Econ 613 - Applied Econometrics - 2022 Spring
Reading 3

Summary of *Do Better Schools Matter?*

Yican Liu

March 23, 2022

Understanding the value of better schools is crucial for economists. To study the effects of school reforms, economists tried to use different methodologies to estimate people's willingness to pay for the houses which are close to good schools. However, such estimations might be inefficient because the neighborhood characteristics are not fully controlled. To solve this problem, the author compares the prices of the houses which are located on the different sides of the attendance boundary in Massachusetts. She finds that parents are willing to pay 2.1 percent more on housing if the test scores of their children increase by 5 percent. In addition, the result passes several robustness checks.

To investigate the relationship between test scores and housing prices, the author uses two regression models as the main empirical strategy. To be specific, in the first regression model, she follows the previous research and uses the log value of household price as the dependent variable. The regressor of interest is the average test scores of the school which is located at the attendance district of this house, and the control variables include the characteristics of the house as well as the neighborhood and school district. However, this model is not efficient because not all relevant characteristics could be observed, so there might be omitted variable problems in this model. Hence, in the second regression model, the author replaces the control variables for neighborhood and school district characteristics with a set of boundary dummies. These dummies, which are used as the new control variables in this regression, indicate whether the house is located on either side of an attendance district boundary or not. Thus, any unobserved characteristics for the houses located on the same boundary could be controlled, and omitted variable problems are solved by this strategy.

In the empirical analysis part, the author uses the housing trading information of Middlesex, Essex, and Norfolk counties for analysis. For easy comparison, the author only focuses on single-family residences. In addition, for each school district, there should be at least two elementary schools which have overlap grades. The author also excludes the school districts where parents are allowed to choose the school programs from the other districts as well as the districts whose boundaries are hard to define. As a result, the full sample is consisted by 39 school districts, including 22,679 residences and 181 attendance district boundaries. To measure school quality, the author uses the average grade of reading and math tests of the Massachusetts Educational Assessment Program (MEAP) for 1988, 1990, and 1992. As for control variables, she includes several census block group variables and school district characteristics in her regressions.

The author starts her empirical analysis by comparing the regression results from the two models. In the first model where she controls the household, neighborhood, and school district variables, she gets the conclusions that are similar to previous literature, such as the positive

relationship between bedrooms and house prices, etc. Moreover, the model shows that parents would like to pay 4.9 percent more on housing for a 5 percent increase in their children's test scores. However, such regression may cause serious omitted variable biases because of the unobserved characteristics. Hence, she focuses on the second model, where the boundary fixed effects are controlled. By restricting the sample of residences located within 0.35 miles from attendance district boundaries, she finds that the coefficient of test scores in this regression shrinks a lot compared to the previous regression model. This regression coefficient does not vary a lot when she changes the distance from 0.35 miles to 0.20 miles or 0.15 miles. To verify that the results are not driven by the sample size, she also runs the first regression model on the residences which are located within 0.15 miles from the boundary, which generates similar results as the first regression.

Furthermore, the author shows similarity in houses when we restrict the sample close to the attendance district boundary. She estimates the difference in means of the house and neighborhood characteristics between the sides with higher and lower MEAP test scores of the boundary. In general, as the sample of residences becomes closer to the boundaries, the difference in means for most characteristics becomes less significant, which indicates that the houses are similar if they are located close enough to the same boundaries.

Finally, the author compares the magnitude of test score coefficients which are estimated in the prior regressions. If we limit the houses which are located within 0.15 miles from the boundary and look at the second regression model which controls for the boundary fixed effects, people are willing to pay 2.1 percent more for 5 percent increase in test scores. The value is half of the value estimated in the first model, which indicates that omitted variable problems exist in the first regression model.

At the end of this paper, the author runs several sensitivity tests to show the robustness of her empirical results. First, attendance district boundaries might be the boundary for neighborhoods. To solve this problem, she excludes the boundaries of railroads, highways, and major streets. The results show little difference from the previous results. Additionally, better schools might be in better neighborhoods. To address this problem, she creates artificial attendance district boundaries. In her regression, the test score is replaced by the dummy variable which indicates the better side with higher average test scores of these artificial boundaries. The regression yields insignificant coefficients, which indicates that this hypothesis is not true. Second, she includes neighborhood characteristics as control. The results are close to the previous regressions. Third, the author tests the hypothesis that there might be some unobservable difference in the quality of residences. She runs the regression with internal square footage or lot size as dependent variables with the average test score and boundary dummies as regressors. These two regressions generate insignificant test score coefficients, which indicates that observable variables do not vary a lot on different sides of the boundaries. Fourth, the author investigates the effect of test scores on the prices of housing with three- or more bedrooms, because people who purchase large houses with more than three bedrooms are more likely to have children. In this regression, she replaces the test score in the second regression model with two interaction terms: one for test score and a dummy of three or more bedrooms, and the other for test score and a dummy of fewer than three bedrooms. She finds that the regression coefficient for the first interaction term is larger and more significant than the first interaction term, which reveals that the empirical results are driven by the difference in school quality.

In conclusion, the author uses a novel approach to estimate how much parents would like to pay for a better education. By controlling for the fixed effects of attendance district boundaries, she solves the problem of omitted variable biases which is generally existed in the previous works. In general, a 2.1 percent increase in housing price is associated with a 5 percent increase in children's test scores. There is one limitation in this paper: the author could use some other datasets, for example, the housing information from California, as robustness checks.