Measuring Production Efficiency Using Aggregate Data

Francisca G.-C. Richter
Department of Agricultural Economics
Oklahoma State University

Contact Author: B. Wade Brorsen
Department of Agricultural Economics
Oklahoma State University
Stillwater, OK 74078-6026
Phone: (405) 744-6836
FAX: (405) 744-8210
e-mail: **brorsen@okstate.edu**

Annual Meeting of the American Agricultural Economics Association
Chicago, Illinois August 5-8, 2001

Measuring Production Efficiency Using Aggregate Data

Abstract

This paper develops a measure of efficiency when data have been aggregated. Unlike the most commonly used efficiency measures, our estimator handles the heteroskedasticity created by aggregation appropriately. Our estimator is compared to estimators currently used to measure school efficiency. Theoretical results are supported by a Monte Carlo experiment. Results show that for samples containing small schools (sample average may be about 100 students per school but sample includes several schools with about 30 students), the proposed aggregate data estimator performs better than the commonly used OLS and only slightly worse than the multilevel estimator. Thus, when school officials are unable to gather multilevel or disaggregate data, the aggregate data estimator proposed here should be used. When disaggregate data is available, standardizing the value-added estimator should be considered.

Measuring Production Efficiency Using Aggregate Data

1. Introduction

Over the last three decades, resources devoted to education have continuously increased while student performance has barely changed (Odden and Clune 1995). In response to this fact, several states now reward and provide incentives for public schools that perform better than others, based on their own measures of school quality (Ladd 1996). Test scores are used not only by policymakers in reward programs but are also presented in state report cards issued to each school. Already more than 35 states have comprehensive report cards reporting on a variety of issues including test scores and a comparison of school variables with district and state averages. But often the information presented is misleading or difficult to interpret. Accurate information on school performance is needed if report cards and reform programs are to succeed in improving the public school system.

Hierarchical linear modeling (HLM), a type of multilevel modeling, has been recognized by most researchers as the appropriate technique to use when ranking schools by effectiveness. As Webster argues, HLM recognizes the nested structure of students within classrooms and classrooms within schools, producing a different variance at each level for factors measured at that level. Multilevel data, also called disaggregate data is needed to implement HLM. For example, two-level data could consist of variables for students within schools. The value-added framework within the HLM methodology has become popular among researchers (Hanushek, Rivkin, and Taylor 1996; Goldstein 1997; Woodhouse and Goldstein 1998). Value-added regressions are able to isolate school's effect on test scores during a given time period, by using regressors such as previous test scores, and student and school characteristics. But as of 1996, among the 46 out of 50 states that have accountability systems with some type of assessment, only 2 had used value-added statistical methodology in implementing such systems (Webster et al. 1996). Multilevel analysis has been said to involve complicated statistical analyses that school officials are unable to understand (Ladd 1996).

A common approach is to use aggregate data. As opposed to having data for each student within each school, aggregate data refers to having only averages of these data over all students, within a school. School administrators may be able to obtain records of each student's individual test score but may not be able to match them with their parents' income, for example. Therefore, average test scores in a school are matched to the average income in the respective school district.

To obtain a measure of school quality with aggregate data, it is common to regress school mean outcome measures on the means of several demographic and school variables The residuals from this regression are totally attributed to the school effect, and thus, are used to rank schools. Although the use of aggregate data has been widely criticized in the literature (Webster et al. 1996; Woodhouse and Goldstein 1998), many states use aggregate data. This paper purpose proposes a new and more efficient estimator of quality based on aggregate data, and then it compares it with the commonly used OLS estimator as well as with the value-added-disaggregate estimator. Evidently, estimators based on disaggregate data will perform better than any estimator based on aggregate data. The questions that arise are: by how much will their performances differ? Should schools be using OLS, when they can use a more efficient aggregate estimate at no extra cost?

One of Goldstein's main oppositions to aggregate data models is that they say nothing about the effects upon individual students. Also, aggregate data does not allow studying differential effectiveness, which distinguishes between schools that are effective for low achieving students and schools that are effective for high achieving students. The inability to handle differential effectiveness is a clear disadvantage of aggregate as compared to disaggregate data. However, when aggregate data are all that schools have, is it still possible to detect the extreme over and under performing schools? When using OLS on aggregate data, it has been observed that small schools are disproportionately rewarded (Clotfelter and Ladd 1996). The estimator proposed here eliminates that bias.

Woodhouse and Goldstein (1998) argue that residuals from aggregate level regression analysis are highly unstable and therefore, unreliable measures of school efficiency. Woodhouse and Goldstein analyze an aggregate model used in a previous study and show how small changes in the independent variables as well as the inclusion of non-linear terms will change the rank ordering of regression residuals. However, their data set is small and they do not examine whether disaggregate data would have also lead to fragile conclusions.

As of today, most of the research has focused on criticizing the commonly used aggregate data model, which uses OLS residuals to estimate school quality. Goldstein (1995), for example, illustrates the instability of aggregate data models with an example in which he compares estimates coming from an aggregate model versus estimates from several multilevel models showing they are different. The aggregate model, however, does not provide an estimate of the between-student variance, which suggests that the author does not use MLE residuals to estimate school effects. Maximum likelihood estimation is possible since the form of heteroskedasticity for the aggregate model is known (Dickens 1990).

While it is expected that aggregation will attenuate the bias due to measurement error, few researchers have compared aggregate data models versus multilevel models while considering measurement error. Hanushek, Rivkin, and Taylor (1996) analyze the impact of aggregation on specified models aimed at measuring school resource effects on student learning, and find that aggregation produces an ambiguous bias on the estimated regression parameters. Thus they suggest an empirical examination of the effects of aggregation in the presence of measurement error.

Although it has become conventional wisdom that aggregate data should not be used to measure school quality, the literature on which this argument is based on, is insufficient to support the claim. Research comparing aggregate with disaggregate models have used ordinary least squares rather than maximum likelihood estimators so the validity of their criticism is unclear. Standardized efficient estimators of school quality based on aggregate data, as well as their confidence intervals will be developed here and compared to multilevel estimators with and without measurement error. In the process, a standardized version of the value-added multilevel estimator is also proposed and compared. Since many states either continue to use aggregate data or use other less accurate measures to rank and reward schools, the relevance of this issue cannot be denied.

2. Theory

Estimators for the effect of schools on student achievement based on disaggregate data have been developed and reviewed extensively in the education literature, and will be presented only briefly here. However, since aggregate data have been disregarded due to the loss of information that aggregation implies, little effort has been devoted to develop appropriate estimators for aggregate data.

This section consists of three parts. The first part will show how aggregation of a 2-level error components model, with heterogeneous number of first-level units within second-level units, leads to a model with heteroskedastic error terms. Therefore, for estimators of the parameters of the model to be efficient, ML or GLS estimation is required. The aggregate data estimator is presented as well as its standardized version.

The second part derives confidence intervals for the aggregate data estimator and presents the confidence intervals commonly used for disaggregate data. The third part introduces measurement error in the model and derives the bias when estimating the parameters of the explanatory variables in both the disaggregate and aggregate models.

2.1. Aggregation of a Simple 2-Level Error Components Model

Consider the following model:

$$Y_{ij} = (\boldsymbol{X\beta})_{ij} + u_j + e_{ij}, \quad i = 1, \ldots, n_j \quad j = 1, \ldots J, \tag{1}$$

where Y_{ij} is the test score of the ith student in the jth school, $(\mathbf{X}\boldsymbol{\beta})_{ij}$ is the fixed part of the model, likely to be a linear combination of student and school characteristics, such as previous test score (for a value added measure), parents' education, and average parents' income for each school, u_j is the random effect for school, that we are trying to estimate, and e_{ij} is the unexplained portion of the test score, with distributions given by

$u_j \sim iid\ N(0, \sigma_u^2), e_{ij} \sim iid\ N(0, \sigma_e^2), \quad \mathrm{cov}(u_j, e_{ij}) = 0$.

In matrix notation the model is:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z}\,\boldsymbol{u} + \boldsymbol{e}, \tag{1.a}$$

where

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{1}_{\boldsymbol{n_1}} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{1}_{\boldsymbol{n_J}} \end{bmatrix},$$

$$\boldsymbol{Z}\,\boldsymbol{u} + \boldsymbol{e} \sim N(\boldsymbol{0},\, \boldsymbol{V}),$$

$$\boldsymbol{V} = \begin{bmatrix} \sigma_e^2 \boldsymbol{I}_{\boldsymbol{n_1}} + \sigma_u^2 \boldsymbol{J}_{\boldsymbol{n_1}} & & 0 \\ & \ddots & \\ 0 & & \sigma_e^2 \boldsymbol{I}_{\boldsymbol{n_J}} + \sigma_u^2 \boldsymbol{J}_{\boldsymbol{n_J}} \end{bmatrix}.$$

The random effect u_j represents the departure from the overall mean effect of schools on students' scores. While the intercept contains the overall mean effect of schools, u_j measures by how much school j deviates from this mean.

The shrinkage estimator of u_j is (Goldstein 1995):

$$\hat{u}_j = (\sigma_u^2 / (\sigma_u^2 + \sigma_e^2 / n_j))(\textstyle\sum_{i=1}^{n_j} \hat{y}_{ij}) / n_j \qquad (2)$$

$$\hat{y}_{ij} = Y_{ij} - (\boldsymbol{X}\hat{\beta})_{ij},$$

where the $\hat{y}_{ij}$'s are called raw residuals and $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$. So the school effect for school j is estimated by the raw residuals, averaged over all students, and 'shrunken' by a factor that is a function of the variance components and the number of students in the school. The larger the number of students in a school, the closer this factor is to one. But if school size is small, there will be less information to estimate the school effect. Thus, the shrinkage factor becomes smaller, making the estimate of the school effect deviate less from the overall mean.

Now let us see how the model changes with aggregation. Adding over all students within each school,

$$\textstyle\sum_{i=1}^{n_j} Y_{ij} = \sum_{i=1}^{n_j} (\boldsymbol{X}\beta)_{ij} + n_j u_j + \sum_{i=1}^{n_j} e_{ij}$$

and dividing by the number of students in each school, leads to the following model:

$$Y_{.j} = (\boldsymbol{X\beta})_{.j} + u_j + e_{.j}, \quad j = 1, \ldots, J \tag{3}$$

$$u_j \sim iid\ N(0, \sigma_u^2), e_{.j} \sim N(0, \sigma_e^2 / n_j), \quad \operatorname{cov}(u_j, e_{.j}) = 0,$$

where the dot is the common notation to denote that the variable has been averaged over the corresponding index; students in this case. The error term for the aggregated model will be $v_j \sim (0, \sigma_u^2 + \sigma_e^2 / n_j)$.

Again, in matrix notation the model is:

$$\boldsymbol{Y}_a = \boldsymbol{X}_a \beta + \boldsymbol{u} + \boldsymbol{e}_a, \tag{3.a}$$

$$\boldsymbol{X_a} = \begin{bmatrix} \frac{1}{n_1}\boldsymbol{1'_{n_1}} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \frac{1}{n_J}\boldsymbol{1'_{n_J}} \end{bmatrix} \boldsymbol{X}, \quad \boldsymbol{Y_a} = \begin{bmatrix} \frac{1}{n_1}\boldsymbol{1'_{n_1}} \\ \vdots \\ \frac{1}{n_J}\boldsymbol{1'_{n_J}} \end{bmatrix} \boldsymbol{Y}, \quad \boldsymbol{e_a} = \begin{bmatrix} \frac{1}{n_1}\boldsymbol{1'_{n_1}} \\ \vdots \\ \frac{1}{n_J}\boldsymbol{1'_{n_J}} \end{bmatrix} \boldsymbol{e}$$

$$\boldsymbol{u} + \boldsymbol{e}_a \sim N(\boldsymbol{0}, \boldsymbol{V}_a),$$

$$\boldsymbol{V_a} = \begin{bmatrix} \sigma_u^2 + \sigma_e^2 / n_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_u^2 + \sigma_e^2 / n_J \end{bmatrix}$$

We are interested in estimating the random effects u_j's. For this, we estimate the MLE residuals of the error term v_j. We define our estimator as the conditional mean of u_j given v_j, i.e., $\tilde{u}_j = \hat{E}(u_j / v_j)$, This value can be shown to be (see appendix):

$$\tilde{u}_j = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma_e^2 / n_j)} (Y_{.j} - (\boldsymbol{X}\hat{\beta})_{.j}), \tag{4}$$

where $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$ for the aggregate model. Notice that this estimator has the same shrinkage factor as the disaggregate estimator.

However, the school effects in (4) are heteroskedastic, while the true school effects are not. Thus, to correct for heteroskedasticity, we divide the estimator by its standard deviation obtaining the standardized estimator of school effect:

$$\breve{u}_j = \frac{1}{\sqrt{\sigma_u^2 + \sigma_e^2 / n_j}} (Y_{.j} - (\boldsymbol{X}\hat{\beta})_{.j}) \tag{5}$$

Thus, the set of $\breve{u}_j$'s may also be used to rank schools. Similarly, the multilevel estimator in (2) can also be standardized to obtain:

$$\hat{u}_j = (1/(\sqrt{\sigma_u^2 + \sigma_e^2 / n_j}))(\sum_{i=1}^{n_j} \hat{y}_{ij})/n_j \tag{2.a}$$

2.2. Confidence Intervals for the Estimates of School Quality

A confidence interval for school effects is: $\hat{u}_j \pm t_{1-\alpha/2}\sigma_{u|\hat{u}}$. Thus, it is necessary to obtain the conditional variance of the random effect given its estimator; that is, $Cov(\boldsymbol{u} \mid \hat{\boldsymbol{u}})$.

For both, disaggregate and aggregate estimators, the covariance matrix is derived similarly. First it is necessary to obtain the joint distribution of the vector of school effects $\boldsymbol{u}$ and its estimator. For this, notice that in both cases, the estimator is a linear combination of the vector of dependent variables, test scores in our case. Thus the joint distribution can be derived from the joint of $\boldsymbol{u}$ and **Y**. Then, using a theorem from Moser (theorem. 2.2.1, page 29), the conditional covariance matrix of school effects is obtained. A derivation of this covariance matrix is given in the appendix.

The conditional covariance matrix based on the disaggregate estimator is:

$$Cov(\boldsymbol{u} \mid \hat{\boldsymbol{u}}) = \sigma_u^2 \boldsymbol{I} - \sigma_u^4 \boldsymbol{Z'V^{-1}}\left(\boldsymbol{V} - \boldsymbol{X}\left(\boldsymbol{X'V^{-1}X}\right)^{-1}\boldsymbol{X'}\right)\boldsymbol{V^{-1}Z}, \tag{6}$$

The conditional covariance matrix based on the aggregate estimator is:

$$Cov(\boldsymbol{u} \mid \widetilde{\boldsymbol{u}}) = \sigma_u^2 \boldsymbol{I} - \sigma_u^4 \boldsymbol{V}_a^{-1} \left(\boldsymbol{V}_a - \boldsymbol{X}_a (\boldsymbol{X}_a' \boldsymbol{V}_a^{-1} \boldsymbol{X}_a) \boldsymbol{X}_a' \right) \boldsymbol{V}_a^{-1}. \tag{7}$$

2.3. Bias in Estimation Introduced by Measurement Error

Let us consider a two-level model with measurement error. The model is:

$$y_{ij} = (\mathbf{x}\boldsymbol{\beta})_{ij} + u_j + e_{ij}, \qquad i = 1, \ldots, n_j \quad j = 1, \ldots, J \tag{8}$$

$$Y_{ij} = y_{ij} + q_{ij}$$

$$X_{hij} = x_{hij} + m_{hij}, \quad h = 1, \ldots, H$$

$$\text{cov}(q_{ij}, q_{i'j}) = \text{cov}(m_{hij}, m_{hi'j}) = 0$$

$$E(q_{ij}) = E(m_{hij}) = 0$$

$$\text{cov}(m_{h_1 ij}, m_{h_2 ij}) = \sigma_{(h_1, h_2)m}$$

where y_{ij} is the real test score for the i^{th} student in the j^{th} school, q_{ij} is the measurement error for y_{ij}, $q_{ij} \sim N(0, \sigma_q^2)$, Y_{ij} is the observed test score, x_{hij} is the true measure of the h^{th} student or school characteristic corresponding to the i^{th} student in the j^{th} school, m_{hij} is the measurement error for x_{hij}, u_j is the random component for school j, e_{ij} is the residual, and $\sigma_{(h_1, h_2)m}$ is the covariance of measurement errors from two explanatory variables, h_1 and h_2, for the same student. The covariance of measurement errors from any two variables is assumed to be equal for all students regardless of the school they attend.

Following Goldstein (1995), it can be seen that without measurement error, $\boldsymbol{\beta}$ could be estimated by the FGLS estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{x}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{x})^{-1}(\boldsymbol{x}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y})$. But measurement error as defined by model (8) implies that $E(\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}) - E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$; so an unbiased estimator for $\boldsymbol{\beta}$ in the presence of measurement error is proposed by Goldstein (1995) to be:

$$\hat{\beta} = [\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X} - E(\boldsymbol{m}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{m})]^{-1}(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{Y}) . \tag{9}$$

When measurement error is not taken into account, the matrix $E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$ is omitted. Using Goldstein's derivation of $E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$ and realizing that the inverse of V is also a block diagonal with elements $\frac{(n_j - 1)\sigma_u^2 + \sigma_e^2}{\sigma_e^2(n_j\sigma_u^2 + \sigma_e^2)}$ in the diagonal, each element (h_1, h_2) of the $H \times H$ matrix $E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$ can be expressed as

$$\sum_{j=1}^{J} n_j \left\{ (n_j - 1)\frac{\sigma_u^2}{\sigma_e^2} + 1 \right\} \frac{\sigma_{(h_1,h_2)m}}{n_j\sigma_u^2 + \sigma_e^2} . \tag{10}$$

Now let us see how does this omitted matrix, $E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$, compares with the one to be obtained when aggregating the model. Aggregating the true disaggregate model, we obtain:

$$y_{.j} = (\mathbf{x}\boldsymbol{\beta})_{.j} + u_j + e_{.j},\ j = 1, \ldots, J \tag{11}$$

$$Y_{.j} = y_{.j} + q_{.j}$$

$$X_{h.j} = x_{h.j} + m_{h.j}$$

$$\mathrm{cov}(q_{.j}, q_{.j'}) = 0$$

$$E(q_{.j}) = E(m_{h.j}) = 0$$

$$\mathrm{cov}(m_{h_1ij}, m_{h_2ij}) = \sigma_{(h_1,h_2)m} / n_j$$

where notation is as in model (8).

Notice how the covariance of measurement error between any two fixed explanatory variables is reduced in the aggregate model. Now the covariance matrix of the true model is a diagonal matrix with elements defined in the first part of this section; and which will be denoted by $\boldsymbol{V}_a$. Following a procedure analogous to Goldstein's derivation for the disaggregate model, one can obtain the following unbiased estimator of $\boldsymbol{\beta}$ for the aggregate model:

$$\hat{\beta}_a = [\boldsymbol{X}'_a \hat{\boldsymbol{V}}_a^{-1} \boldsymbol{X}_a - E(\boldsymbol{m}'_a \hat{\boldsymbol{V}}_a^{-1} \boldsymbol{m}_a)]^{-1} (\boldsymbol{X}'_a \hat{\boldsymbol{V}}_a^{-1} \boldsymbol{Y}_a), \qquad (12)$$

where the subscript *a* denotes aggregate data. As can be seen, the bias now will depend on $E(\boldsymbol{m}'_a \boldsymbol{V}_a^{-1} \boldsymbol{m}_a)$, an $H \times H$ matrix whose (h_1, h_2) element is

$$\sum_{j=1}^{J} \frac{\sigma_{(h_1,h_2)m}}{n_j \sigma_u^2 + \sigma_e^2}. \qquad (13)$$

As can be seen by comparing values in (10) and (13), the bias in $\boldsymbol{\beta}$ due to measurement error is attenuated in the aggregate model. Bias in the estimation of $\boldsymbol{\beta}$ without accounting for measurement error, is likely to affect the estimators of school effects, as suggested in (2) and (4). This result is worth considering since adjustments for measurement error are seldom made and, as Woodhouse et. al. (1996) argue, different assumptions about variances and covariances of measurement error may lead to totally different conclusions (when ranking schools, for example). Therefore, when not correcting for measurement error, gains from aggregation may somewhat offset the negative consequences of aggregation. Then, at least asymptotically, aggregate estimates of school effects may be less inaccurate than what researchers have claimed.

However, to examine the properties of our aggregate and disaggregate estimators of school effects in small samples, a Monte Carlo study will be necessary. Also, from the study we will be able to compare the estimators' asymptotic and small sample behavior.

3. Data and Procedures

A Monte Carlo study was used to compare aggregate and disaggregate estimates of school effects with their true values. These values were also compared to OLS estimates with aggregate data since this is what is most often done. The model on which the data generating process was based, was taken from Goldstein's 1997 paper, table 3, page 387, since it was simple, and provided estimates of the random components for school and student, based on real data.

This model regresses test scores of each student against a previous test score, a dummy variable for gender, and a dummy for type of school (boys', girls', or mixed school). Test scores were transformed from ranks to standard normal deviates. The random part consists of the school effect and the student effect.

According to Goldstein, multilevel analysis provides the following estimated model:

$$\hat{T}score_{ij} = \text{-}0.09 + 0.52 Pscore_{ij} + 0.14 Girl_{ij} + 0.10 GirlsSch_j + 0.09 BoysSch_j ,$$

$$i = 1, \ldots, n_j \quad j = 1, \ldots J . \tag{14}$$

The estimated variance of school effects, also called between-school variance, is $\hat{\sigma}_u^2 = 0.07$, and the variance of student effects, also called within-school variance, is $\hat{\sigma}_e^2 = 0.56$. These values and the estimates of the fixed part of the model were used to generate the disaggregate data using SAS. At each replication a number of n_j observations were generated for each school, where n_j was a random realization of a lognormal distribution with mean equal to 100 and variance equal to 50000. Lagged test scores were generated from a standard normal. Dummy variables were generated from binomial distributions. The random components of the model for school and student were generated using a normal with zero mean and variance

$\hat{\sigma}_u^2 = 0.07$ and $\hat{\sigma}_e^2 = 0.56$ respectively, and the actual test score was obtained as in equation (2). Then measurement error was introduced to the previous and actual test scores. Measurement error was assumed to be a normal random variable with a zero mean and a standard deviation of 0.2. All dummy variables are assumed measured without error.

Once a disaggregate data set is generated, estimates for school effects and variance components are obtained using multilevel analysis as provided by the Mixed procedure in SAS. Then, the disaggregate data set is aggregated by schools. Residuals as well as the two components of the variance of the error term are estimated using NLMIXED in SAS. At this point, we will have a set of 100 true school effects (since the number of schools in the sample is 100), and two sets of estimated school effects using aggregate and disaggregate data. Each of these sets generates a ranking of the schools in the sample. The greater the school effect, the better the school's performance, and therefore, the higher its position will be in the ranking. We will also have standardized rankings for each estimate and the OLS estimate of school effects to see how this set compares to the alternative estimators and to the true ranking. Finally, we compute the estimated variance components under both approaches and compare them with the true values.

A comparison of the school effect estimators is done in several different ways. Spearman's correlation coefficient is calculated for all estimators in order to measure the degree of correlation of each ranking with the true schools ranking. Another measure used for comparison is the root mean squared error of the estimates, and finally we compare the top-ten set of schools obtained with each estimator, with the true top-ten set. The whole process described above constitutes a single iteration of the Monte Carlo study. As many as 1000 iterations were conducted.

As many iterations as needed can be performed for each set of parameter values of interest. In particular, outcomes with and without measurement error are compared in order to see if the aggregate estimator is in fact more robust to errors in measurement than the disaggregate estimator. The parameters used to randomly generate the number of students in each school are also changed, to corroborate the theory's suggestion that as schools in the sample grow larger, the difference in the estimators' performance will narrow[1].

4. Results

Table 1 shows the first set of results for 1000 samples, each of 100 schools whose size is distributed lognormal with mean 100 and variance 50000. As expected, the disaggregate estimator performs best on almost all measures. The aggregate estimator's performance, however, is surprisingly good, and clearly above the OLS estimator's performance. OLS in fact tends to reward small schools. The average school size for the top ten schools as estimated by OLS is about 76, while the true average for this group is about 99. However, table 1 also shows that both the aggregate and disaggregate estimators tend to reward large schools. This can be explained as follows: OLS estimators are based on residuals whose variance is $\sigma_u^2 + \sigma_e^2 / n$. So, small schools will have a larger variance and will be more likely to be either at the bottom or top of the rankings.

The aggregate and disaggregate estimators have a shrinkage factor that compensates for these large residuals by reducing the residuals of small schools. Recall the shrinkage factor is $\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 / n}$. This factor is always less than one, but decreases with school size, bringing down the

[1] This is because the shrinkage factor tends to one and also because the larger the sample, the closer averages are to their true means.

absolute value of small school residuals. Results in table 1 suggest that the shrinkage factor may over-compensate for the residuals effect, and thus, leave only large schools in the extremes. Estimators with a smaller shrinkage factor (the factor is $\frac{1}{\sqrt{\sigma_u^2+\sigma_e^2/n}}$) such as the standardized aggregate (equation 5) and standardized disaggregate estimators seem to alleviate this problem. Table 1 shows how the average size for the top ten schools according to the standardized estimators only differs by two or three students from the true top-ten group size average. These standardized estimators also seem to give a somewhat better match than their non-standardized versions when determining how many of the real top ten schools are selected by the estimators.

When measuring the root mean squared error (RMSE) of the estimators with the true ranking we find again that the disaggregate estimator performs only slightly better than the aggregate estimator. For the standardized estimators, the RMSE's were calculated using the standardized true rankings, and thus, cannot be compared to the non-standardized versions. Since we are measuring the performance of the estimators by their ability to match the true ranking and not the true values of the school effects, the RMSE might not be as good of a measure as all the others presented in the table.

The between- and within-school variance estimates are presented in Table 1. Although the aggregate point estimates are very close to the true variances, by looking at the standard deviations of these estimates, it is clear that aggregation will always reduce the ability to estimate the within schools variance as compared to the disaggregate estimator.

Table 2 introduces measurement error as 20% of the highest possible test score. We had hypothesized that measurement error would have less effect on the aggregate estimators. This is true but almost unperceivable, considering that a 20% measurement error is high. Thus, measurement error is relatively unimportant in this case.

Finally, table 3 shows the results for ranking estimates when schools have on average 350 students. As school size increases, the variation in averaged residuals due to students (σ_e^2 / n) becomes insignificant. This implies that aggregation becomes less of a concern for estimating school effects (thus, the aggregate and disaggregate estimators should perform more alike now), and heteroskedasticity is almost insignificant (thus OLS is not as bad of a choice as before). In fact, table 3 shows differences among ranking measures have narrowed for all estimators, and that the problem with small or large schools being consistently rewarded, has almost disappeared. However, aggregate data will no longer be able to estimate the variance components of the model with any accuracy.

5. Conclusions

Researchers argue that value-added multilevel models provide the most accurate measures of school quality. But most states continue to use aggregate data (usually not in a value added framework) to rank and reward schools. Research criticizing aggregate models, by comparing them with disaggregate models, have used ordinary least squares rather than maximum likelihood estimators so part of their criticism is uncertain. States need to know the correct way to handle aggregate data and how much accuracy is lost by using aggregate data. Efficient estimators of school quality based on aggregate data and confidence intervals are derived here and compared to multilevel and OLS estimators with and without measurement error. A Monte Carlo study is used in order to perform this comparison that includes measuring the correlation of aggregate versus disaggregate estimates with the true values of school effects.

Results show that when many small schools are present in the data, the proposed aggregate data estimator performs better than OLS on aggregate data, and only slightly worse than the disaggregate data estimator. However, as school size increases, the three estimates perform more alike.

Even though the aggregate data estimator is only slightly worse than the disaggregate data estimator for ranking schools based on efficiency, we still want to encourage the collection of disaggregate data because of their many uses in understanding school quality and student learning.

Also, OLS estimators do tend to reward small schools over bigger ones, as the empirical literature has shown, while the shrinkage disaggregate estimator unexpectedly rewards large schools. A standardized version of this estimate is presented that eliminates this problem.

Thus, when school officials are able to collect multilevel data, this study suggests they consider standardizing the estimates of school quality before ranking schools. However, when disaggregate data are not available, and small schools are present in the sample the standardized aggregate estimator proposed here should be used over the OLS approach.

The methods proposed and evaluated here provide a one-dimensional measure that can be used to understand school quality. However, an efficiency measure based on standardized test scores is not the only measure that should be considered when evaluating schools. This study provides new information about the strengths and weaknesses of alternative methods and data. Our application is to schools, but these results are applicable to measuring efficiency in any industry where aggregate data may be the only data available.

5. References

Clotfelter, C. T., & Ladd, H. F. (1996) Recognizing and Rewarding Success in Public Schools in H. F. Ladd (Ed.), Holding Schools Accountable. Performance-Based Reform in Education Washington, D.C.: The Brookings Institute.

Dickens, W. (1990). Error Components in Grouped Data: Is it Ever Worth Weighting? Review of Economics and Statistics, 328-333.

Goldstein, H. (1995). Multilevel Statistical Models. London: Edward Arnold.

Goldstein H. (1997). Methods in School Effectiveness Research.. School Effectiveness and School Improvement, 8, 369-395.

Hanushek, E. A., Rivkin S. G. & Taylor L. L. (1996). Aggregation and the Estimated Effects of School Resources. The Review of Economics and Statistics, 2, 611-627.

Ladd, H. F. (1996). Catalysts for Learning. Recognition and Reward Programs in the Public Schools. Brookings Review, 3, 14-17.

Moser, B. K. (1996) Linear Models: A Mean Model Approach. California: Academic Press.

Odden, A., & Clune W. (1995) Improving Educational Productivity and School Finance. Educational Researcher, 9, 6-10.

Webster, W. J., Mendro, R. L., Orsak, T. H., & Weerasinghe, D. (1996) The Applicability of Selected Regression and Hierarchical Linear Models to the Estimation of School and Teacher Effects. Paper presented at the annual meeting of the National Council of Measurement in Education, New York, NY.

Woodhouse, G., & Goldstein, H. (1998). Educational Performance Indicators and LEA League Tables. Oxford Review of Education, 14, 301-320.

Woodhouse, G., Yang, M., Goldstein, H., & Rasbash, J. (1996) Adjusting for Measurement Error in Multilevel Analysis. Journal of the Royal Statistical Society A, 159, 201-212.

Table 1. Comparison of estimates of school quality using aggregate vs. disaggregate data with no measurement error.

Measure	Type of estimator	Mean	Std.Dev.
Spearman	Disaggregate	0.8527	0.0341
	Std. disaggregate	0.8444	0.0375
	Aggregate	0.8431	0.0367
	Std. aggregate	0.8373	0.0399
	OLS	0.8167	0.0451
RMSE	Disaggregate	0.1332	0.0136
	Std. disaggregate	0.5378	0.0573
	Aggregate	0.1416	0.0164
	Std. aggregate	0.0591	0.0592
	OLS	0.1873	0.0284
Top Ten	Disaggregate	6.50	1.203
	Std. disaggregate	6.53	1.178
	Aggregate	6.33	1.284
	Std. aggregate	6.42	1.207
	OLS	6.01	1.258
School Size Avg.	Real Group	98.96	69.27
In Top Ten Group	Disaggregate	126.13	85.78
	Std. disaggregate	101.60	73.26
	Aggregate	126.93	82.45
	Std. aggregate	102.14	73.58
	OLS	76.34	60.19
Variance Estimates	Dis. Within Sch.	0.560	0.008
	Dis. Between Sch.	0.070	0.013
	Agg.Within Sch.	0.572	0.342
	Agg Between Sch.	0.067	0.016

Note: Results are for 1000 simulations, each including 100 schools. The number of students per school is a lognormal random variable with mean 100 and variance 50000. Mean is the average over all simulations, RMSE is root mean squared error, Top Ten is the average number of schools ranked in the top ten with the estimator, that belong to the true top ten set. Estimators compared are the disaggregate estimator, its standardized version, the aggregate estimator, its standardized version, and the OLS estimator of school effects. Variance estimates are also presented for the disaggregate and aggregate methods.

Table 2. Comparison of estimates of school quality using aggregate vs. disaggregate data with measurement error.

Measure	Type of estimator	Mean	Std.Dev.
Spearman	Disaggregate	0.8445	0.0346
	Std. disaggregate	0.8362	0.0381
	Aggregate	0.8391	0.0363
	Std. aggregate	0.8330	0.0394
	OLS	0.8119	0.0455
RMSE	Disaggregate	0.1364	0.0135
	Std. disaggregate	0.5513	0.0575
	Aggregate	0.1433	0.0165
	Std. aggregate	0.0561	0.0590
	OLS	0.1925	0.0302
Top Ten	Disaggregate	6.43	1.236
	Std. disaggregate	6.42	1.192
	Aggregate	6.30	1.260
	Std. aggregate	6.35	1.205
	OLS	5.94	1.244
School Size Avg.	Real Group	103.40	86.84
In Top Ten Group	Disaggregate	131.58	96.26
	Std. disaggregate	104.83	89.76
	Aggregate	132.50	96.79
	Std. aggregate	107.39	90.93
	OLS	78.79	75.18
Variance Estimates	Dis. Within Sch.	0.610	0.009
	Dis. Between Sch.	0.071	0.013
	Agg.Within Sch.	0.611	0.361
	Agg Between Sch.	0.067	0.016

Note: Results are for 1000 simulations, each including 100 schools. The number of students per school is a lognormal random variable with mean 100 and variance 50000. Measurement error is 20% in actual and previous scores. Mean is the average over all simulations, RMSE is root mean squared error, Top Ten is the average number of schools ranked top ten with the estimator, that belong to the true top ten set.

Table 3. Comparison of estimates of school quality using aggregate vs. disaggregate data for large schools.

Measure	Type of estimator	Mean	Std.Dev.
Spearman	Disaggregate	0.9620	0.0140
	Std. disaggregate	0.9620	0.0140
	Aggregate	0.9588	0.0178
	Std. aggregate	0.9606	0.0168
	OLS	0.9558	0.0185
RMSE	Disaggregate	0.0693	0.0114
	Std. disaggregate	0.2718	0.0440
	Aggregate	0.0854	0.0330
	Std. aggregate	0.2794	0.0521
	OLS	0.0746	0.0140
Top Ten	Disaggregate	8.17	0.967
	Std. disaggregate	8.16	0.974
	Aggregate	7.94	1.139
	Std. aggregate	8.09	1.029
	OLS	8.05	1.043
School Size Avg.	Real Group	348.73	73.10
In Top Ten Group	Disaggregate	351.07	71.59
	Std. disaggregate	347.13	70.57
	Aggregate	375.02	71.02
	Std. aggregate	359.47	67.65
	OLS	345.76	70.64
Variance Estimates	Dis. Within Sch.	0.610	0.005
	Dis. Between Sch.	0.071	0.011
	Agg.Within Sch.	2.344	3.447
	Agg Between Sch.	0.060	0.016

Note: Results are for 100 simulations, each including 100 schools. The number of students per school is a lognormal random variable with mean 350 and variance 50000. Measurement error is 20% in actual and previous scores. Mean is the average over all simulations, RMSE is root mean squared error, Top Ten is the average number of schools ranked top ten with the estimator, that belong to the true top ten set.

Appendix

Derivation of the aggregate estimators of school effects

Recall equation (3.a) which shows the aggregate model:

$$\boldsymbol{Y}_a = \boldsymbol{X}_a \beta + \boldsymbol{u} + \boldsymbol{e}_a$$

However, the aggregate model has no way of differentiating among its random terms, thus we rewrite the model as:

$$\boldsymbol{Y}_a = \boldsymbol{X}_a \beta + \boldsymbol{w}.$$

We are to obtain the conditional mean of $\boldsymbol{u}$ given the total residual $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{e}_a$ based on the distributions of $\boldsymbol{u}$ and $\boldsymbol{e}$.

Since $\boldsymbol{u}$ and $\boldsymbol{e}$ are independent normal random vectors, its distribution is given by:

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} \sim N(0,\ \boldsymbol{V}_{\boldsymbol{u,e}})$, where $\boldsymbol{V}_{\boldsymbol{u,e}} = \begin{bmatrix} \sigma_u^2 \boldsymbol{I}_{\boldsymbol{J}} & 0 \\ 0 & \sigma_e^2 \boldsymbol{I}_N \end{bmatrix}$, $\boldsymbol{N}$ being the total number of students.

But $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_a \end{pmatrix}$ is a linear combination of $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix}$, this is:

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_a \end{pmatrix} = \boldsymbol{A_1} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_J & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \frac{1}{n_1}\boldsymbol{1}'_{\boldsymbol{n_1}} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \frac{1}{n_J}\boldsymbol{1}'_{\boldsymbol{n_J}} \end{bmatrix} \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix}$, where $\boldsymbol{1}_{\boldsymbol{n_j}}$ is an n_j vector of 1's. Thus, its distribution will be as follows:

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_a \end{pmatrix} \sim N\left(\boldsymbol{0}, \boldsymbol{A_1}\, \boldsymbol{V}_{\boldsymbol{u,e}}\, \boldsymbol{A}_1'\right).$$

From this random vector, we construct $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{pmatrix}$ pre-multiplying $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_a \end{pmatrix}$ by $\boldsymbol{A_2} = \begin{pmatrix} \boldsymbol{I}_J & \boldsymbol{0} \\ \boldsymbol{I}_J & \boldsymbol{I}_J \end{pmatrix}$. Then, its distribution will be:

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{pmatrix} \sim N\left(\boldsymbol{0}, \boldsymbol{A_2}\ \boldsymbol{A_1}\ \boldsymbol{V_{u,e}} \boldsymbol{A_1}'\ \boldsymbol{A_2}' \right).$$

Having the joint distribution of $\boldsymbol{u}$ and $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{e}_a$, our estimator is easily derived (Moser, theorem 2.2.1) as:

$$E(\boldsymbol{u} \mid \boldsymbol{w}) = Cov(\boldsymbol{u}, \boldsymbol{w})\ Cov(\boldsymbol{w})^{-1} (\boldsymbol{w}) = \begin{bmatrix} \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 / n_1} w_1 \\ \vdots \\ \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 / n_J} w_J \end{bmatrix}$$

Derivation of the conditional covariance matrix $Cov(\boldsymbol{u} / \hat{\boldsymbol{u}})$

Disaggregate data: Recall equation (1.a):

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{1}_{n_1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{1}_{n_J} \end{bmatrix}$$

$$\boldsymbol{Zu} + \boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{V})$$

The shrinkage estimator of school effects (equation 2) in matrix notation is:

$$\hat{\boldsymbol{u}} = \sigma_u^2 \boldsymbol{Z}' \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \text{ or } \hat{\boldsymbol{u}} = \sigma_u^2 \boldsymbol{Z}' \boldsymbol{V}^{-1} (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}' \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{V}^{-1}) \boldsymbol{Y} \qquad (*)$$

This shows clearly that the shrinkage estimator is a linear combination of the independent variable vector.

Thus, we can derive the joint distribution of $(\boldsymbol{u}, \hat{\boldsymbol{u}})'$ by knowing the distribution of $(\boldsymbol{u}, \boldsymbol{Y})'$.

The distribution of $(\boldsymbol{u}, \boldsymbol{Y})'$ is: $\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{Y} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X\beta} \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \sigma_u^2 \boldsymbol{Z}' \\ \sigma_u^2 \boldsymbol{Z} & \boldsymbol{V} \end{bmatrix} \right).$

In general, $\boldsymbol{u}$ and any linear combination of $\boldsymbol{Y}$ of the form $\hat{\boldsymbol{u}} = \boldsymbol{AY}$, will be jointly distributed as follows:

$$\begin{bmatrix} \boldsymbol{u} \\ \hat{\boldsymbol{u}} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}\beta \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \sigma_u^2 \boldsymbol{Z}'\boldsymbol{A}' \\ \sigma_u^2 \boldsymbol{AZ} & \boldsymbol{AVA}' \end{bmatrix}\right)$$

Then, by Moser's theorem 2.2.1, the conditional covariance is:

$$Cov(\boldsymbol{u} \mid \hat{\boldsymbol{u}}) = \sigma_u^2 \boldsymbol{I} - \sigma_u^4 \boldsymbol{Z}'\boldsymbol{A}'(\boldsymbol{AVA}')^{-1}\boldsymbol{AZ} .$$

Equation (6) is obtained by replacing $\boldsymbol{A}$ with $\sigma_u^2 \boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1})$, from (*), in the expression above.

Aggregate data: Again, we will use the same argument. First, re-express the aggregate estimators of school quality in matrix notation:

$\tilde{\boldsymbol{u}} = \sigma_u^2 \boldsymbol{V}_a^{-1}(\boldsymbol{Y}_a - \boldsymbol{X}_a \hat{\beta})$, or $\tilde{\boldsymbol{u}} = \sigma_u^2 \boldsymbol{V}_a^{-1}(\boldsymbol{I} - \boldsymbol{X}_a(\boldsymbol{X}_a'\boldsymbol{V}_a^{-1}\boldsymbol{X}_a)^{-1}\boldsymbol{X}_a'\boldsymbol{V}_a^{-1})\boldsymbol{Y}_a$ (**)

The distribution of $(\boldsymbol{u}, \boldsymbol{Y}_a)'$ is:

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{Y}_a \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}_a\beta \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \sigma_u^2 \boldsymbol{I} \\ \sigma_u^2 \boldsymbol{I} & \boldsymbol{V} \end{bmatrix}\right)$$

So, the distribution of $\boldsymbol{u}$ and $\boldsymbol{AY_a}$, a linear combination of $\boldsymbol{Y_a}$ is:

$$\begin{bmatrix} \boldsymbol{u} \\ \tilde{\boldsymbol{u}} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}_a\beta \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \sigma_u^2 \boldsymbol{A}' \\ \sigma_u^2 \boldsymbol{A} & \boldsymbol{AV}_a\boldsymbol{A}' \end{bmatrix}\right),$$

and the conditional covariance matrix is:

$$Cov(\boldsymbol{u} \mid \tilde{\boldsymbol{u}}) = \sigma_u^2 \boldsymbol{I} - \sigma_u^4 \boldsymbol{A}'(\boldsymbol{AVA}')^{-1}\boldsymbol{A} .$$

When $\boldsymbol{A} = \sigma_u^2 \boldsymbol{V}_a^{-1}(\boldsymbol{I} - \boldsymbol{X}_a(\boldsymbol{X}_a'\boldsymbol{V}_a^{-1}\boldsymbol{X}_a)^{-1}\boldsymbol{X}_a'\boldsymbol{V}_a^{-1})$, we obtain equation (7).