A Monte Carlo Study of Migration and Child Educational Production: Aggregated vs. Disaggregated Resource Modeling

Thomas A. Mroz and Guo Li*

February 26, 2013

Abstract

This paper studies the sensitivity of estimates on various assumptions about aggregation in modeling the school's effect in child educational production. Building a structural model to control the endogeneity of school qualities in the production function, the authors uses Monte Carlo simulations to evaluate the performance of a "correct" aggregation educational production model versus simple aggregation educational production model in estimating school resources' effect on academic outcome. Comparion of both specifications to the benchmark model without aggregation shows that the simple aggregation of school resources over a geographic area causes serious specification errors, and thus generate biased estimates for the marginal contribution of the school resources to test scores. Fortunately, such biasedness can be minimized by using the "correct" aggregation specification.

1 Introduction

Research on the estimated effect of additional resources to local schools has direct implications for tax policies and government budgets. If school

*NIH/NICHD grant (1R01HD047213) provides partial support for this research.We thank Palmetto Cluster at Clemson University, Ithaca and Hokieone Cluster at Virginia Tech. for their help to run large Fortran jobs crucial to the success of conducting this research.

policies are undertaken with a biased estimate of school inputs, they can cause an enormous waste of society's resources. For example, it is mentioned in Hanushek (1997) that pupil-to-teacher ratio has been found as a major force driving student outcome, which leads the average pupil-to-teacher ratio in US public schools to fall from 28 percent to less than 16 percent over the 1940 to 1990 period. Even with this drop in the pupil-to-teacher ratio, the test scores of primary and secondary school students show no improvement across the nation during this time period.

In studying the effect of schools on educational reform, the literature inputs a wide range of efforts. According to a recent review by Hanushek, Rivkin and Taylor (1996), there were 277 separate investigations of the school quality indicator: pupil-to-teacher ratio, and 163 studies about one of the other quality measure of schools - expenditure per pupil. Endless research efforts notwithstanding, little consensus has been reached about the magnitude or even the direction of the school's role in a child's education. Some studies in school performance yield a simple conclusion that there is no strong or consistent relationship between school resources and student performance (Childs and Shakeshaft (1986), Glass and Smith (1979)). Conversely, the widely publicized findings of Card and Krueger (1990), together with several other studies (Johnson and Stafford (1973), Link and Ratledge (1975), Rizzuto and Wachtel (1980)) indicate that variations in school resources are related to returns to education.

To explain the discrepancies of findings about school effects, several studies proposes a number of reasons. Firstly, Todd and Wolpin (2007) posit that most of the previous estimations of the child educational production function failed to accommodate the fact that educational policies and household behavior interact to determine student outcomes. The positive raw correlation between school spending and test outcome disappears when family background is controlled in the estimation. Secondly, Heckman, Layne-Farrar and Todd (1996) points out that endogeneity biases take effect when there are correlations among the different inputs from school, family, and student preparations. For example, parents make systematic school choices through migration, meaning that "good" parents self-select themselves into better school districts. This belief confounds the task of sorting out the school's effect from the parents' effect. Thirdly, Loeb and Bound (1995)[1], and Hanushek et al. (1996) show the importance of data characteristics in explaining the conflict in the findings of school effectiveness studies. They point out that significant positive school effects frequently appear in educational outcome studies using aggregate data, while strong school effects are not found when micro level data are used. As suggested in Hanushek et al. (1996), aggregation can alter the magnitude of omitted variable bias and implies upward bias of estimated school resource effects.

This paper set up a structural model to simultaneously address the endogeneity problem of school choices together with the aggregation problems in measuring school resources. Folding the migration decisions of parents into the child educational production function, we specify a random parameters model for the collective decisions about school choices and parental input - i.e., in the joint estimation of school choices and the educational production function, we allow the educational outcome to depend on the same unmeasured family preference factor that affects the family's school choice. Parents are assumed to choose a school district based on the quality of that particular school district, and student outcome is achieved based on the quality of the chosen school district. In order to show the aggregation bias caused by assuming both the location choice outcome and student outcome are based on the school qualities aggregated to the county level, we compare our "correct" aggregation model with a simple aggregation model in a Monte Carlo study based on school district information extracted from 1994 Common Core of Data[2]. We find that the widely used production function model with aggregated resources can cause serious specification errors that increase the endogeneity bias. We experiment with different assumptions about the heterogeneity pattern of the unobserved preference factor on the three models and find that specification errors generated by simple

[1] Loeb and Bound (1995) found large school effects using division level aggregation data. They argue that the difference in data characteristics could crucially affect estimation "more than the differences in outcome measures or biases from labor market influences".

[2] Survey conducted by National Center for Educational Statistics (NCES).

aggregation can be overcome by "correctly" aggregating the resources over a geographic area.

The paper is organized as follows: Section II provides a brief review of the relevant literature. Section III lists the three different modeling specifications of the joint distribution of school choice and educational production function, and section IV applies the three models to a Monte Carlo experiment. Section V provides concluding remarks.

2 Background

Early studies of school effectiveness, e.g. Oates (1969) and Kain and Quigley (1970) focus on the linear relationship between school resources and child educational achievement or local housing value outcomes[3]. This group of studies, which is recognized as "hedonic pricing" model or "linear test outcome projections", has a number of shortcomings. The first is related to the fact that school inputs and family inputs are both important factors affecting child academic achievement. Estimations of school effects could be severely biased if there is no control for family background, the models implicitly assume that the marginal impact of school effects is the same across all socioeconomic background or academic skill of the students. The second problem proven to be more crucial in the estimation arises from the fact that school input and parental inputs are inter-correlated. The school resources a child is receiving is endogenous because they are not "given", but "chosen" by his/her parents through sorting on locations. Tiebout (1956) and Bayer (2000) emphasizes the importance of the systematic location choice of families and the possible impact of that on the performance of the family members. This unobserved preference factor is quite important, as it not only affects location choice of the family, but also affects the choices concerning their children's education, including choices among various complementary programs that aid in the learning and choices of helping children with their daily studies.

[3]Studies of housing values shows that school inputs have been capitalized into housing price (Blackburn, Bloom and Freeman (1991)).

Hanushek et al. (1996) use a linear model under omitted variable and measurement error assumptions to demonstrate that the aggregation of all school characteristics to the state level in the production function unambiguously bias the schooling parameters upward if some of the state level impact factors are omitted in the estimation. For example, some states may improve teacher competence regulations that ensure the better quality of teachers while maintain the same state average key school quality metrics[4]. The estimated effects of schools would undoubtedly bias upwards if the teacher competence regulations are neglected in the estimation. Hanushek et al. (1996) highlighted the reasons why a perfect control of between-state school characteristics variations is hard to achieve: Schools in the United States are organized by different states and counties and thus follow different types of policies depending on state resources/preference. For example, "Some 37 states have forms of teacher competency testing, while others do not, and details about the requirements for teachers vary a lot", "Policies are different across states for teacher tenures", "States also vary in terms of requirement for graduating with high school diploma". While controlling for these unmeasured factors is difficult, these unmeasured state regulatory heterogeneity bias the estimates upward to a certain level.

By jointly modeling the location/school choice of parents and the student outcome, our analysis has a number of advantages Firstly, it emphasizes the importance of the endogeneity of school characteristics on the estimation of the effect of schools on student academic achievement. Since parents valuing better schools raise the other (unmeasured) parental inputs to their child's education, the unmeasured family preference on school choice affects the student test outcome. Secondly, this model has the ability to account for the local/state level unobserved school quality indicators since families look at both the observed (e.g., pupil-to-teacher ratio, teacher salary, dropout rate, etc.) and unobserved (e.g., state regulations on teacher competency test, etc.) factors when choosing a school district for their children. With this endogeneity issue controlled for in the model, we access the performance

[4] Pupil-to-teacher ratio, Expenditure per pupil, and Dropout rate are frequently used in the estimation of the production function.

of different modeling specifications (aggregated versus disaggregated) on the estimation of child educational production function. While studies usually focused on county level or state level measures of schools, we suspect that parents in their decision making almost certainly focus on school characteristics at a finer level of geographical detail. We explore a plausible modeling specification that "correctly" aggregates school resources over particular geographic areas when exact school choices of parents are unknown in the study. This model successfully deals with two aspects of aggregation bias in the typical examples of school effective studies: 1) the bias caused by the aggregated local public goods that a family is comparing from the choice set when they make their residential decision. 2) The bias caused by linking the aggregated school resources to an individual in the educational production function.

3 Modeling Specifications

The model follows a simple time line: at the first stage, parents make a school choices based on the local offer of public goods. At the second stage, the student's educational outcome is achieved based on the quality of schools, family background and other preparation factors. The aim is to control the non-random sorting of households across locations and schools, and correctly aggregate the district information if aggregation over a certain geographic region is required.

3.1 Full Location Information Model

Families are assumed to make their location choices within a school district's boundary. Each family with a child over the age of five is faced with a choice from all school districts in the United States. Since about 34% of US counties contain more than one school district, consider the following model. Suppose that the nation has K counties where $k = \{1, ..., K\}$. For each k labeled county, there are J_k school districts where each school district can be described as j_k =$\{1, ..., J_k\}$. Families make a location choice within a typical school district. In particular, they consider the school quality of the destination Ω_{j_k}, the expected income from the location, which is idiosyncratic to the parents' characters, and the travel distance between the current location of the family and the possible location from the choice set (Z_{i,j_k}). There is also an unobserved preference factor, μ_i, showing how much the parents care about the school qualities, that affects the decision. The expected utility of choosing each location is specified as:

$$U_{i,j_k} = U_{i,j_k}(\Omega_{j_k}, Z_{i,j_k}, \mu_i) + \varepsilon_{i,j_k} \tag{1}$$

Assuming the error term ε_{i,j_k} follows an *i.i.d.* extreme value distribution, the possibility of this family i choosing school district j_k for their children, controlling the unobserved preference factor μ_i specific to the family, is specified as a standard conditional logit formula:

$$\Pr(j_k = j_k^* | \Omega_{j_k}, Z_{i,j_k}, \mu_i) = \frac{\exp[U_{i,j_k}(\Omega_{j_k^*}, Z_{i,j_k^*}, \mu_i)]}{\sum_{j_k=1}^{J} \exp[U_{i,j_k}(\Omega_{j_k}, Z_{i,j_k}, \mu_i)]} \tag{2}$$

The test score outcome for the children is achieved conditional on the chosen school district's qualities. Therefore, the student outcome is a function of the characteristics of the quality of school his/her family chooses (Ω_{i,j_k^*}), personal specific characteristics and student preparations (X_i) such as the age and race of the child, the mother's AFQT score, the mother's working status, and the same family specific preference factor (μ_i) that affects the utility function (Equation 1). For example, parents who care much about the academic performance of the children reside in their best affordable school district. At the same time, they spend as much time as possible helping their children with homework and improve the test score outcome.

$$test_i = f(test_i = test_{i,j_k^*} | \Omega_{j_k^*}, X_i, \mu_i) = f(\Omega_{i,j_k^*}, X_i, \mu_i) + \nu_i \tag{3}$$

Assuming error term ν_i follows an *i.i.d.* normal distribution, the likelihood of observing an achieved test score can be specified as the standard normal distribution formula conditional on the unobserved heterogeneity factor μ_i:

$$g(\nu_i) = \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp[\frac{test_{i,j_k^*} - f(\Omega_{j_k^*}, X_i, \mu_i)}{2\sigma_\nu^2}] \tag{4}$$

From simple probability theory, the joint distribution of achieving student outcome and choosing the school district can be written as the product of Equations 2 and 4 if heterogeneity factor μ_i is observed. However, μ_i is unobserved, making school characteristics endogenous in the above child educational production function (i.e., the unmeasured factors that affect the location choice of the family could also affect the student test outcome).

We use a random parameters specification for the utility function determining the optimal school choice ($\Omega_{j_k^*}$) and the production function determining the test outcome ($test_i$) to solve the endogeneity problem. Our functional specification is based on the Discrete Factor Approximation method detailed in Mroz (1999). It is similar to McFadden and University of California (1973) and Train (2003)'s mixed logit random parameters specifications that allows the preference of the family on school characteristics to vary across individuals. Based on this specification, we estimate the value of the heterogeneity points that are drawn from a discrete multinomial distribution of together with their probabilities.

We specify a linear utility function that links school quality indicators Ω_{j_k} and location specific characteristics idiosyncratic to each family (Z_{i,j_k}) to the family's utility. So that Equation 1 can be rewritten as:

$$V_{i,j_k} = \alpha_1 Z_{i,j_k} + \alpha_2(\mu_i)\Omega_{j_k} + \varepsilon_{i,j_k}$$

where

$$\alpha_2(\mu_i) = \alpha_2^0 + \alpha_2^1 \mu_i \tag{5}$$

ε_{i,j_k} follows *i.i.d.* extreme value distribution. Note that in this utility function specification, parents are differentiated in terms of their attitude on how important schools are to the educational outcome.

In the child educational production function (Equation 3), test score is assumed to depend linearly on the chosen school characteristics ($\Omega_{j_k^*}$) and

X_i, while the family preference factor also contributes to the test outcome:

$$test_i = f(test_i = test_{i,j_k^*}|\Omega_{j_k^*}, X_i, \mu_i) = \beta_0 + \beta_1 X_i + \beta_2 \Omega_{j_k^*} + \rho\mu_i + \nu_i \quad (6)$$

whereas in Equation 3, ν_i is assumed to be *i.i.d.* normally distributed. The parental preference over school characteristics could also be an unobserved input to the child educational production function.

After controlling for the common factor μ_i in both the utility and the production function, we assume the two error terms ε_{i,j_k} from 5 and ν_i from 6 are mutually independent. The unconditional likelihood function can be obtained by integrating over the distribution of the common factor μ_i. With finite supporting points for the distribution ($h = 1, ...H$), we discretely approximate the true likelihood function by specifying the following likelihood function:

$$L_i = \sum_{h=1}^{H} \Pr(\mu_h)\{\Pr(j_{k,i} = j_{k,i}^*|\Omega_{j_k}, Z_{i,j_k}, \mu_h) f(test_i = test_{i,j_k^*}|\Omega_{j_k^*}, X_i, \mu_i) \quad (7)$$

In equation 7, $\Pr(\mu_h)$ is the probability of observing the preference factor having the value μ_h. Assuming the exact school choice of the family is known, this is what we call the Full Location Information Model and it will serve as the benchmark of comparison of the following two models under aggregation assumptions.

3.2 Correct Aggregation Model

If the data contains information about the exact school district choices, then the Full Location Information model can be implemented empirically. However, due to data limitations on nationwide school choice studies like this one, economists frequently need to aggregate school resource data into some higher level to enter into the production function. For example, because of the confidentiality concerns from the NLSY national survey, we only know the county where the surveyed family is located. Under this circumstance, there are two pieces of information missing from the full location information model. First, only the county of residence of the child is known to the researcher, not the exact school district; second, while we do observe the test score for each child, we can not link the exact school input the child receives to the test outcome. To follow this reality, we assume the family considers the optimal school district characteristics to make a location choice so that the county choice is the sum of the probabilities of choosing each school district within the county. To obtain the probability of observing a chosen county of residence and a child's test score outcome, we integrate $\Pr(j_{k,i} = j^*_{k,i}) * f(test_i = test_{i,j^*_k})$ over all of the school districts in county k:

$$\Pr(k_i = k^*_i \ \& \ test_{i,k} = test_{i,k^*}) = \sum_{j_k=1}^{J_k} [\Pr(j_{k,i} = j^*_{k,i}) * f(test_i = test_{i,j^*_k})] \tag{8}$$

Note that in this specification 8, school quality information is detailed at the school district level even though it is not directly linked to a student in that district. Under discrete heterogeneity assumptions about the unobserved preference factor , if estimating both the possibility of choosing the county of residence and achieving the test score conditional on the chosen county we model, the likelihood function of observing the chosen county and test score can be written as:

$$L_i = \sum_{h=1}^{H} \Pr(\mu_h)\{\sum_{j_{k^*=1}}^{J_{k^*}} \Pr(j_{k,i} = j^*_{k,i} | \Omega_{j_k}, Z_{i,j_k}, \mu_h) f(test_i = test_{i,j^*_k} | \Omega_{j^*_k}, X_i, \mu_i) \tag{9}$$

This model is specified differently from the Full Location Information Model, but is derived directly from Equation 7. This model is a summation of joint probabilities of observing the chosen county and the test score within each county. Even though only one test score is observed and location choice is known at the county level, this model maintains as much information as possible by using school district level quality information and "correctly" specifies the family's incentive to move. This property of the model is consistent with the individually observed Full Location Information Model presented in the last section. It should yield estimations with less bias than the commonly used simple aggregation model we are about to discuss in the next section. Thus, we call this model the "correct" aggregation model.

3.3 Ad Hoc Simple Aggregation Model

When location choice is partially observable to the researcher, the simple aggregation model conveniently assumes that a family makes their county choice through considering the county level average school qualities. The child educational production function inputs will be simply aggregated to the county level. This is a widely used method in the literature and is served here as a substitute for the "correct" aggregation model.

Similarly, assuming the errors are *i.i.d.* extreme value distributed, the probability that person i choosing county k is a function of the average school characteristics in county k and other factors idiosyncratic to the person and location (dependent on the preference factor) can be specified as:

$$\Pr(k_i = k_i^* | \overline{\Omega_k}, Z_{i,k^*}, \mu_i) = \frac{\exp[U_{i,j_k}(\overline{\Omega_{k^*}}, Z_{i,k^*}, \mu_i)]}{\sum_{k'=1}^{K} \exp[U_{i,j_k}(\overline{\Omega_{k^*}}, Z_{i,k^*}, \mu_i)]} \tag{10}$$

Note that aggregation causes some information loss for the production function through the loss of the variations in $\overline{\Omega_{k^*}}$.

Test score outcome, similarly, is based on the average school quality of the chosen county $\overline{\Omega_{k^*}}$ along with other background factors (X_i) of the student. Therefore, the likelihood functions for the aggregation case can be written by specifying a discrete approximation of the distribution of the unobserved heterogeneity:

$$L_i = \sum_{h=1}^{H} \Pr(\mu_h) \{ \Pr(k_i = k_i^* | \overline{\Omega_k}, Z_{i,k^*}, \mu_h) f(test_i = test_{i,k^*} | \Omega_{k^*}, X_i, \mu_i) \tag{11}$$

Unlike the Correct Aggregation Model, this model cannot be derived from the full location information model, because it is based on the assumption that parents choose a school district considering the county average school characteristics. Test score outcome is not linked to the school district characteristics but aggregated to the county level. In the next sections, we use a Monte Carlo method to calibrate the information loss, or specification errors of this Ad Hoc Simple Aggregation Model (Equation 11), compared with the Full Location Information Model (Equation 7), and the Correct Aggregation Model (Equation 9).

4 Data Generation Process in Monte Carlo

This section details the data generation process in the Monte Carlo Study for the comparison of the models. To best approximate the geographic distributions of labor markets and the quality of school districts across locations, we extract a sample of school districts that best matches with the distribution of school districts across US counties. We also obtain a mean wage measure for each of the counties in the sample to serve as an exogenous factor impacting the parents' school district choices (Z_{i,j_k} in Equation 5). To approximate the student preparation in the test score function, we generate the "previous period math score" outcomes based on the distribution of 1992 PIAT math score from the National Longitudinal Survey of Youth (NLSY) serving as the control variable X_i in the educational production function (Equation 6). In the data-generating process, we specified a number of distribution assumptions for the unobservable heterogeneity factor μ_i, to accommodate the fact that the distribution of this factor can be rather arbitrary in the real world. Details come in the last section. For each distribution assumption we generate 100,000 observations to approximate the asymptotic characteristics of the data.

4.1 Data Generation for Location Choice Outcomes

School district characteristics considered to be the major factors affecting the utility function of families are Pupil-to-teacher ratio, Expenditure per Pupil and Dropout Rate, all at the school district level. We obtain these three measurements of the school quality from the National Center for Education Statistics (NCES) survey Common Core Data files (CCD) for year 1994[5]. We use the 1990 Census Geofiles [6] to get the geographic occupation of these school districts and match them into their prospective counties. Figure 1 shows the distribution of school districts among US counties in 1994. 1,065 counties have only one school district which accounts for approximately 34% of the county file. There are 1,100 counties that have two to five school districts per county and about 600 having 6 to 16 school districts per county. Many counties have a number of school districts with very different quality levels within the county. For example, "Cook County" in Illinois has 95 school districts where the pupil-to-teacher ratio of these school districts ranges from 8 students per teacher to 21 students per teacher; "Bergen County" in New Jersey has 70 school districts, and the expenditure among these school districts goes from 1,400 dollars to 9,000 dollars per pupil. Figure 1 also shows the distribution of school districts in our extracted sample of counties. In order to best approximate the distribution of school districts across counties in the real world, we extract a sample of 200 counties that has 872 school districts to serve as the choice set in the Monte Carlo study. In our sample, the number of counties that has only one school districts takes about 34% of the counties, and there are also counties that have more than one school district, up to 19.

Table 1 provides the summary statistics for the local public school dataset, and the mean wage in the county that is used in the Monte Carlo study. Generally, class size has a mean of 16 students in 1994, expenditures per

[5]The CCD from the NCES has three major survey categories for the local school district level data in year 1994: Local Education Agency (School District) Universe Survey Data: 1986–Present; Local Education Agency (School District) Finance Survey (F-33) Data: 1990–Present; Local Education Agency (School District) Universe Survey Dropout and Completion Data (1991–Present). For every local school district agency, there is a consistent ID code called LEAID (Local Education Agency ID) assigned by NCES to the agency.

[6]The Map has 15,512 local education agencies in year 1994 where 12,920 of the 15,512 are defined as the “Unified” or primary/secondary school districts that offer a degree up to high school diploma. “Standalone” school districts account for 3,582 of the total school districts where only partial degrees (not up to 12) are offered.

pupil averages at 3934 dollars per student, and overall dropout rate over the first to 12th grade ranges from zero to 39.28% per grade, averages at 7.66% per grade. The county average wage is having a mean of $11.26 per hour. We show the division of within and between county standard deviation of all school districts as well. The difference in the standard deviations of all quality indicators shows that the school district qualities are not uniform within counties. For dropout rate, the between county standard deviation is very similar to the overall standard deviation even though this variable experience a lot of variations across individual school districts. This probably shows that the dropout rate has little variations within county. This has a direct impact on the performance of the Ad hoc Aggregation model.

Other school effects studies mostly have a number of regressors, but we specify the location choice decision to be made upon the Pupil-to-teacher ratio, Expenditure per Pupil, Dropout Rate and the county mean wage. This is to make the comparison of aggregated and disaggregated models more straightforward and to simplify the data generating in the Monte Carlo experiment. For the location choice, the utility function is simulated by the sum of 1) a known part that is a linear function of the above four characteristics, 2) a heterogeneity part (which represents the unobserved family preference factor for school inputs) that interacts with the school characteristics, and 3) error term that distributed i.i.d. extreme value in all choices. More specifically, we define equation 5 to be:

$$\begin{aligned} U_{i,j_k} &= -0.5(1+\mu_i)Pupil\_Teacher_{j_k} + 0.5(1+\mu_i)Exp\_Pupil_{j_k} \\ &\quad -0.1(1+\mu_i)Dropout_{j_k} + 0.5(1+\mu_i)Mean\_wage_{j_k} + \varepsilon_{i,j_k} \end{aligned} \tag{12}$$

Most previous studies (Loeb and Bound (1995), and Hanushek (1997)) found the marginal contribution of the three quality factors of schools are in the range of zero to 0.50 for the metrics we study here, so a coefficient scale of -0.5 to 0.5 is used as "true" coefficients in this study. We also parameterize the value of σ_ν in Equation 2 to be 10. The school district outcome is generated as the one that gives the maximum utility from a sorting through the utilities that each location gives according to equation 1. There are 100,000 families in our sample, all having only one child age 5 to 15.

4.2 Data Generation for Test Score Outcomes

We simulate the test score outcome to be the student PIAT percentile math score in 1994 which is mostly used in the empirical study as the indicator of the children's educational achievement. We adopt the value added approach (see Loeb and Bound (1995)) by controlling for the student's last period (1992) PIAT math score. To best approximate the distribution of real world math performance of children in 1994, the value of this explanatory variable in our sample is an expanded version of the 1992 PIAT math score (values 1-10) of the surveyee of NLSY Child Supplement (There are 2584 data points in that sample, we expand it to be 100,000 via random re-sorting). The generation of test score outcome is based on 1) the chosen (optimal) school district characteristics, and the previous time period PIAT math score, 2) the heterogeneity preference component, and 3) an *i.i.d.* standard normal random error term. Therefore, we define equation 6 in the Monte Carlo data generating to be:

$$\begin{aligned} test_i &= f(test_i = test_{i,j_k^*}|\Omega_{j_k^*}, X_i, \mu_i) = -2.5 Pupil\_Teacher_{j_k^*} \\ &\quad +2.5 Exp\_Pupil_{j_k^*} - 0.25 Dropout_{j_k^*} + 0.5 \Pr e\_Math_i + 1.0\mu_i + \nu_i \end{aligned} \tag{13}$$

Note that the test score outcome is generated based on the chosen school district's characteristics, the person i's own characteristics, and the unobserved heterogeneity factor, together with the normal random error.

4.3 Data Generation for Heterogeneity Factor μ_i

The common factor that appears in both equation 12 and equation 13 is unobservable. To not to lose generosity in the Monte Carlo study, we allow the heterogeneity factor μ_i to follow different distribution assumptions. We have the following distribution assumptions on the heterogeneity factor:

1. *Discrete distribution with 5 points of support. Each point takes a equal probability for five discrete values in the scale 0 to 1. In this distribution, the value of $Pr(\mu_i)$ used in the data generating is a random draw from values 0, 0.2, 0.4, 0.6, 0.8 and 1;*

2. *Uniform distribution in the range 0 to 1;*

3. *Normal distribution with mean 0 and variance 1. In this distribution assumption. The value of μ_i used in the data generating is a random draw from a standard normal distribution;*

4. *Log-normal distribution with mean $e^{0.5}$ and variance $(e^1 - 1)e^1$. The value of μ_i used in the data generating is a random draw from this log_normal distribution;*

5. *Chi-square distribution with 4 degrees of freedom. The chi-square distribution has mean $= \deg rees\ of\ freedom$ and variance $= 2 * \deg rees\ of\ freedom$.*

We normalize the scale of the value of μ_i to fall in (0,...,1), and the variance of the disturbance of the function (either 12 or 13) to be 10. To do this, we are setting the fraction of the total error variance that is due to heterogeneity to be 0.3, and 0.7 is due to the normal (or gumbo) error ν_i. we first normalize both error components so they are variance 1, then define a weighted sum of these two variances so that the whole variance of the disturbances is 10 (std=10).

$$\begin{aligned} Error &= std * \sqrt{\frac{0.3}{var(\mu_i)}} * \mu_i + std * \sqrt{1 - 0.3} * \nu_i \\ &\Longrightarrow var(Error) = std\ if\ var(v_i) = 1 \end{aligned}$$

The number of observation in our sample is 100,000. The number of school district in the choice set is 872 districts in 200 counties. For each model to consider comparison, we generate 100 sets of random outcomes depending on the distribution of the heterogeneity factor listed above, the results are based on the 100 repeats of the same estimation process for each model.

5 Parameterization and Optimization Process

There are several difficulties in estimating parameters in the quasi-maximum likelihood factor models like this one (see Mroz (1999)). The first difficulty involves the parameterization and the estimation of the heterogeneity factor μ_i. One can choose to impose the "uniform" distribution that all heterogeneity points have an equal share at the range of [0,1]. But letting the data decide the shape of the heterogeneity distribution, and estimate both the location and the scale of the points might be much better for real world problems. In this study, we parameterize the points of support and the probabilities for the discrete factor approximation estimation of the structural model.

Suppose we chose H points of support for the approximation of this rather continuous distribution of family heterogeneity. In this case, imagine there are H different types of families in the country, each type would have a different opinion on the importance of school inputs to their children's educational output on a scale of 0 to 1. It is possible for a family to fall on any of the h point μ_h. We restrict the μ_h's to lie on [0,1], with $\mu_1 = 0$ and $\mu_H = 1$. For this purpose, for the values of points in between we use standard logit to obtain the sub optimization over the $h - 2$ parameters :

$$\mu_h = \frac{\exp(\theta_h)}{1 + \exp(\theta_h)}, h = 2, ...H - 1$$

For the probability of each point of support for μ_h, we consider the constraint that those h probabilities have to sum up to 1. Also each probability has to be non negative. We use an easy parameterization as the following to satisfy these two constraints. Define:

$$\tau_h = \begin{cases} 1+\sin(\frac{3\pi}{2}+\theta_h), \ for \ h=1,2,..H-1 \\ 1+\sin(\sum_{h'=1}^{H-1}\theta_{h'}), \ for \ h=H \end{cases}$$

and let

$$\Pr(\mu_h) = \frac{\tau_h}{\sum_{h'=1}^{H-1}\tau_{h'}}$$

In the maximum likelihood estimation with $H = 5$ points discrete support, we obtain 3 estimations for μ_h, h= 2, 3, 4 and 4 estimations of $\Pr(\mu_h)$, h= 1, 2, 3, 4. According to equation 5, for the utility function we have the marginal contribution of Ω_{j_k} to the probabilities of the family choosing a particular school district should be an average of $\alpha_2(\mu_i)$ across all estimated points of support for the unobserved discrete factor. For example, in estimating the Full Location Info. Model with 5 points of heterogeneity support in the Monte Carlo, the marginal contribution of the regressor $Pupil\_Teacher_{j_k}$ to the utility function is calculated by:

$$\alpha_{Pupil\_Teacher_{j_k}}(\mu) = \tag{14}$$
$$Pupil\_Teacher_{j_k} + \sum_{h'=1}^{5}(Pupil\_Teacher_{j_k}\_MIU)_{h'} * \mu_{h'} * \Pr(\mu_{h'})$$

Similarly is true for other parameters in the utility function. Results presented in the next section are simulated marginal contributions in the utility function calculated this way. Another difficulty in the estimation of this type of simulated likelihood models, is the fact that a true global maximum is harder to achieve (see Mroz (1999)). We have put several efforts to avoid the local optimums that frequently exists when the number of discrete points is large and the number of parameters grows bigger when one have more regressors: 1). We choose a fairly extensive grid for starting values. We use more than 100 starting values for each optimization problem; 2) We programed the first derivatives of the joint likelihood functions (Equation 7, 9, and 11) into the optimization process, so that the optimization is faster with more accurate analytical first derivatives to consider in each iteration, instead of imprecise numerical first derivatives that is embedded in the optimization program; 3). We adopt a "two step" optimization process to guarantee that the gradient of the first order derivatives to be close to zero. For all the models with number of heterogeneity factors >1, we start with 100 random set of starting values for each case, process when the good looking function values are larger than 20. We choose the best 5 out of the 20 good starting values, optimize for 500 iterations for each of the 5 sets of starting values, and choose the one with the largest likelihood function value, save the results as the basis for second round optimization. In the second step, we start by gathering all 100 cases of first round optimization results, calculate the function values associated with that case and choose the best 5 function values as the starting value for this second round, and optimize each case for 100 more iterations. Usually this would make the normality of the first partial derivatives to be very close to zero. And our experience suggests this is adequate for eliminating most local optima.

6 Monte Carlo Results

The results from the Monte Carlo evaluations of the three models (i.e., estimations of Equation 7, 9, and 11) are discussed in this section. Since the production function itself (the estimate of marginal contributions of school inputs to the test score outcome) is of concern to most researchers/policy makers, we present the production function estimation results in the first section. The second section presents the utility function results.

6.1 Comparison of Models: Production Function Estimations

6.1.1 μ_i Follows Discrete Distribution with 5 Points of Support

Although it is more practical to assume the heterogeneity factor μ_i follows continuous distributions, it is undoubtedly true that the discrete factor approximation method works best if the true distribution of this unobserved factor is indeed discrete. Readers would also want to know how different each model performs, by estimating the likelihood functions using discrete factor approximation method compared with the popular OLS estimation for the educational production function. In Table 2 , we present the mean, variance, and Mean Square Error (MSE) of the production function estimators out of the 100 Monte Carlo cases based on data generation process in Equation 13, and the assumption that μ_i follows discrete distribution with 5 points of support. Each of the 100 cases here is composed of a sample of 100,000 observations to approximate the asymptotic characteristics of the data. The results listed in Table 2 strongly support the superiority of Discrete Factor Approximation method over OLS. Specifically, OLS gives average estimations further to the true, and higher MSE, for the bench mark Full Location Info. Model. This shows that OLS is not the ideal tool for estimating structural models with unobserved heterogeneity in people's decision making process. Comparing the estimation results for the three models listed at the top portion of Table 2, it can be seen that the Ad Hoc model gives more biased estimation results, though it has given more consistent estimates over the 100 random cases due to aggregation. Note that the Correct Aggregation Model follows the benchmark model very closely, and the MSE is considerably smaller (about 3 to 4 fold) than that of the Ad Hoc model.

6.1.2 μ_i Follows Various Continuous Distribution Assumptions

Table 3 presents the performance of all three models under continuous distribution assumptions for the unobserved heterogeneity factor μ_i. Although the Ad Hoc Simple Aggregation Model performs poorly compared with the other two models across all the distribution assumptions, which conforms to our expectations, there are some interesting differences in the scale of its MSE across the distribution assumptions and the parameter estimates. Uniform distribution is the one continuous distribution that is most similar to our discrete case (recall that each of the 5 points takes equal probability in the [0,1] scale for our discrete assumption). Under this distribution assumption, the aggregation model using discrete factor approximation method maintains a MSE of 0.892 for Pupil-to-teacher Ratio, and 0.567 for Expenditure-per-pupil. Mean estimates for these two variables are 1% higher than the true value, which shows moderate aggregation bias. However, in other non-uniform, or asymmetric distribution assumptions, such as the Lognormal and Chi-square distribution cases, the simple aggregation gives much more serious MSE with an average estimates sometimes doubles (100%) the true value. It is hard to describe the true world using one distribution, but it could be imagined that the simple aggregation would only perform worse when the error distribution has heteroskedasticity or autocorrelation characteristics, or with a lot of local optima and asymmetry. On the other hand, the correct aggregation model is not affected by the distribution assumptions at all. Its estimates continues to stay closely to that of the benchmark model, and its average estimates for all of the parameters are only 0.1% or 0.2% away from the true value. This finding strongly suggest the justification of using the correct aggregation model when aggregation is needed and endogeneity bias are present. Drop-out rate is the variable that experiences little within county variations according to Table 1, which is the main reason that the estimates from the simple aggregation model has a relatively low MSE for this variable. But the mean estimates from the 100 cases is still considerably off the track for drop-out rate using the Ad Hoc model.

6.2 Comparison of Models: Utility Function Estimations

6.2.1 μ_i Follows Discrete Distribution with 5 Points of Support

The task of estimating the utility function under discrete distribution assumption for μ_i is different from continuous distribution cases in terms of evaluating Equation 12. Since we could fully identify every one of the 5 possible values of μ_i and its probability in the discrete distribution, it is possible to compare the true value of the utility function parameters with the mean estimates through calculations of Equation 14 [7].Table 4 shows

[7] The "true" values can not be obtained for the continuous cases since the continuous distributions can only be approximated.

the comparison of using discrete factor approximation method versus using conditional logit estimation without endogeneity control. Note that both the full location information model and the correct aggregation models perform well regardless of which method one chooses. This is mainly due to the fact that the heterogeneity factor μ_i does not appear as an independent variable in the utility function, so that the effect of μ_i is achieved through the aggregation of the 5 discrete points. Since each of the 5 discrete points takes equal possibility in the data generating process, aggregating them is not more than taking the mean value. Therefore, using conditional logit method here is asymptotically equivalent to using the discrete factor approximations for the utility function estimations. Even under this extreme condition, the simple aggregation model still yield sizable bias for most of the parameter estimates.

6.2.2 μ_i Follows Various Continuous Distribution Assumptions

Table 5 shows the utlity function estimations under the assumptions that μ_i follows various continuous distributions. Comparing the mean and variances across all four sections of the table, we found that the correct aggregation model closely follows the benchmark model in parameter estimates regardless of the distribution of the heterogeneity factor. On the other hand, the performance of the ad hoc simple aggregation model is quite worrysome. The uniform distribution is the one continous distribution that is most similar to the discrete distribution, and the ad hoc model gives estimates 1/3 to 1/4 that of the benchmark model for the variables Pupil-to-teacher ratio and the mean wage. For assymetric distributions like the lognormal distribution and the Chi-square distributions, this model provide parameter estimates that is unrealistically high in magnitude and variances, which is due largely to the inprecise estimate of locations of the heterogeneity points.

7 Conclusion

A trend in the literature to empirically estimate the educational production function, under the data restrictions either in the choice set side or in the school input side, is to frequently link the individual academic outcome to the aggregated resources of school inputs. This Monte Carlo experimental condition enable us to provide some insight into the effect of aggregation on the evaluation of importance of school inputs in the USA, both in terms of family's migration choices and the children's academic outcomes. The simulations of two structural models under the aggregation conditions are compared with the benchmark model without aggregation. The estimation results prove that the correct aggregation model could provide fairly unbiased estimations of school effect, while simply aggregating the resources to the county level will cause strong bias brought by the wrong specification of incentives of parents choosing a school district.

References

Bayer, P.: 2000, Tiebout sorting and discrete choices: A new explanation for socioeconomic differences in the consumption of school quality, *Unpublished manuscript* .

Blackburn, M., Bloom, D. and Freeman, R.: 1991, Changes in earnings differentials in the 1980s: concordance, convergence, causes, and consequences, *Technical report*, National Bureau of Economic Research.

Card, D. and Krueger, A.: 1990, Does school quality matter? returns to education and the characteristics of public schools in the united states, *Technical report*, National Bureau of Economic Research.

Childs, T. and Shakeshaft, C.: 1986, A meta-analysis of research on the relationship between educational expenditures and student achievement, *Journal of Education Finance* **12**(2), 249–263.

Glass, G. and Smith, M.: 1979, Meta-analysis of research on class size and achievement, *Educational Evaluation and Policy Analysis* **1**(1), 2–16.

Hanushek, E.: 1997, Assessing the effects of school resources on student performance: An update, *Educational evaluation and policy analysis* **19**(2), 141–164.

Hanushek, E., Rivkin, S. and Taylor, L.: 1996, Aggregation and the estimated effects of school resources, *Technical report*, National Bureau of Economic Research.

Heckman, J., Layne-Farrar, A. and Todd, P.: 1996, Human capital pricing equations with an application to estimating the effect of schooling quality on earnings, *The Review of Economics and Statistics* pp. 562–610.

Johnson, G. and Stafford, F.: 1973, Social returns to quantity and quality of schooling, *Journal of Human Resources* pp. 139–155.

Kain, J. and Quigley, J.: 1970, Measuring the value of housing quality, *Journal of the American Statistical Association* pp. 532–548.

Link, C. and Ratledge, E.: 1975, Social returns to quantity and quality of education: A further statement, *Journal of Human Resources* pp. 78–89.

Loeb, S. and Bound, J.: 1995, The effect of measured school inputs on academic achievement: Evidence from the 1920s, 1930s and 1940s birth cohorts, *Technical report*, National Bureau of Economic Research.

McFadden, D. and University of California, B. I. o. U. . R. D.: 1973, Conditional logit analysis of qualitative choice behavior.

Mroz, T.: 1999, Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome, *Journal of Econometrics* **92**(2), 233–274.

Oates, W.: 1969, The effects of property taxes and local public spending on property values: An empirical study of tax capitalization and the tiebout hypothesis, *The Journal of Political Economy* **77**(6), 957–971.

Rizzuto, R. and Wachtel, P.: 1980, Further evidence on the returns to school quality, *Journal of Human Resources* pp. 240–254.

Tiebout, C.: 1956, A pure theory of local expenditures, *The journal of political economy* **64**(5), 416–424.

Todd, P. and Wolpin, K.: 2007, The production of cognitive achievement in children: Home, school, and racial test score gaps, *Journal of Human capital* **1**(1), 91–136.

Train, K.: 2003, *Discrete choice methods with simulation*, Cambridge Univ Pr.

Figure 1: Number of School Districts within Counties, year 1994

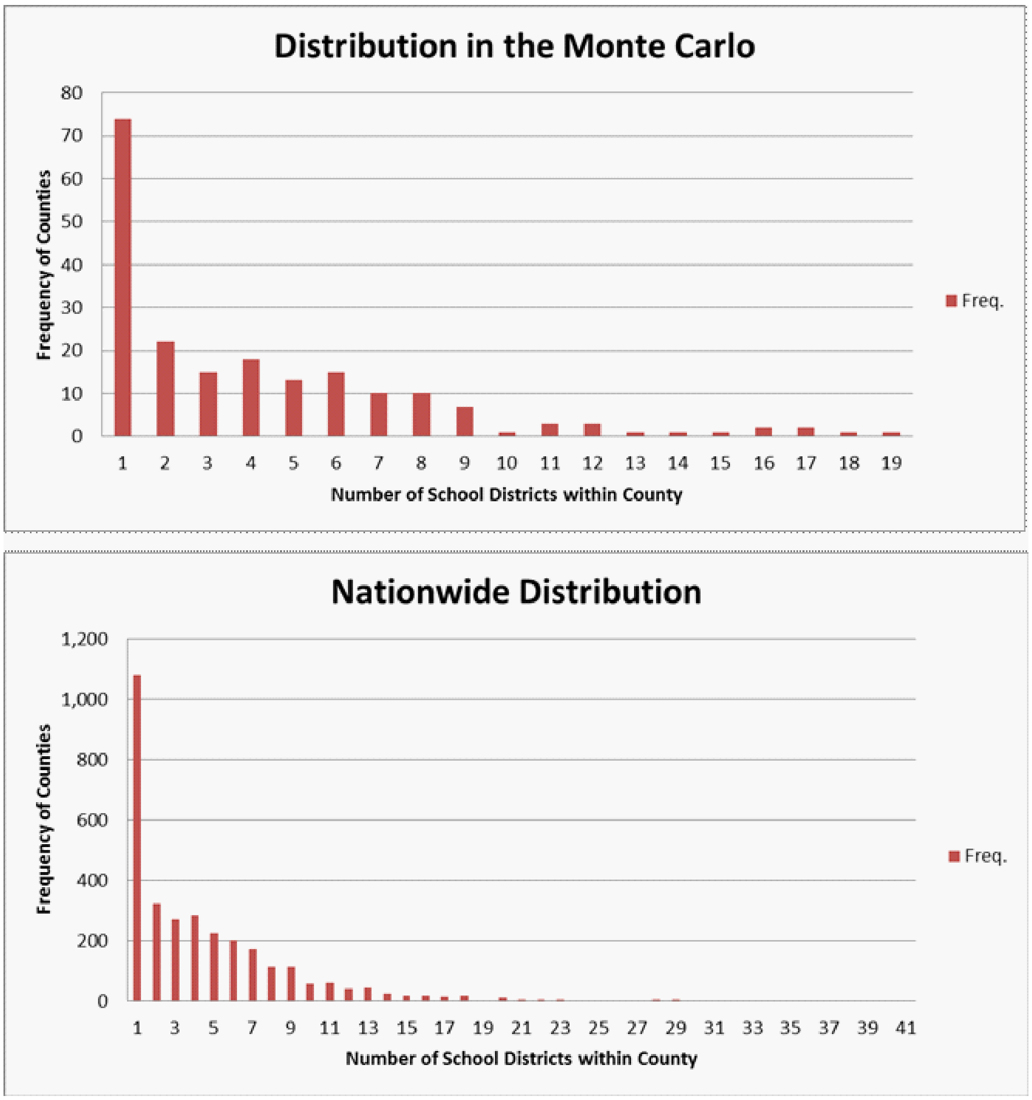

Table 1: Summary Statistics for School Charcteristics in MonteCarlo

		Mean	Std.
Ln Pupil-to-teacher Ratio	Overall	2.75	0.27
	Between Counties		0.21
	Within Counties		0.17
Ln Expenditure per Pupil ($1000)	Overall	1.37	0.35
	Between Counties		0.29
	Within Counties		0.24
Dropout Rate(%)	Overall	7.66	7.49
	Between Counties		7.41
	Within Counties		4.33
Ln Mean Wage ($1000)	Overall	2.42	0.12
	Between Counties		0.09
	Within Counties		0

Observations: N = 872, n = 200, T-bar = 4.36

Note: The mean is calculated using all school districts in the data.
The within county standard deviation is an average of the within county std. across all 3130 counties, and the between county standard deviation is the std. of the mean of each county across all 3130 counties. The small difference in the standard deviations of all quality indicators shows that the school district qualities are not unanimous within a county.

Table 2: Estimation Results for Production Function under Discrete Distribution Assumption: Discrete Factor Approximation vs. OLS

		Full Location Information Model	Ad Hoc Aggregation Model	Correct Aggregation Model
Discrete Factor Approximation				
Ln Pupil-to-teacher Ratio	TrueValue	-2.500	-2.500	-2.500
	Mean	-2.417	-3.564	-2.322
	Std	0.330	0.161	0.477
	MSE	**0.115**	**1.157**	**0.256**
Ln Expenditure per Pupil ($1000)	TrueValue	2.500	2.500	2.500
	Mean	2.471	3.438	2.414
	Std	0.357	0.132	0.352
	MSE	**0.127**	**0.897**	**0.130**
Dropout Rate(%)	TrueValue	-0.250	-0.250	-0.250
	Mean	-0.242	-0.305	-0.224
	Std	0.049	0.007	0.043
	MSE	**0.002**	**0.003**	**0.002**
Ln Mean Wage $1000	TrueValue	0.500	0.500	0.500
	Mean	0.500	0.500	0.500
	Std	0.002	0.002	0.002
	MSE	**0.000**	**0.000**	**0.000**
OLS Estimation				
Ln Pupil-to-teacher Ratio	TrueValue	-2.500	-2.500	-2.500
	Mean	-3.165	-3.854	-3.106
	Std	0.118	0.153	0.139
	MSE	**0.457**	**1.857**	**0.387**
Ln Expenditure per Pupil ($1000)	TrueValue	2.500	2.500	2.500
	Mean	3.549	3.262	3.602
	Std	0.107	0.127	0.130
	MSE	**1.113**	**0.596**	**1.231**
Dropout Rate(%)	TrueValue	-0.250	-0.250	-0.250
	Mean	-0.434	-0.298	-0.439
	Std	0.006	0.007	0.008
	MSE	**0.034**	**0.002**	**0.036**
Ln Mean Wage $1000	TrueValue	0.500	0.500	0.500
	Mean	0.500	0.500	0.500
	Std	0.002	0.002	0.002
	MSE	**0.000**	**0.000**	**0.000**

Table 3: Estimation Results for Production Function: Continuous Distribution Assumptions

		Full Location Information Model	Ad Hoc Aggregation Model	Correct Aggregation Model			Full Location Information Model	Ad Hoc Aggregation Model	Correct Aggregation Model
µi ˜ N(0,1)					**µi ˜ Lognormal Distribution**				
Ln Pupil-to-teacher Ratio	TrueValue	-2.500	-2.500	-2.500	Ln Pupil-to-teacher Ratio	TrueValue	-2.500	-2.500	-2.500
	Mean	-2.561	-4.742	-2.597		Mean	-2.663	-5.682	-2.705
	Std	0.227	0.370	0.279		Std	0.121	0.193	0.146
	MSE	**0.055**	**5.161**	**0.086**		**MSE**	**0.041**	**10.159**	**0.063**
Ln Expenditure per Pupil ($1000)	TrueValue	2.500	2.500	2.500	Ln Expenditure per Pupil ($1000)	TrueValue	2.500	2.500	2.500
	Mean	2.573	4.869	2.620		Mean	2.641	4.108	2.616
	Std	0.283	0.163	0.359		Std	0.089	0.137	0.108
	MSE	**0.085**	**5.637**	**0.141**		**MSE**	**0.028**	**2.603**	**0.025**
Dropout Rate(%)	TrueValue	-0.250	-0.250	-0.250	Dropout Rate(%)	TrueValue	-0.250	-0.250	-0.250
	Mean	-0.265	-0.598	-0.272		Mean	-0.286	-0.190	-0.310
	Std	0.043	0.006	0.055		Std	0.011	0.009	0.016
	MSE	**0.002**	**0.121**	**0.003**		**MSE**	**0.001**	**0.004**	**0.004**
µi ˜ Uniform in [0,1]					**µi ˜ Chi_Square Distribution**				
Ln Pupil-to-teacher Ratio	TrueValue	-2.500	-2.500	-2.500	Ln Pupil-to-teacher Ratio	TrueValue	-2.500	-2.500	-2.500
	Mean	-2.427	-3.432	-2.416		Mean	-2.525	-7.089	-2.529
	Std	0.401	0.152	0.476		Std	0.139	0.096	0.157
	MSE	**0.165**	**0.892**	**0.232**		**MSE**	**0.020**	**21.071**	**0.025**
Ln Expenditure per Pupil ($1000)	TrueValue	2.500	2.500	2.500	Ln Expenditure per Pupil ($1000)	TrueValue	2.500	2.500	2.500
	Mean	2.425	3.243	2.438		Mean	2.516	4.164	2.505
	Std	0.303	0.123	0.421		Std	0.142	0.090	0.163
	MSE	**0.096**	**0.567**	**0.179**		MSE	**0.020**	**2.777**	**0.026**
Dropout Rate(%)	TrueValue	-0.250	-0.250	-0.250	Dropout Rate(%)	TrueValue	-0.250	-0.250	-0.250
	Mean	-0.234	-0.280	-0.232		Mean	-0.252	-0.093	-0.251
	Std	0.036	0.007	0.040		Std	0.019	0.016	0.026
	MSE	**0.002**	**0.001**	**0.002**		**MSE**	**0.000**	**0.025**	**0.001**

Table 4: Estimation Results for Utility Function under Discrete Distribution Assumption: Discrete Factor Approximation vs. OLS

		Full Location Information Model	Ad Hoc Aggregation Model	Correct Aggregation Model
Discrete Factor Approximation				
Ln Pupil-to-teacher Ratio	TrueValue	-0.750	-0.750	-0.750
	Mean	-0.749	-0.158	-0.746
	Std	0.014	0.015	0.022
	MSE	**0.000**	**0.351**	**0.000**
Ln Expenditure per Pupil ($1000)	TrueValue	0.750	0.750	0.750
	Mean	0.750	0.658	0.751
	Std	0.011	0.008	0.016
	MSE	**0.000**	**0.008**	**0.000**
Dropout Rate(%)	TrueValue	-0.150	-0.150	-0.150
	Mean	-0.150	-0.144	-0.150
	Std	0.001	0.001	0.002
	MSE	**0.000**	**0.000**	**0.000**
Ln Mean Wage $1000	TrueValue	0.750	0.750	0.750
	Mean	0.751	0.233	0.748
	Std	0.029	0.085	0.032
	MSE	**0.001**	**0.274**	**0.001**
Conditional Logit Estimation				
Ln Pupil-to-teacher Ratio	TrueValue	-0.750	-0.750	-0.750
	Mean	-0.765	-0.268	-0.772
	Std	0.013	0.015	0.018
	MSE	**0.000**	**0.233**	**0.001**
Ln Expenditure per Pupil ($1000)	TrueValue	0.750	0.750	0.750
	Mean	0.758	0.611	0.766
	Std	0.010	0.007	0.013
	MSE	**0.000**	**0.019**	**0.000**
Dropout Rate(%)	TrueValue	-0.150	-0.150	-0.150
	Mean	-0.144	-0.146	-0.144
	Std	0.001	0.001	0.001
	MSE	**0.000**	**0.000**	**0.000**
Ln Mean Wage $1000	TrueValue	0.750	0.750	0.750
	Mean	0.752	3.782	0.761
	Std	0.030	0.029	0.032
	MSE	**0.001**	**9.193**	**0.001**

Table 5: Estimation Results for Utility Function: Other Continuous Distributions

		Full Location Information Model	Ad Hoc Aggregation Model	Correct Aggregation Model			Full Location Information Model	Ad Hoc Aggregation Model	Correct Aggregation Model
N(0,1) Distribution					**Lognormal Distribution**				
Ln Pupil-to-	Mean	-0.498	0.597	-0.493	Ln Pupil-to-	Mean	-1.327	-1209.588	-1.336
teacher Ratio	Std	0.018	0.735	0.028	teacher Ratio	Std	0.016	313.138	0.022
Ln Expenditure	Mean	0.500	0.874	0.502	Ln Expenditure	Mean	1.318	-125.884	1.328
per Pupil ($1000)	Std	0.012	0.268	0.016	per Pupil ($1000)	Std	0.011	34.937	0.014
Dropout Rate(%)	Mean	-0.100	-0.095	-0.100	Dropout Rate(%)	Mean	-0.285	-19.695	-0.282
	Std	0.001	0.043	0.002		Std	0.013	5.046	0.017
Ln Mean Wage	Mean	0.501	1.453	0.498	Ln Mean Wage	Mean	1.327	3937.360	1.340
$1000	Std	0.032	1.672	0.043	$1000	Std	0.041	1017.283	0.047
Uniform in [0,1]					**Chi-Square Distribution**				
Ln Pupil-to-	Mean	-0.750	-0.153	-0.752	Ln Pupil-to-	Mean	-2.505	-150.369	-2.505
teacher Ratio	Std	0.016	0.017	0.023	teacher Ratio	Std	0.017	44.302	0.017
Ln Expenditure	Mean	0.747	0.652	0.746	Ln Expenditure	Mean	2.503	-21.463	2.503
per Pupil ($1000)	Std	0.010	0.008	0.014	per Pupil ($1000)	Std	0.019	6.307	0.020
Dropout Rate(%)	Mean	-0.150	-0.145	-0.150	Dropout Rate(%)	Mean	-0.503	-2.112	-0.507
	Std	0.001	0.001	0.002		Std	0.019	0.622	0.028
Ln Mean Wage	Mean	0.747	0.229	0.749	Ln Mean Wage	Mean	2.510	468.854	2.512
$1000	Std	0.030	0.096	0.037	$1000	Std	0.055	140.246	0.060