

Within-District School Lotteries, District Selection, and the Average Partial Effects of School Inputs*

Eleanor Jawon Choi** · Hyungsik Roger Moon*** · Geert Ridder****

This study proposes an econometric framework to consistently estimate the average partial effects (APE) of school inputs on academic achievement when students are randomly assigned to schools within each school district but endogenously sort across school districts. We illustrate our method by estimating the APE of single-sex schooling and class size on standardized test scores using data from Seoul, Korea in the period of 2008-2009. Our APE estimates are smaller than the estimates from conventional linear regressions with school district fixed effects, which do not fully correct for endogenous district choice and thus suffer from selection bias.

JEL Classification: C21, I21

Keywords: Average Partial Effect, Within-District School Lottery, Selection, Treatment Effect Heterogeneity, Education Production

Received: June 16, 2018. Revised: Aug. 13, 2018. Accepted: Sept. 30, 2018.

* We thank two referees, Jae-Young Kim, Alan Krueger, David Lee, Nayoung Lee, Heonjae Song, Jungmo Yoon, and seminar participants at KLEA, Sogang University, and SKKU, as well as the Hanyang-Renmin, KAEA-KEA, KES, and SOLE-EALE conferences for helpful comments and suggestions; the Korea Education and Research Information Service, the Korean Educational Development Institute, and the Korean Ministry of Education for providing data; and Jaewoo Choi and Sae Won Om for research assistance. Choi acknowledges that this work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A8010685) and by the research fund of Hanyang University (HY-2017). Moon thanks the research award from the USC Lusk Center for Real Estate. All remaining errors are our own. The supplementary materials are available at <https://sites.google.com/view/eleanorjchoi/research>.

** Corresponding and First Author, Associate Professor, College of Economics and Finance, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea. Email: choiej@hanyang.ac.kr. Phone: +82-2-2220-1019. Fax: +82-2-2296-9587.

*** Second Author, Professor, Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089-0253, USA. School of Economics, Yonsei University, Seoul, Korea. Email: moonr@usc.edu. Phone: +1-213-740-2109. Fax: +1-213-740-8543.

**** Third Author, Professor, Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089-0253, USA. Email: ridder@usc.edu. Phone: +1-213-740-7432. Fax: +1-213-740-8543.

I. Introduction

The effect of various school inputs on student achievement has long been debated among researchers, policy makers, and parents. Nevertheless, credible evidence is limited, and the results from a handful of high-quality studies are often mixed.¹ This is mainly due to difficulties in measuring the causal effect of school inputs using observational data on student performance and school characteristics. When students are free to choose their school, observed and unobserved student and family attributes influence both the academic achievement and school choice of the students; this situation renders spurious correlation between student performance and school characteristics. Thus, addressing endogenous selection into schools and neighborhoods has been a key econometric issue in studies of school effectiveness.

This study provides an econometric framework to identify the average partial effects (APE) of school inputs when school districts are endogenously chosen but school assignment is random within each school district. We show that even a modest difference in school characteristics across districts can induce a substantial degree of sorting and that school input effects can be inconsistently estimated when endogenous district choice is ignored. We illustrate our econometric framework by estimating the effects of single-sex schooling and class size, which are time-invariant and time-varying inputs, respectively, and have been major policy variables in the education production literature.²

We exploit quasi-experimental variation from within-district high school lotteries in Seoul, South Korea (Korea hereinafter). Students who entered high school in Seoul before 2010 were randomly assigned to high schools within school districts. Under this assignment rule, students and their parents can only choose the average quality³ of the schools to which they can potentially be assigned by choosing a residential neighborhood. Thus, self-selection bias arises from endogenous sorting across school districts but not from endogenous selection into schools within each school district.⁴

¹ A review of the literature on education production and relevant econometric methods can be found in Hanushek (2006) and Meghir and Rivkin (2011).

² The class size effect has been extensively studied over the past three decades. Well-known studies include Angrist and Lavy (1999), Krueger (1999), and Hoxby (2000b) to name a few. The effect of single-sex education has become an active area of economic research in recent years. See Jackson (2012, 2016), Park et al. (2013), Lee et al. (2014), and Dustmann et al. (2017) for effects of single-sex schooling and Hoxby (2000a), Whitmore (2005), and Lavy and Schlosser (2011) for gender peer effects.

³ Or more generally the distribution of school quality.

⁴ Altonji and Mansfield (2014) suggest that, with rich data on student and family characteristics, school district averages of observed individual characteristics can control for sorting on (un)observables if district choice is independent of school inputs given these averages. However, this type of control function approach cannot be applied to many studies, including ours, that use quasi-experimental variation from within-district random school assignment in Seoul. The reason is that individual attributes are unobserved in most of the test score data currently available in Korea.

We construct a model of heterogeneous treatment effects that considers within-district random assignment and across-district sorting. Given data on the College Scholastic Ability Test (CSAT) scores and high school characteristics from 2008–2009, we specify a random coefficient model. In the model, each student's potential outcomes depend on observed and unobserved school inputs (either time varying or time invariant) and unobserved time-invariant district characteristics that interact with the student's unobserved productivity for these inputs. Based on the individual-level education production, we derive a district-specific (random coefficient) linear regression model of individual test scores on school inputs. The APE of school inputs can be consistently estimated by a weighted average of the district-specific estimates with the fraction of students in each district as the weight.

The APE estimation based on our econometric model finds no evidence of better student performance at single-sex schools than mixed-gender schools; this finding is unlike those of most previous studies using quasi-experimental variation from within-district school lotteries in Seoul.⁵ Meanwhile, our APE estimates of class size are small and negative and similar to the numbers reported in the literature. We also find that the district-specific estimates are more heterogeneous for single-sex effects than for class size effects. This result implies a substantial degree of endogenous sorting across school districts, especially based on the likelihood of getting into a single-sex school.

Apart from proposing a robust weighted average method, we also point out the possible inconsistency of the commonly used district(-year) fixed effects estimation under endogenous district selection. This point is neglected in empirical studies on school quality effects that rely on within-district random school assignment (Park et al., 2012, 2013, 2018; Lee et al., 2014; Choi et al., 2015; Sohn, 2016; Dustmann et al., 2017; Hahn et al., 2018). This point is also relevant in studies that use the same design as Project STAR and school fixed effects (e.g., Krueger, 1999) and studies that use lottery fixed effects in comparing students who win the lottery for their chosen school with students who lose (Cullen et al., 2006; Deming, 2011; Deming et al., 2014).

Our econometric framework can be applied to any (quasi-)experiment with a two-stage assignment where selection among experiment sites is endogenous and assignment between treatment and control groups within sites is random.⁶

⁵ Park et al. (2013), Lee et al. (2014), Sohn (2016), and Dustmann et al. (2017) study the effect of single-sex education using various datasets on standardized test scores and secondary school characteristics in Seoul. Except for Sohn (2016) whose estimates are not directly comparable, these studies find a positive association between single-sex education and student performance on standardized test scores.

⁶ Examples include the Work Incentive training experiments in four locations (Hotz et al., 2005), Opower energy conservation experiments in 111 locations (Allcott, 2015), STAR experiment in 80 schools (Krueger, 1999), and school choice programs over a large number of school choice lotteries (Cullen et al., 2006; Deming, 2011; Deming et al., 2014).

We also contribute to the empirical literature on education production by providing quasi-experimental estimates on the effects of class size reduction and single-sex schooling in upper secondary education. Studies on the class size effect in high schools are few (Chingos, 2013; Jepsen, 2015). Although the effect of single-sex schooling has been explored, high-quality works are still limited (Jackson, 2012, 2016).

The remainder of the paper is organized as follows. Section 2 outlines the institutional background on the high school assignment lotteries in Seoul. Section 3 describes the data on CSAT scores and school characteristics. Section 4 presents the potential outcomes model of education production under within-district random school assignment. Section 5 discusses estimation and statistical inference. Empirical results are given in Section 6. Section 7 concludes.

II. Institutional Background

Before 2010, middle school seniors in Seoul were randomly assigned to general academic high schools within their school districts. The high school assignment lotteries were conducted as a part of the High School Equalization Policy (HSEP) and covered general academic high schools in 10 school districts (districts 1-4 and 6-11) in Seoul. Those general academic high schools were subject to the lottery-based assignment system regardless of their resource levels or types. The random assignment system excluded vocational high schools; selective high schools specialized in math and science, foreign languages, arts, or sports; and academic high schools near the city center, which are mostly in district 5 and some in districts 1, 2, 10, and 11.⁷ Online Appendix B.1 presents details on the HSEP and high school assignment in Seoul.

We indirectly test whether high school random assignment was properly implemented within school districts using data from the Korean Education Longitudinal Study (KELS). By comparing those who attended the same middle school but randomly assigned to different high schools, we show that students' characteristics before entering high schools, including parental education, household income, expenditures on private out-of-school education, and 9th-grade standardized test scores, are similar between single-sex and coed high schools and across high schools with different class sizes.⁸ We cannot reject at the 5%

⁷ In 2008, 295 high schools, namely, 203 academic, 78 vocational, 12 specialized, and 2 other, were established in 11 school districts of Seoul. Forty of the academic high schools did not participate in the assignment lotteries. The academic high schools outside the random assignment system accepted applications from students in all school districts of Seoul mainly because those schools were located in the business area that is near the city center and has a small number of residents.

⁸ We use the KELS data for the balance test because of the unavailability of the baseline

significance level the null hypothesis that the baseline characteristics of students are unrelated to school characteristics (single-sex schooling or class size) when we restrict our analysis to students attending high schools subject to the random assignment lotteries. However, when we repeat this exercise using students attending high schools not subject to the random assignment lotteries, we strongly reject (at any conventional level of significance) the null hypothesis. Online Appendix B.2 and Table B1 present further details (tables and figures labeled with B are in the Online Appendix).

Noncompliance to the random high school assignment is not a major issue in this setting. Under the random assignment system, students had very little incentive to move to another school district after the initial high school assignment. Once assigned to a high school, students were prohibited to transfer to another school within the same school district. When students (and their family) moved to another school district, their high school assignment was again determined by a random lottery in the new district. Transfers and dropouts after the initial high school assignment were indeed rare (less than 3%) and unrelated to school characteristics (Sohn, 2016; Dustmann et al., 2017). Senior enrollment in each school is approximately 98% of the freshman enrollment from two years before on average, and the across-school standard deviation of this proportion is less than 4 percentage points in our data.

III. Data and Analysis Sample

We use data on the CSAT scores and high school characteristics provided by the Korean Ministry of Education. The CSAT score data include the entire population of CSAT takers in Korea but contain no individual characteristics other than gender, whether or not enrolled in high school, and the name of the school currently attending. Data on high school characteristics come from the school information database maintained by the Korea Education and Research Information Service. The database contains information on all primary and secondary schools in Korea, including school type, number of students by gender, class size, number of teachers by gender and employment type, and school budget.⁹ We link individual test scores and school characteristics by matching school names.

The CSAT is the standardized test used for college admissions in Korea. The test

characteristics in our main data on CSAT scores. We use the middle school that the student graduated from as a proxy for his or her residential school district due to the lack of school district identifiers in the KELS data. Online Appendix B.2 describes a justification for and a limitation of the balance check exercise using the proxy school districts.

⁹ Information on schools for the most recent four years is publicly available at <https://www.schoolinfo.go.kr>.

is administered by the Korea Institute of Curriculum and Evaluation and is offered once a year in November. Approximately 600,000 individuals, including high school seniors, high school graduates, and GED holders, took the CSAT each year during the analysis period.¹⁰ The CSAT is a high-stakes exam because the CSAT score is a major factor that determines college admission outcomes and thus potentially affects future earnings. In this study, we use the Korean and English CSAT scores as the educational outcome. The sum of the Korean and English CSAT scores (hereafter referred to as the CSAT score) is normalized to have zero mean and unit variance across all observations within each cohort.¹¹

We restrict our analysis to the 2008 and 2009 cohorts of high school seniors (12th graders) attending academic high schools in school districts 3, 4, and 6-9 to focus on students randomly assigned to academic high schools in Seoul. In the six school districts, all the general academic high schools admitted students using assignment lotteries. Our analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The analysis sample covers 54% of high school seniors taking the CSAT in Seoul. Online Appendix B.3 provides further details on sample restriction and construction of the outcome variable.

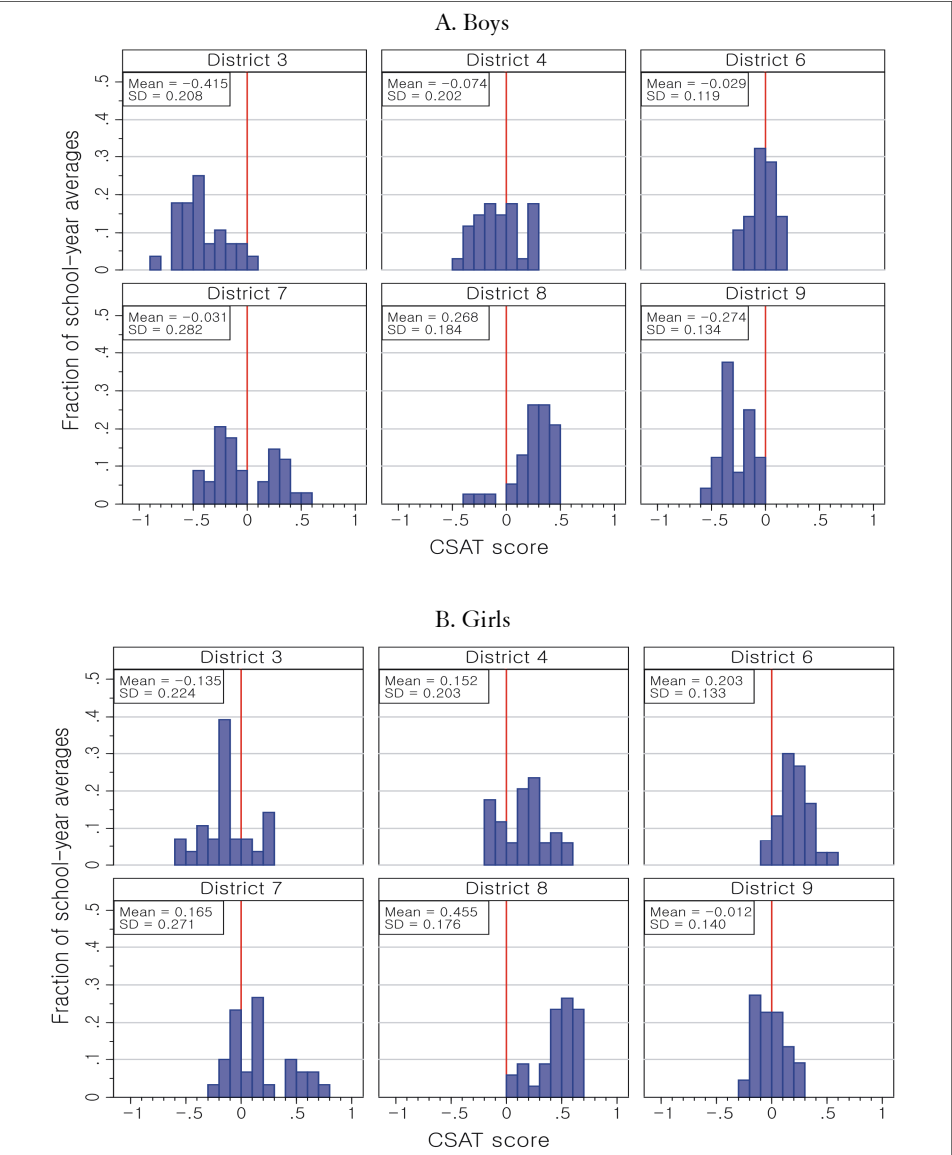
Using the analysis sample, Figure 1 plots the within-district distribution of school-year averages of the combined CSAT scores. That is, the observations are the school-level average scores in 2008 or 2009 (each school contributes two average scores). The district-specific distribution of the average test score is plotted for boys and girls separately.¹² Each bar in Figure 1 shows the fraction of observations in an interval with a width equal to 0.1 standard deviation of the distribution of the average score. Although we pool the two years of observations on each school in Figure 1, the graph looks similar when we plot the distribution for 2008 and 2009 separately. The mean and the standard deviation of the average test score are also reported in the top left corner of each bar graph. Similar graphs for Korean and English CSAT scores are shown in Figures B1 and B2, respectively. The English score is slightly more dispersed within and across districts than the Korean score.

¹⁰ In our analysis sample, 93% of high school seniors (12th graders) took the CSAT. Given that the test taking rate is very high, nonrandom selection of CSAT takers is less of an issue than in the US.

¹¹ In the raw data, the Korean and English CSAT scores are standardized to have a mean of 100 points and a standard deviation of 20 points.

¹² For mixed-gender schools, average test scores are computed for boys and girls separately, and the gender-specific averages are included in the graphs for each gender.

[Figure 1] School-Year Average CSAT Scores by School District



Notes: Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The unit of observation is school-year because the CSAT score is averaged at the school-year level. The CSAT score refers to the sum of Korean and English CSAT scores and is standardized to have zero mean and unit variance across all test takers in a given year. For CSAT applicants who were absent during the exam, missing raw scores on Korean or English are imputed with zeros. The mean and the standard deviation in the top left box in each bar graph are computed using school-year average test scores from each school district.

The school-level average test score varies substantially across districts. According to the district means, the six school districts can be grouped into three categories: high (district 8), middle (districts 4, 6, and 7), and low (districts 3 and 9). The three categories roughly correspond to the income levels in the districts: wealthy, middle income, and poor.¹³

Table 1 shows the descriptive statistics of student- and school-level variables used in our empirical analysis. In addition to the mean and the standard deviation, we report the proportion of the within-district sum of squares (in deviation from the district mean) in the total sum of squares (in deviation from the overall mean).¹⁴ The mean and the standard deviation of the CSAT scores are reported at individual (panel A) and school (panel B) levels. The within-district variation accounts for 69-99% of the total variation in school characteristics (except for class size), but only 44-57% of the total variation in school-level averages of Korean and/or English CSAT scores can be explained by within-district variation. Therefore, a major portion of the test score variation across schools is likely due to student-level sorting across school districts. Meanwhile, the large proportion of within-district variation in school characteristics implies that school characteristics should be controlled to disentangle the APE of a specific school input from many other confounding factors. The district-specific mean and standard deviation of school characteristics are shown in Table B2.

Although we do not observe student characteristics in our main dataset, Table B3 reports the mean of students' baseline characteristics across school districts from the KELS data. Given that school district identifiers or school names are not provided in the KELS data, we impute school districts in Seoul by grouping middle schools that send their students to an overlapping set of high schools. This procedure finds 11 school districts (11 school districts are indeed established in Seoul), but we cannot map these districts into the school districts in our main data. However, we find a considerable degree of sorting of students among the imputed districts. As shown in Section 4.1, even a small degree of sorting results in a substantial variation in the APE of school characteristics by district.

¹³ The income levels can be inferred from the 2000-2006 average housing price (in 10,000's of 2009 KRW/ m^2): 280.0, 229.1, 520.4, 375.4, 676.9, and 315.0 in districts 3, 4, 6, 7, 8, and 9, respectively. The housing price statistics are from the database maintained by the Mirae Asset Real Estate, which is a leading real estate investment company in Korea (<https://www.rl14.com>).

¹⁴ For time-varying variables, we compute the within-district-year sum of squares.

[Table 1] Descriptive Statistics

	Boys			Girls		
	Mean [SD]	% within-dist. variation ^a	Obs. ^b	Mean [SD]	% within-dist. variation ^a	Obs. ^b
A. Student-level CSAT scores^c						
Combined	-0.023 [1.001]	95.1	57,443	0.205 [0.894]	95.5	52,271
Korean	-0.082 [1.015]	97.0	57,443	0.166 [0.898]	97.5	52,271
English	0.032 [0.976]	94.3	57,443	0.214 [0.890]	94.4	52,271
B. School-level average CSAT scores^c						
Combined	-0.066 [0.294]	44.2	186	0.155 [0.270]	51.9	178
Korean	-0.119 [0.240]	47.5	186	0.126 [0.211]	57.1	178
English	-0.012 [0.309]	42.8	186	0.162 [0.293]	49.1	178
C. School-level characteristics						
Single-sex school	0.409 [0.494]	97.2	93	0.382 [0.489]	98.7	89
Private school	0.495 [0.503]	91.1	93	0.461 [0.501]	91.6	89
School establishment year	1,977.0 [25.0]	88.9	93	1,978.3 [25.9]	85.6	89
Class size	35.2 [2.5]	42.1	186	35.5 [2.6]	40.6	178
Total enrollment	1,422.8 [279.2]	76.1	186	1,398.7 [288.6]	68.6	178
Number of teachers	83.6 [13.5]	79.5	186	82.8 [15.0]	78.9	178
Fraction of female teachers	0.400 [0.190]	90.5	186	0.533 [0.120]	88.5	178
Fraction of regular teachers	0.895 [0.067]	92.0	186	0.892 [0.065]	89.8	178
Annual school spending ^d (in 1,000's of 2009 KRW)	2,146,877 [1,052,292]	88.4	186	2,014,230 [893,447]	87.1	178

Notes: Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. Standard deviations are in brackets.

^a For time-invariant variables, the proportion of the within-district sum of squares in the total sum of squares is reported. For time-varying variables, the proportion of the within-district-year sum of squares in the total sum of squares is reported.

^b The unit of observation is individual for student-level variables, school for time-invariant school-level variables, and school-year for time-varying school-level variables.

^c The combined score is the sum of Korean and English scores. The CSAT scores are standardized to have zero mean and unit variance. For CSAT applicants who were absent during the exam, missing raw scores on Korean or English are imputed with zeros.

^d Annual school spending measures the school operating budget per academic year excluding teacher or staff salaries. Approximately 1,100 KRW is worth 1 USD.

IV. Potential Outcomes Framework

We construct a potential outcomes model of education production tailored to the institutional setting of our data, that is, random assignment of students to schools within each school district.

Let $\mathcal{D} = \{1, \dots, N_D\}$ denote the collection of school districts in Seoul where high school assignment is random. Let $\mathcal{I} = \{1, \dots, N_I\}$ be the population of high school seniors in these school districts who were randomly assigned to their schools. $\mathcal{S}_d = \{1, \dots, N_S(d)\}$ denotes the collection of high schools in school district $d \in \mathcal{D}$, where $N_S(d)$ is the number of high schools in district d . The analysis period is denoted by $\mathcal{T} = \{2008, 2009\}$.

Our econometric framework is based on the *potential outcomes* of student $i \in \mathcal{I}$ attending school $s \in \mathcal{S}_d$ in district $d \in \mathcal{D}$ and year $t \in \mathcal{T}$.¹⁵ We assume a linear random coefficient model of education production as follows:

$$Y_i(s, t, d) = \alpha'_i \mathbf{X}_s + \beta'_i \mathbf{Z}_{s,t} + \gamma_i v_s + \delta_i u_{s,t} + \eta_i c_d. \quad (1)$$

The potential outcome $Y_i(s, t, d)$ is the potential CSAT score of student i if (s)he attends school s (that is in district d) in year t . Observed school inputs are denoted by \mathbf{X}_s and $\mathbf{Z}_{s,t}$, which are vectors of time-invariant and time-varying school characteristics, respectively. v_s and $u_{s,t}$ are time-invariant and time-varying unobserved school inputs, respectively. c_d represents unobserved district characteristics.¹⁶ We allow for educational productivity (or sensitivity to school inputs) to vary across individuals. Heterogeneous effects of school inputs and district characteristics on test scores are represented by a vector of coefficients,

$$\theta_i = (\alpha'_i, \beta'_i, \gamma_i, \delta_i, \eta_i)'.$$

The potential outcomes model in equation (1) assumes that a student's potential test scores are determined by the interaction between the student's productivity θ_i .¹⁷

¹⁵ We use "high school senior" and "student" interchangeably.

¹⁶ The time effects are excluded in equation (1) because each year's CSAT scores are normalized.

¹⁷ We do not restrict the relationship between productivity θ_i and characteristics or quality of students. A high-achieving student can benefit more or less than a low-achieving student from a

and school and district characteristics,

$$\mathbf{F}_{s,t,d} = (\mathbf{X}'_s, \mathbf{Z}'_{s,t}, v_s, u_{s,t}, c_d)'.$$

Thus,

$$Y_i(s, t, d) = \boldsymbol{\theta}'_i \mathbf{F}_{s,t,d}.$$

The potential outcomes model in equation (1) is motivated by the textbook potential outcomes model that defines $Y_{1i} - Y_{0i}$ as the treatment effect for person i when the treatment is binary.¹⁸ Note that the textbook potential outcomes model is nonparametric and naturally allows for treatment effect heterogeneity. We assume the linear and interactive functional form in equation (1) to take into account the multidimensional nature of the treatment because each school is a package of various educational inputs. Assuming linearity is convenient when we analyze the effect of a subset of the multiple treatment components. Other than the functional form assumption, our potential outcomes model can be considered a natural extension of the textbook potential outcomes model.

In our data, we observe (Y_i, S_i, T_i, D_i) at the individual level and $(\mathbf{X}'_s, \mathbf{Z}'_{s,t})$ at the school level. S_i denotes the school that student i attends. T_i denotes the senior year of student i and also represents the birth cohort. D_i is the school district where student i lives and school S_i is located. Notably, student i 's CSAT score observed in the dataset is determined by the realized school assignment, birth cohort, and district choice as follows:

$$Y_i = Y_i(S_i, T_i, D_i). \quad (2)$$

The sampling scheme of the observed and unobserved variables is described in the following assumptions.

Assumption 1 (Individual Productivity). *Individual productivity $\boldsymbol{\theta}_i$ is randomly drawn as $\boldsymbol{\theta}_i \sim i.i.d. (\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$.*

particular school input.

¹⁸ When treatment takes more than two values, the potential outcomes can be expressed by a function $f_i(k)$ for person i and treatment value k , as described in Angrist and Pischke (2009). A well-known example is when the treatment is years of education and the outcomes are earnings. In this case, $f_i(m) - f_i(n)$ is the earnings effect of person i obtaining m instead of n years of education. In a regression model, a single error at all levels of education implies that the joint distribution of potential outcomes at all levels of education is identified. Because this joint distribution is not identified, we do not use the implications of the one-dimensional error. Nevertheless, we maintain the single error to facilitate the exposition.

Assumption 1 assumes that the productivities of individual students θ_i are randomly drawn from a superpopulation of students. Note that our data have all high school seniors in two cohorts who were randomly assigned to their schools. Thus, we do not have a random sample from the population of these high school seniors.

Assumption 2 (Random School Assignment within Districts and No Cohort Effects). *We assume that school assignment S_i and cohort effects T_i are independent of individual characteristics θ_i within each school district D_i , that is,*

$$\theta_i \perp (S_i, T_i) \mid D_i.$$

The orthogonality between school choice S_i and individual productivity θ_i within districts is based on within-district random school assignment. However, potential threats to the validity of the assumption remain.

First, students may sort within school districts (e.g., across neighborhoods smaller than a school district) if they believe they can increase their chances of being assigned to a particular school by living nearby.¹⁹ If commuting distance is considered in the school assignment process as many people conjectured, then (quasi-)experiment sites will be the neighborhoods within a school district and not the entire district. Then, the estimates on school input effects based on misspecified site boundaries will not be entirely free of selection bias. However, our proposed method is still valid for correcting bias from endogenous site selection as long as we have information on the exact boundaries between sites where the orthogonality assumption holds.

Second, students (and their parents) can adjust out-of-school educational inputs (e.g., private tutoring, parental involvement, and student effort) in reaction to school quality, and the non-school inputs may partly account for students' academic performance. In our potential outcomes model, θ_i represents innate productivity unaffected by school assignment and thus excludes such behavioral responses. This condition is not unrealistic in our case given that we find little effect of school environment (single-sex status and class size) on parental and student behaviors, including expenditures on private out-of-school education, parental support and supervision of student learning, and student's homework or self-study hours, from the KELS data. We will need a multi-period model as in Todd and Wolpin (2003) to rigorously consider the aforementioned responses as mediating outcomes in

¹⁹ We abstract from the possibility of within-district sorting following the majority of previous studies that exploit random high school assignment in Seoul: Park et al. (2012, 2013, 2018), Choi et al. (2015), Sohn (2016), Dustmann et al. (2017), and Hahn et al. (2018). See Online Appendix B.1 for a discussion of high school assignment in Seoul.

education production. Our data contain no information about family inputs to implement such a full-fledged model, and this limitation is shared with many of the existing studies on school input effects.

The assumption that the distribution of θ_i is the same in two different cohorts does not directly follow from the random assignment. However, given that we are considering cohorts born in two consecutive years, the random coefficients can be assumed to have the same distribution in 2008 and 2009 within a school district.

The sampling variation of our estimators is due to the sampling of θ_i and the random assignment to school S_i and cohort T_i . We condition on school and district characteristics and district choice.

Assumption 3 (School and District Characteristics). *All of the school- and district-level variables, $\mathbf{F} = \{\mathbf{F}_{s,t,d} : s \in \mathcal{S}_d, t \in \mathcal{T}, d \in D\}$, are fixed.*

We treat school and district characteristics as predetermined, that is, we assume that school and district characteristics do not depend on student characteristics θ_i .²⁰ If this assumption holds, treating school and district characteristics as fixed does not bias our estimates.

In the inference, we also condition on the district choices (D_1, \dots, D_{N_I}) of the students (or their parents). We do not have enough information on the students or their parents to construct and estimate a model of district choice. However, district choice likely depends on θ_i . Thus, the distribution of θ_i will be different across districts due to self-selection into districts.

In Section 4.1, we explore the effect of self-selection into districts on the district average of θ_i . The average productivity of students in district d is the district-specific APE denoted by

$$\theta_d = (\alpha'_d, \beta'_d, \gamma_d, \delta_d, \eta_d)' = \mathbb{E}[(\alpha'_i, \beta'_i, \gamma_i, \delta_i, \eta_i)' | D_i = d]. \quad (3)$$

We show in Section 5 that the population APE, which is the parameter of interest, is unaffected by (endogenous) district choice unlike the district-specific APE.

We introduce notations that will be used in the following sections. Let $N_I(s, t, d) = \sum_{i \in \mathcal{I}} \mathbb{I}\{S_i = s, T_i = t, D_i = d\}$ denote the number of high school seniors at school s (in district d) in year t .²¹ $N_I(s, d) = \sum_{t \in \mathcal{T}} N_I(s, t, d)$ is the number

²⁰ This assumption is not unrealistic given our institutional setting. The school budget is determined before the beginning of an academic year and thus cannot be immediately adjusted to changes in student characteristics. In addition, even private schools were heavily regulated by the government during the analysis period.

²¹ $\mathbb{I}\{\cdot\}$ denotes an indicator function that takes a value of 1 if the statement in the curly brackets is true and 0 otherwise.

of students at school s (in district d) in all years. Similarly, $N_I(d) = \sum_{s \in \mathcal{S}_I} N_I(s, d)$ is the number of students in district d in all years. Note that $\hat{w}_d^s = \frac{N_I(s, d)}{N_I(d)}$ and $\hat{w}_{s, d}^t = \frac{N_I(s, s, d)}{N_I(s, d)}$ are consistent estimators of $w_d^s = \mathbb{P}[S_i = s \mid D_i = d]$ and $w_{s, d}^t = \mathbb{P}[T_i = t \mid S_i = s, D_i = d]$, respectively, if $N_I \rightarrow \infty$.

4.1. A Simple Numerical Example

In this subsection, we provide a simple numerical example to illustrate that differences in students and district characteristics can cause endogenous sorting into districts and subsequently result in heterogeneous district-specific APE, $\{(\alpha'_d, \beta'_d)', d \in \mathcal{D}\}$. For simplicity, we assume a setting with a single binary school characteristic (e.g., single-sex school), one time period, and two school districts.

In the simplified setting, the potential outcomes model in equation (1) becomes

$$Y_i(s, d) = \alpha_i X_s + \eta_i c_d, \quad (4)$$

where X_s is a binary school characteristic, $d \in \{1, 2\}$, and $s \in \mathcal{S}_d$.

Under random assignment to schools within each district, the expected outcome if district d is chosen is

$$U_i(d) = \alpha_i \mathbb{P}[X_s = 1 \mid D = d] + \eta_i c_d = \alpha_i P_d + \eta_i c_d,$$

where the uncertainty is over X_s . $P_d = \sum_{s \in \mathcal{S}_d} X_s w_d^s$ denotes the probability of being assigned to a school with $X_s = 1$ in district d .

Suppose that student i chooses his or her school district by maximizing the expected outcome, that is, student i locates in district 1 over 2 if and only if $U_i(1) \geq U_i(2)$. In this case,

$$D_i = 2 - I\{\alpha_i p + \eta_i c \geq 0\},$$

where $p = P_1 - P_2$ and $c = c_1 - c_2$. p is the difference in the probability of being assigned to a school with $X_s = 1$ and c is the difference in other characteristics between two districts.

We generate the data by setting $\alpha_i \sim N(0, 1)$, $\eta_i \sim N(0, 1)$, and $(p, c) = \{(0, 0.2), (0.2, 0), (0.2, 0.2)\}$. When $(p, c) = (0, 0.2)$, district choice depends only on η_i . When $(p, c) = (0.2, 0)$, district selection depends only on α_i . Selection depends on both types of individual heterogeneity in the more general case of $(p, c) = (0.2, 0.2)$. Notably, the APE is 0, and half of the students live in either district in all cases. Thus, the students are indifferent between the two districts on average.

[Table 2] Simulated Data

	$p = 0, c = 0.2$		$p = 0.2, c = 0$		$p = 0.2, c = 0.2$	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
A. DGP						
$\bar{\alpha}$	-0.000	[-0.011, 0.012]	-0.000	[-0.011, 0.012]	-0.000	[-0.011, 0.012]
B. Selection						
$\bar{\alpha}$ in district 1	-0.000	[-0.017, 0.016]	0.798	[0.788, 0.808]	0.564	[0.551, 0.578]
$\bar{\alpha}$ in district 2	-0.000	[-0.016, 0.016]	-0.798	[-0.808, -0.788]	-0.564	[-0.578, -0.551]
Fraction in district 1	0.500	[0.494, 0.506]	0.500	[0.494, 0.506]	0.500	[0.494, 0.506]
Fraction in district 2	0.500	[0.494, 0.506]	0.500	[0.494, 0.506]	0.500	[0.494, 0.506]
C. Estimation						
$\hat{\alpha}^{\text{APE}}$	-0.000	[-0.023, 0.023]	-0.000	[-0.029, 0.028]	-0.000	[-0.023, 0.023]
$\hat{\alpha}^{\text{DFE}}$	-0.000	[-0.023, 0.023]	0.069	[0.040, 0.098]	0.049	[0.026, 0.072]
$\hat{\alpha}^{\text{OLS}}$	-0.000	[-0.031, 0.030]	0.199	[0.169, 0.229]	0.400	[0.368, 0.431]

Notes: Data are generated from 10,000 replications of the model in equation (4) with 20,000 observations each. $\bar{\alpha}$ denotes the sample mean of α_i in each simulated dataset. $\hat{\alpha}^{\text{APE}}$, $\hat{\alpha}^{\text{DFE}}$, and $\hat{\alpha}^{\text{OLS}}$ denote the treatment effect estimator computed by our weighted average method, a linear regression with district fixed effects, and a linear regression without district fixed effects, respectively.

Table 2 presents the results for 10,000 repetitions with 20,000 observations each. We report the sample average of α_i in the entire sample (panel A) and by district (panel B). The simulation results show that, even if students are indifferent between the two districts on average, small differences in the probability of being assigned to $X_s = 1$ between the districts result in substantial deviations (half the standard deviation) of the district-specific APE from the population APE. This result can be observed from the positive selection in district 1 and the negative selection in district 2 with respect to α and/or η . This finding supports our interpretation of the differences in APE across districts as a consequence of self-selection into districts and not of omitted school inputs.

V. Estimation of and Inference on Average Partial Effects

The parameters of interest are the APE of the observed school inputs \mathbf{X}_s and $\mathbf{Z}_{s,t}$:

$$(\alpha', \beta')' = \sum_{d \in \mathcal{D}} \mathbb{E}[(\alpha'_i, \beta'_i)' | D_i = d] w_d = \sum_{d \in \mathcal{D}} (\alpha'_d, \beta'_d)' w_d, \quad (5)$$

where $w_d = \frac{N_I(d)}{N_I}$ is the fraction of individuals who choose school district d . Note that the APE is invariant to the district choice of students (or their parents).

If the district choice is independent of individual productivity, the average productivity is identical across districts, that is, $(\alpha', \beta')' = (\alpha'_d, \beta'_d)'$. This condition holds even if productivity levels are heterogeneous across individuals. Thus, we take into account the possibility of district selection being endogenous, that is, dependent on θ_i , by explicitly allowing that the average productivity varies across school districts.

5.1. Identification

To identify the APE $(\alpha', \beta')'$, we need to identify the district-specific school input effects $(\alpha'_d, \beta'_d)'$ for each school district $d \in \mathcal{D}$. This subsection provides sufficient conditions under which $(\alpha'_d, \beta'_d)'$ is identified.

Based on the potential outcomes in equation (1) and the observed outcome in equation (2), the observed test score of student i attending the assigned school S_i in the district of his or her choice $D_i = d$ in year T_i is expressed as

$$\begin{aligned} Y_i &= \alpha'_i \mathbf{X}_{S_i} + \beta'_i \mathbf{Z}_{S_i, T_i} + \gamma_i v_{S_i} + \delta_i u_{S_i, T_i} + \eta_i c_d \\ &= \alpha'_d \mathbf{X}_{S_i} + \beta'_d \mathbf{Z}_{S_i, T_i} + \gamma_d v_{S_i} + \delta_d u_{S_i, T_i} + \eta_d c_d + \varepsilon_i \end{aligned} \quad (6)$$

where

$$\varepsilon_i = (\alpha_i - \alpha_d)' \mathbf{X}_{S_i} + (\beta_i - \beta_d)' \mathbf{Z}_{S_i, T_i} + (\gamma_i - \gamma_d) v_{S_i} + (\delta_i - \delta_d) u_{S_i, T_i} + (\eta_i - \eta_d) c_d.$$

The independence of θ_i and (S_i, T_i) conditional on $D_i = d$ in Assumption 2 implies that

$$\mathbb{E}[\varepsilon_i \mid S_i, T_i, D_i = d] = 0. \quad (7)$$

The expectation is the average over the conditional distribution of productivities $(\alpha'_i, \beta'_i, \gamma_i, \delta_i, \eta_i)'$ for students in school S_i , district d , and year T_i .

Note that equation (6) is a linear regression model for data on students and schools in district d . In the district-specific regression, \mathbf{X}_{S_i} and \mathbf{Z}_{S_i, T_i} are regressors, $\gamma_d v_{S_i} + \delta_d u_{S_i, T_i} + \varepsilon_i$ is a composite regression error, and $\eta_d c_d$ is the intercept.

We need to impose restrictions on the relation between observed and unobserved school characteristics to identify the regression coefficients $(\alpha'_d, \beta'_d)'$ in equation (6).

Assumption 4 (Orthogonality Conditions). *For each $d \in \mathcal{D}$, we assume that*

$$\begin{aligned}
(i) \quad & \sum_{s \in \mathcal{S}_d} w_d^s \left(\mathbf{X}_s - \sum_{s \in \mathcal{S}_d} w_d^s \mathbf{X}_s \right) \nu_s = \mathbf{0}, \\
(ii) \quad & \sum_{s \in \mathcal{S}_d} w_d^s \sum_{t \in \mathcal{T}} w_{s,d}^t \left(\mathbf{X}_s - \sum_{s \in \mathcal{S}_d} w_d^s \mathbf{X}_s \right) u_{s,t} = \mathbf{0}, \\
(iii) \quad & \sum_{s \in \mathcal{S}_d} w_d^s \sum_{t \in \mathcal{T}} w_{s,d}^t \left(\mathbf{Z}_{s,t} - \sum_{t \in \mathcal{T}} w_{s,d}^t \mathbf{Z}_{s,t} \right) \left(u_{s,t} - \sum_{t \in \mathcal{T}} w_{s,d}^t u_{s,t} \right) = \mathbf{0}.
\end{aligned}$$

In Assumption 4 (i), we assume no correlation between the observed and unobserved time-constant school characteristics across schools within a district. In (ii), we assume that the time-constant observed school inputs are also uncorrelated with the time-average of the time-varying unobserved school input. In (iii), we assume that the time-varying part of \mathbf{Z} , that is, \mathbf{Z}_{j_t} minus its time average, is uncorrelated with the time-varying unobserved school input u_{st} . The time average of \mathbf{Z}_{j_t} and u_{st} can be correlated.

We have one time-varying and one time-invariant school characteristic of interest, namely, class size and single-sex education, respectively. Thus, we have to include additional school characteristics as controls to ensure that Assumption 4 holds. Accordingly, the assumption is that (i), (ii), and (iii) hold for the variables of interest, conditional on the control variables.

Time-varying control variables include total enrollment, number of teachers, fraction of regular teachers, fraction of female teachers, and log of annual school spending.^{22,23} Time-invariant control variables include a private school indicator, school establishment year, and the interaction between the two variables.²⁴

If Assumption 4 is violated, our estimates of the class size and single-sex effects are subject to omitted variable bias (OVB). We evaluate the robustness of our estimates to potential OVB by applying the method proposed by Altonji et al. (2005)

²² Annual school spending in our data is the school operating budget in an academic year excluding salaries for school personnel. The school operating budget covers expenditures on supplies, utilities, facilities, and academic or athletic programs. In Korea, teachers at public and private schools are paid according to a salary schedule set by the government. However, this information is unavailable in our data. Thus, by including spending on non-personnel items, number of teachers, and total enrollment in the covariate list, we effectively control for school resources per student.

²³ The annual school spending is measured in thousands of 2009 KRW. Approximately, 1,100 KRW is worth 1 USD.

²⁴ The private status and the age of school control for unobserved teacher quality, which is known to have a considerable impact on students' academic achievement (Rivkin et al., 2005). The reasons are twofold: 1) public and private schools have different teacher hiring processes, and 2) schools with longer history and stronger alumni network usually try to recruit better teachers to maintain their alumni power than their counterparts. Given that single-sex schools are three times more likely to be private and 23 years older than coed schools on average, unobserved teacher quality is likely to be systematically different between the two types of schools.

and Oster (forthcoming) (AET-Oster hereinafter). The idea is to gauge the magnitude of the bias from omitted unobserved confounders under the assumption that selection on unobservables is proportional to selection on observables. Table B4 shows that the bias-adjusted estimates from the AET-Oster method are similar to our APE estimates. Online Appendix B.4 presents further details.

Assumption 5 (Rank Conditions). *For each $d \in \mathcal{D}$, we assume that the matrices*

$$\sum_{s \in \mathcal{S}_d} w_d^s \left(\mathbf{X}_{s,d} - \sum_{s \in \mathcal{S}_d} w_d^s \mathbf{X}_{s,d} \right) \left(\mathbf{X}_{s,d} - \sum_{s \in \mathcal{S}_d} w_d^s \mathbf{X}_{s,d} \right)'$$

and

$$\sum_{s \in \mathcal{S}_d} w_d^s \sum_{t \in \mathcal{T}} w_{s,d}^t \left(\mathbf{Z}_{s,t} - \sum_{t \in \mathcal{T}} w_{s,d}^t \mathbf{Z}_{s,t} \right) \left(\mathbf{Z}_{s,t} - \sum_{t \in \mathcal{T}} w_{s,d}^t \mathbf{Z}_{s,t} \right)'$$

are positive definite.

Assumption 5 precludes perfect multicollinearity of the observed school inputs for the time-constant school input in deviation from its district average and for the time-varying input in deviation from its time average. α_d and α are identified from the variation in time-constant school resources across schools within each district. β_d and β are identified from the variation in time-varying school resources within schools over time.

Theorem 1. *Under Assumptions 2, 4, and 5, the district-specific APE $(\alpha'_d, \beta'_d)'$ in equation (3) and the population APE $(\alpha', \beta')'$ in equation (5) are identified.*

Proof. See Appendix A.1. □

5.2. Estimation and Statistical Inference

A natural estimator of the APE is the sample analog of equation (5), that is, we estimate the APE of school inputs by taking the weighted average of the estimated coefficients on the time-invariant and time-varying school characteristics in regression equation (6),

$$(\hat{\alpha}', \hat{\beta}')' = \sum_{d \in \mathcal{D}} (\hat{\alpha}'_d, \hat{\beta}'_d)' w_d. \quad (8)$$

We use the fraction of CSAT takers in each school district as the weight.²⁵

We implement a two-step estimation procedure using data on individual test scores and school characteristics from school district d to estimate the district-specific effects of school inputs $(\alpha'_d, \beta'_d)'$.²⁶ We first estimate the effect of time-varying school inputs β_d and then estimate the effect of time-invariant school inputs α_d using the estimates obtained in the first step. The estimators $\hat{\alpha}_d$ and $\hat{\beta}_d$ obtained from the two-step estimation are consistent. The asymptotic properties of the school input effect estimators are described in Online Appendix B.5.

Step 1. Estimation of β_d : The estimator that solves the sample version of the identifying moment condition in equation (A.1) is

$$\hat{\beta}_d = \left[\sum_{s \in S_d} \sum_{i: S_i = s} \left(\mathbf{Z}_{S_i, T_i} - \frac{1}{N_I(s, d)} \sum_{i: S_i = s} \mathbf{Z}_{S_i, T_i} \right) \left(\mathbf{Z}_{S_i, T_i} - \frac{1}{N_I(s, d)} \sum_{i: S_i = s} \mathbf{Z}_{S_i, T_i} \right)' \right]^{-1} \\ \times \sum_{s \in S_d} \sum_{i: S_i = s} \left(\mathbf{Z}_{S_i, T_i} - \frac{1}{N_I(s, d)} \sum_{i: S_i = s} \mathbf{Z}_{S_i, T_i} \right) \left(Y_i - \frac{1}{N_I(s, d)} \sum_{i: S_i = s} Y_i \right). \quad (9)$$

Note that $\hat{\beta}_d$ is the within estimator obtained from the individual-level ordinary least squares (OLS) regression of Y_i on time-varying school inputs \mathbf{Z}_{S_i, T_i} and school fixed effects using data from school district d .

Step 2. Estimation of α_d : The estimator that solves the sample version of the identifying moment condition in equation (A.2) is

$$\hat{\alpha}_d = \left[\sum_{i: D_i = d} \left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)} \sum_{i: D_i = d} \mathbf{X}_{S_i} \right) \left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)} \sum_{i: D_i = d} \mathbf{X}_{S_i} \right)' \right]^{-1} \\ \times \sum_{i: D_i = d} \left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)} \sum_{i: D_i = d} \mathbf{X}_{S_i} \right) (Y_i - \hat{\beta}'_d \mathbf{Z}_{S_i, T_i}). \quad (10)$$

Using data from school district d , $\hat{\alpha}_d$ is computed by the individual-level OLS regression of $Y_i - \hat{\beta}'_d \mathbf{Z}_{S_i, T_i}$ on \mathbf{X}_{S_i} with a constant term.

²⁵ The APE estimates are robust to using the fraction of high school seniors (12th graders) in each district or the fraction of high school freshmen (10th graders) at random assignment in each district as the weight.

²⁶ The codes used for the APE estimation are available upon request.

In estimation of $\text{Var}(\hat{\alpha}_d)$ and $\text{Var}(\hat{\beta}_d)$, we use White's robust standard errors clustered at the school-by-cohort level to account for heteroskedastic errors in the random coefficient specification and the correlation of individual outcomes within each school and cohort in a school district. The sampling variation of $\hat{\beta}_d$ estimated in the first step is taken into account when the standard error of $\hat{\alpha}_d$ in the second step is computed. Appendix A.2 shows the formula of the standard errors of $\hat{\alpha}_d$ and $\hat{\beta}_d$, and Online Appendix B.5 shows the derivations.

Based on the APE estimator in equation (8) and Assumption 3, the vector of estimated variances of $\hat{\alpha}$ and $\hat{\beta}$ are expressed as

$$\hat{\sigma}_{\alpha}^2 = \sum_{d \in \mathcal{D}} \widehat{\text{diag}(\text{Var}(\hat{\alpha}_d))} w_d^2, \quad \hat{\sigma}_{\beta}^2 = \sum_{d \in \mathcal{D}} \widehat{\text{diag}(\text{Var}(\hat{\beta}_d))} w_d^2,$$

where $\widehat{\text{diag}(\text{Var}(\hat{\alpha}_d))}$ and $\widehat{\text{diag}(\text{Var}(\hat{\beta}_d))}$ are the diagonal elements of $\widehat{\text{Var}(\hat{\alpha}_d)}$ and $\widehat{\text{Var}(\hat{\beta}_d)}$, respectively.

VI. Empirical Results

6.1. Average Partial Effect Estimates

The main estimation results are presented in Table 3. The estimated effects of single-sex schooling and class size on the CSAT score are reported for boys and girls separately.²⁷ As mentioned above, the CSAT score and the estimates are in standard deviation units. Panel A presents our preferred estimates controlling for time-varying and time-invariant school characteristics that are potentially confounded with variables of interest, namely, single-sex indicator and class size. The estimates without controls are reported in panel B.

The APE estimates in the first row of panel A show that single-sex schools have nearly zero or insignificant negative effects on the academic performance of male and female students (-0.003 on boys' CSAT scores and -0.041 on girls'). Controlling for other confounding factors is important when estimating the effect of single-sex schooling. The reason is that we rely on the "selection on observables" in Assumption 4 to identify the APE of a time-invariant school input within each school district. Notably, the large positive estimates of the single-sex school effect in panel B (0.148 for boys and 0.106 for girls) disappear when the school characteristics listed in Section 5.1 are controlled in the regression. A robustness check using the

²⁷ We conduct a separate analysis by gender because high school assignment lotteries are separate for male and female students.

[Table 3] Effects of School Inputs on the CSAT Score

	Boys		Girls	
	Single-sex	Class size	Single-sex	Class size
A. With controls				
<i>All districts</i>				
APE	-0.003 (0.084)	-0.014 (0.003)***	-0.041 (0.068)	-0.006 (0.002)**
<i>By district</i>				
District 3	0.472 (0.171)***	0.000 (0.009)	0.041 (0.123)	0.002 (0.006)
District 4	0.068 (0.294)	-0.016 (0.008)*	0.294 (0.137)**	-0.009 (0.004)**
District 6	-0.004 (0.049)	-0.022 (0.006)***	-0.077 (0.096)	0.000 (0.005)
District 7	-0.115 (0.218)	-0.020 (0.006)***	-0.672 (0.279)**	-0.016 (0.006)***
District 8	0.116 (0.111)	-0.007 (0.005)	0.162 (0.080)**	-0.004 (0.006)
District 9	-0.727 (0.361)**	-0.017 (0.009)*	0.086 (0.168)	-0.004 (0.008)
<i>Testing school input effects identical across districts</i>				
χ^2 -statistic	12.61 [0.027]	7.50 [0.186]	13.48 [0.019]	7.53 [0.184]
B. No controls				
<i>All districts</i>				
APE	0.148 (0.027)***	-0.010 (0.002)***	0.106 (0.029)***	0.005 (0.001)***
<i>By district</i>				
District 3	0.068 (0.078)	-0.004 (0.007)	0.085 (0.077)	0.003 (0.005)
District 4	0.214 (0.061)***	-0.014 (0.005)***	0.200 (0.052)***	0.010 (0.003)***
District 6	0.034 (0.045)	-0.015 (0.002)***	0.018 (0.053)	0.004 (0.002)*
District 7	0.190 (0.090)**	0.001 (0.004)	0.029 (0.109)	0.008 (0.003)***
District 8	0.186 (0.047)***	-0.008 (0.005)*	0.146 (0.044)***	0.011 (0.003)***
District 9	0.189 (0.045)***	-0.024 (0.006)***	0.192 (0.058)***	-0.013 (0.005)***
<i>Testing school input effects identical across districts</i>				
χ^2 -statistic	10.43 [0.064]	18.71 [0.002]	8.76 [0.119]	21.75 [0.001]

Notes: Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The CSAT score refers to the sum of Korean and English CSAT scores and is standardized to have zero mean and unit variance. For CSAT applicants who were absent during the exam, missing raw scores on Korean or English are imputed with zeros. Time-varying control variables include total enrollment, number of teachers, fraction of regular teachers, fraction of female teachers, and log of annual school spending (in thousands of 2009 KRW). Time-invariant control variables include a private indicator, school establishment year, and the interaction between the two variables. Robust standard errors clustered at the school-year level are in parentheses. Section 5.2 and Appendix A.2 provide additional details on standard error computation. p -values associated with χ^2 -statistics are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

AET-Oster method shows that OVB is unlikely to be a major threat to the APE estimates in panel A (see Online Appendix B.4 and Table B4).²⁸

²⁸ By contrast, the specification used in Choi et al. (2014) seems subject to OVB. Online Appendix B.6 and Table B5 present a comparison with Choi et al. (2014).

Smaller classes appear to slightly increase the test scores for both genders. For boys, a one-student decrease in class size is associated with an increase in the CSAT score by 0.014 standard deviations, and the effect is statistically significant at the 1% level. The effect of class size reduction for girls is also positive and significant at the 5% level but is less than half in magnitude than that for boys. The APE estimate of the class size effect is very robust to different sets of control variables possibly because most of the confounding factors are absorbed by school fixed effects. This result can also be observed from the similarity between the estimated APE of class size reported in panels A (with controls) and B (without controls).²⁹

The next six rows in panel A of Table 3 show district-specific coefficients on single-sex schooling and class size, which include district selection effects. The estimated effects of single-sex schooling on the CSAT score vary substantially across school districts. By contrast, the district-specific estimates of class size effects are considerably more homogeneous across school districts. This finding is confirmed by the χ^2 tests on the null hypothesis that the school input effects are uniform across school districts. The null hypothesis is rejected at the 5% significance level for single-sex effects on boys and girls but not rejected at any conventional level of significance for class size effects on both genders.

The heterogeneous effects of a school input imply a substantial degree of sorting on the individual productivity of that input across school districts.³⁰ The results suggest that the chance of getting into a single-sex school is an important factor for students and their parents in choosing their district in Seoul. However, we need a model for district choice to understand the exact mechanism of sorting, and our data are not rich enough to identify such a model.

Our estimation results are consistent with findings from previous studies on the effect of class size reduction or single-sex schools in secondary education. Benefits from smaller classes are found to be small (not more than 2% of a standard deviation of standardized test scores for a one-student change) or nonexistent in American, European, and Korean secondary schools (Rivkin et al., 2005; Wößmann and West, 2006; Chingos, 2012; Han and Ryu, 2017).³¹ The results are mixed on whether or how class size effects vary by student characteristics (Chingos, 2013; Jepsen, 2015). Meanwhile, the effect of single-sex schooling seems more heterogeneous than that of class size. Jackson (2012, 2016) finds single-sex schools

²⁹ Class size is positively correlated with the CSAT score for girls when no control variables are applied, but the magnitude of the estimate is very small.

³⁰ We present the estimation results for Korean and English CSAT scores in Tables B6 and B7, respectively. When the CSAT score on each section is used as an outcome variable, we normalize the test score to have zero mean and unit variance within each section. The results are qualitatively similar to the main findings using the total score in Table 3. However, the effects on the English score are more heterogeneous across districts than those on the Korean score.

³¹ Additional references can be found in Jepsen (2015).

ineffective in enhancing academic performance for most students but beneficial for girls with strong preferences for single-sex schools or students with disadvantaged backgrounds.

6.2. Comparison with Estimates from Conventional Linear Regressions

We compare the estimated APE of single-sex schooling and class size based on our econometric framework with estimates from conventional linear regression models that ignore endogenous district selection. The latter includes a linear regression with district fixed effects that has been commonly used in previous studies to exploit random assignment within districts.

We describe the main differences between our weighted average method for the estimation of the APE and the commonly used linear regression method first by using the numerical example with a single binary treatment X_i in Section 4.1. We report the sampling mean and 95% confidence interval of three estimators in panel C of Table 2. Our method computes $\hat{\alpha}^{\text{APE}}$ by regressing Y_i on X_i using observations in each district and taking the average of the district-specific coefficient estimates on X_i weighted by the fraction of students in each district. $\hat{\alpha}^{\text{DFE}}$ is obtained from a linear regression of Y_i on X_i with district fixed effects and $\hat{\alpha}^{\text{OLS}}$ from a linear regression of Y_i on X_i without district fixed effects.

We compare the performance of the three estimation methods in three illustrative cases with different values of (p, c) . Our weighted average method recovers the true average treatment effect of zero in all cases. If the district choice does not depend on α_i (i.e., $p = 0$), then all three methods estimate the APE well. However, the regression model with or without district fixed effects does not effectively remove the selection bias generated by the dependence of the district choice on α_i (i.e., $p > 0$). Note that the district fixed effects regression assumes district selection is based only on η_i , whereas the simple linear regression model assumes no endogenous selection (neither on α_i nor on η_i). The bias in $\hat{\alpha}^{\text{DFE}}$ is smaller than that in $\hat{\alpha}^{\text{OLS}}$ because district-specific intercepts partially remove the selection bias.

Next, we extend the comparison to our real-world data. Since the students in each cohort are randomly assigned to schools within each school district, we report results from a specification that includes district-year fixed effects, as in Lee et al. (2014), Sohn (2016), Dustmann et al. (2017), and Hahn et al. (2018). When we use district fixed effects instead of district-year fixed effects, the estimation results change little. A linear regression with district-year fixed effects (DFE) estimates the following equation:

$$Y_i = \alpha^{\text{DFE}'} \mathbf{X}_{S_i} + \beta^{\text{DFE}'} \mathbf{Z}_{S_i, T_i} + \sum_{t \in T, d \in D} \lambda_{t,d} \mathbb{I}\{T_i = t, D_i = d\} + \chi_i.$$

[Table 4] Comparison with Other Empirical Strategies

	Boys			Girls		
	Our approach	DFE	OLS	Our approach	DFE	OLS
A. With controls						
Single-sex	-0.003 (0.084)	0.012 (0.047)	-0.070 (0.053)	-0.041 (0.068)	0.061 (0.052)	0.041 (0.055)
Class size	-0.014 (0.003)***	-0.008 (0.011)	-0.033 (0.009)***	-0.006 (0.002)**	0.003 (0.010)	-0.019 (0.008)**
B. No controls						
Single-sex	0.148 (0.027)***	0.148 (0.027)***	0.196 (0.041)***	0.106 (0.029)***	0.099 (0.029)***	0.145 (0.040)***
Class size	-0.010 (0.002)***	0.009 (0.009)	-0.009 (0.009)	0.005 (0.001)***	0.031 (0.008)***	0.013 (0.008)*

Notes: Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The CSAT score refers to the sum of Korean and English CSAT scores and is standardized to have zero mean and unit variance. For CSAT applicants who were absent during the exam, missing raw scores on Korean or English are imputed with zeros. Time-varying control variables include total enrollment, number of teachers, fraction of regular teachers, fraction of female teachers, and log of annual school spending (in thousands of 2009 KRW). Time-invariant control variables include a private indicator, school establishment year, and the interaction between the two variables. Robust standard errors clustered at the school-year level are in parentheses. Section 5.2 and Appendix A.2 present additional details on standard error computation for our approach. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 presents the estimated effects of single-sex schooling and class size from our approach and those from the DFE regression. In panel A with controls, the DFE model overestimates the effect of single-sex schools. Compared with the negative APE estimates from our approach, the estimates from the DFE regression are opposite in sign (positive) and larger in magnitude (-0.003 vs. 0.012 for boys and -0.041 vs. 0.061 for girls). However, the estimates from both methods are not significantly different from zero. For the class size effects, estimates from the DFE regressions are closer to zero (less negative or positive with a smaller magnitude) than our APE estimates. However, the difference is small (-0.014 vs. -0.008 for boys and -0.006 vs. 0.003 for girls).^{32,33}

³² In panel B without controls, the two approaches yield very similar estimates on the single-sex school effect. On the contrary, estimates from the DFE regressions are substantially larger than our estimates for the class size effect.

³³ We also report estimates from pooled OLS regressions without district or district-year fixed effects in Table 4. The OLS estimates of single-sex and class size effects are likely biased and indeed quite

Both in the numerical example and the empirical application, the estimate of the APE is upwardly biased if district fixed effects are used to control for selection into districts. Therefore, it is not surprising that previous studies using the DFE model and within-district random assignment in Seoul report a large positive estimate (around 0.1 standard deviation of a standardized test score) of the single-sex effect.³⁴ However, the bias in the DFE estimator can go either way depending on the sign and the magnitude of district-specific APE and the within-district variance of the school input. If there is more variation in the school input in districts with positive selection than in districts with negative selection, the DFE estimate will be biased upwards (and vice versa).

VII. Conclusion

In this study, we propose an econometric framework to estimate the average partial effects (APE) of school inputs on students' academic performance by exploiting within-district school lotteries. Our estimation method is implemented using data on CSAT scores and high school characteristics from Seoul, Korea in the period of 2008-2009. Identification relies on the fact that self-selection occurs across school districts but not into schools within each district. We focus on the effects of single-sex schooling and class size to illustrate the APE estimation of time-invariant and time-varying school inputs.

In our potential outcomes model, each student's academic outcome is determined by observed and unobserved school inputs that interact with the student's productivity. From the random coefficient model of individual-level education production, we derive a district-specific linear regression equation of individual academic outcomes on school inputs. The district-specific effects of school inputs are estimated in two steps within each district: the coefficients on time-varying school inputs are estimated first and then the coefficients on time-invariant school inputs are estimated using the estimates from the first step. The APE estimates are the weighted average of the district-specific estimates.

The Monte Carlo simulations show that even a modest difference in school characteristics across districts can lead to a substantial degree of sorting. In the empirical analysis on Seoul high schools, we find a small positive effect of smaller classes but no evidence on benefits of single-sex schools. A substantial degree of across-district heterogeneity is observed especially in the effect of single-sex schools. This finding is consistent with the results from the simulation analysis. We also

different from our APE estimates or the DFE estimates. A systematic pattern cannot be found in the sign or magnitude of the bias.

³⁴ See Park et al. (2013), Lee et al. (2014), and Dustmann et al. (2017).

show that estimates from a conventional linear regression with district(-year) fixed effects, which only partially corrects for endogenous sorting across school districts, are biased.

Although we focus on the case of within-district high school lotteries and education production in Seoul, the multi-site experiment framework can be applied to a broader range of randomized controlled trials and natural experiments conducted in multiple locations or groups. Our work complements studies on multi-site experiments that focus on the external validity with endogenous site selection (Hotz et al., 2005; Allcott, 2015).

A. Appendix

A.1. Proof of Theorem 1

From equations (6) and (7), we have

$$Y_i - \mathbb{E}[Y_i | S_i, D_i = d] - \boldsymbol{\beta}'_d \left(\mathbf{Z}_{S_i, T_i} - \sum_{t \in T} w'_{S_i, d} \mathbf{Z}_{S_i, t} \right) = \boldsymbol{\delta}'_d \left(u_{S_i, T_i} - \sum_{t \in T} w'_{S_i, d} u_{S_i, t} \right) + \varepsilon_i.$$

In this expression, $\mathbb{E}[Y_i | S_i, D_i = d]$ is the time-average outcome in school S_i and district d , that is,

$$\mathbb{E}[Y_i | S_i, D_i = d] = \sum_{t \in T} w'_{S_i, d} \mathbb{E}[Y_i | S_i, T_i = t, D_i = d].$$

Under Assumption 4, the right-hand side error is uncorrelated with

$$\mathbf{Z}_{S_i, T_i} - \sum_{t \in T} w'_{S_i, d} \mathbf{Z}_{S_i, t}.$$

Thus, if the inner \mathbb{E} denotes an average over $T_i | D_i = d$, we have the moment condition that identifies $\boldsymbol{\beta}_d$.

Identification of $\boldsymbol{\beta}_d$: Under Assumptions 2, 4, and 5, the regression coefficients $\boldsymbol{\beta}_d$ can be identified by the following moment condition:

$$\begin{aligned} \mathbb{E} \left[\left\{ \begin{array}{c} Y_i - \mathbb{E}[Y_i | S_i, D_i = d] \\ -\mathbf{b}'_d (\mathbf{Z}_{S_i, T_i} - \mathbb{E}[\mathbf{Z}_{S_i, T_i} | S_i, D_i = d]) \end{array} \right\} (\mathbf{Z}_{S_i, T_i} - \mathbb{E}[\mathbf{Z}_{S_i, T_i} | S_i, D_i = d]) | D_i = d \right] &= \mathbf{0} \\ \Leftrightarrow \mathbf{b}_d &= \boldsymbol{\beta}_d \end{aligned} \quad (\text{A.1})$$

where uniqueness follows from Assumption 5.

Identification of $\boldsymbol{\alpha}_d$: Once $\boldsymbol{\beta}_d$ is identified, we can treat the parameter as if it is known. We have

$$\begin{aligned} Y_i - \mathbb{E}[Y_i | D_i = d] - \boldsymbol{\alpha}'_d (\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i} | D_i = d]) - \boldsymbol{\beta}'_d (\mathbf{Z}_{S_i, T_i} - \mathbb{E}[\mathbf{Z}_{S_i, T_i} | D_i = d]) \\ = \gamma_d (v_{S_i} - \mathbb{E}[v_{S_i} | D_i = d]) + \delta_d (u_{S_i, T_i} - \mathbb{E}[u_{S_i, T_i} | D_i = d]) + \varepsilon_i. \end{aligned}$$

Under Assumptions 2 and 4 (i) for the error term on the right-hand side, we have

$$\mathbb{E}[(v_{S_i} - \mathbb{E}[v_{S_i} | D_i = d])(\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i} | D_i = d]) | D_i = d] = 0.$$

From Assumption 4 (ii), we have

$$\mathbb{E}[(u_{S_i, T_i} - \mathbb{E}[u_{S_i, T_i} | D_i = d])(\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i} | D_i = d]) | D_i = d] = 0.$$

Therefore, from Assumption 5, we can identify the regression coefficients α_d by the following moment condition.

$$\begin{aligned} \mathbb{E} \left[\begin{pmatrix} Y_i - \mathbb{E}[Y_i | D_i = d] - \beta_d'(\mathbf{Z}_{S_i, T_i} - \mathbb{E}[\mathbf{Z}_{S_i, T_i} | D_i = d]) \\ -\mathbf{a}_d'(\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i} | D_i = d]) \end{pmatrix} (\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i} | D_i = d]) | D_i = d \right] = \mathbf{0} \\ \Leftrightarrow \mathbf{a}_d = \alpha_d \end{aligned} \quad (\text{A.2})$$

A.2. Standard Errors of $\hat{\alpha}_d$ and $\hat{\beta}_d$

The OLS residuals from the first-step estimation are

$$\hat{\xi}_{i, S_i, T_i, d} = Y_i - \frac{1}{N_I(S_i, d)} \sum_{j: S_j = S_i} Y_j - \hat{\beta}_d' \left(\mathbf{Z}_{S_i, T_i} - \frac{1}{N_I(S_i, d)} \sum_{k: S_k = S_i} \mathbf{Z}_{S_k, T_k} \right).$$

The OLS residuals from the second-step estimation are

$$\hat{\zeta}_{i, S_i, T_i, d} = Y_i - \widehat{\eta_d c_d} - \hat{\alpha}_d' \mathbf{X}_{S_i} - \hat{\beta}_d' \mathbf{Z}_{S_i, T_i},$$

where $\widehat{\eta_d c_d}$ is the estimated constant term.

We estimate the asymptotic variance of $\hat{\beta}_d$ by using White's robust standard errors approach with clustering to account for heteroskedastic errors in the random coefficient specification and correlation of individual outcomes within each school and cohort in a school district:

$$\widehat{\text{Var}}(\hat{\beta}_d) = \hat{\mathbf{A}}_{\beta}^{-1} \hat{\Sigma}_{\beta} \hat{\mathbf{A}}_{\beta}^{-1}$$

with

$$\hat{\mathbf{A}}_{\beta} = \sum_{s \in \mathcal{S}_d} \hat{\mathbf{w}}_d^s \sum_{t \in \mathcal{T}} \hat{\mathbf{w}}_{s, d}^t \left(\mathbf{Z}_{s, t} - \sum_{t' \in \mathcal{T}} \hat{\mathbf{w}}_{s, d}^{t'} \mathbf{Z}_{s, t'} \right) \left(\mathbf{Z}_{s, t} - \sum_{t' \in \mathcal{T}} \hat{\mathbf{w}}_{s, d}^{t'} \mathbf{Z}_{s, t'} \right)',$$

$$\hat{\Sigma}_{\beta} = \sum_{s \in \mathcal{S}_d} \sum_{t \in T} (\hat{w}_d^s \hat{w}_{s,d}^t)^2 \bar{\xi}_{s,t,d}^2 \left(\mathbf{Z}_{s,t} - \sum_{t' \in T} \hat{w}_{s,d}^{t'} \mathbf{Z}_{s,t'} \right) \left(\mathbf{Z}_{s,t} - \sum_{t' \in T} \hat{w}_{s,d}^{t'} \mathbf{Z}_{s,t'} \right)',$$

and the average residual in cluster s, t, d ,

$$\bar{\xi}_{s,t,d} = \frac{1}{N_I(s, t, d)} \sum_{i: S_i = s, T_i = t} \hat{\xi}_{i, S_i, T_i, d}.$$

The variance of $\hat{\alpha}_d$ is estimated by

$$\widehat{\text{Var}}(\hat{\alpha}_d) = \hat{\mathbf{A}}_{\alpha}^{-1} (\hat{\Sigma}_{11} + \hat{\Sigma}_{22} - 2\hat{\Sigma}_{12}) \hat{\mathbf{A}}_{\alpha}^{-1}$$

with

$$\begin{aligned} \hat{\mathbf{A}}_{\alpha} &= \sum_{s \in \mathcal{S}_d} \hat{w}_d^s \left(\mathbf{X}_{s,d} - \sum_{s' \in \mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s',d} \right) \left(\mathbf{X}_{s,d} - \sum_{s' \in \mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s',d} \right)', \\ \hat{\Sigma}_{11} &= \sum_{s \in \mathcal{S}_d} \sum_{t \in T} (\hat{w}_d^s \hat{w}_{s,d}^t)^2 \bar{\xi}_{s,t,d}^2 \left(\mathbf{X}_s - \sum_{s' \in \mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s'} \right) \left(\mathbf{X}_s - \sum_{s' \in \mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s'} \right)', \\ \hat{\Sigma}_{22} &= \hat{\mathbf{B}} \widehat{\text{Var}}(\hat{\beta}_d) \hat{\mathbf{B}}', \\ \hat{\Sigma}_{12} &= \sum_{s \in \mathcal{S}_d} \sum_{t \in T} (\hat{w}_d^s \hat{w}_{s,d}^t)^2 \bar{\xi}_{s,t,d}^2 \left(\mathbf{X}_s - \sum_{s' \in \mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s'} \right) \left(\mathbf{Z}_{s,t} - \sum_{t' \in T} \hat{w}_{s,d}^{t'} \mathbf{Z}_{s,t'} \right)' \hat{\mathbf{A}}_{\beta}^{-1} \hat{\mathbf{B}}', \\ \hat{\mathbf{B}} &= \left[\sum_{s \in \mathcal{S}_d} \sum_{t \in T} \hat{w}_d^s \hat{w}_{s,d}^t \left(\mathbf{X}_s - \sum_{s' \in \mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s'} \right) \mathbf{Z}_{s,t}' \right], \end{aligned}$$

and the cluster average of the residuals

$$\bar{\xi}_{s,t,d} = \frac{1}{N_I(s, t, d)} \sum_{i: S_i = s, T_i = t} \hat{\xi}_{i, S_i, T_i, d}.$$

When computing the standard error of $\hat{\alpha}_d$ in the second step, we take into account the sampling variation of $\hat{\beta}_d$ estimated in the first step. Note that omitting $\hat{\Sigma}_{22} - 2\hat{\Sigma}_{12}$ yields White's robust standard errors without considering the sampling variation of $\hat{\beta}_d$.

References

- Allcott H. (2015), "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 130, 1117–1165. DOI: 10.1093/qje/qjv015.
- Altonji J. G., T. E. Elder, and C. R. Taber (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151–184. DOI: 10.1086/426036.
- Altonji J. G., and R. K. Mansfield (2014), "Group-Average Observables as Controls for Sorting on Unobservables When Estimating Group Treatment Effects: the Case of School and Neighborhood Effects," Working Paper 20781, National Bureau of Economic Research. DOI: 10.3386/w20781.
- Angrist J. D., and V. Lavy (1999), Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533–575. DOI: 10.1162/003355399556061.
- Angrist J. D., and J. S. Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.
- Chingos M. M. (2012), "The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida's Statewide Mandate," *Economics of Education Review*, 31, 543–562. DOI: 10.1016/j.econedurev.2012.03.002.
- _____ (2013), "Class Size and Student Outcomes: Research and Policy Implications," *Journal of Policy Analysis and Management*, 32, 411–438. DOI: 10.1002/pam.21677.
- Choi E. J., H. R. Moon, and G. Ridder (2014), "Estimation of an Education Production Function under Random Assignment with Selection," *American Economic Review: Papers and Proceedings*, 104, 206–211. DOI: 10.1257/aer.104.5.206.
- Choi J., H. Park, and J. R. Behrman (2015), "Separating Boys and Girls and Increasing Weight? Assessing the Impacts of Single-Sex Schools Through Random Assignment in Seoul," *Social Science & Medicine*, 134, 1–11. DOI: 10.1016/j.socscimed.2015.03.053.
- Cullen J. B., B. A. Jacob, and S. Levitt (2006), "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica*, 74, 1191–1230. DOI: 10.1111/j.1468-0262.2006.00702.x.
- Deming D. J. (2011), "Better Schools, Less Crime?" *Quarterly Journal of Economics*, 126, 2063–2115. DOI: 10.1093/qje/qjr036.
- Deming D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014), "School Choice, School Quality, and Postsecondary Attainment," *American Economic Review*, 104, 991–1013. DOI: 10.1257/aer.104.3.991.
- Dustmann C., H. Ku, and D. W. Kwak (2017), "Why Are Single-Sex Schools Successful?" Working Paper 6535, CESifo.
- Hahn Y., C. L. Wang, and H. S. Yang (2018), "Does Greater School Autonomy Make a Difference? Evidence from a Randomized Natural Experiment in South Korea," *Journal of Public Economics*, 161, 15–30. DOI: 10.1016/j.jpubeco.2018.03.004.
- Han J., and K. Ryu (2017), "Effects of Class Size Reduction in Upper Grades: Evidence from Seoul, Korea," *Economics of Education Review*, 60, 68–85. DOI:

- 10.1016/j.econedurev.2017.07.004.
- Hanushek E. A. (2006), "School Resources," In Hanushek E, Welch F (eds.) *Handbook of the Economics of Education*, Volume 2, Chapter 14, Elsevier, 865–908. DOI: 10.1016/S1574-0692(06)02014-9.
- Hotz V. J., G. W. Imbens, and J. H. Mortimer (2005), "Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations," *Journal of Econometrics*, 125, 241–270. DOI: 10.1016/j.jeconom.2004.04.009.
- Hoxby C. (2000a), "Peer Effects in the Classroom: Learning from Gender and Race Variation," Working Paper 7867, National Bureau of Economic Research. DOI: 10.3386/w7867.
- Hoxby C. M. (2000b). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 115, 1239–1285. DOI: 10.1162/003355300555060.
- Jackson C. K. (2012), "Single-Sex Schools, Student Achievement, and Course Selection: Evidence from Rule-Based Student Assignments in Trinidad and Tobago," *Journal of Public Economics*, 96, 173–187. DOI: 10.1016/j.jpubeco.2011.09.002.
- _____ (2016), "The Effect of Single-Sex Education on Test Scores, School Completion, Arrests, and Teen Motherhood: Evidence from School Transitions," Working Paper 22222, National Bureau of Economic Research. DOI: 10.3386/w22222.
- Jepsen C. (2015), "Class Size: Does It Matter for Student Achievement?" *IZA World of Labor*. DOI: 10.15185/izawol.190.
- Krueger A. B. (1999), "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 114, 497–532. DOI: 10.1162/003355399556052.
- Lavy V., and A. Schlosser (2011), "Mechanisms and Impacts of Gender Peer Effects at School," *American Economic Journal: Applied Economics*, 3, 1–33. DOI: 10.1257/app.3.2.1.
- Lee S., L. J. Turner, S. Woo, and K. Kim (2014), "All or Nothing? The Impact of School and Classroom Gender Composition on Effort and Academic Achievement," Working Paper 20722, National Bureau of Economic Research. DOI: 10.3386/w20722.
- Meghir C., and S. Rivkin (2011), "Econometric Methods for Research in Education," In E. A. Hanushek, S. Machin, and L. Woessmann (eds.) *Handbook of the Economics of Education*, Volume 3, Chapter 1. Elsevier, 1–87. DOI: 10.1016/B978-0-444-53429-3.00001-6.
- Oster E. forthcoming, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics*. DOI: 10.1080/07350015.2016.1227711.
- Park H., J. R. Behrman, and J. Choi (2012), Single-Sex Education: Positive Effects," *Science*, 335, 165–166. DOI: 10.1126/science.335.6065.165-b.
- _____ (2013), "Causal Effects of Single-Sex Schools on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools," *Demography*, 50, 447–469. DOI: 10.1007/s13524-012-0157-1.
- _____ (2018), "Do Single-Sex Schools Enhance Students' STEM (Science, Technology, Engineering, and Mathematics) Outcomes?" *Economics of Education Review*, 62, 35–47. DOI: 10.1016/j.econedurev.2017.10.007.

- Rivkin S. G., E. A. Hanushek, and J. F. Kain (2005), "Teachers, Schools, and Academic Achievement," *Econometrica*, 73, 417–458. DOI: 10.1111/j.1468-0262.2005.00584.x.
- Sohn H. (2016), "Mean and Distributional Impact of Single-Sex High Schools on Students' Cognitive Achievement, Major Choice, and Test-Taking Behavior: Evidence from a Random Assignment Policy in Seoul, Korea," *Economics of Education Review*, 52, 155–175. DOI: 10.1016/j.econedurev.2016.02.007.
- Todd P. E., and K. I. Wolpin (2003), "On the Specification and Estimation of the Production Function for Cognitive Achievement," *The Economic Journal*, 113, F3–F33. DOI: 10.1111/1468-0297.00097.
- Whitmore D. (2005), "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment," *American Economic Review: Papers and Proceedings*, 95, 199–203. DOI: 10.1257/000282805774670158.
- Wößmann L., and M. West (2006), "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS," *European Economic Review*, 50, 695–736. DOI: 10.1016/j.eurocorev.2004.11.005c.