]See Primo, et. al (2006).

Of interest as well is how this model relates to a one way fixed effects model - where the groups (states, in this case) are modeled using dummy variables for each (N-1) groups. The performance of the standard error using this approach is not investigated here. This is because with state level variables such as employment growth rates and other 'irrelevant' state level variables included in the regression, introducing dummy variables would induce a collinearity. The multilevel approach therefore has an advantage over the dummy variables approach in the sense that the *"fixed effects"* can be partially captured while at the same time the effects of group level covariates can also be estimated. As discussed by Gelman and Hill (2007, pp. 245-246), multilevel models are an alternative to the complete pooling model (such as OLS where one regression equation is run for the entire sample) and the no pooling model (where a separate regression equation is run for each group).

Related to the fixed effects model is the random effects model which is prevalent in the econometrics of panel data. The formulation in equation 1 is similar to the formulation of random effects model in panel data. In panel data the group j is replaced with time t. Multilevel models are more general in the sense they apply not only to panel data models. As to when "fixed" or "random" effects models should be used (multilevel models are sometimes referred to as "mixed" models), Gelman and Hill (2007, pg. 246) note: "Our advice (elaborated upon in the rest of this book) is to *always* use multilevel modeling ("random effects"). Because of the conflicting definitions and advice we avoid the terms "fixed" and "random" entirely, and focus on the description of the model itself (for example, varying intercepts and constant slopes), ..." Gelman and Hill's advice is to exploit the natural description of the data and to model the data itself instead of "fixing" the standard errors of an OLS estimate.

3.2 The relationship between survey sampling and clustered standard errors

A simple formula for the standard error of a clustered estimate is derived from the true variance-covariance matrix C and is given by Moulton as

$$C = \sigma^2(X'X)^{-1}[1 + \rho(m-1)]$$

where$\rho$is the intra-class correlation coefficient that measures the dependence of units within a cluster and m is the number of units within a cluster (assuming that all clusters have the same number of units).

Under the assumption that errors of a regression are i.i.d, the true covariance matrix is $C_I = \sigma^2(X'X)^{-1}$. Restating Moulton's formula and with some abuse of notation, the ratio of the covariance matrix under clustering C and under the usual i.i.d. assumption $C_I$is simply:

$$\frac{C}{C_I} = [1 + \rho(m-1)]$$

In survey sampling, the right hand side term is simply called the design effect of a cluster sample (mean) where all clusters are of equal size, m. (Kalton (1983), pp. 30-31) The design effect measures the ratio of the variance of a statistic from cluster sampling to the variance of the statistic (e.g. the mean) from simple random sampling (SRS). In sampling, the average size of the cluster, m depends also on the number of clusters. For fixed sample size, even with a small value of ρ, the design effect can be large if the number of clusters is small. When the number of clusters is small, each cluster will necessarily be larger for a fixed sample size.

Section 5 exploits the equivalence between survey sampling and clustered standard errors to extend the investigation to multiway clustering, i.e. when units are in turn clustered within other units such as students within classrooms, and classrooms within schools.[11] Multiway clustering also provides a framework for exploring the role played by the number of clusters as well as the number of individuals within a cluster. The rate of rejection when the number of clusters is reduced will also be explored.

4 Results from revisiting Moulton

The multilevel version of Moulton's model is estimated using assuming only random intercepts. This is because the focus is on the possibility of over-rejection of the state level variable, in particular the state level random variable. A random variable is generated at the state level and the number of times the coefficient on the random variable is rejected is counted. The results using 1000 simulations for each method of estimation are summarized in Figure 1.

As was previously shown by Moulton, the rejection rate given by the OLS t-statistic is much higher than that of clustered standard errors. However, the rejection rate of clustered standard errors is still much higher than that of multilevel models (about 25 percent versus 6 percent). The clustered standard errors over-reject the null hypothesis of no relationship between a state level random variable and the dependent variable by four times.

Figure 2 investigates the how the degree of clustering as measured by the number of individuals within a state affects the probability of rejection. With a large number of observations as is the case here, 100 draws of a random variable is sufficient to detect over-rejection for this part of the exercise. The average number of individuals are noted in parenthesis for each subset drawn from the entire data set. Subset 0 is the full sample and there are an average of 1,600 persons in each state. Beginning with subsets 1 through 8, 10 percent of the sample is incrementally dropped using a simple random sample, so that by subset 8, only 20 percent of the data set remains. The average number of persons in each state in subset 8 is 320. As the degree of clustering falls, the proportion

[11] This equivalence implies that estimates using clustered standard errors from STATA are of the same order of magnitude as those using STATA's *svy* estimation methods. The appendix gives the code to replicate the standard errors estimated using clustering and using the survey estimation methods. Note that this does *not* imply that that a fully specified model using survey weights and the information in the survey design should be ignored. The code is purely for demonstration purposes only.

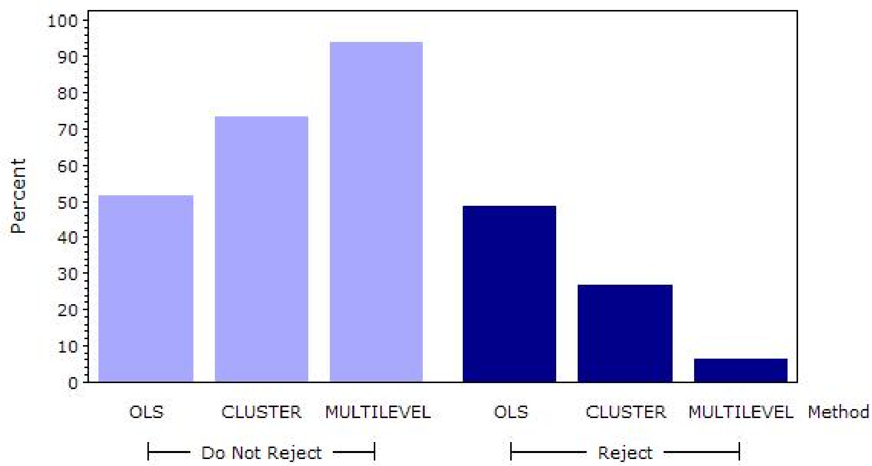

Figure 1: Percent of rejection when the null hypothesis of no relationship is true, by method of estimation (1000 replications per method)

of over-rejection by the OLS estimate approaches that of the clustered standard error which remains somewhat constant. The rejection rate for the multilevel model remains more or less constant regardless of the degree of clustering but is always lower than either the OLS or clustered standard error. Thus even with with a small number of replications (100 in this case), the tendency to over-reject can be detected.

5 Multiway Clustering

In multiway clustering, there are (at least) two levels of clustering to explore - the number of individuals within a cluster, and the number of clusters within a larger group. The data used is from West et. al. (2006). The original data was a study of math achievement scores of 1190 first and third graders in randomly selected classrooms from a national sample of elementary schools. Students are clustered within classrooms and classrooms are clustered within schools. In contrast to Moulton's study, the number of individuals is smaller - ranging from 1 to 10 students per classroom.

The dependent variable in this case is *mathgain*. All the other variables except for *mathprep* are included in the regression equation.[12] The correction for the standard error in an OLS regression assumes that multiway clustering can be approximated by applying another survey sampling concept: stratification with cluster sampling. In stratified and clustered sampling, the population is divided

[12] Except for OLS, the IDs are used by the statistical program (e.g. SAS, STATA) to describe the structure of the data.

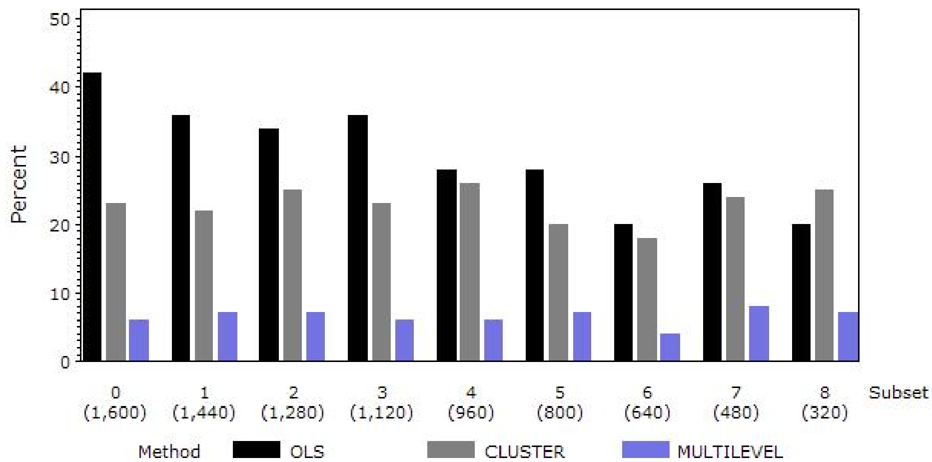

Figure 2: Percent of rejection when the null hypothesis of no relationship is true: Effects of number of individuals within a cluster (100 replications for each method per subset)

Variable	Description
sex	Indicator variable (0 = boys, 1 = girls)
minority	Indicator variable (0 = non-minority students, 1 = minority students)
mathkind	Student math score in the spring of their kindergarten year
mathgain	Student gain in math achievement score from the spring of kindergarten to the spring of first grade (the dependent variable)
ses	Student socioeconomic status
yearstea	First grade teacher years of teaching experience
mathknow	First grade teacher mathematics content knowledge: based on a scale composed of 30 items (higher values indicate higher content knowledge)
housepov	Percentage of households in the neighborhood of the school below the poverty level
mathprep	First grade teacher mathematics preparation: number of mathematics content and methods courses
classid	Classroom ID number
schoolid	School ID number
childid	Student ID number

Table 2: List of variables in classroom level data set

into different groups from which elements within each group are then sampled (possibly at different rates). The difference between strata and clusters is that every strata appears in the sample while only some clusters are selected. For the purposes of this investigation however, it is adequate to utilize this concept since it captures the fact that the OLS standard errors will be corrected by assuming that classrooms are nested within schools. In addition, the performance of the estimator when clustering is only at the classroom level is also examined.[13]

Similar to the analysis of a stratified cluster sample, a multilevel model also allows the analyst to capture the effects of the school and classroom level variables at the same time. Random numbers are generated at the school and classroom level and both are included in the regression equation. The structure of the data is used to ask the following question: What happens to the rate of rejection when the number of classrooms become smaller? Classrooms are randomly selected from the original data set so that in the first subset the entire data set of classrooms and students are used. In the second subset, only half the classrooms are selected (and all the students in the classrooms are used), and so on until the 10th subset when only 1/10th of the classrooms are randomly selected.[14] For subsets 2 through 10, classrooms are randomly drawn 50 times and the equations are estimated 100 times for each draw. In the full data set, 1000 replications of the equation is estimated. The rate of rejection at the classroom and school level are then tabulated.

Figures 3 and 4 show the results of the simulation. The random number at the school level is rejected more often under OLS and clustered standard errors than in the multilevel models. (CLUSTER refers to classrooms nested within schools while CLUSTER-2 refers to clustering only at the classroom level.) When the classroom level random number is considered, clustered standard errors perform as well as multilevel models until the number of clusters falls to 104. However, multilevel models are less likely to reject the null hypothesis of no significance than clustered standard errors although the difference in rejection rates between the two can be small and in some cases all methods reject more than 5 percent of the time. The standard error correction using stratified clusters which assume that classrooms are nested within schools does not always lead to less rejection than clustering only at the classroom level.

Taken together, the rejection rates at the school and classroom level imply that it is better to model the data as a multilevel model than to fix the standard errors by adjusting for clustering. A school and/or classroom level effect is less likely to be rejected when the null hypothesis of no effect is true.

What if the number of children in the classrooms are increased so that they more resemble the "typical enrollment" of a classroom. The average number of children per classroom is increased so that it ranges from 10 to 30 instead of

[13]In SAS, the former is implemented by assuming that schools are the strata and classrooms are the cluster while the later is implemented by assuming that there is no strata and the classroom is the cluster.

[14]This method of selection is fairly arbitrary in the sense that it does not allow the number of schools to stay fixed. Schools with only one or multiple classrooms may end up entirely out of the sample by chance.

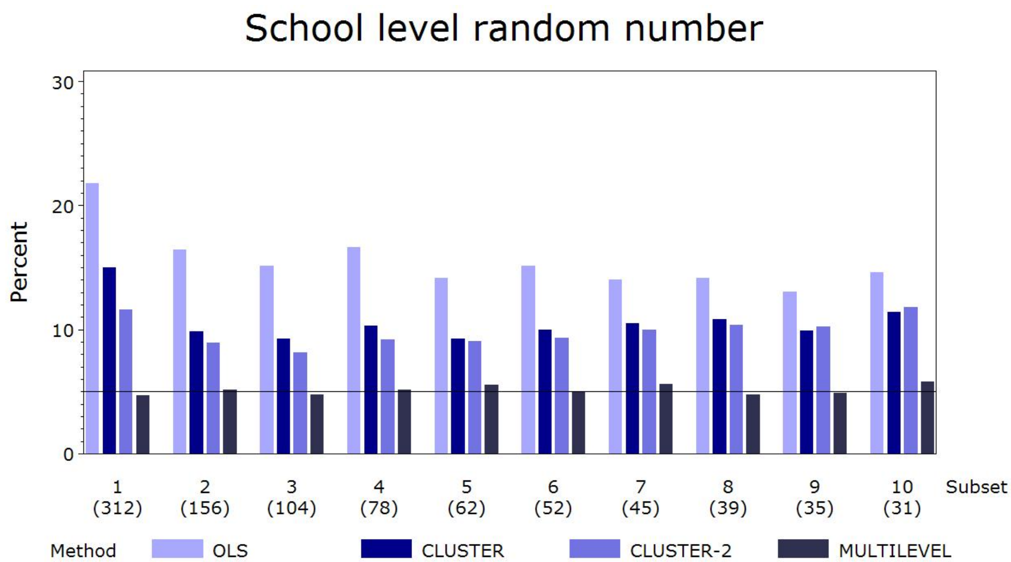

Figure 3: School level rejection rates in multiway clustering when number of clusters (classrooms) are reduced. (Number of classrooms are in parentheses. 100 replications for each method per subset with each sub-sample drawn 50 times except for 312 classrooms with 1000 replications. CLUSTER refers to classrooms nested within schools while CLUSTER-2 refers to clustering only at the classroom level.

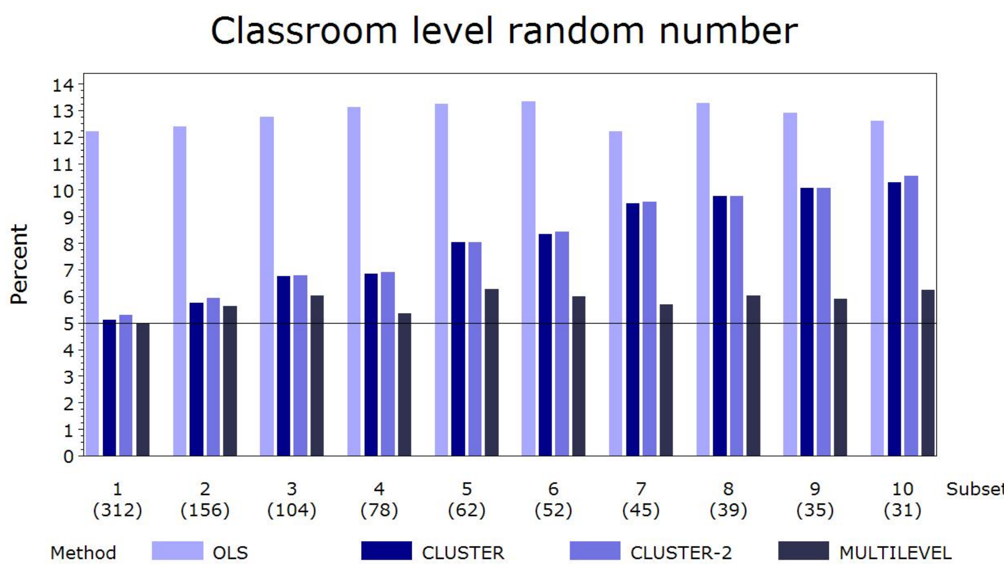

Figure 4: Classroom level rejection rates in multiway clustering when number of clusters (classrooms) are reduced. (Number of classrooms are in parentheses. 100 replications for each method per subset with each sub-sample drawn 50 times except for 312 classrooms with 1000 replications). Reference horizontal line is 5 percent. CLUSTER refers to classrooms nested within schools while CLUSTER-2 refers to clustering only at the classroom level.

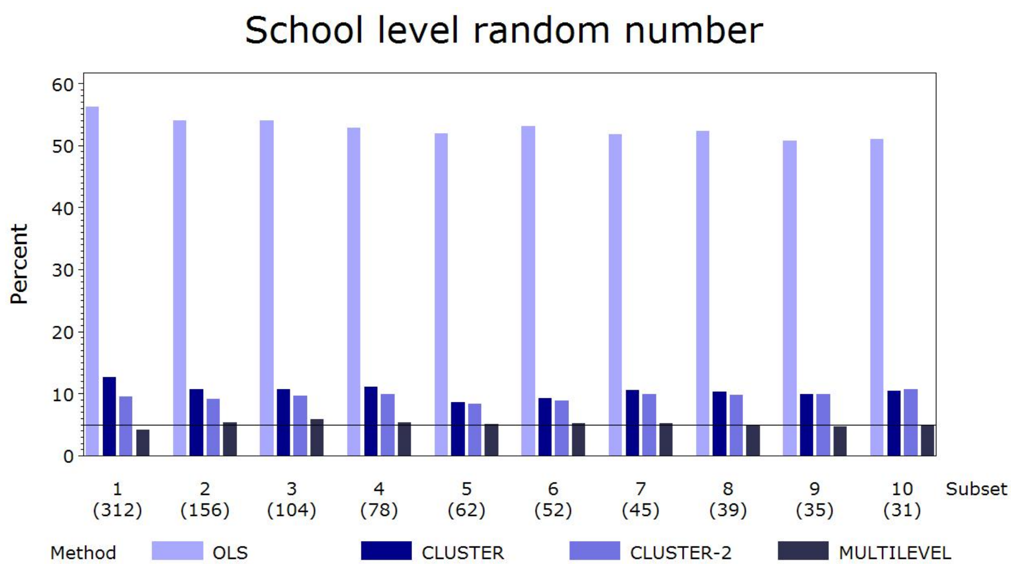

Figure 5: Number of students in the classrooms are increased. School level rejection rates in multiway clustering when number of clusters (classrooms) are reduced. (Number of classrooms are in parentheses. 100 replications for each method per subset with each sub-sample drawn 50 times except for 312 classrooms with 1000 replications. CLUSTER refers to classrooms nested within schools while CLUSTER-2 refers to clustering only at the classroom level.

from 1 to 10.[15] The same exercise is then performed: For subsets 2 through 10, classrooms are drawn 50 times and 100 replications is estimated for each draw.[16] The full data set (subset 1) is estimated 1000 times. Figure 5 and 6shows the results of the estimation. For the school level random variable, the results are similar to those with the smaller classroom size. For the classroom level random number, the tendency to over-reject by clustered standard errors is not as large although it is still larger than multilevel models when the number of clusters fall to 78. Again, the results suggest that it is better to model the data with multilevel models than to cluster the standard error.

[15]To be precise, if the number of children is less than 3 then the number of is multiplied by 10, and if the number of children is between 3 and 8 then the number is multiplied by 4, while those classrooms with 9 or 10 kids are multiplied by 3. The children are identical to those already in the classroom in every respect except that SES and MATHKIND are randomly perturbed to increase the variation between children.

[16]Again, for the full data set, 1000 replications are estimated. Results do not vary when the experiment is repeated with 1000 replications drawn 50 times or 100 replications drawn 100 times.

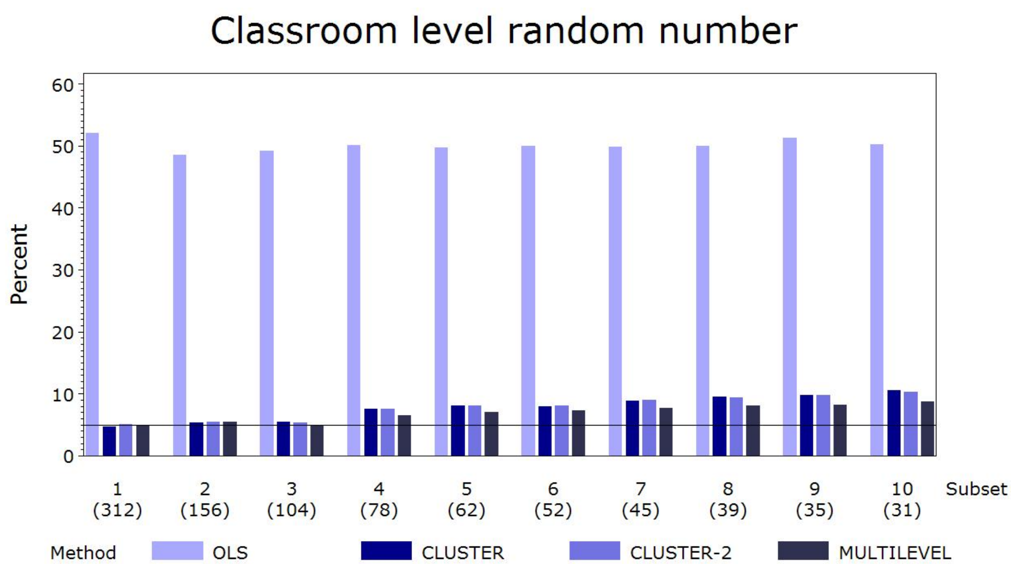

Figure 6: Number of students in the classrooms are increased. School level rejection rates in multiway clustering when number of clusters (classrooms) are reduced. (Number of classrooms are in parentheses. 100 replications for each method per subset with each sub-sample drawn 50 times except for 312 classrooms with 1000 replications. CLUSTER refers to classrooms nested within schools while CLUSTER-2 refers to clustering only at the classroom level.

6 Conclusion

This note introduces the use of multilevel models as an alternative to clustered standard errors that followed Moulton's (1990) recommendation. It extends Moulton's (1990) analysis by using multilevel models as an additional method of estimation. While clustered standard errors outperform OLS standard errors in terms of over-rejection of the null hypothesis of no effects, standard errors from multilevel models outperform both types of estimation methods. Drawing from existing work on survey sampling, the investigation is extended into multiway analysis where students are clustered within classrooms and classrooms are clustered within schools. Clustering is considered at the classroom level only as well as when classrooms are *nested* within schools. Random numbers at the school level are over rejected by all types of standard error correction except for multilevel models. At the classroom level, the over rejection of the null hypothesis of no effect is about the same for multilevel models as for the two methods of correction for clustering when the number of clusters (classrooms) is large. When the number of clusters fall, the over rejection rate for clustered standard errors rises. These simulations suggest that modeling data using multilevel models is a better approach than attempting to fix the standard errors.

7 Appendix: STATA commands to demonstrate equivalence of survey estimation and clustered standard errors

```
webuse highschool
svyset, clear /// clear existing sampling information
svyset school /// artificially assume that school is the
              /// cluster and set the survey sample accordingly
svy: regress weight height
regress weight height, vce(cluster school)
```

References

[1] Bafumi, Joseph and Andrew Gelman (2006), *Fitting Multilevel Models When Predictors and Group Effects Correlate*, presented at the 2006 Annual Meeting of Midwest Political Science Association, Chicago, IL.

[2] Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan (2004), *How Much Should We Trust Differences-in-Differences Estimates?,* The Quarterly Journal of Economics (Vol. 119(1)), pages 249-275, February.

[3] Gelman, Andrew (2006), *Multilevel (Hierarchical) Modeling: What It Can and Cannot Do*, Technometrics (Vol. 48(3), August), pp. 432-435.

[4] Gelman, Andrew and Jennifer Hill (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

[5] Kalton, Graham (1983). *Introduction to Survey Sampling*, Sage University Press.

[6] Moulton, Brent (1990), *An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units*, The Review of Economics and Statistics, (Vol. 72(2)), pp. 334-338.

[8] Primo, David M., Matthew L. Jacobsmeier, and Jeffrey Milyo (2006), Estimating the Impact of State Policies and Institutions with Mixed-Level Data, manuscript.

[7] Topel, Robert (1986). *Local Labor Markets*, Journal of Political Economy, Vol. 94(3) (June 1986), pp. S111-S143.

[8] West B., Welch, K. & Galecki, A (2006). *Linear Mixed Models: A Practical Guide Using Statistical Software*, Chapman Hall / CRC Press.