Making Causal Claims in Educational Studies:
The Effect of After-school Programs on Private Education Cost in South Korea

Junok Kim
junokkim@ucla.edu
Social Research Methodology
Graduate School of Education

INTRODUCTION

- In educational studies, researchers are often interested in estimating the effect of interventions, programs or policies on student outcomes. When there are confounders which can affect the relationship between treatment and outcome, statistical techniques based on correlational methods might be problematic in that the estimates are likely to be biased and imprecise.
- In order to deal with these confounders and draw a causal inference, a set of analytical strategies have been developed across disciplines under the causal inference framework. The Instrumental Variable (IV) approach is one of these efforts, providing strong evidence for causal relationships by introducing an exogenous variable, IV, into the model to eliminate the influence of confounders. When participants are randomly assigned to treatment conditions, the assignment indicator is used as IV to obtain Complier Average Causal Effect (CACE), which means the average treatment effect on participants who actually received that treatment.
- This study aims to estimate the causal effect of after-school programs on costs of private tutoring in Korean middle schools, using the IV approach. Since 2006, after-school programs have been expanded nationwide and diversified by three different regimes, with an explicit goal of decreasing the dependency of students on private education by improving the quality of public education. Despite the large amount of government investment in this policy, research results are mixed in terms of the effect of after-school programs on private tutoring. Given this inconsistency in the findings across the literature, this study aspires to provide strong evidence from the IV analysis in order to contribute to the continuing discussion around the effect of after-school programs.

METHOD

➢ **Participants and data**

- This study analyzed a part of a large-scale educational data set from the Korean Education Longitudinal Study, a longitudinal follow-up study conducted from 2005 to 2010 with a nationally representative sample of students, schools and parents.
- The analytic sample included 1,183 9th graders in 2007, attending 25 middle schools in Seoul, South Korea.
- For the major variables used in this study, the observed outcome was private education cost based on 18 items of parent survey responses. The outcome was log-transformed to adjust its non-normality. The observed assignment indicator was the current status of running after-school programs from school survey responses and the observed treatment indicator was participation in after-school programs from student survey responses.

➢ **Analysis model: Complier Average Causal Effect (CACE)**

- In order to estimate the causal effect of attending after-school programs on private education cost, this study used the assignment indicator as IV. The main analysis followed two steps: first, estimate the effect of the assignment on the treatment; second, estimate the effect of the treatment on the outcome.
- The analysis model is described as follows:

$$1^{st} \text{ stage: } Y_i = \beta_0 + \beta_1 \cdot D_i + \varepsilon_i$$

$$2^{nd} \text{ stage: } D_i = \alpha_0 + \alpha_1 \cdot Z_i + v_i$$

Y_i: the observed outcome for person i (i.e. private education cost in 2009)

D_i: the observed treatment for person i (i.e. attending after-school program in 2009)

Z_i: the observed assignment status for person i, IV (i.e. randomly assigned to schools with or without after-school programs)

➢ The overall analysis process consisted of three steps: first, check the required assumptions to use the IV model based on Angrist, Imbens and Rubin (1996); second, obtain a naïve estimate of the program effect using simple linear regression; third, obtain CACE using the IV approach. For the software program, R 3.40 was used with the AER package (Kleiber and Zeileis, 2017) in the analysis.

RESULT 1. Assessment of Assumptions

✓ **Assumption 1: Stable Unit Treatment Value Assumption (SUTVA)**

Attending after-school programs was not affected by the assignment of other students to the treatment conditions and the cost for private tutoring was not affected by that assignment.

✓ **Assumption 2: Ignorable Assumption**

Assignment to schools with after-school programs was based on the lottery process, and thus was random.

✓ **Assumption 3: Exclusion restriction**

Private education cost was not affected by the treatment assignment after treatment was taken into account.

✓ **Assumption 4: Nonzero Average Causal Effect of Z on D**

Allocation to the schools with after-school programs increased the probability of attending those programs.

✓ **Assumption 5: Monotonicity Assumption**

There were no students who took after-school classes in control conditions where their schools didn't provide such programs.

RESULT 2. Regression Analysis Result

✓ In the simple regression analysis, the intercept was the average cost for private tutoring for students who didn't take after-school classes. The parameter estimate was 2.8723 with standard error of 0.0641, which was significant at $\alpha = 0.05$. When exponentiated, the intercept was $e^{2.8723} = 17.6776$ and it meant that on average, students spent about 166 dollars for private education when they were not in after-school programs.

✓ For the estimated treatment slope, the coefficient was -0.6842 with standard error of 0.1271, which was also significant. The exponentiated slope was $e^{-0.6842} = 0.5045$, which further implied that the average cost of private education for participants of after-school programs was $e^{2.8723} * e^{-0.6842}$ = 17.6776 * 0.5045 = 8.9183, 83.68 in dollars.

✓ The regression result suggested that there was 0.5045 * 100 = 50.45% reduction in private education cost for the 9th graders who attended after-school programs provided by their schools.

RESULT 3. IV Analysis Result

✓As represented in Table 2, the analysis result from the IV model indicated that after-school programs were effective in reducing private education cost for the students who took those programs.

✓The meaning of the intercept was the average cost spent in private tutoring for control-group students who didn't have a chance to take after-school classes since their schools didn't provide those programs. The parameter estimate for the intercept was 3.0549 with standard error of 0.1631, which was significant at α=0.05. The exponentiated coefficient was $e^{3.0549} = 21.9191$, which was 205.43 dollars.

✓ For the students who actually received the treatment, the estimated slope was -1.4009 with standard error of 0.6011, which was statistically significant. When exponentiated, the after-school program effect was $e^{-1.4009} = 0.2464$, which suggested that the average cost of private education for the students who attended after-school programs provided by their schools was $e^{3.0549} * e^{-1.4009}$ = 21.9191 * 0.2464 = 5.2278, corresponding to 49.00 in dollars.

✓The analysis result from the IV model showed that the private education cost for compliers to the treatment was only 0.2464 * 100 = 24.64% of the students in control condition.

➢ [Table 1] Regression Analysis Result

	Estimate	Std.Error	t	p
Intercept	2.8723	0.0641	44.782	<0.0001
Treatment	-0.6842	0.1271	-5.384	<0.0001

➢ [Table 2] IV Analysis Result

	Estimate	Std.Error	t	p
Intercept	3.0549	0.1631	18.727	<0.0001
Treatment	-1.4009	0.6011	-2.331	0.02

CONCLUSION

- **SUMMARY**

✓ This study capitalized on the random allocation of students to middle schools by lottery process to estimate the causal effect of after-school programs on private education cost in Seoul, South Korea. The critical assumptions required to proceed to the IV analysis were all satisfied as described in Result 1. After assessing the assumption, naïve estimate of the effect of after-school programs was obtained from linear regression. The regression analysis result suggested that the average cost of private education was significantly lower in the treatment group compared to the control group. Specifically, the average private education cost for the treated students was half of the amount which control students spent in private tutoring. The IV analysis result appeared more dramatic in that the CACE indicated that the average cost spent in private education for the treatment group was only 24% of that for the control group.

- **LIMITATION AND FURTHER DIRECTION**

✓ Although the IV model provided a strong evidence that after-school programs were effective in reducing the cost for private education, the result should be interpreted with caution since the IV estimate was sensitive to the proportion of compliers in treatment group. For this sample, the proportion of students who attended after-school programs in the treatment group was 0.2957, which might cause the overestimation of the CACE.

✓ This study was also limited since the nested structure of data was not considered. Given the school-based characteristic of the treatment, future study should focus on the heterogeneous treatment effect across schools, as described in Raudenbush, Reardon and Nomi (2012).

✓ In addition, covariates which might have a substantial influence on the outcome (i.e. Socio-economic status) were not included in the analysis. Adding proper covariates into the model would improve the precision of the causal estimate as well as adjust the differences between treatment and control condition to make them more comparable. Further studies should consider these limitations to better estimate the causal effect of after-school programs to evaluate whether this policy achieved the desired result of decreasing the dependency of students on private education and narrowing the achievement gap.

References

[1] Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American statistical Association, 91(434), 444-455.

[2] Raudenbush, S. W., Reardon, S. F., & Nomi, T. (2012). Statistical analysis for multisite trials using instrumental variables with random coefficients. Journal of research on Educational Effectiveness, 5(3), 303-332.