ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-756

The Causal Impact of Charter Schools on Private Tutoring Prevalence

Edward J. Kim
Bentley University

Luke W. Miratrix
Harvard University

Greater school choice leads to lower demand for private tutoring according to various international studies, but this has not been explicitly tested for the U.S. context. To estimate the causal effect of charter school appearances on neighboring private tutoring prevalence, we employ a comparative event study model combined with a longitudinal matching strategy to accommodate differing treatment years. In contrast to findings from other countries, we estimate that charter schools increase, rather than decrease, tutoring prevalence in the United States. We further find that the effect varies considerably based on the characteristics of the treated neighborhood: areas with the highest income, educational attainment, and proportion Asian show the greatest treatment impacts, while the areas with the least show null effects. Moreover, methodologically this investigation offers a pipeline for flexibly estimating causal effects with observational, longitudinal, geographically located data.

VERSION: April 2023

Suggested citation: Kim, Edward J., and Luke W. Miratrix. (2023). The Causal Impact of Charter Schools on Private Tutoring Prevalence. (EdWorkingPaper: 23-756). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/qs5q-ga02

The Causal Impact of Charter Schools on Private Tutoring Prevalence*

Edward J. Kim[1] and Luke W. Miratrix[2]

[1]Department of Mathematical Sciences, Bentley University
[2]Harvard Graduate School of Education

March 27, 2023

Abstract

Greater school choice leads to lower demand for private tutoring according to various international studies, but this has not been explicitly tested for the U.S. context. To estimate the causal effect of charter school appearances on neighboring private tutoring prevalence, we employ a comparative event study model combined with a longitudinal matching strategy to accommodate differing treatment years. In contrast to findings from other countries, we estimate that charter schools increase, rather than decrease, tutoring prevalence in the United States. We further find that the effect varies considerably based on the characteristics of the treated neighborhood: areas with the highest income, educational attainment, and proportion Asian show the greatest treatment impacts, while the areas with the least show null effects. Moreover, methodologically this investigation offers a pipeline for flexibly estimating causal effects with observational, longitudinal, geographically located data.

*This work was supported by the U.S. Department of Education, Institute for Education Sciences, through Grant R305D200010. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

1 Introduction

A properly functioning educational marketplace can improve family satisfaction with schools. With multiple schools to choose from, families can act on dissatisfaction by enrolling at a competitor without moving to another enrollment zone. At the same time, competition motivates schools to improve and respond to family preferences in order to attract students the way businesses complete to atttract customers (Hirschman, 1970; Chubb and Moe, 1990). Charter schools, as a free, alternative, mainstream schooling option, fulfill this competitor role, and marketization theory has motivated the expansion of charter schools across the U.S. over the last few decades. (Renzulli and Roscigno, 2005). To the theory's credit, research has found that families are indeed more satisfied when offered a choice in mainstream schooling, particularly with charter schools (Oberfield, 2020; Rhinesmith, 2017). Though charter schools were originally imagined in the 1970's and 1980's as a site to experiment with innovative pedogogical techniques, by the time the first charter school opened in 1992, charter schools were solidly part of the broader school choice movement (Renzulli and Roscigno, 2005). As of 2019, charter schools successfully "competed" for 7% of all K12 students in the U.S (National Center for Education Statistics, 2022).

Across a similar timeframe, the private tutoring industry rose in prominence as well. Defined as "tutoring in an academic school subject, which is taught in addition to mainstream schooling for financial gain" (Bray and Silova, 2006), private tutoring offers educational services like a mainstream school but does not fully substitute for one. Researchers have found private tutoring is more common in societies with greater potential returns for education or greater social inequality (Bray, 2003; Ireson, 2004), and also places with high competition for educational opportunities (Baker and LeTendre, 2005). But private tutoring practices and industries vary tremendously across, and even within, countries. Tutoring centers can be local and small or part of large national chains. Lessons range across all subject matter, with varied adherence to mainstream school curricula or standardized exam content. Classes can be taught at an individual, small group, or large group scale. And, tutors differ widely in age and qualifications (Bray and Silova, 2006). Exactly how a tutoring firm will operate and what sort of services families will pursue is the result of the cultural and contextual factors facing the business and the clients.

In theory, both private tutoring industry and school choice operates off of family satisfaction: families dissatisfied with their mainstream school can turn to private tutoring to supplement their child's education, just as dissatisfied families might choose to enroll in an alternative charter school. So, how do private tutoring and school choice interact with each other? This paper attempts to answer that question for the United States context.

Private tutoring in the U.S. is relatively understudied (Bray et al., 2010); nearly all research on private tutoring has taken place outside of the United States, leaving details about the burgeoning U.S. market largely a mystery. The bulk of research on private tutoring in the U.S. consists of program evaluations of Supplementary Educational Service providers, recruited under No Child Left Behind to teach students in low-performing school districts (Ascher, 2006). However, these remedial service providers, specifically targeted at low-income students in underperforming schools, do not represent the larger industry. Buchmann et al. (2010) suggest that upper-income families are the most likely to engage in private tutoring, and Kim et al. (2021) similarly report private tutoring is most prevalent in areas with high income, high educational attainment, high proportion Asian, and high proportion foreign born populations.

Notably absent from Kim, Goodman, and West's list of prominent covariates was *school choice*. Despite international evidence suggesting a negative direct relationship between school choice and private tutoring (Davies, 2004; Kim and Lee, 2010), Kim et al. (2021) find that school choice in the US was a relatively poor predictor of private tutoring prevalence, and, if anything, positively associated. The underlying causal relationship may still in fact be negative: their analysis was descriptive in nature, and other variables could have confounded the results. At the same time, the U.S. school choice environment and private tutoring industry may be sufficiently unique to manifest an altogether different interaction than what has been observed previously in other countries.

We investigate the causal impact of charter school availability on private tutoring across the united states using a mix of date science and causal inference tools to identify geographical regions differently impacted by charter schools, and seeing how those regions evolve across time. In particular, would the availability of a charter school depress the demand for tutoring centers, and would this trend vary across different geographical regions?

A causal investigation into the relationship between school choice and private tutoring in the United States context would broaden our understanding of family satisfaction with respect to schooling and the role that private tutoring firms potentially play. Charter schools, as a public, secular, and free schooling option, present an opportunity to study school choice using an alternative school more comparable to the default option than even private schools with vouchers. Using data from the U.S., we contribute to a nascent but important body of work examining specifically the U.S. private tutoring industry and its relationships to other educational institutions. Along the way, we demonstrate a pipeline for conducting a causal analysis on system-level data using a mix of data science, machine learning, and causal inference tools:

- Define school neighborhoods as geospatial zones surrounding non-charter schools
- Assess charter school availability over time, and tutoring prevalence over time, for each of these neighborhoods
- Construct a propensity score model using a LASSO shrinkage procedure with hundreds of candidate variables
- Match neighborhoods using a mix of baseline trend in tutoring center prevalence and calculated propensity score
- Use an event study model on the resulting dataset to investigate how the new availability of a charter schools could plausibly impact tutoring prevalence for the years following

Ideally, this overall structure could provide a platform for similar analyses.

We next discuss our sources of data, and describe how we make our unit-level (neighborhood) data from nationwide datasets on schools and tutoring centers. We then describe our analytic approach of first matching and then conducting an event study analysis on the resulting data. We present results, with an emphasis on treatment impact heterogeneity, looking at how the relationship between charter school and tutoring center tends to appear in only some types of neighborhoods. We finally conclude with a discussion of our findings, their limitations, and suggestions for future investigations.

2 Data and Data Preparation

Our data come from three sources: the Common Core of Data, the U.S. Census Bureau, and a proprietary data set on businesses in the U.S compiled by InfoGroup.

From the Common Core of Data (CCD) we use the school-level files for public schools, and the school-district-level files for public school districts, from 1997 to 2016. We extract longitude, latitude, charter school status, and grades served, in addition to demographic and size variables, from the school-level files. The school-district-level file provided additional demographic and size variables, as well urbanicity codes and fiscal data.[1]

The U.S. Census Bureau collects rich demographic data from individuals and households and compiles it at various levels of geography. Through the work of the National Historical Geographic Information System and the Urban Institute's Education Data Portal, we obtain school-district-level variables from the 2000 Census and the American Community Survey (ACS) five-year data sets. We use ACS data sets between 2004-2009 and 2011-2016. Note that in this paper, we identify ACS information by the final year of each five-year interval (e.g., data from the 2006-2011 ACS is matched to other data from 2011).[2]

Measures of private tutoring prevalence come from Infogroup's U.S. Historical Business dataset. Specifically, we subset to firms registered as either "Tutoring" (SIC Code 829909) or "Test Preparation Instruction" (SIC Code 874868). This data set spans 1997 to 2016 and comprises approximately 20,000 unique firms across almost 35,000 locations. Across our years of data, "Tutoring" firms were approximately 40 times more numerous than "Test Preparation Services" firms, however we could determine no meaningful distinction between the categories (e.g., the Kaplan tutoring franchise had firms

[1] The variables latitude, longitude, charter school status, and grades served, had missing values that easily lent themselves to imputation. These values can be missing for various reasons (e.g., data on the first three variables were explicitly not collected in earlier CCD surveys). For each school, we extrapolate the chronologically first non-missing value to prior years if we observe the same non-missing value consecutively for three years. We repeat a similar process with the last non-missing value. Next, if, given a missing value, the subsequent and preceding values are identical, we interpolate the missing middle value as the same.

[2] For continuous variables from either the CCD or the Census we linearly impute missing values. This is particularly useful to bridge the gap in observations between the 2000 Census and the earliest available ACS in 2009. For each variable, any values the imputation suggested that were greater than the observed maximum, we maintained as missing. We treated minimum values similarly. We only use these imputed values for the LASSO and propensity score model, which we detail in the next section. The eventual event study models estimating causal effect sizes do not use imputed covariate values.

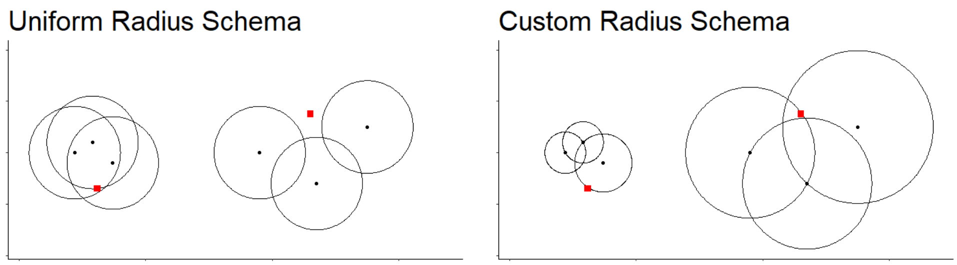

Figure 1: A comparison of a uniform radius versus a custom radius to determine proximal relevance

in both categories), and thus combined them for the outcome measure. Variables utilized from this data set are location and year of establishment.

Our unit of analysis are school "neighborhoods," defined by the location of public schools, and "treatment" is defined by a change in the number of charter schools within those neighborhoods. Charter school enrollment zones do not match up with well documented geographic boundaries such as school districts, which complicates clear definitions of treated and untreated units. To determine whether a charter school or private tutoring firm opening is proximally relevant to a neighborhood, we use geographic density to calculate a school neighborhood-specific radius for each non-charter school in our dataset. For school i in year t, we define the distance to the nearest non-charter school with overlapping serviced grades as $d_{i,t}$. We then set the neighborhood radius of school i, δ_i, as twice the median of $d_{i,t}$ across all years t. All radii smaller than 0.1 miles or greater than 50 miles are arbitrarily set to 0.1 miles and 50 miles, respectively. The middle half of our radii are between 1.4 and 9.7 miles, which aligns with suggestions from related charter school literature that implement a uniform radius schema across schools of between 1.5 and 2.5 miles for urban settings (Slungaard Mumma, 2022).[3]

We calculate, for school i in year t, the number of charter schools that are both within the radius and also have an overlap in serviced grades. We define a charter school appearance, our treatment condition, in year t as a positive

[3]We considered numerous alternative specifications, including scaling the median δ_i by 1.5x, 2.5x and 3x, as well as using the unscaled distance based on the second nearest school with overlapping serviced grades, and the unscaled average of the distance to the nearest and the distance to the second nearest school. The results for all alternative specifications did not contradict those presented in our main analysis.

change in the charter school count between year $t-1$ and year t. Note this definition does not limit treatment to newly opened charter schools, but also includes school relocations, reopenings, and changes to serviced grades. For parsimony, we restrict our sample to schools that observed one positive change in charter count (i.e., received treatment in exactly one year), or observed no change in charter school presence (i.e., never received treatment), during the period of observation. This caused us to drop 23% of the schools that had originally met our data requirements. We derive our outcome measure similarly, by counting the number of tutoring centers within the relevant neighborhood for each year.

3 Methods

Our focus is to estimate the impact of a charter school opening in a neighborhood on tutoring prevalence, as measured by the number of tutoring centers in that neighborhood. We do this via a combination of two quasi-experimental tools: a matching procedure and an event study model.

We use the matching procedure to identify, for each of our treated neighborhoods, a neighborhood that never experienced a charter school opening, but which is similar to the treated neighborhood in terms of growth in tutoring center prevalence and the chance of an increase in the charter school count for the same year as the treated unit (i.e., a propensity score). We will then use these selected control neighborhoods to estimate the counterfactual trajectories we might have seen for the treated neighborhoods, had they not had an increase in charter school prevalence.

We then estimate our impacts via an event study model. One advantage of an event study model for our context is that it does not impose a parametric form on the treatment effect; immediate or delayed, sustained or temporary, growing or shrinking, a non-parametric model can accurately capture any such effect. Our model spans 10 years before treatment onset, to consider possible pretreatment differences between the treated and control units, to 10 years after onset, to consider long term effects. The matching procedure increases the comparability of the treated and control neighborhoods, and thus the credibility of our underlying assumptions behind this model.

We next detail our matching procedure and then our analytic model in the following two subsections.

3.1 Matching

Preprocessing a data set by filtering the potential controls based on similarity to treated units (i.e., by matching) increases the comparability of the treated group and control group and makes subsequent analyses less dependent on modeling choices and specifications (Ho et al., 2007). We determine similarity between a treatment unit and potential control units along two dimensions: similarity on the trajectory of pretreatment outcomes, and similarity on propensity score (i.e., probability that a unit receives treatment in a given year). Matching on propensity score is used to create balance across potentially confounding covariates, resulting in a treatment and control group that can be treated as essentially randomized (Rosenbaum and Rubin, 1983). Matching on outcome trends directly checks whether two units are comparable in the pretreatment period so that the change or growth of the identified control units is a plausible counterfactual for how the treated neighborhoods would have changed, absent an increase in charter schools. Importantly, we do not need to directly match on pre-treatment level, as our event study model adjusts for levels via two-way fixed effects.

Specifically, we calculate a match distance $D_{a,b}$ between a treated unit a and each of its potential control units b as

$$D_{a,b} = (PS_{a,t_a^*} - PS_{b,t_a^*})^2 + (\delta_{a,\mathbf{t}_{a,b}} - \delta_{b,\mathbf{t}_{a,b}})^2, \tag{1}$$

where t_a^* is the treatment onset year for treated unit a, $\mathbf{t}_{a,b}$ is the set of pretreatment years of treatment unit a that are also observed for potential control unit b, $PS_{i,t}$ is the standardized propensity score (in logits) for unit i receiving treatment in year t, and $\delta_{i,\mathbf{t}}$ is the pre-treatment trend of unit i in our outcome, estimated via a least squares regression of the outcome on time for unit i across the time points in set $\mathbf{t}$. We standardize the $\delta_{i,\mathbf{t}}$ based on the standard deviation of the pretreatment slopes of treated units. Our match distance matches on the propensity score for treatment for a particular *year*; once we match, we then take the year of the treated unit as the effective year of non-treatment of the matched control. We can then calculate year post-treatment for all units, treatment and control, in the event study model discussed below. A demonstration for deriving $D_{a,b}$ explicitly is available in Appendix C.

Once we obtain all pairwise match distances, we use a full matching strategy via the "optmatch" package (Hansen and Klopfer, 2006), which creates n:1 and 1:m treatment-control matched groups such that within group distance and total distance across units are minimized. For any candidate match we set a maximum allowable distance (i.e., "caliper") of 0.1 standard deviations for both the left-hand and right-hand term in Equation (2). We further require exact matching on state to control for structural differences between states in unobserved characteristics such as strength of charter school movements or sentiment towards private tutoring. For analytic simplicity, we assign every treated unit a weight of 1 in the final model which orients our perspective towards the treatment effect on treated units. Each control unit is assigned the weight $\frac{n}{m}$, where n:m is the treatment-control unit ratio within its matched group.

We estimate propensity scores via a LASSO procedure with 5-fold cross-validation, optimizing the RMSE of the logistic regression model. The LASSO (Least Absolute Shrinkage and Selection Operator) identifies a set of linear predictors that explains the most variation in the outcome variable subject to a penalty for possible overfitting (Tibshirani, 1996). In our context, the outcome variable is propensity for treatment assignment. Every year of available data for each of the school units, including post treatment years for treated units, was used in the estimation procedure. Some of the utilized variables were identical across some schools (e.g., district level funding); a list of all variables used in the LASSO can be found in Appendix A. To account for structural differences between states, we conduct a separate LASSO for each state. We drop the 11 states[4] from our sample that had fewer than 15 treated units (4 of whom had none), to ensure at least three treated units in each cross-validation fold. Full details on the LASSO results are available upon request.[5]

We finally note that alternate definitions of $D_{a,b}$ are certainly possible. In fact, for a prior iteration of this analysis, we replaced the right-hand term in Equation (2) with the squared differences of pretreatment outcome values, rather than pretreatment outcome trajectories, attempting to units by their entire sequence of pre-treatment outcomes. Though the main results from this version were largely similar to the prior iteration, for some robustness

[4]Alabama, Delaware, Kentucky, Montana, Nebraska, North Dakota, South Dakota, Vermont, Washington, West Virginia, Wyoming

[5]We implement the LASSO procedure using the R package 'glmnet' (Friedman et al., 2010)

[7]Covariates standardized by "Before" variation.

[7]Due to weighting the effective n for control units after matching is equal to the n for treated units after matching

	Average in Treatment		Average in Control		Tx - Co (Standardized)[6]	
	Before	After	Before	After[7]	Before	After
Income per Capita	25,397	25,237	26,095	25,782	-0.075	-0.059
Prop. At Least Bachelor's	0.163	0.162	0.158	0.167	0.065	-0.073
Prop. Asian Student	0.0386	0.0406	0.0381	0.0432	0.007	-0.031
Prop. Foreign Born	0.109	0.117	0.0864	0.199	0.227	-0.023
N	5,782	4,267	33,199	19,004		

Table 1: Covariate Means by Treatment Status Before and After Matching.

checks (described in a later section) the matching process using the prior distance definition failed to produce matches with comparable pretreatment outcome periods. Also, results from Daw and Hatfield (2018) underscore that matching on pretreatment level with difference-in-difference approaches can introduce bias; matching on trend avoids this danger. For simplicity we only present the trajectory matched version, but for transparency acknowledge this original attempt here. Future analysts should ideally declare from the outset an intent to consider both options before proceeding with whichever produces, e.g., better pretreatment balance.

The matching process returned a set of 4,267 treated units and 19,004 control units across 3,982 matched groups, which represents 73.8% of candidate treated units and 57.2% of candidate control units (some treated units were dropped if no comparable control unit existed, e.g., propensity score close to 1, and vice versa). To assess covariate balance, we consider each unit in 2009, a time point near the middle of our period of observation with minimal missingness. As shown in Table 1, matching yielded mixed results for shrinking group mean differences on the four covariates of interest identified by Kim et al. (2021), which were available at the district level. Standardized differences reveal that group averages do not necessarily get closer, which is particularly true of the proportion Asian covariate. That being said, all imbalances are below 0.10σ, generally considered a reasonable level of imbalance, post matching. However, Figure 2 reveals that the under-

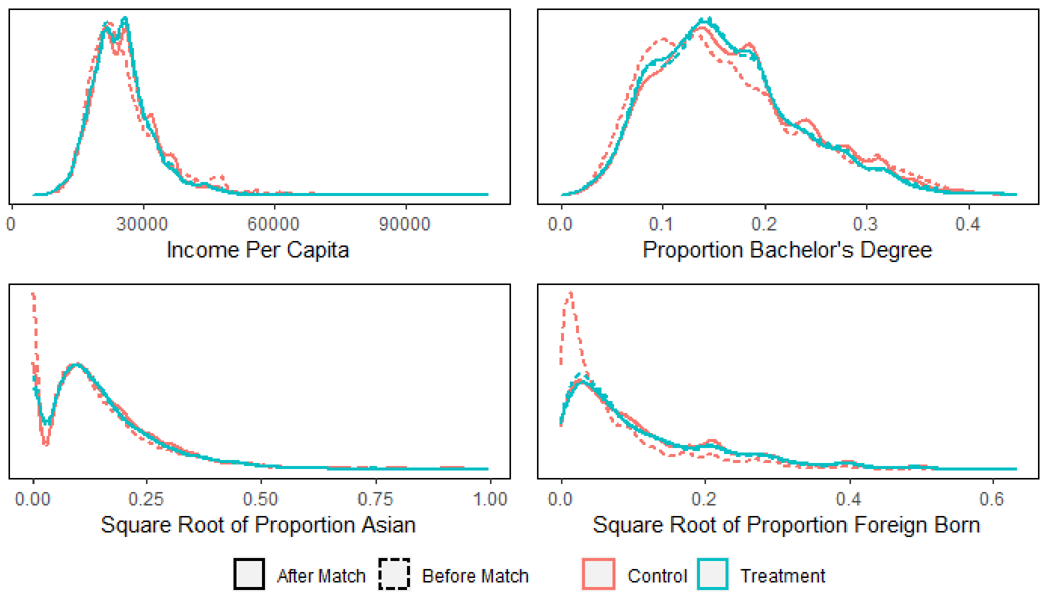

Figure 2: Density Plots by Matched and Treatment Status

lying distributions did align substantially as a result of matching. Further, Figure 3, which centers the outcome of matched units by the unit average to approximate the unit fixed effects employed in the model, illustrates that the average pretreatment trends within the matched sample are similar.

3.2 The Event Study Model

Our analytic model for neighborhood i in time t is:

$$
\begin{aligned}
Y_{it} = & \sum_{k=-10}^{-1} (\gamma_k \mathbf{1}_k(t - TxYr_i) + \beta_k \mathbf{1}_k(t - TxYr_i) * EverTx_i) \\
& + \sum_{k=1}^{10} (\gamma_k \mathbf{1}_k(t - TxYr_i) + \beta_k \mathbf{1}_k(t - TxYr_i) * EverTx_i) \\
& + \alpha_i + \delta_t
\end{aligned} \tag{2}
$$

[8]Note that we cannot derive an equivalent of Figure 3 before matching, since control units' year-relative-to-treatment value is defined by their match.

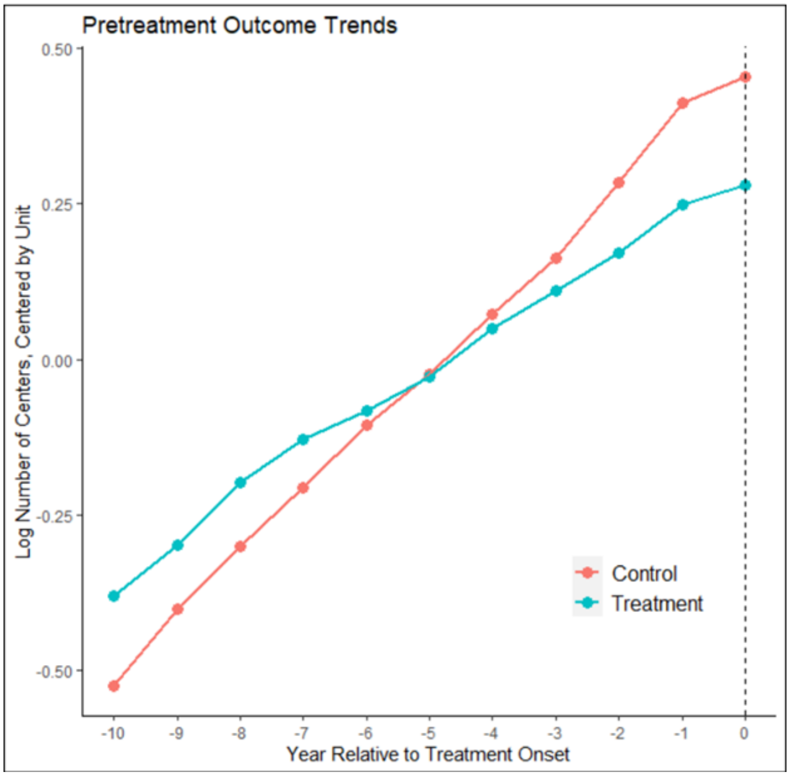

Figure 3: Pretreatment Trends Across Treatment Assignment [8]

where Y_{it} is the log number of tutoring centers plus one[9], t_i^* is the treatment onset year (or the treatment onset of the matched treatment unit in the case of a control unit), $\mathbf{1}_k(x)$ is an indicator function that equals 1 when x equals k and 0 otherwise, γ_k describes the counterfactual (i.e., untreated) trend, $EverTx_i$ indicates treatment group status, and α_i and δ_t are unit and year fixed-effects, respectively. The coefficients of interest are the β_k terms, each representing the relative increase in the outcome for each year before or following an increase in charter school availability for those locations that had such increases. Specifically, given comparable treatment and control units and a nonzero treatment effect, we would expect β_k to be indistinguishable from zero when k is negative (i.e., during the pretreatment period), and

[9]We add one to our outcome before applying the log function to avoid dropping observations with no tutoring centers, as is standard practice.

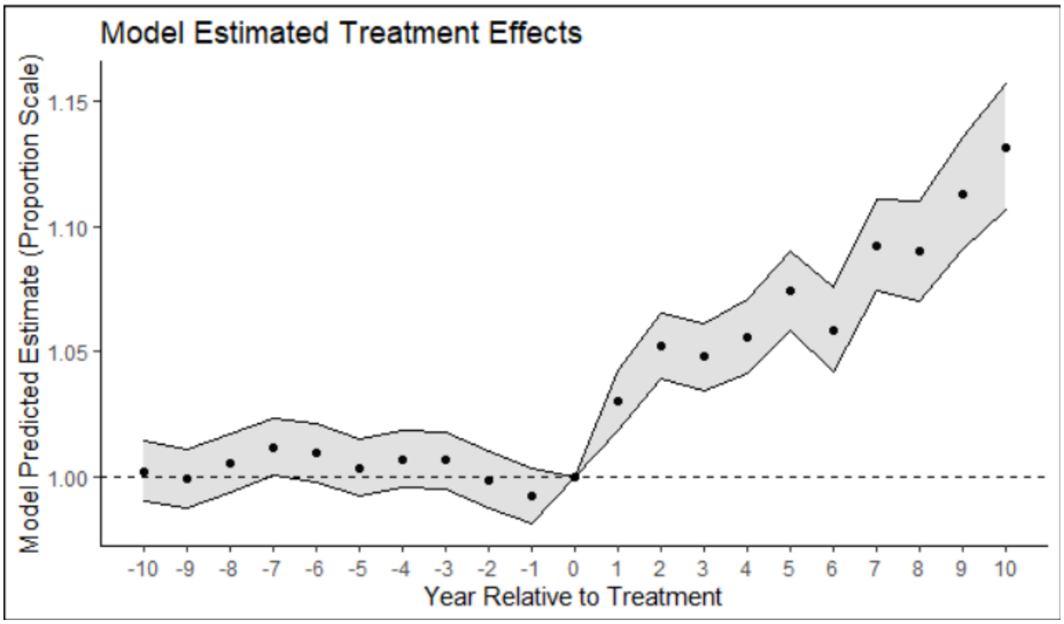

Figure 4: Model Estimated Group Differences

significantly different from zero when k is positive (i.e., the posttreatment period). See Appendix B for further details on data coding for this analysis.

Our analytic strategy relies on several assumptions to warrant a causal interpretation of the impact estimate. In addition to the typical requirements of least squares regression, we require that the onset of treatment does not affect control units, that the changes in outcome of control units serve as a valid counterfactual for the treatment treated units in the absence of treatment, and that no other event systematically occurs at the time of treatment that would differentially affect the outcome of treated and control units. With these assumptions in place, our model can statistically account for variations in the outcome not due to treatment and thus yield a quasi-experimental estimate of the effect charter school openings on private tutoring prevalence.

4 Results

The model estimates, displayed in Figure 4, suggest a consistent, positive, and growing treatment impact on private tutoring prevalence in response to charter school openings. For ease of interpretation we discuss the results, which were estimated using the log scale, as proportional differences, an absolute scale. After one year post treatment, the treatment units had, on average, 3.1% additional growth in tutoring centers as compared to control units. By five years post treatment, the treatment group had an average additional 7% of growth. Also evident from Figure 4 is the alignment between the treatment and control groups before treatment onset (or hypothetical treatment onset), which we take as an encouraging signal that the control units served as a valid counterfactual. We present a wide period of observation to clearly illustrate trends over time, but we caution the reader from generalizing or interpreting too closely the results at the distal time points. Due to the different years of treatment onset, the sample represented by each post treatment period estimate changes; only units treated early in the sample would have data for 10 years post treatment available, as evident in Figure 5.

We conceived of three primary threats to the validity of our results: cohort effects, skewing from the log transformation, and multiply counted tutoring centers. The first concern is that charter schools that opened in a specific year may be driving the observed results and we cannot generalize beyond this cohort. To investigate this, we separate the analytic sample by treatment year and directly plot the outcome values. As shown in Figure 5, though some cohorts show minimal separation posttreatment, and some seem to exhibit negative effects, the overall results do not seem driven by any one cohort and multiple cohorts demonstrate positive effects.

Our second concern is that the decision to apply a log transformation to the outcome may have skewed the results given the preponderance of zero valued outcomes (i.e., schools with no tutoring centers in the surrounding neighborhood) in the data set. Adding 1 before applying the log function is necessary to retain units with outcome value 0 in the analysis, but this complicates interpretations of the model results. Consider that from the model's perspective, which operates on the log scale, units that change from 0 to 1 tutoring centers (i.e., $\log(2) - \log(1)$) exhibit about twice the magnitude of difference as units that change from 1 to 2 tutoring centers (i.e., $\log(3) - \log(2)$). Though subsequent results are not inaccurate, we would hesitate

[10] The cohort of treatment year 2016 (14.34% of treated units) was used in the model estimation but dropped from Figure 5 as it contained no posttreatment period data. The pretreatment alignment was comparable to that of other cohorts.

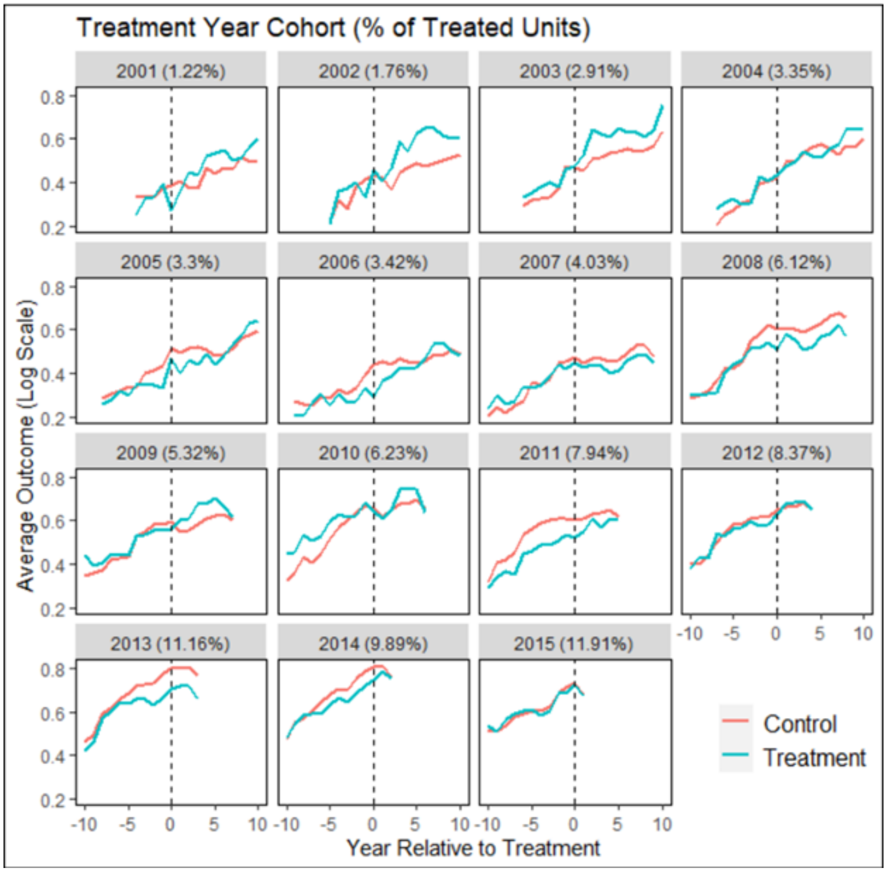

Figure 5: Raw Outcomes by Treatment Year[10]

to interpret them broadly if primarily driven by a preponderance of zeroes, which make up 30% of the unit-by-year outcome values in the analytic sample. We therefore conduct an alternative version of the analysis where we recode units with values of 0 such that, from the perspective of the model, changes from 0 to 1 tutoring centers exhibit the same magnitude difference as changes from 1 to 2 tutoring centers. The results, presented in Appendix D, are largely comparable, which suggests our results are not the product of the log transformation process.

The last concern we check is the fact that by our data coding schema the same tutoring center may be counted multiple times for different schools if the respective neighborhood radii overlapped. Insofar as multiple charter school appearances could have a compounding effect on tutoring center openings our main results are not necessarily invalid, but they could be incorrectly generalizing or inflating the direct causal effect of one charter school appearance event. To investigate this, we rerun the entire main analysis, including the LASSO and matching steps, but code the outcome in three different ways: (1) tutoring centers are only counted towards the outcome of the nearest school in the data set (*absolute*), (2) every one tutoring center counts $1/n$ towards the outcome of each of the n school neighborhoods it resides in (*split*), and (3) every tutoring center j counts $w_{i,j}/N_j$ towards the outcome of school i whose neighborhood the center j resides in, where $w_{i,j}$ is school i's neighborhood radius divided by the distance between school i and center j, and N_j is the sum of all $w_{\cdot,j}$'s (*weigh*). Figure 6, which depicts Figure 4 for each of the three schemas, shows a similar pattern as the main results for two of these specifications, though the scale of the outcome is smaller and some pretreatment periods are significantly different between treatment and control groups. Notably, the *absolute* schema shows almost entirely different treatment and control trends in the pretreatment period indicating poor matches and invalidating a causal interpretation of posttreatment period patterns. We could not remedy this by tightening the caliper in the matching process, and so we discount this specification. Uniformly counting private tutoring firms is a coarse measure, unable to capture how much tutoring, increasing or decreasing, takes place in a given firm. However, the correspondence of our main results and two of the sensitivity checks here adds evidence to the plausibility of an underlying positive, immediate, and growing effect.

4.1 Heterogeneity Results

Research on charter schools shows that charter schools differ vastly in their quality, student body served, and interaction with other schooling options (Buddin and Zimmer, 2005). Private tutoring firms similarly vary based on the client demographics. We attend to this variation by considering effect heterogeneity across the four covariates identified by Kim et al. (2021) as the most prominent predictors of private tutoring prevalence in the United States: income per capita, proportion bachelor's degree holders, proportion Asian student population, and proportion foreign born population. For each variable, we divide the treated units into a bottom tercile, middle tercile, and top tercile, according to their covariate value in the year 2009. Control units

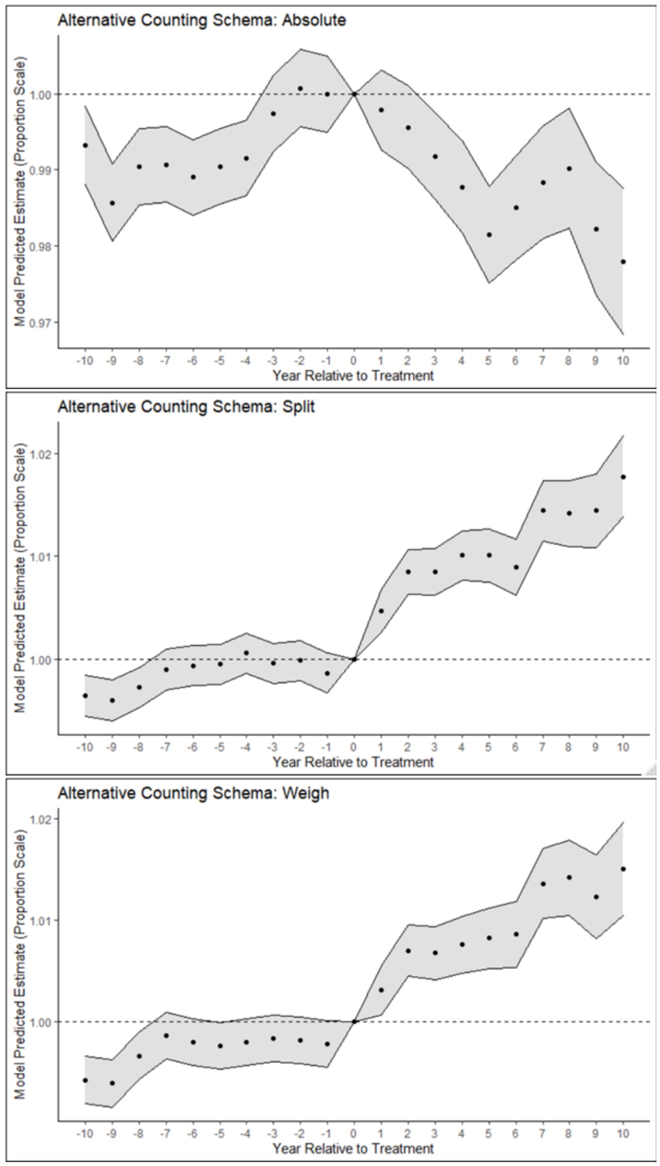

Figure 6: Main Results with Alternative Outcome Coding Schemas

are assigned the same tercile as the treated units to which they were matched. In the case of matched groups with multiple treated units assigned to a single control unit, the control unit is assigned fractional tercile assignments in proportion to the tercile assignments of its matched treated units. We operationalize these categories by including interactions with all the previous model covariates as follows:

$$
\begin{aligned}
Y_{it} = & \sum_{k=-10}^{-1} (\gamma_k \mathbf{1}_k(t - TxYr_i) + \beta_k \mathbf{1}_k(t - TxYr_i) * EverTx_i) \\
& + LowGrp_i \sum_{k=-10}^{-1} \left(\gamma_k^{(L)} \mathbf{1}_k(t - TxYr_i) + \beta_k^{(L)} \mathbf{1}_k(t - TxYr_i) * EverTx_i\right) \\
& + HighGrp_i \sum_{k=-10}^{-1} \left(\gamma_k^{(H)} \mathbf{1}_k(t - TxYr_i) + \beta_k^{(H)} \mathbf{1}_k(t - TxYr_i) * EverTx_i\right) \\
& + \sum_{k=1}^{10} (\gamma_k \mathbf{1}_k(t - TxYr_i) + \beta_k \mathbf{1}_k(t - TxYr_i) * EverTx_i) \\
& + LowGrp_i \sum_{k=1}^{10} \left(\gamma_k^{(L)} \mathbf{1}_k(t - TxYr_i) + \beta_k^{(L)} \mathbf{1}_k(t - TxYr_i) * EverTx_i\right) \\
& + HighGrp_i \sum_{k=1}^{10} \left(\gamma_k^{(H)} \mathbf{1}_k(t - TxYr_i) + \beta_k^{(H)} \mathbf{1}_k(t - TxYr_i) * EverTx_i\right) \\
& + \alpha_i + \delta_t
\end{aligned}
\tag{3}
$$

where covariates are defined similarly as before in Equation (2). Note that we set the middle tercile as the reference category. Our interest is both which of the terciles indicate differences between treatment and control groups, as well as whether the terciles demonstrate substantially different patterns from one another. We run the model in Equation (3) four times, one for each of our variables of interest.

The results in Figure 7 suggest that the treatment impact of charter school appearances has considerable heterogeneity across our variables of interest. The treated and control groups do occasionally show significant differences in the pretreatment period, but not in a manner that would explain the post treatment patterns. In relation to income per capita, proportion with bachelor's degree, or proportion Asian student population, we surmise that treatment impact has a positive association: the top terciles demonstrate the

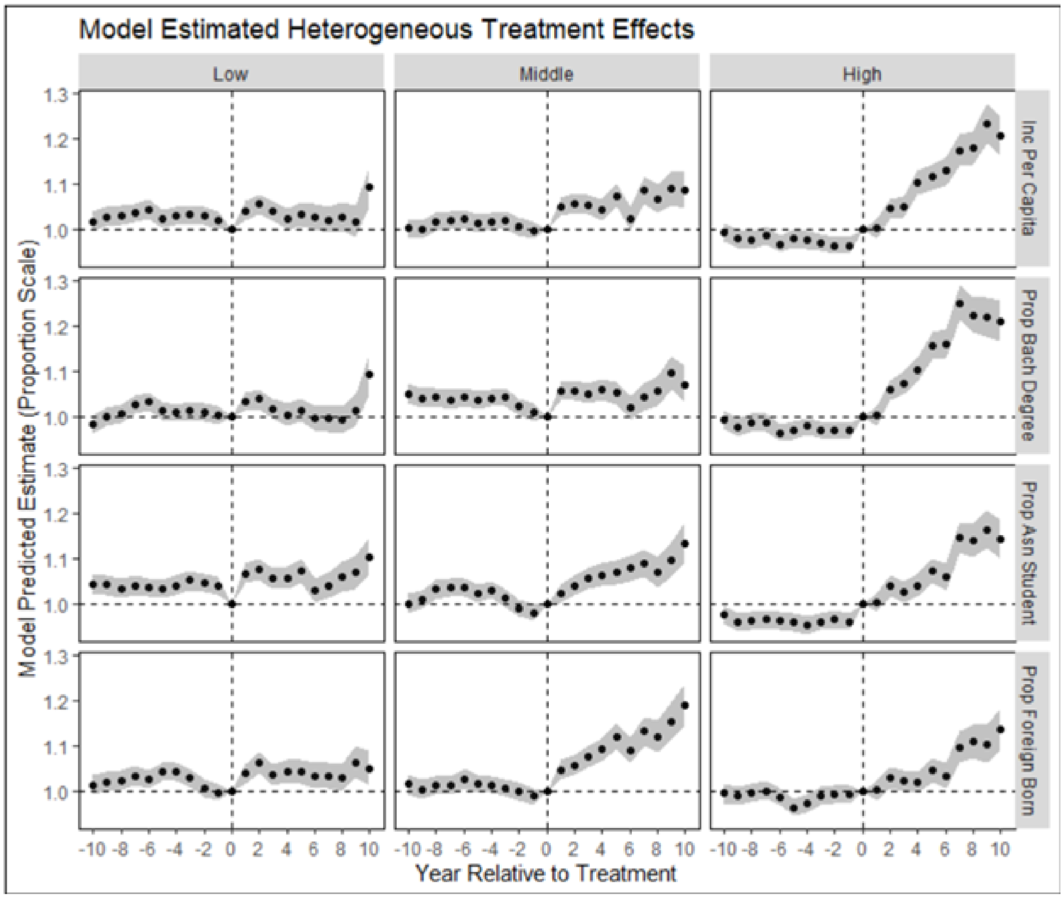

Figure 7: Model Estimated Heterogeneous Treatment Effects

largest treatment impacts, while the bottom terciles showed essentially no significant differences and the middle tercile was somewhere in between. Given all these variables are positively associated with private tutoring prevalence (Kim et al., 2021), we might conclude that treatment impacts are greater in areas with predisposed interests in private tutoring. However, the variable proportion foreign born showed a somewhat different pattern where the middle tercile observed the largest treatment impacts, followed by the top tercile.

We caution the reader from interpreting the pattern of heterogeneous treatment effects too closely. By dint of our analytic strategy, we can observe that charter school appearances do appear to cause greater changes to private tutoring prevalence in, say, higher income neighborhoods, but we could not attribute this greater change in treatment effect to income levels; our implementation of heterogeneity was non-causal. Take for example the finding that the middle tercile of proportion foreign born demonstrated the largest treatment effects. In our sample, units at the middle of the proportion foreign born distribution also demonstrate the highest income per capita.[11] Thus, we cannot distinguish which, if any, of our four covariates induced the variation in treatment effect.

5 Discussion

Our paper explores the causal relationship between school choice and private tutoring in the United States. Contrary to research from other national contexts, the model results suggest a positive treatment impact, one that presented immediately after treatment onset and grew in magnitude over time. That is, relative to control, tutoring prevalence in treated neighborhoods increased, and continued to increase, after the appearance of a charter school. We further find evidence for heterogeneous treatment impacts. In general, neighborhoods with more income, educational attainment, proportion Asian students, and proportion foreign born demonstrated the largest treatment impacts, while the neighborhoods with the least showed no significant differences.

The existing literature on private tutoring in the United States, limited as it is, offers a few suggestions to explain the observed pattern of impacts. For example, the United States context may present as a parallel to the Canadian context: greater school choice encourages families to bundle together charter schools and private tutoring, and the effects we observe are substitutions for private school enrollment that would have occurred in the absence of treatment. This mechanism would be most salient in high educational attainment and high-income communities, who tend to have higher expectations for education and could afford to consider both choices. Alternatively, if high-achieving students turn to private tutoring in response to pressures of increasing competition for elite higher education (Bound et al., 2009), the appearance of a charter school, competing with the other schooling

[11] This pattern is described in more detail in Kim et al. (2021), which uses a similar data set.

options, may induce families to perceive more competition and react accordingly. Or treatment impact may be a function of preexisting private tutoring infrastructure. Private tutoring businesses may be reacting to charter school appearances directly, interpreting them as signals of interest in greater educational resources, rather than reacting to demand from families (Davies and Aurini, 2008). Asian American communities, given a cultural familiarity with private tutoring (Bray and Silova, 2006), could foster the requisite critical mass, and the positive correlation between proportion Asian with the covariates for income and education level would explain the rest of the observed treatment heterogeneity.

More research is necessary to place these results definitively within a narrative context. The present investigation suggests that charter schools have a positive causal effect on private tutoring prevalence in the United States but is not well equipped to describe why this relationship exists or whether this pattern extends outside of the available sample or to school choice more broadly. At the least, this study reveals an unexpected dynamic between private tutoring and mainstream schooling and prompts further inquiry, as an understanding of family satisfaction with schooling underpins education spending policies and school choice initiatives.

From a methodological standpoint, this paper uses a number of innovative analytic strategies to overcome obstacles in the data. To compensate for ambiguous treatment status, we defined geographic radii around each school based on school density, and within that radius measured treatment status via charter school appearances. To match units with multiple years of data and varying treatment onset years, we customized a match distance formula that could accommodate all available pretreatment outcome data and used propensity scores derived from a LASSO model. Finally, without adequate theory or prior research to inform our choice of model, we opted for a comparative event study framework to flexibly describe any possible outcome patterns. In summary, this combination of methods leverages observational data of geographically located, longitudinal units into a workable causal inference strategy, and we hope this study serves as an illustrative example for researchers working with similarly complex data.

A List of LASSO Variables

Calendar year, school operational status, school level, urbanicity code, school enrollment size*, proportion free or reduced-price lunch*, tutoring center count*, tutoring sales*, proportion white student*, proportion black student*, proportion Hispanic student*, proportion Asian student*, proportion special education*, proportion English language learner*, preschool serviced, kindergarten serviced, 1^{st} grade serviced, 2^{nd} grade serviced, 3^{rd} grade serviced, 4^{th} grade serviced, 5^{th} grade serviced, 6^{th} grade serviced, 7^{th} grade serviced, 8^{th} grade serviced, 9^{th} grade serviced, 10^{th} grade serviced, 11^{th} grade serviced, 12^{th} grade serviced, student teacher ratio*, number of schools*~, preschool through 12^{th} grade enrollment*~, students enrollment*~, private school payment*~, charter school payment*~, FRPL dissimilarity index*~, proportion population aged 5 through 19*~, proportion population white*~, proportion population black*~, proportion population Hispanic*~, proportion population Asian*~, proportion population American Indian*~, proportion population native Hawaiian*~, proportion families with children*~, proportion families married*~, proportion with high school degree*~, proportion with at least some college*~, proportion with at least bachelor's degree*~, proportion with bachelor's degree*~, proportion with graduate degree*~, proportion under poverty line*~, proportion under half the poverty line*~, proportion over twice the poverty line*~, proportion with occupation in management, business, sciences, or arts*~, proportion with occupation in service industry*~, proportion with occupation in sales or office*~, proportion with occupation in resources*~, construction, or maintenance*~, proportion with occupation in production, transportation, or moving, median household income*~, income per capita*~, proportion enrolled in private school*~, proportion foreign born*~, school to student ratio*~, student teacher ratio*~, student guidance counselor ratio*~, student administrator ratio*~, total revenue per student*~, total federally sourced revenue per student*~, total state sourced revenue per student*~, total locally sourced revenue per student*~, proportion revenue from local sources*~, proportion revenue from state sources*~, state sourced revenue to locally sourced revenue ratio*~, total expenditures per student*~, total expenditures for elementary and secondary per student*~, total expenditures for instruction per student*~, total expenditures for support services per student*~, enrollment as summed by individually reported racial groups*~, proportion white student*~ , proportion black student*~ , proportion Hispanic student*~ , proportion Asian*~, proportional free or reduced price lunch*~, proportion magnet schools*~, proportion charter schools*~[12] [13]

---

[12] ~ indicates the variable is at the school district level, rather than the school level.

[13] * indicates the LASSO also included the 1-year and 2-year lagged versions of the variables when possible.

B Defining Indicator Functions for Control Units

Equation (1) requires a year of treatment onset to define the pre- and post-treatment indicators $\mathbf{1}_k$. In the case of a 1:n or 1:1 matched group, the control units are assigned the same indicator variable values as their matched treated unit, and can be intuitively understood as contemporaneous counterfactuals. For n:1 matched groups, though, if the treated units differ in year of treatment onset, we require some mathematical cunning. Consider control unit b, which was matched to treated units a_1, a_2, ..., a_n. The treatment indicator variable for control unit b in year t, $\mathbf{1}_k(t-t_b^*)$, is the average of $\mathbf{1}_k(t-t_{a_1}^*)$, $\mathbf{1}_k(t-t_{a_2}^*)$, ..., $\mathbf{1}_k(t-t_{a_n}^*)$, i.e., the average across the matched treated units in year t.

For example, in a matched group with one control unit and three treated units, if a given year T is one year post treatment for two of the matched treated units, and two years post treatment for a third matched treated unit, then the control unit's one year post treatment indicator $\mathbf{1}_1(t-t_b^*)$ would be coded as 2/3, and its two year post treatment indicator $\mathbf{1}_2(t-t_b^*)$ would be coded as 1/3. This produces the intended effect in our analytic model, which would assign a weight of 3 to this control unit. Intuitively, the weight clones our control unit three times, and the fractional values assign each clone to correspond to one of the three treated units.

Year T	Treatment: a_1	Treatment: a_2	Treatment: a_3	Control: b
$\mathbf{1}_1(t-t_i^*)$	1	1	0	2/3
$\mathbf{1}_2(t-t_i^*)$	0	0	1	1/3
$\mathbf{1}_3(t-t_i^*)$	0	0	0	0
$\mathbf{1}_4(t-t_i^*)$	0	0	0	0
Year $T+1$	Treatment: a_1	Treatment: a_2	Treatment: a_3	Control: b
$\mathbf{1}_1(t-t_i^*)$	0	0	0	0
$\mathbf{1}_2(t-t_i^*)$	1	1	0	2/3
$\mathbf{1}_3(t-t_i^*)$	0	0	1	1/3
$\mathbf{1}_4(t-t_i^*)$	0	0	0	0

Table B: Example coding of post treatment indicators for 3:1 matched group

C Calculating Match Distance

Let n and m be the number of treated units and control units respectively. Further, let M be the number of unit-by-year observations across all control units (note the number of available years of observation may differ between units as schools open and close at varying times). The entire distance matrix would then be of dimensions n x M, where each row represents a treated unit in the year of treatment onset, and each column represents an observed year of a given control unit.

Consider row i and column j in the distance matrix, which correspond to treated unit a and year t of control unit b respectively. If t is not equal to t_a, the treatment onset year for treated unit a, then the distance is set to infinity, due to our exact matching on year. Thus, each row has at most m potential matches (i.e., non-infinite values), corresponding to the relevant year across each of the m control units. Similarly, the distance is set to infinity if the states for treated unit a and control unit b are not equal, due to our exact matching on state.

As shown in Figure C, the match distance between two units depends largely on the treatment year of the treated unit, as this determines which propensity score value to compare, and how many years of outcome data to consider. We can thus interpret the match distance in the following way: given a pair of treated unit a with treatment onset year t_a and control unit b, compare the propensity scores in year t_a, and then compare the outcome values for all years available prior to and including t_a.

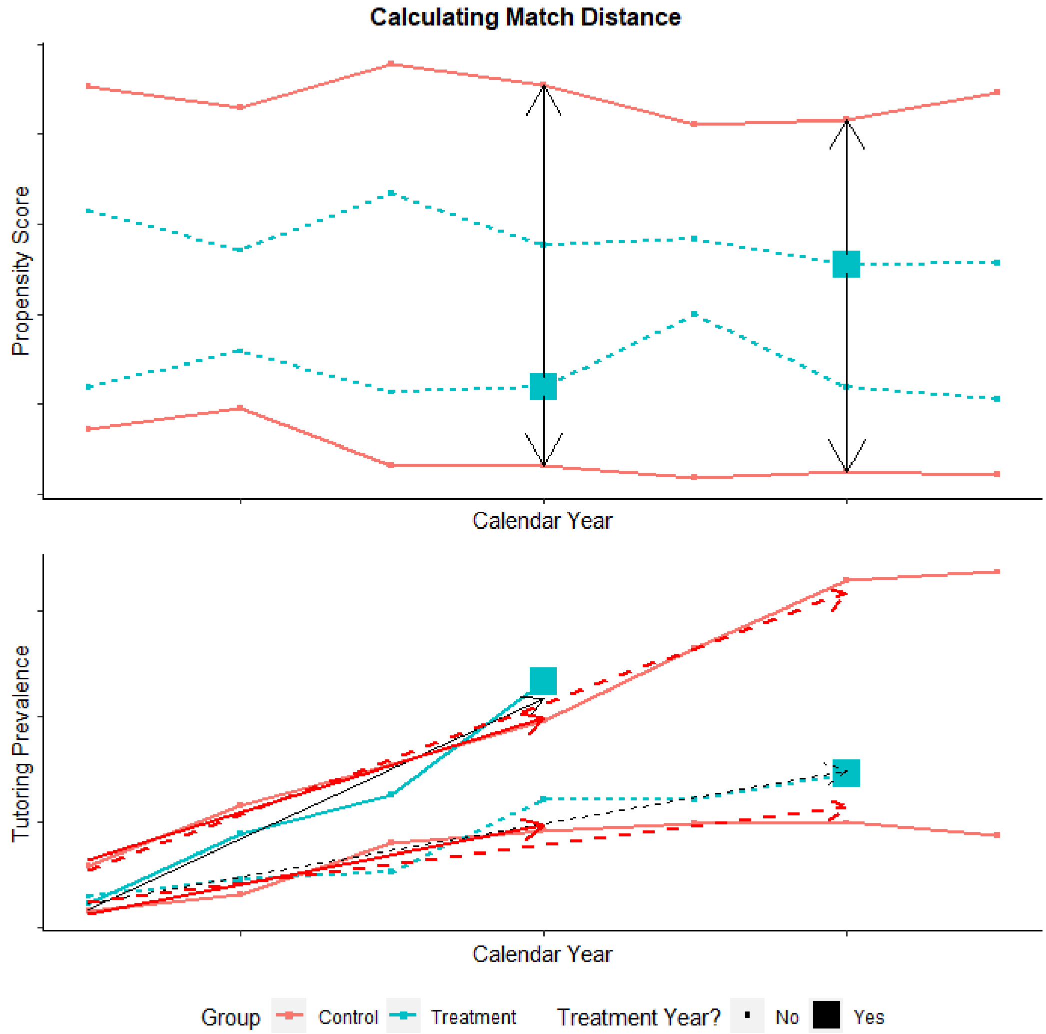

Figure C: Match Distance Between Two Treatment Units and One Control Unit

D Diminished Impact Zeroes

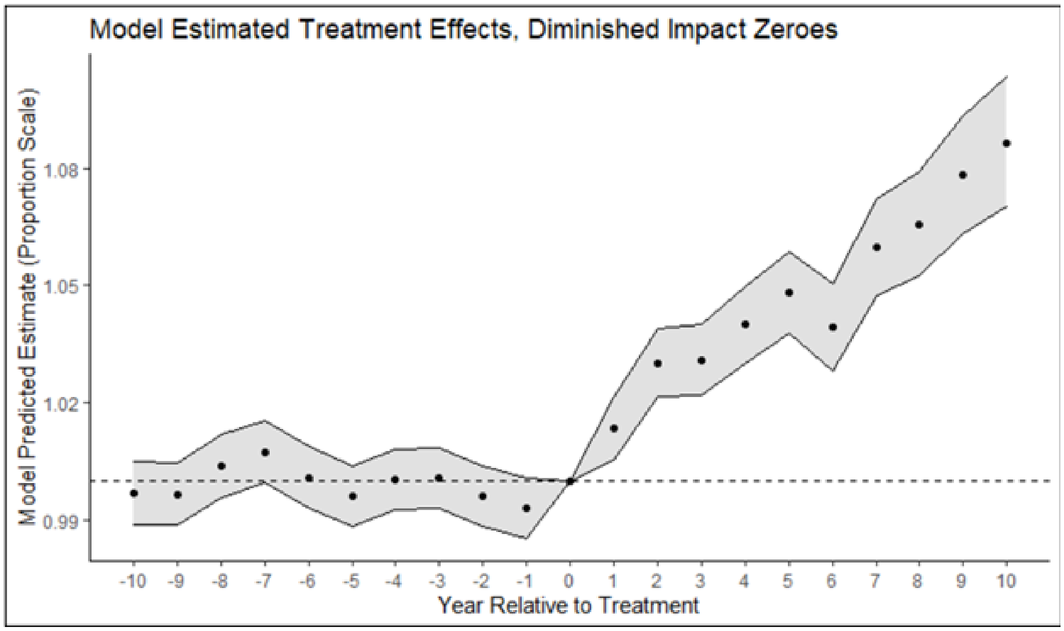

Figure D: Model Estimated Group Differences with Diminished Impact Zeroes.

Units with outcome values of 0 after the log transformation were recoded as having value 0.5, which reduces the magnitude of changes in outcome perceived by the model.

References

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

Baker, D. and LeTendre, G. K. (2005). *National differences, global similarities: World culture and the future of schooling*. Stanford University Press.

Bound, J., Hershbein, B., and Long, B. T. (2009). Playing the admissions game: Student reactions to increasing college competition. *Journal of Economic Perspectives*, 23(4):119–46.

Bray, M. (2003). *Adverse effects of private supplementary tutoring: Dimensions, implications and government responses*. UNESCO. Instituto Internacional de Planeamiento de la Educacion.

Bray, M. et al. (2010). Blurring boundaries: The growing visibility, evolving forms and complex implications of private supplementary tutoring. *Orbis scholae*, 4(2):61–73.

Bray, M. and Silova, I. (2006). The private tutoring phenomenon: International patterns and perspectives. *Education in a hidden marketplace: Monitoring of private tutoring*, pages 27–40.

Buchmann, C., Condron, D. J., and Roscigno, V. J. (2010). Shadow education, american style: Test preparation, the sat and college enrollment. *Social forces*, 89(2):435–461.

Buddin, R. and Zimmer, R. (2005). Student achievement in charter schools: A complex picture. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 24(2):351–371.

Chubb, J. E. and Moe, T. M. (1990). *Politics, markets, and America's schools*. Brookings Institution Press.

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. and Aurini, J. D. (2008). *School choice as concerted cultivation: The case of Canada*. na.

Daw, J. R. and Hatfield, L. A. (2018). Matching and regression to the mean in difference-in-differences analysis. *Health services research*, 53(6):4138–4156.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Hansen, B. B. and Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of computational and Graphical Statistics*, 15(3):609–627.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*, volume 25. Harvard university press.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236.

Ireson, J. (2004). Private tutoring: How prevalent and effective is it? *London Review of Education*.

Kim, E., Goodman, J., and West, M. R. (2021). Kumon in: The recent, rapid rise of private tutoring centers. *Brown University Annenberg EdWorkingPaper*, (21-367).

Kim, S. and Lee, J.-H. (2010). Private tutoring and demand for education in south korea. *Economic development and cultural change*, 58(2):259–296.

National Center for Education Statistics (2022). Public charter school enrollment.

Oberfield, Z. W. (2020). Parent engagement and satisfaction in public charter and district schools. *American Educational Research Journal*, 57(3):1083–1124.

Renzulli, L. A. and Roscigno, V. J. (2005). Charter school policy, implementation, and diffusion across the united states. *Sociology of Education*, 78(4):344–366.

Rhinesmith, E. (2017). A review of the research on parent satisfaction in private school choice programs. *Journal of School Choice*, 11(4):585–603.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Slungaard Mumma, K. (2022). The effect of charter school openings on traditional public schools in massachusetts and north carolina. *American Economic Journal: Economic Policy*, 14(2):445–74.

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