ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 21-367

Kumon In: The Recent, Rapid Rise of Private Tutoring Centers

Edward Kim
Harvard University

Joshua Goodman
Boston University

Martin R. West
Harvard University

The growing phenomenon of private tutoring has received minimal scholarly attention in the United States. We use 20 years of geocoded data on the universe of U.S. private tutoring centers to estimate the size and growth of this industry and to identify predictors of tutoring center locations. We document four important facts. First, from 1997-2016, the number of private tutoring centers grew steadily and rapidly, more than tripling from about 3,000 to nearly 10,000. Second, the number and growth of private tutoring centers is heavily concentrated in geographic areas with high income and parental education. Nearly half of tutoring centers are in areas in the top quintile of income. Third, even conditional on income and parental education, private tutoring centers tend to locate in areas with many immigrant and Asian-American families, suggesting important differences by nationality and ethnicity in demand for such services. Fourth, we see little evidence that prevalence of private tutoring centers is related to the structure of K-12 school markets, including the prevalence of private schools and charter or magnet school options. The rapid rise in high-income families' demand for this form of private educational investment mimics phenomena observed in other spheres of education and family life, with potentially important implications for inequality in student outcomes.

VERSION: March 2021

Suggested citation: Kim, Edward, Joshua Goodman, and Martin R. West. (2021). Kumon In: The Recent, Rapid Rise of Private Tutoring Centers. (EdWorkingPaper: 21-367). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/z79x-mr65

Kumon In: The Recent, Rapid Rise of Private Tutoring Centers

Edward Kim*
Harvard University

Joshua Goodman
Boston University

Martin R. West
Harvard University

March 2021

Abstract

The growing phenomenon of private tutoring has received minimal scholarly attention in the United States. We use 20 years of geocoded data on the universe of U.S. private tutoring centers to estimate the size and growth of this industry and to identify predictors of tutoring center locations. We document four important facts. First, from 1997-2016, the number of private tutoring centers grew steadily and rapidly, more than tripling from about 3,000 to nearly 10,000. Second, the number and growth of private tutoring centers is heavily concentrated in geographic areas with high income and parental education. Nearly half of tutoring centers are in areas in the top quintile of income. Third, even conditional on income and parental education, private tutoring centers tend to locate in areas with many immigrant and Asian-American families, suggesting important differences by nationality and ethnicity in demand for such services. Fourth, we see little evidence that prevalence of private tutoring centers is related to the structure of K-12 school markets, including the prevalence of private schools and charter or magnet school options. The rapid rise in high-income families' demand for this form of private educational investment mimics phenomena observed in other spheres of education and family life, with potentially important implications for inequality in student outcomes.

* Corresponding author: Harvard Graduate School of Education, 13 Appian Way, Cambridge, MA 02138 (edwardjkim@g.harvard.edu).

1. Introduction

Private tutoring has grown increasingly popular across the world in recent decades (Bray & Lykins, 2012). National surveys suggest the United States is no exception to this pattern. In 1992, 10 percent of high school seniors reported taking a private class to prepare for the SAT.[1] That number quadrupled by 2012, with 40 percent reportedly taking a course to prepare for a college admissions exam.[2] In 2016, approximately 6-7% of US families with children between ages 6 and 17 had paid for tutoring in the past year, paying an average of about $230 in months with such a purchase. The top 10% and 5% of reported monthly spending was closer to $500 and $750, respectively.[3] Despite such high levels of and marked growth in student participation, the tutoring industry in the U.S. has received minimal scholarly attention (Bray, 2010).

We focus on the private tutoring industry, defined as "tutoring in an academic school subject, which is taught in addition to mainstream schooling for financial gain" (Bray and Silova, 2006). As a supplementary resource, private tutoring occupies a different role in the education marketplace than private schools, which are full substitutes for mainstream schooling. Families can combine private tutoring with any schooling arrangement, whether private or public. Though informal arrangements such as hiring a neighborhood teenager qualify as private tutoring, research suggests the growing popularity of private tutoring is due to the rise of larger scale private tutoring firms (Aurini, 2004). Such firms run physical locations outside of schools called tutoring centers, which students attend to receive services, often in small groups. Lessons can focus on mainstream

[1] Authors' calculations based on the 1992 wave of the U.S. Department of Education's National Education Longitudinal Study.
[2] Authors' calculations based on the 2012 wave of the U.S. Department of Education's High School Longitudinal Study.
[3] Authors' calculations based on the 2016 Consumer Expenditure Survey, which asks about spending habits in each month.

school curricula or standardized exam content, and tutors differ widely in age and qualifications (Bray & Silova, 2006). Some private tutoring firms run single locations while others are large national chains, such as Kumon, Sylvan, and Huntington Learning Centers.

In this paper, we use two decades of geocoded data on the universe of U.S. private tutoring centers to measure the size and growth of this industry, as well as to explore the economic and demographic predictors of tutoring center locations. We document four important facts. First, from 1997-2016, the number of private tutoring centers in the U.S. grew steadily and rapidly, more than tripling from about 3,000 to nearly 10,000. Second, the number and growth of private tutoring centers is heavily concentrated in geographic areas with high incomes and high levels of parental education. As of 2016, 44 percent of tutoring centers were in areas representing the top fifth of the income distribution, and 55 percent of tutoring centers newly opened between 2000 and 2016 opened in such areas. Third, even conditional on income and educational attainment, private tutoring centers tend to locate in areas with many immigrant and Asian-American families, suggesting important differences by nationality and ethnicity in demand for their services. Fourth, we see little evidence that the prevalence of private tutoring centers is related to the structure of K-12 school markets, including the prevalence of private schools and charter or magnet school options.

Our work contributes to three strands of the research literature. First, empirical research on private tutoring enrollment in the United States is limited, with notable exceptions documenting its prevalence among Asian-American immigrants and communities (Byun & Park, 2012; Lee & Zhou, 2014; Zhou & Kim, 2006). Studies of its effects on participating students have focused on the approximately 2,000 firms that were approved to provide supplemental educational services under the federal No Child Left Behind Act; a meta-analysis of twenty-eight evaluations of such

providers documented an overall small positive effect on state test scores with considerable heterogeneity (Ascher, 2006; Chappell et al., 2011). However, this NCLB policy targeted low-income students in underperforming schools, a very different population than those served by the typical private tutoring center. While the effectiveness of one-on-one or very small group tutoring is well documented (Cohen, Tulik, & Tulik, 1982; Wasik & Slavin, 1993; Fryer, 2014; Kraft, 2014), we have little evidence on the effectiveness of the specific forms of tutoring offered by these centers. We also know little about which students and families have demand for these services, something our descriptive evidence attempts to remedy.

Second, our work documents a rapid rise in high-income families' demand for private education investments that mimics phenomena observed in other spheres of family life. High-income families are increasing their investments in early childcare (Ramey and Ramey, 2010), parental time spent with children (Guryan, Hearst and Kearny, 2008), and extracurricular activities (Levey Friedman, 2013), while viewing kindergarten increasingly as a time for academic focus rather than play and socializing (Bassok, Latham and Rorem, 2016). Such families are also demanding more intensive and competitive secondary education, pushing for dual enrollment programs, Advanced Placement classes, and International Baccalaureate programs (Davies & Hammack, 2005). This increased competition and pressure among high-ability students leads them to search for ways to maximize their chances for success, including the use of private tutoring (Bound, Hershbein, & Long, 2009), which does seem to moderately increase SAT scores and rates of selective college enrollment (Buchmann, Condron, and Roscigno, 2010).

Third, our findings help place the U.S. private tutoring industry into a broader international context. Some characteristics, such as a tight linkage with consequential exam systems, are common to private tutoring industries across the world, but each country's context can shape the

industry significantly (Bray & Silova, 2006). In South Korea, although upper-income families exhibit the greatest demand for private tutoring, the practice is widespread (Kim & Park, 2010). Household expenditures on private tutoring rival government spending on primary and secondary schooling, perhaps because homogenization of secondary school quality and a hierarchical higher education system drive students to use private tutoring to distinguish themselves for college admission processes (Kim & Lee, 2001; Kim & Lee, 2010). Research also suggests that demand for private tutoring in South Korea is greater in areas with lower local school quality and fewer school choice options (Kim, 2004; Kim & Lee, 2010).

In Canada, private tutoring serves primarily as a financial middle ground for families who are dissatisfied with their public schooling options but cannot afford the tuition of private schools (Davies, 2004). With neither university entrance exams nor a strict hierarchy of university prestige, the Canadian private tutoring industry advertises itself in response to perceived shortcomings of public schools, emphasizing small class sizes, personalized curricula, and individual attention (Aurini, 2004; Aurini & Davies, 2004). Middle-income families appear to be the target market for private tutoring, as upper-income families dissatisfied with public school options can afford private schools, while for lower-income families private tutoring may be unaffordable. As in South Korea, demand is tied to desire among families dissatisfied with mainstream schooling to provide additional educational resources to their children. However, the type of family associated with that dissatisfaction differs between the two countries.

Our results suggest that private tutoring centers in the U.S. are closer to the South Korean model in targeting high-income families, though they are not (yet) nearly as widespread as in that country. Like South Korea, the U.S. has a hierarchical higher education system with intense competition for admission to the most elite institutions. The prevalence and growth in tutoring

centers we document may be related to the perception by U.S. parents of the high stakes associated with their children's educational achievement given this postsecondary landscape.

2. Data

We combine multiple data sets to conduct our investigation. Measures of private tutoring prevalence come from Infogroup's Historical Business Data. Information on a business's identification and location are sourced via yellow page directories. Each business included in the data represents a different physical location, with identifiers that allow the business to be tracked across years. We observe businesses' names, addresses, and industry codes, as well as some measures of size. We identify businesses registered as either "Tutoring" (SIC Code 829909) or "Test Preparation Instruction" (SIC Code 874868), ultimately finding about 20,000 unique firms with almost 35,000 locations. "Tutoring" firms are 40 times more numerous than "Test Preparation Services" firms. Some franchises have branches in both categories, however, so we combine them for our primary outcome measure. We then combine the tutoring center location data with school district boundary files to locate tutoring centers within school districts. We successfully match 98.8 percent of our business observations to a school district.

We also use two levels of data from the Common Core of Data (CCD) for the years 1997 to 2016: school-level files for public schools and school-district-level files for public school districts. We use the school-level data to calculate, for each school district, the proportion of students enrolled in charter schools and a segregation metric (dissimilarity index) calculated from the share of students eligible for free or reduced-price lunch (FRPL). At the district level we use the ratio of students to various staff (e.g., student-to-teacher, student-to-administrator), fiscal data, and urbanicity designation (i.e., rural, town, suburb, or urban).

We also merge demographic information at the school-district level from the 2000 U.S. census and American Community Survey (ACS) five-year data sets for elementary, secondary, and unified school districts. We use the earliest and latest ACS data sets accessible at the time of analysis (2004-2009 and 2011-2016) and in this paper will refer to each ACS data set by the final year of each five-year interval. Variables we draw from the census and the ACS include per capita income, proportion of adults with at least a bachelor's degree, proportion of individuals foreign born, and number of children enrolled in private or public school. We substitute values from the 2004-2009 ACS for mobility and fertility variables unreported in the 2000 census at the school-district level. We define our main outcome variable, tutoring centers per 1,000 students, based on student totals from the census and ACS, rather than the CCD, as the former capture student enrollment in both public and private school. After aggregating all data sources, our final analytic data set includes approximately 13,000 unique school districts.

3. Methods

We perform three sets of analyses, all of which use as a primary outcome the number of tutoring centers per 1,000 children enrolled in public or private schools in a given school district and year. First, we simply document the rise in tutoring center prevalence over time and by location, with the goal of understanding the magnitude and geographic spread of this phenomenon. We believe this is the first analysis to measure how many tutoring centers exist, where they locate, and how this has changed over time.

Second, we ask which school-district-level covariates predict the prevalence of tutoring centers in 2016, the most recent available year of data. To do this, we run regressions of the form:

$$\text{Tutoring}_{\text{d,2016}} = \beta_0 + \beta'_{\text{char}} \times \text{Characteristics}_{\text{d,2016}} + \epsilon_{\text{d}} \quad (1)$$

where the outcome is the number of tutoring centers per 1,000 school-age children in 2016 in school district *d* and *Characteristics* represents a vector of district-level demographic and economic characteristics. The coefficients β'_{char} estimate the association between such characteristics and tutoring center prevalence, helping us demonstrate which contemporaneous variables most strongly predict tutoring center location.[4]

Third, we ask which school-district-level covariates predict the change over time in prevalence of tutoring centers between 2000 and 2016, the period over which we can observe all potential predictors in our data. To do so, we run regressions of the form:

$$\text{Tutoring}_{\text{d,2016}} = \beta_0 + \beta'_{\text{char}} \times \text{Characteristics}_{\text{d,2000}} + \alpha \times \text{Tutoring}_{\text{d,2000}} + \epsilon_{\text{d}} \quad (2)$$

which differs from Equation 1 by controlling for the number of tutoring centers in the year 2000 and by using school district characteristics as of 2000. The coefficients β'_{char} estimate the association between baseline school district characteristics and growth in tutoring center prevalence between 2000 and 2016.

We choose the set of school district characteristics to use as predictors based on the prior literature and a machine learning approach. Existing research suggests a number of potentially important correlates of tutoring center prevalence and growth, including: income (Lee, 2005; Tansel & Bircan, 2006), income inequality (Dang & Rogers, 2008; Atalmis, Yilmaz, & Saatcioglu 2016), racial demography (Byun & Park, 2012; Bray & Lykins, 2012; Shin, 2012), immigration

[4] The results generated by using data from the most recent year (2016) are quite similar to those using earlier years of data, so we omit the latter for simplicity.

status (Sriprakash, Proctor, & Hu, 2016), school quality (Kim, 2004), and availability of school choice (Kim & Lee, 2001; Kim & Lee 2010). We generate various measures of these constructs, resulting in over 50 potential covariates.

Given the large number of potential covariates and our desire not to select among them based on our own priors, we use the LASSO (Least Absolute Shrinkage and Selection Operator) variable selection method. LASSO uses within-sample cross-validation to identify a set of predictors that explains the most variation in the outcome subject to a penalty for overfitting (Tibshirani, 1996). From a pool of more than 50 candidate covariates, we identified 5 that performed consistently well across model specifications. We describe the results of this procedure in more detail below.

4. Results

4.1 Tutoring Center Growth and Geographic Spread

The number of private tutoring centers in the U.S. roughly tripled between 1997 and 2016, as seen in Figure 1. In 1997, there were just over 3,000 private tutoring centers in the United States. That number increased steadily and roughly linearly over time, averaging 6.2% growth each year, so that by 2016 there were over 9,000 such firms.

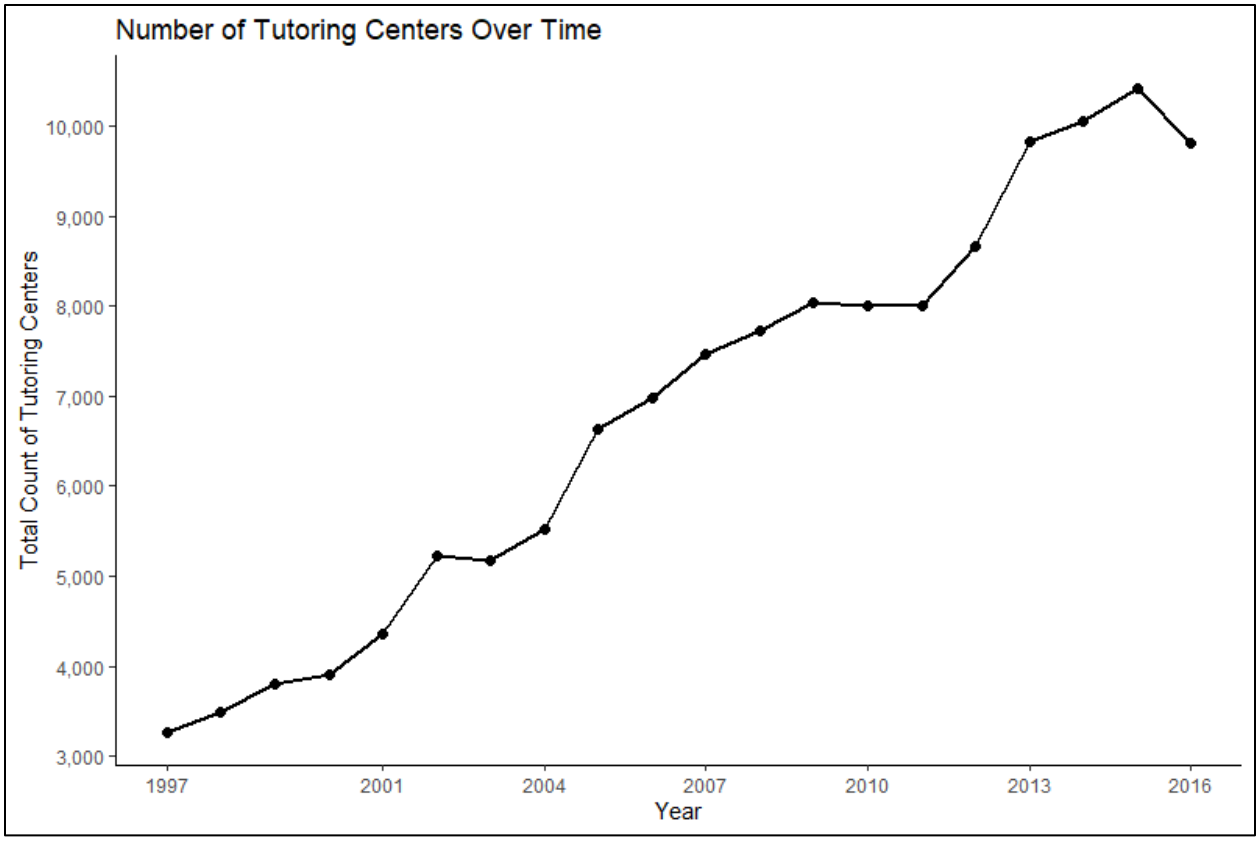

Figure 1: Private Tutoring Industry Growth Over Time.

Tutoring centers are more prevalent on the East and West coasts of the U.S., but growth over time has occurred throughout the country. Figure 2 shows the number of tutoring centers per school-age child by county in both 2000 (the first year for which we have student counts by county; panel A) and 2016 (panel B). Though most counties still have no tutoring centers, the industry has both expanded to new areas of the country and become more densely concentrated. Table 1 shows that the percent of counties without any tutoring centers decreased from 77.5% to 72.2% from 2000 to 2016. The share of counties with ratios less than 1:10,000 also decreased during this period, from 13.0% to 7.5% Meanwhile, the share of counties with ratios of at least 1:5,000 nearly quadrupled from 2.7% to 9.3%.

This pattern of greater prevalence and greater density is evident for all four major regions of the United States but is most pronounced for the Northeast: between 2000 and 2016 the percent

of counties in that region without any tutoring centers dropped by 7.3 percentage points, while the percent of counties with a ratio greater than 1:5,000 increased nearly tenfold from 2.7% to 22.3%.

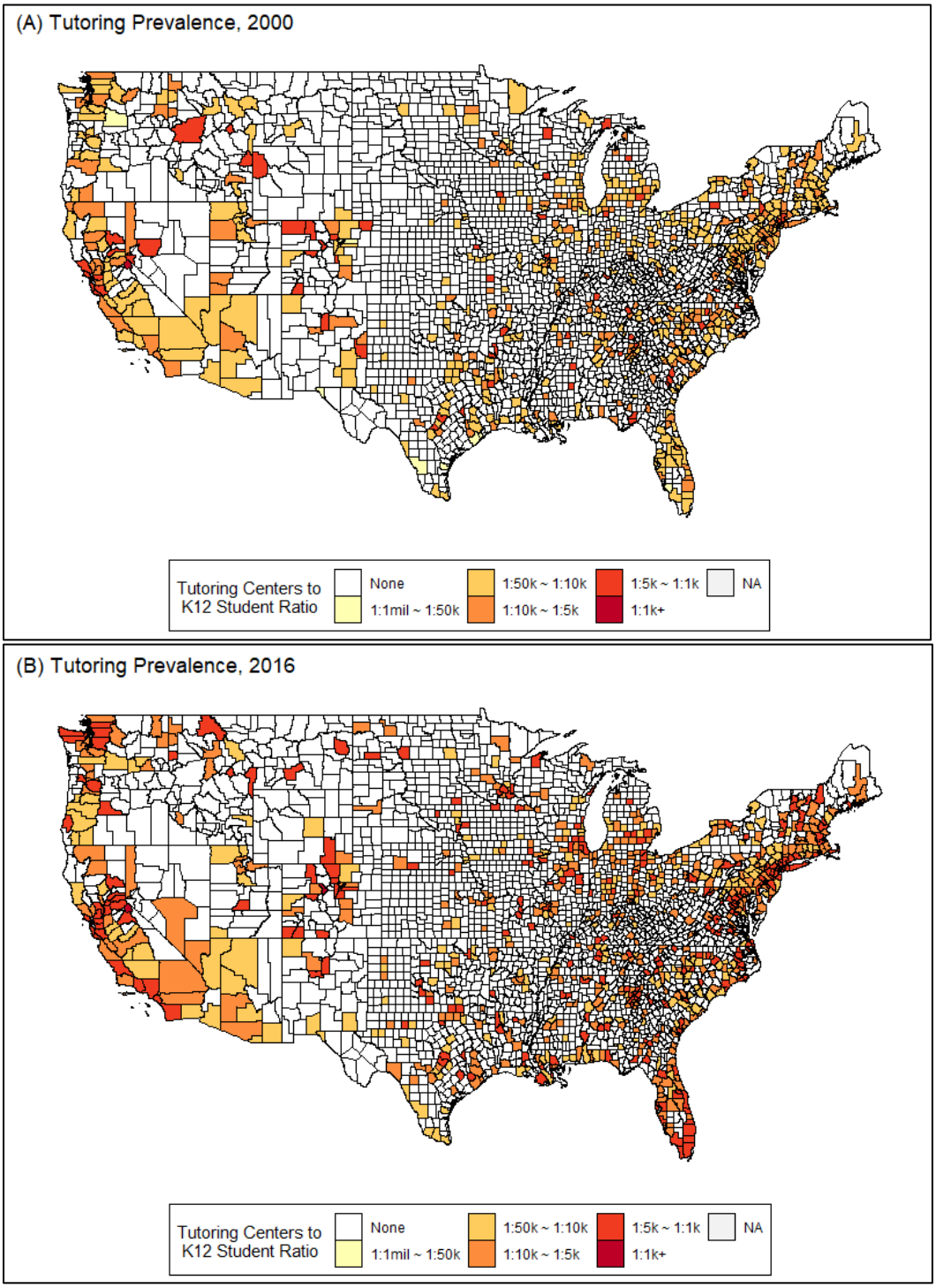

Figure 2: Tutoring center to K12 student ratio per county in 2000 and 2016.

Table 1: Geographic Location and Growth

2000 Percent 2016 Percent	None	1:1mil to 1:50k	1:50k to 1:10k	1:10k to 1:5k	1:5k to 1:1k	1:1k or more
Overall	77.55% 72.17%	0.45% 0.00%	12.55% 7.46%	6.72% 11.06%	2.61% 9.16%	0.13% 0.16%
Northeast	47.27% 40.00%	0.91% 0.00%	33.64% 16.36%	15.45% 21.36%	2.73% 22.27%	0.00% 0.00%
South	78.96% 72.09%	0.42% 0.00%	10.98% 7.61%	6.83% 10.92%	2.60% 9.16%	0.21% 0.21%
Midwest	85.33% 80.37%	0.31% 0.00%	9.61% 5.68%	3.51% 9.19%	1.24% 4.75%	0.00% 0.00%
West	69.37% 67.34%	0.45% 0.00%	14.86% 8.72%	9.01% 10.96%	6.08% 12.53%	0.23% 0.45%

4.2 Choosing Predictors

We employed a LASSO procedure to guide our investigation of associations between school-district characteristics and private tutoring. This procedure identifies the optimal combination of predictors from a set of variables by balancing predictive accuracy against model parsimony. We began with over 50 variables describing wealth, education, age, race and ethnicity, immigration, mobility, occupation, family structure, school and district staff, and district funding and expenses. We also included private tutoring prevalence in 2000 for the predicted-change model. The full list of candidate variables is given in Appendix A.

The variables we identified from this procedure were: (1) proportion of students who identify as Asian, (2) proportion bachelor's degree holders, (3) income per capita, (4) proportion foreign born, and (5) urbanicity.

The LASSO procedure suggested similar covariates between the cross-sectional and predicted-change models, which informed our variable selection decisions. Most prominent at the highest level of model parsimony in both the cross-sectional and predicted-change model were proportion of students who identify as Asian and proportion bachelor's degree holders. Income

per capita was also included in the highest level of parsimony for the cross-sectional model and was nearly as prominent in the predicted-change model. Both proportion foreign born and urbanicity followed closely after income per capita.

Though the LASSO results identified other variables of potential interest (e.g., income segregation and income inequality), we restrict our attention to these most prominent variables and leave the rest for future investigation. We also omit some variables to avoid repeating domains (e.g., proportion with bachelor's degree, proportion with high school degree, and proportion with graduate degree). Full details of which variables the LASSO procedure identified at each level of model parsimony can be found in Appendix B.

Though theory and prior research suggest a relationship between school choice and private tutoring, the proportion of children enrolled in private school was not identified as a salient predictor of tutoring center prevalence or growth; nor was the prevalence of charter schools or magnet schools. We nonetheless include a supplementary investigation of private school enrollment at the end of the next subsection in order to relate our findings to this prior literature.

4.3 Predictors of Tutoring Center Prevalence and Growth

We now construct standard OLS models based on the variables identified by the LASSO procedure. Table 2 presents our main findings on the correlates of tutoring center prevalence and growth. In the cross-sectional model (i.e., Column 2), all covariates are statistically significant apart from the indicator for "town" (relative to the reference group "rural"). In 2016, a thousand-dollar difference in per capita income is associated with a 0.00274 difference in tutoring centers per 1,000 students. The 25^{th} and 75^{th} percentile of per capita income in our sample differ by about

$10,000; using the average student count of about 5,000, a 25th- and a 75th-percentile income school district therefore differ by 0.137 tutoring centers on average. (For reference, the average tutoring centers per 1000 in 2016 for our analytic sample was 0.0737.) We can also compare coefficient magnitudes in Column 2. For example, a 1 percentage point difference in either bachelor's degree holders or foreign born is roughly equivalent to a $1000 difference in per capita income, while a 1 percentage point difference in Asian student body is equivalent to more than $2000. All else equal, a rural district would need about $15,000 higher per capita income than a suburban or urban district to have about equal expectation in the outcome.

The predicted-change model (i.e., Column 4) is similar, but with important differences. The coefficient for income is nearly twice as large in the predicted-change model, with every thousand-dollar difference in baseline per capita income predicting school districts added 0.005 more tutoring centers per 1,000 students between 2000 and 2016 on average. The effect of a percentage point difference in Asian student body is now roughly on par with a $1000 difference in per capita income, whereas a percentage point in bachelor's degree holders or foreign born translates to about half as much. Further, only in suburban school districts did tutoring prevalence grow significantly (relative to rural school districts). That is, even though urban school districts had more tutoring centers per student than suburban districts in 2016, the industry grew substantially more in suburban districts over the period of observation.

Table 2: Predictors of Tutoring Center Location and Growth

	Cross Sectional:		Change Predict:	
	(1)	(2)	(3)	(4)
	Bivariate	Fully Controlled	Bivariate	Fully Controlled
Income Per Capita	0.00861***	0.00274***	0.0117***	0.00533***
(Thousands of Dollars)	(0.000304)	(0.000573)	(0.000435)	(0.000838)
Proportion At Least Bachelor's	0.737***	0.351***	0.763***	0.319***
	(0.0234)	(0.0470)	(0.0280)	(0.0539)
Proportion Asian	1.570***	0.630***	1.682***	0.682***
	(0.0591)	(0.0752)	(0.0851)	(0.104)
Proportion Foreign Born	0.810***	0.283***	0.672***	0.218***
	(0.0402)	(0.0492)	(0.0453)	(0.0547)
Town	0.0170	0.0138	0.00716	0.0103
	(0.00900)	(0.00888)	(0.00961)	(0.00957)
Suburb	0.170***	0.0365***	0.136***	0.0375***
	(0.00823)	(0.00956)	(0.00816)	(0.00912)
Urban	0.149***	0.0421**	0.0779***	0.00810
	(0.0146)	(0.0155)	(0.0156)	(0.0161)
Tutoring Centers per 1000 in 2000			Yes	0.399***
				(0.0240)
Observations	—	12931	—	12227

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The binned scatterplots in Figures 3 and 4 reassure us that though the model covariates (i.e., income per capita, proportion bachelor's, proportion Asian, and proportion foreign born) have right-skewed distributions, the model results are not simply the result of high leverage units. Across the spectrum, greater covariate values suggested greater outcome values, though with various levels of concavity. The plots for proportion foreign-born, in particular, suggest a potential non-linear relationship. As we unpack in the following subsections, however, this is not necessarily the case.

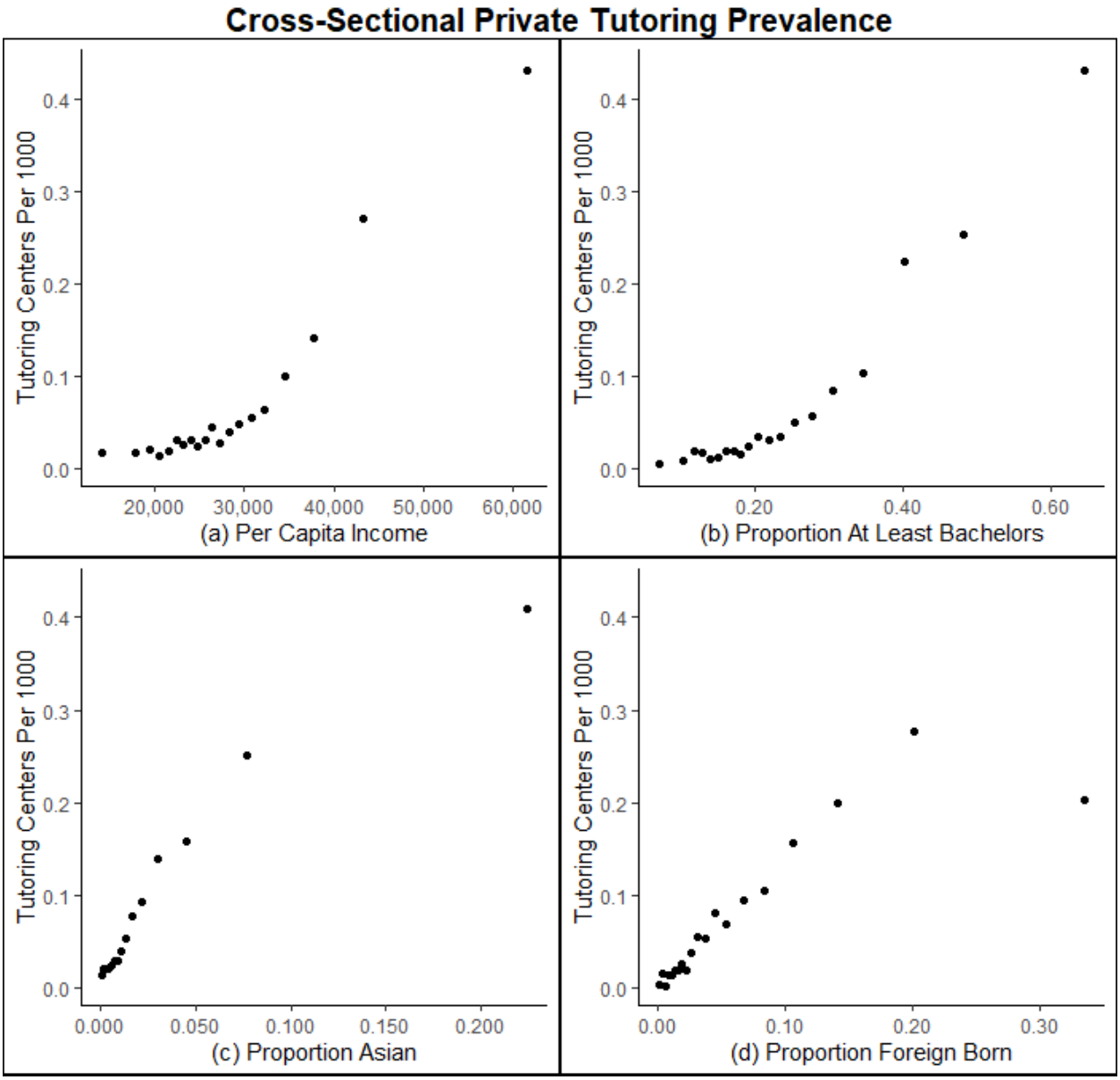

Figure 3: Each dot represents a vigintile (one twentieth) of the distribution of the demographic variable shown. In panel C, due to the uneven distribution of the covariate, the leftmost dot contains more than five percent of observations.

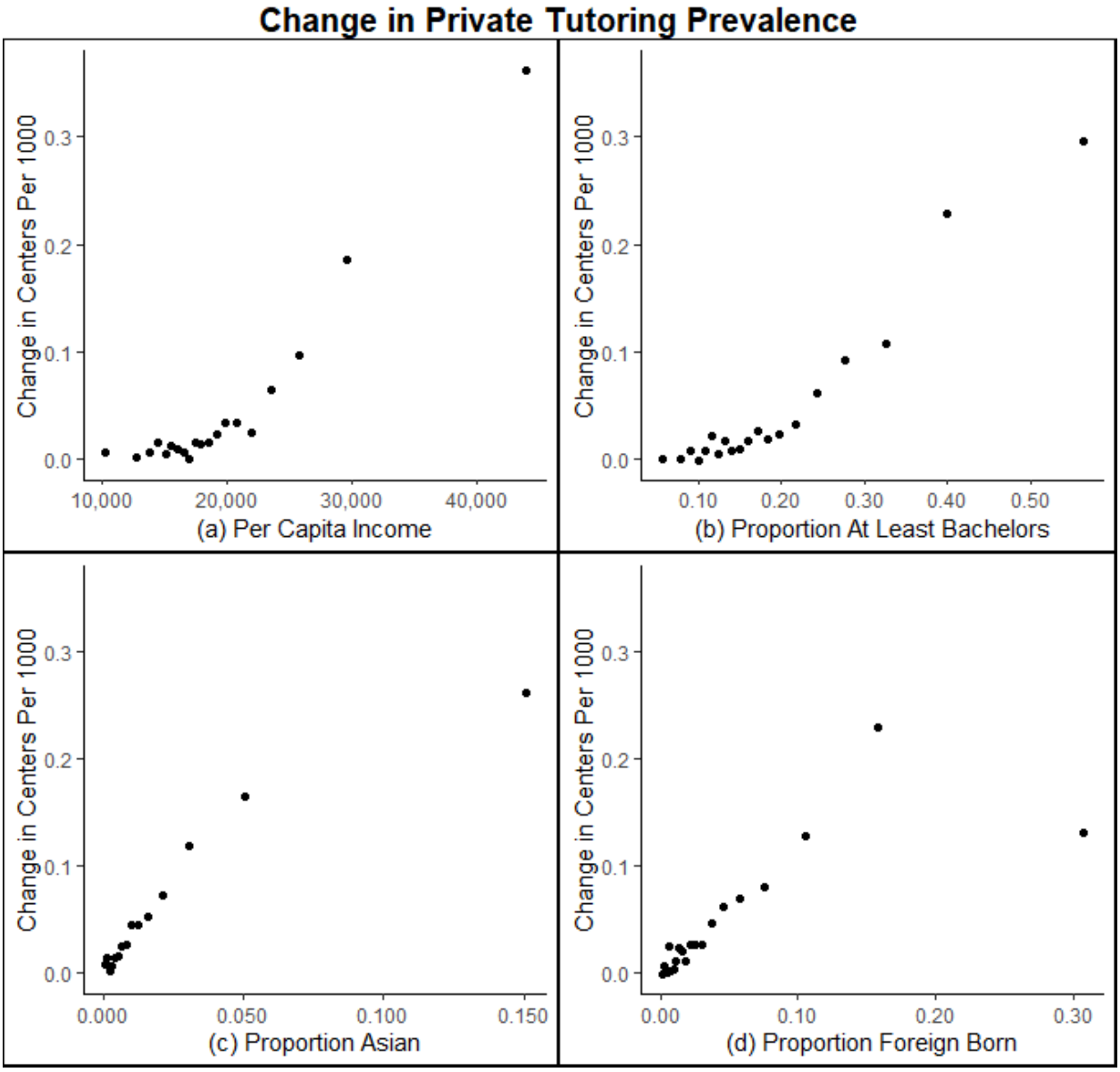

Figure 4: Each dot represents a vigintile (one twentieth) of the distribution of the demographic variable shown. In panel C, due to the uneven distribution of the covariate, the leftmost dot contains more than five percent of observations.

Income

Income and wealth variables were drawn directly from the census and ACS compiled at the school-district level. We use one such variable, per capita income, in our multivariate models based on its performance in the LASSO procedure. We expected some association between income levels and private tutoring since enrollment requires disposable income. However, as demonstrated by the comparison of South Korea and Canada, it is less clear which families among those who can afford private tutoring show the most interest.

Table 2 shows an overall positive association, with and without other covariates, between income per capita and private tutoring prevalence. Controlling for other covariates, a school district with one thousand dollars higher per capita income in 2016 on average has 0.00274 more tutoring centers per 1000 students in 2016, or 3.7% of the average tutoring prevalence in 2016. The same wealth difference in 2000 suggests a district can expect 0.00533 more tutoring centers per 1000 students by 2016, or 20% of the average tutoring prevalence in 2000. Figure 3a illustrates that the cross-sectional relationship is monotonic and convex, with the highest income school districts demonstrating increasingly greater private tutoring prevalence. Figure 5 replicates Figure 1 separately for three groups of per capita income percentiles, each capturing a third of the observations. A disproportionate amount of the overall growth took place in the highest income brackets, with the lower income group showing minimal growth.

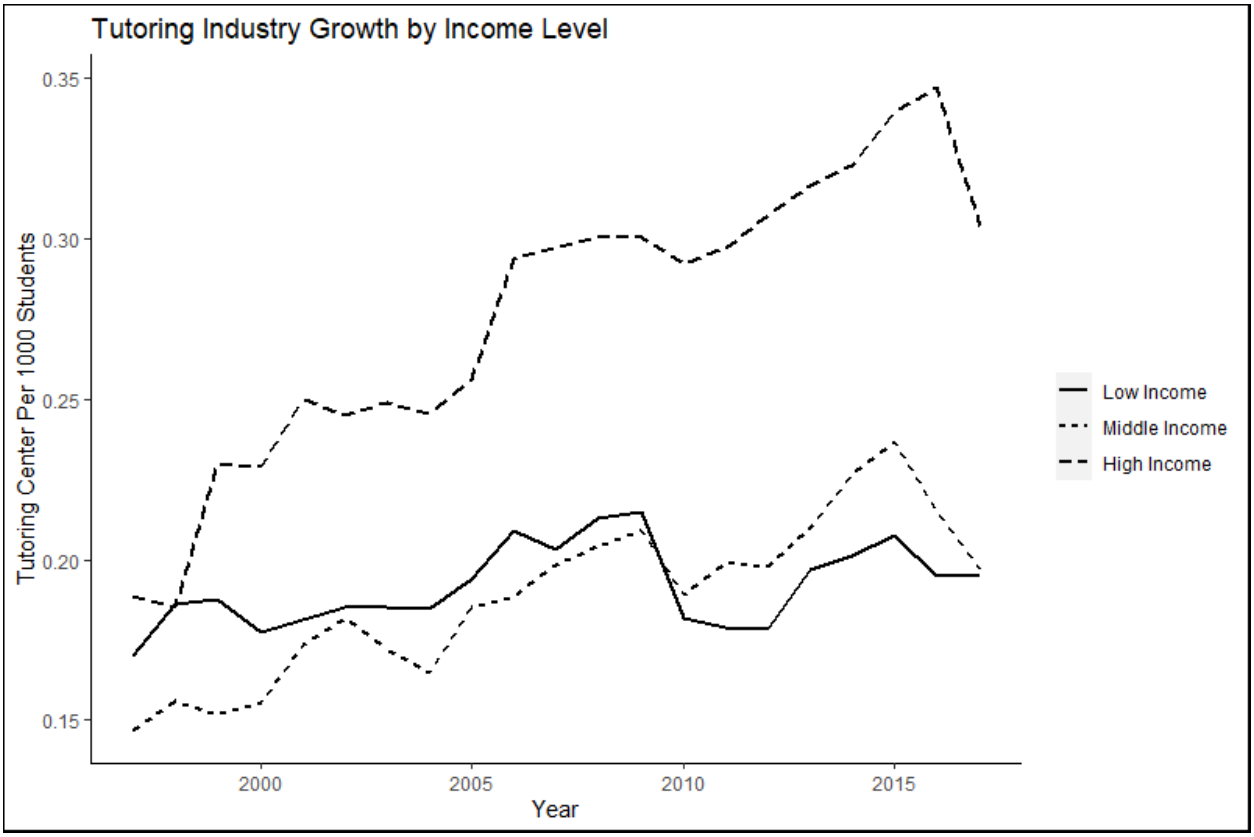

Figure 5: Private Tutoring Industry Growth Over Time by Per Capita Income Group

To the extent that private tutoring demand is driven by a dissatisfaction with the educational resources offered in mainstream schooling, the large degree of segregation by income across U.S. school systems (Bischoff & Reardon, 2014) would suggest that high-income families would be the least interested in private tutoring, as their children arguably already receive the best educational experience (e.g., well-funded schools through local tax revenue, access to high social capital networks, etc.). Our descriptive results suggest this is not the case. Private tutoring is disproportionately concentrated in higher income neighborhoods, and higher income families seem the most interested in private tutoring.

Bound, Hershbein, and Long (2019) offer a potential explanation: the highest performing and most advantaged students perceive the most pressure to succeed due to increasingly intense

competition, and, as one response, turn towards private tutoring to supplement their educational resources. To maximize educational resources, high-income families would enroll in private tutoring in addition to high-quality mainstream schooling, which offers a similar experience to one's peers and therefore one's immediate competition. Consistent with this theory, school segregation by income (as measured by within-district dissimilarity index by FRPL status) appeared in the LASSO procedure with a positive association. That is, controlling for other covariates, districts with schools that are more highly segregated by income tend to have higher private tutoring prevalence. Future investigations could empirically confirm the direct relationship between interest in private tutoring and relative economic standing among immediate peers.

Educational Attainment

The patterns in Figures 3b and 4b, depicting the outcome across proportion of population over 25 with at least a bachelor's degree, closely resemble those for per capita income: The cross-sectional figure shows a clear, upward sloping trend; the predicted-change figure is less consistent for lower values of the predictor but suggests the same. But the relationship between tutoring prevalence and educational attainment is not merely a proxy for an association with income. Proportion with at least a bachelor's degree remained significant after controlling for income per capita, and the LASSO procedure arguably identified educational attainment, not wealth, as the most important predictor of private tutoring prevalence in the United States (see Appendix B).

The relationship between private tutoring enrollment and parent educational attainment levels appears consistently positive across research settings (Tansel & Bircan, 2006; Nath, 2008; Kim & Park, 2010; Zhang & Xie, 2016). Part of this observed effect is likely due to the high

coincidence between educational attainment and income, the latter being a prerequisite for private tutoring enrollment. But parent educational attainment may additionally reflect higher parental expectations of children (Bray & Kwok, 2003), and children reared in communities of well-educated adults may pursue similar outcomes via transmission of cultural capital (Lareau, 2001).

Asian and Foreign Born

Previous studies have looked at the relationship between private tutoring and racial/ethnic groups in the United States, particularly for Asian Americans (Shrake, 2010), and some have even suggested private tutoring as an explanation for Asian American communities' exceptional academic performance (Byun & Park, 2012; Zhou & Kim, 2006). We build on these observations by considering demographic composition with respect to proportion Asian and proportion foreign born, both of which were identified as relevant in the LASSO procedure. Note the proportion Asian variable was calculated with respect to the school district's student population, but proportion foreign born was calculated with respect to the entire population within the school district's geographic boundaries.

Proportion of Asian students was identified as an important predictor by the LASSO procedure at every level of parsimony and demonstrates a large coefficient in the multivariate regressions. A percentage point higher in Asian student composition in 2016 predicts 0.0063 more private tutoring centers per 1000, and the same difference in 2000 predicts a 0.00682 greater increase in the outcome between 2000 and 2016. Though the covariate's distribution is right skewed, Figures 3c and 4c show that proportion Asian consistently predicts greater tutoring prevalence across the distribution.

Proportion foreign born population was among the most important predictors according to the LASSO procedure, but it demonstrated relatively small coefficients in both models. In Figures 3c, 4c, 3d, and 4d we see that in contrast to income and educational attainment, proportion Asian and proportion foreign born show concave relationships. The highest data point for proportion foreign born in both figures appears lower than expected based on a linear projection, but closer examination revealed those units were particularly low income; controlling for income revealed a more monotonic pattern.

Based on relevant research documenting the behavior of Asian and foreign-born communities toward schooling in the U.S., we posit some portion of the observed relationship comes from a cultural familiarity with private tutoring in families' countries of origin. Research conducted in common countries of origin for Asian Americans reports substantial amounts of private tutoring, for example: China (Kwok, 2010), India (Bhorkar & Bray, 2018), Philippines (de Castro & de Guzman, 2014), Vietnam (Dang, 2007) and South Korea (Kim & Lee, 2002). Immigrant parents in the United States, particularly those of Asian origin, are also relatively optimistic and hold high expectations of their children with regard to educational opportunities (Duong, Badaly, & Liu, 2016, Kao & Tienda, 1995, Schneider & Lee, 1990; Goyette & Xie, 1999; Raleigh & Kao, 2010), a perspective which could encourage interest in supplemental educational resources. However, Sriprakash, Proctor, and Hu (2016), in their study of Chinese immigrants in Australia, warn against essentializing these communities' demand for private tutoring as a cultural phenomenon. They suggest that private tutoring enrollment can instead be understood as a "considered, strategic response" from families with disposable income, but less social and cultural capital, to education systems that appear to highly weigh exam results while minimally tailoring curricula to exam preparation. While our investigation cannot confirm this theory, the factors that

Sriprakash, Proctor, and Hu (2016) describe in the Australian context seem present in the U.S. context, too.

Private School Enrollment

The theoretical connection between private tutoring and school choice consists of multiple facets. Research on school choice suggests more options help families find schools that match their preferences, and competition between schools can increase school quality. Both these dynamics would theoretically reduce demand for private tutoring. Further, private tutoring markets overlap with mainstream schooling competition, insofar as private tutoring provides similar goods without offering a full substitute. Families can substitute a higher quality but more expensive mainstream schooling option with a cheaper mainstream schooling choice supplemented by private tutoring, or, given the similarity of goods, choose both the higher cost school and private tutoring.

We calculate proportion private enrollment as the number of children enrolled in private school, out of the total such enrollees at either private or public school, according to the census and ACS. In Table 3, we amend the original model to include private school enrollment and find the original results remain largely unchanged. We do find that private enrollment is positively associated with private tutoring when controlling for other covariates in the cross-sectional model, but the coefficient is relatively small (about as large as the effect for a $500 difference in per capita income). The coefficient on private enrollment is statistically insignificant in the predicted-change model.

Table 3: Predictors of Tutoring Center Location and Growth, Plus Proportion Private Enrollment

	Cross Sectional:		Change Predict:	
	(1)	(2)	(3)	(4)
	Bivariate	Fully Controlled	Bivariate	Fully Controlled
Proportion Private School	0.592***	0.134**	0.665***	0.0766
	(14.71)	(3.06)	(13.24)	(1.38)
Income Per Capita	0.00861***	0.00270***	0.0117***	0.00518***
(Thousands of Dollars)	(28.31)	(4.71)	(26.94)	(6.13)
Proportion At Least Bachelor's	0.737***	0.325***	0.763***	0.314***
	(31.47)	(6.79)	(27.21)	(5.82)
Proportion Asian	1.570***	0.628***	1.682***	0.680***
	(26.56)	(8.36)	(19.77)	(6.55)
Proportion Foreign Born	0.810***	0.295***	0.672***	0.222***
	(20.12)	(5.98)	(14.83)	(4.06)
Town	0.0170	0.0132	0.00716	0.00951
	(1.88)	(1.49)	(0.74)	(0.99)
Suburb	0.170***	0.0336***	0.136***	0.0348***
	(20.63)	(3.49)	(16.71)	(3.72)
Urban	0.149***	0.0394*	0.0779***	0.00477
	(10.20)	(2.54)	(4.98)	(0.29)
Tutoring Centers per 1000 in 2000			Yes	0.399***
				(16.62)
Observations	—	12931	—	12227

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Given that enrollment in private school generally requires greater investment than enrollment in public school, we might expect private-school families to be more secure and satisfied with their child's schooling. And, in fact, survey data indicate that parents of students attending private schools express greater satisfaction with their child's school than do public school parents (Barrows et al., 2019). Why, then, if private tutoring demand supposedly increases with mainstream schooling dissatisfaction, would private tutoring be more popular in areas with greater private school enrollment? A simple explanation, akin to our interpretation for per capita income, is that families who desire maximal educational resources would enroll their children in

both private school and private tutoring. The only barrier would be cost, though for families who can afford private school, private tutoring may not represent a significant burden. But the question remains whether under causal circumstances families would view these options, mainstream schooling choice on the one hand and supplementary schooling on the other, as substitutes or complements.

Conclusion

In this study we combine data on the private tutoring industry and school-district characteristics to describe patterns in the private tutoring industry in the U.S. Private tutoring universally offers families additional resources for their children, though which families enroll in this service varies based on the specific features of a given education system. Beyond tutoring's effectiveness as an educational practice, basic questions about the industry, such as who enrolls in private tutoring, are consequential for understanding its impact. On one hand, providers through NCLB were enlisted to remediate students who were underserved by their mainstream school. On the other, Ochoa's (2013) qualitative study of a California public high school found that private tutoring was so widespread among high-achieving, high-income students that some of the school's teachers adapted the advanced classes' curricula to reflect the supplemental education that so many of their students received. This adaptation made the classes less accessible to high-achieving, low-income students.

Our study is to our knowledge the first to offer a comprehensive analysis of the growth and prevalence of private tutoring in United States. According to our data, private tutoring in the U.S. has grown precipitously in the last two decades, more than tripling both the number of firms

between 2000 and 2016. We selected variables for our multivariate analyses based on a LASSO procedure applied to two types of models: a cross-sectional model using only 2016 data, and a predicted-change model using covariates in 2000 to predict 2016 outcomes. The LASSO results generally aligned with suggestions from relevant literature. Private tutoring exists disproportionately in the highest income and most educated areas, possibly driven by perceived competition among the highest performing students. Communities with a higher proportion Asian and foreign-born population also had greater rates of private tutoring. The availability of private school options, though having been found in some settings to have a negative relationship with demand for tutoring, demonstrated a small positive relationship with tutoring center prevalence and no association with change in tutoring center prevalence over time.

Our study has several limitations. Our primary outcome variable, number of registered private tutoring firms per 1,000 children in a school district enrolled in public or private schools, imperfectly captures firms aimed specifically at K-12 education, assumes a tight relationship between number of firms and demand for private tutoring, and cannot detect individual-level patterns. However, the signal was sufficiently strong to demonstrate clear relationships with our covariates at this aggregate level, and information on the supply side of private tutoring can be valuable in and of itself. Future investigations should endeavor to employ causal estimation strategies to uncover direct relationships between private tutoring and various facets of U.S. education, ideally with student-level data.

Private tutoring represents an increasingly relevant issue for education policy in the U.S. As a private industry it operates outside traditional regulations for educational institutions, but by offering a service that overlaps with mainstream schooling it may still affect students and learning outcomes. The appropriate policy response to a burgeoning private tutoring sector will depend on

private tutoring's effects on American students and schools, a question about which we have minimal information. We hope the patterns documented here serve as motivation and scaffolding for future research to examine this important phenomenon.

Appendix A

Variable	Abbreviation	Data Source
Prop. population between age 5 and 19	Age0519	Census (2000); ACS (2016)
Prop. schools that are charter schools	ChrtrProp	CCD School level
Prop. population with at least a bachelor's degree	EduAtLstBch	Census (2000); ACS (2016)
Prop. population with a graduate degree	EduGrad	Census (2000); ACS (2016)
Prop. population with at most a high school degree or equivalent	EduHS	Census (2000); ACS (2016)
Prop. population with at most some college	EduSomeCol	Census (2000); ACS (2016)
Prop. students in public or private school enrolled in private school	EnrlPropPriv	Census (2000); ACS (2016)
Total district expenditures per student	Exp	CCD Fiscal
Elementary and secondary expenditures per student	ExpElSc	CCD Fiscal
Instructional expenditures per student	ExpInst	CCD Fiscal
Support service expenditures per student	ExpSprt	CCD Fiscal
Prop. families with a child under 18 present	FamChild	Census (2000); ACS (2016)
Prop. families that are married couples	FamMrrd	Census (2000); ACS (2016)
Prop. women age 15-50 that gave birth in last 12 months	FertBirthed	ACS (2009); ACS (2016)
Prop. women age 15-50 who gave birth in last 12 months that are married	FertBirthPropMrrd	ACS (2009); ACS (2016)
Between-school FRPL status dissimilarity index	FRLSegSch	CCD School level
Prop. population foreign born	ImmiForBorn	Census (2000); ACS (2016)
Income inequality GINI coefficient	IncGINI	ACS (2009); ACS (2016)
Median household income	IncMedHH	Census (2000); ACS (2016)
Income per capita	IncPerCap	Census (2000); ACS (2016)
Prop. schools that are magnet	MagnetProp	CCD School level
Prop. population lived abroad in the last 12 months	MbltyDffAbrd	ACS (2009); ACS (2016)
Prop. population lived in different county in the last 12 months	MbltyDffCounty	ACS (2009); ACS (2016)
Prop. population lived in different state in the last 12 months	MbltyDffState	ACS (2009); ACS (2016)
Prop. population lived in different town in the last 12 months	MbltyDffTown	ACS (2009); ACS (2016)
Prop. population lived in same house for the last 12 months	MbltySameHouse	ACS (2009); ACS (2016)
Prop. population in management, business, science, or art occupations	OccuMgmtBsnSciArt	Census (2000); ACS (2016)
Prop. population in production, transportation, moving occupations	OccuProdTransMvng	Census (2000); ACS (2016)
Prop. population occupied in resources, construction, maintenance	OccuRsrcCnstrMntn	Census (2000); ACS (2016)
Prop. population in sales or office occupations	OccuSalesOffice	Census (2000); ACS (2016)
Prop. population in service industry occupations	OccuService	Census (2000); ACS (2016)

Prop. population with income over twice poverty level	PovOvrTwcPov	Census (2000); ACS (2016)
Prop. population with income under poverty level	PovUndr	Census (2000); ACS (2016)
Prop. population with income under half poverty level	PovUndrHlf	Census (2000); ACS (2016)
Total district revenue per student	Rev	CCD Fiscal
Revenue from federal sources per student	RevFed	CCD Fiscal
Revenue from local sources per student	RevLoc	CCD Fiscal
Proportion total revenue from local sources	RevPropLoc	CCD Fiscal
Proportion total revenue from state sources	RevPropSt	CCD Fiscal
Revenue from state sources per student	RevSt	CCD Fiscal
Ratio state source revenue to local source revenue	RevStToLoc	CCD Fiscal
Schools per student	SchPerStd	CCD District level
Prop. students in designated special education	SpecEd	CCD District level
Ratio of students to administrators	StdAdmn	CCD District level
Prop. students that identify as Asian	StdAsian	CCD School level
Prop. students that identify as Black	StdBlack	CCD School level
Prop. students designated free or reduced-price lunch	StdFRL	CCD District level
Prop. students that identify as Hispanic or Latino	StdHisp	CCD School level
Ratio of students to teachers	StdTch	CCD District level
Prop. students that identify as White	StdWhite	CCD School level
Urbanicity locale code	UrbnctyCode	CCD District level

Appendix B

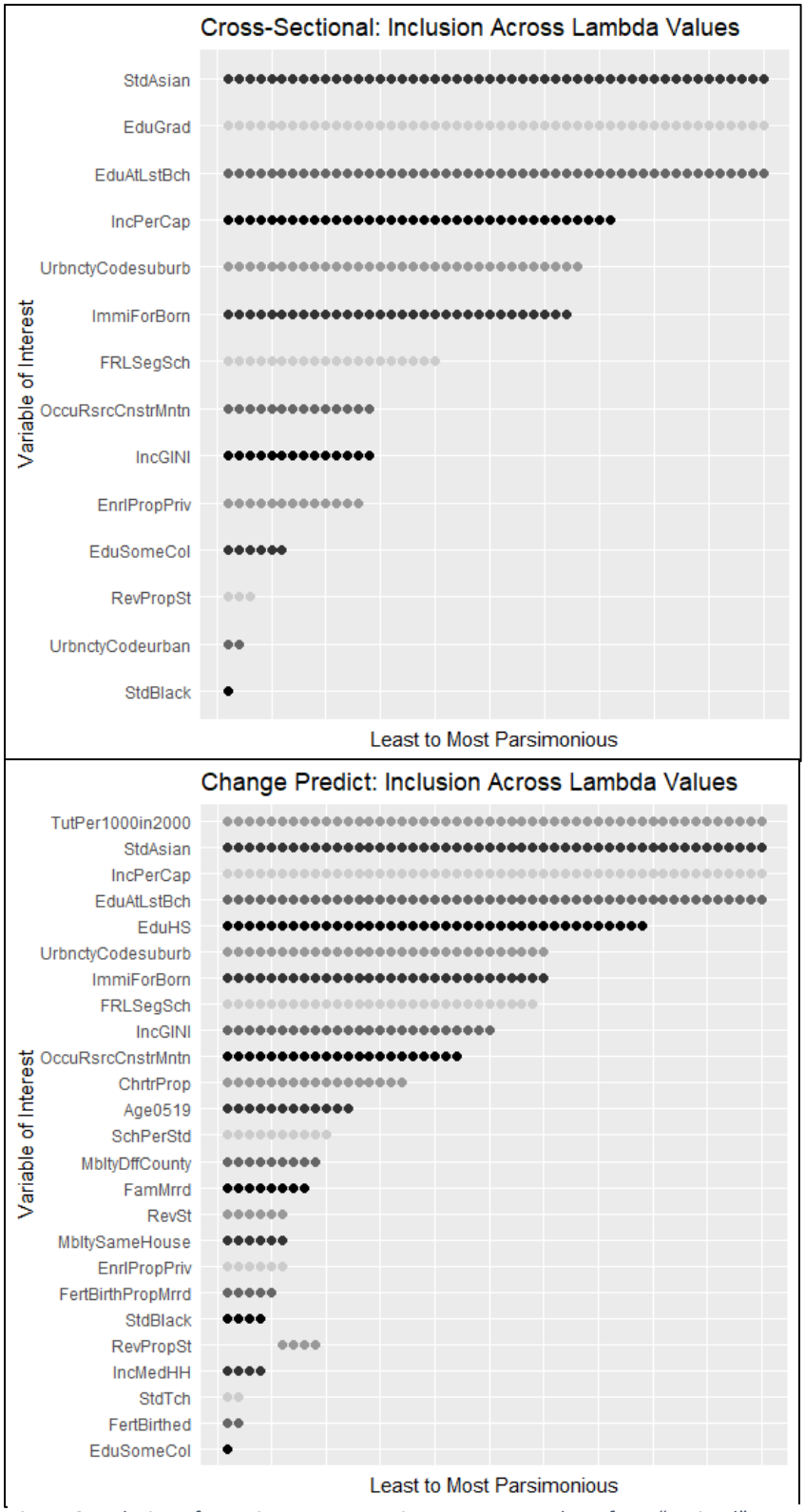

Figure 6: Inclusion of covariates across tuning parameter values, from "optimal" to one standard error away from optimal. Each point indicates a covariate was included in the model at that level of parsimony.

References

Ascher, C. (2006). NCLB's supplemental educational services: is this what our students need?. *Phi Delta Kappan*, *88*(2), 136-141.

Atalmis, E. H., Yilmaz, M., & Saatcioglu, A. (2016). How does private tutoring mediate the effects of socio-economic status on mathematics performance? Evidence from Turkey. *Policy Futures in Education*, 14(8), 1135-1152.

Aurini, J. (2004). Educational Entrepreneurialism in the Private Tutoring Industry: Balancing Profitability with the Humanistic Face of Schooling. *Canadian Review of Sociology*, 41, 475-491.

Aurini, J. & Davies S. (2004). The Transformation of Private Tutoring: Education in a Franchise Form. *Canadian Journal of Sociology*, 29, 419-438.

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open.* https://doi.org/10.1177/2332858415616358

Barrows, S., Cheng, A., Peterson, P. E., & West, M. R. (2019). Do Charters Pose a Threat to Private Schools? Evidence from Nationally Representative Surveys of U.S. Parents. *Journal of School Choice*, *13*(1), 10-32.

Bhorkar, S., & Bray, M. (2018). The expansion and roles of private tutoring in India: From supplementation to supplantation. *International Journal of Educational Development*, *62*, 148-156.

Bischoff, K., & Reardon, S. F. (2014). Residential segregation by income, 1970–2009. *Diversity and disparities: America enters a new century*, *43*.

Bound, J., Hershbein, B., & Long, B. T. (2009). Playing the admissions game: Student reactions to increasing college competition. *Journal of Economic Perspectives*, *23*(4), 119-146.

Bray, M. (2010). Blurring Boundaries: The Growing Visibility, Evolving Forms, and Complex Interactions of Private Supplementary Tutoring. *Orbis Scholae*, 4, 61-73.

Bray, M., & Kwok, P. (2003). Demand for private supplementary tutoring: conceptual considerations, and socio-economic patterns in Hong Kong. *Economics of Education Review*, *22*(6), 611-620.

Bray, M., & Lykins, C. (2012). *Shadow education: Private supplementary tutoring and its implications for policy makers in Asia* (No. 9). Asian Development Bank.

Bray, M. & Silova, I. (2006). The private tutoring phenomena: international patterns and perspectives. In Silova, I., Budiene, V. and Bray, M. (Eds.), *Education in a Hidden Marketplace: Monitoring of Private Tutoring*, (pp. 27-40). Open Society Institute.

Buchmann, C., Condron, D., & Roscigno V. J. (2010). Shadow Education, American Style: Test Preparation, the SAT and College Enrollment. *Social Forces*, *89*(2), 435-461.

Byun, S. Y., & Park, H. (2012). The academic success of East Asian American youth: The role of shadow education. *Sociology of Education*, *85*(1), 40-60.

Chappell, S., Nunnery J., Pribesh S., & Hager J. (2011). A Meta-Analysis of Supplemental Education Services (SES) Provider Effects on Student Achievement. *Journal of Education for Students Placed at Risk*, 16, 1-23.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*(2), 237-248.

Dang, H. A. (2007). The determinants and impact of private tutoring classes in Vietnam. *Economics of Education Review*, *26*(6), 683-698.

Dang, H., & Rogers, F.H. (2008). The Growing Phenomenon of Private Tutoring: Does It Deepen Human Capital, Widen Inequalities, or Waste Resources? *World Bank Research Observer*, 23, 161-200.

Davies, S. (2004). School choice by default? Understanding the demand for private tutoring in Canada. *American Journal of Education*, *110*(3), 233-255.

Davies, S., & Hammack, F. M. (2005). The channeling of student competition in higher education: Comparing Canada and the US. *Journal of Higher Education*, *76*(1), 89-106.

Dawson, W. (2010). Private tutoring and mass schooling in East Asia: Reflections of inequality in Japan, South Korea, and Cambodia. *Asia Pacific Education Review*, *11*(1), 14-24.

de Castro, B. V., & de Guzman, A. B. (2014). From scratch to notch: Understanding private tutoring metamorphosis in the Philippines from the perspectives of cram school and formal school administrators. *Education and Urban Society*, *46*(3), 287-311.

Duong, M. T., Badaly, D., Liu, F. F., Schwartz, D., & McCarty, C. A. (2016). Generational differences in academic achievement among immigrant youths: A meta-analytic review. *Review of Educational Research*, *86*(1), 3-41.

Elbadawy, A., Assaad, R., Ahlburg, D., & Levison, D. (2004, December). Private and group tutoring in Egypt: Where is the gender inequality. In *Economic Research Forum Conference in Beirut*.

Friedman, H. L. (2013). *Playing to win: Raising children in a competitive culture*. University of California Press.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1.

Fryer Jr, R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics*, 129(3), 1355-1407.

Goyette, K., & Xie, Y. (1999). Educational expectations of Asian American youths: Determinants and ethnic differences. *Sociology of Education*, 22-36.

Guryan, J., Hurst, E., & Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic Perspectives*, *22*(3), 23-46.

Kao, G., & Tienda, M. (1995). Optimism and achievement: The educational performance of immigrant youth. *Social Science Quarterly*, 1-19.

Kim, E. M., & Lee, I. (2002). Private Education in South Korea. In C. de Moura Castro & A. E. Verdisco (Eds.), *Making Education Work: Latin American Ideas and Asian Results*, (pp. 115-125). IDB.

Kim, J. H., & Park, D. (2010). The determinants of demand for private tutoring in South Korea. *Asia Pacific Education Review*, *11*(3), 411-421.

Kim, S., & Lee, J. H. (2001). Demand for education and developmental state: Private tutoring in South Korea. *Available at SSRN 268284*.

Kim, S., & Lee, J. H. (2010). Private tutoring and demand for education in South Korea. *Economic development and cultural change*, *58*(2), 259-296.

Kim, T. (1998). The evaluation for the essence and reduction policy of private education expenses. *Seoul: Korea Educational Development Institute*.

Kim, T. (2004). Shadow education: School quality and demand for private tutoring in Korea. *KDI School of Pub Policy & Management Paper*, (4-21).

Kraft, M. A. (2015). How to make additional time matter: Integrating individualized tutorials into an extended day. *Education Finance and Policy*, *10*(1), 81-116.

Kwok, P. L. Y. (2010). Demand intensity, market parameters and policy responses towards demand and supply of private supplementary tutoring in China. *Asia Pacific Education Review*, *11*(1), 49-58.

Lareau, A. (2011). *Unequal childhoods: Class, race, and family life*. University of California Press.

Lee, J. Y. (2013). *Private tutoring and its impact on students' academic achievement, formal schooling, and educational inequality in Korea* (Doctoral dissertation, Columbia University).

Lee, C. J. (2005). Korean education fever and private tutoring. *KEDI Journal of Educational Policy*, *2*(1), 99-108.

Lee, J., & Zhou, M. (2014). The success frame and achievement paradox: The costs and consequences for Asian Americans. *Race and Social Problems*, *6*(1), 38-55.

Nath, S. R. (2008). Private supplementary tutoring among primary students in Bangladesh. *Educational Studies*, *34*(1), 55-72.

Ochoa, G. L. (2013). Academic profiling: Latinos, Asian Americans, and the achievement gap. University of Minnesota Press.

Raleigh, E., & Kao, G. (2010). Do immigrant minority parents have more consistent college aspirations for their children? *Social Science Quarterly*, *91*(4), 1083-1102.

Ramey, G., & Ramey, V. A. (2010). *Brookings Papers on Economic Activity*, Spring, 129-176

Schneider, B., & Lee, Y. (1990). A model for academic success: The school and home environment of East Asian students. *Anthropology & Education Quarterly*, *21*(4), 358-377.

Snyder, T. D., & Dillow, S. A. (2015). Digest of Education Statistics 2013. NCES 2015-011. *National Center for Education Statistics*.

Duffett, A., Johnson, J., Farkas, S., Kung, S., & Ott, A. (2004). All work and no play? Listening to what kids and parents really want from out-of-school time. *Public Agenda*.

Shin, J. C. (2012). Higher education development in Korea: Western university ideas, Confucian tradition, and economic development. *Higher Education*, *64*(1), 59-72.

Shrake, E. (2010). Cram schools. *Encyclopedia of Asian American Issues Today*, *1*, 211-215.

Sriprakash, A., Proctor, H., & Hu, B. (2016). Visible pedagogic work: Parenting, private tutoring and educational advantage in Australia. *Discourse: Studies in the Cultural Politics of Education*, *37*(3), 426-441.

Tansel, A., & Bircan, F. (2006). Demand for education in Turkey: A tobit analysis of private tutoring expenditures. *Economics of Education Review*, *25*(3), 303-313.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics.

U.S. Department of Labor, Bureau of Labor Statistics, Consumer Expenditure Survey, Interview Survey, 2016

Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading research quarterly*, 179-200.

Zhang, Y., & Xie, Y. (2016). Family background, private tutoring, and children's educational performance in contemporary China. *Chinese Sociological Review*, *48*(1), 64-82.

Zhou, M., & Kim, S. (2006). Community forces, social capital, and educational achievement: The case of supplementary education in the Chinese and Korean immigrant communities. *Harvard Educational Review*, *76*(1), 1-29.