------ PRELIMINARY, DO NOT QUOTE, COMMENTS WELCOMED ------

Let 1,000 Flowers Bloom (or Wilt): Heterogeneity in National Market-Level Charter School Effects

Feng Chen Douglas N. Harris* Nick Lacoste

November, 2024

APPAM 2024 Conference, Baltimore, MD

Abstract: More than 1,000 U.S. school districts nationally have had at least one charter school within its boundaries. In prior work, using a generalized difference-in-differences (GDD) method, we found that increasing charter market share in these locations leads to small positive market-level effects on high school graduation and elementary/middle test scores. These reflect the net effects of various mechanisms—participant effects, various competitive effects, and matching effects. The present study focuses on *effect heterogeneity* of these market-level effects, using both GDD and machine learning (causal forests). We find considerable variation in these market-level effects across states and districts. Student demographic measures are, collectively, more important than measures related to market context, school funding, or other charter-related policies. Specifically, charter effects are generally larger in districts with more racial minorities and low-income students (especially when test scores are the outcome). School funding plays a complex role because market-level effects are generated by the responses of both charter schools and traditional public schools, whose finances are interconnected. Aside from the above, the results are inconsistent when we switch from test scores to high school graduation rates as the dependent variable. This may suggest that the forces making charter schools effective in raising test scores on average also make them ineffective in raising high school graduation.

Keywords: Charter schools; competition; high school graduation; student achievement
JEL Codes: H75, I21, I28

Acknowledgements: This research was carried out under the auspices of the National Center for Research on Education Access and Choice (REACH) based at Tulane University, which is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C100025 to The Administrators of the Tulane Educational Fund. The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, Tulane University, or any other organization. For useful comments, we thank Joshua Cowen, Ron Zimmer All remaining errors are those of the authors.

Author Information: Feng Chen is an assistant professor at the Li Anmin Institute of Economic Research at Liaoning University, China (fchen9@tulane.edu). Douglas Harris (corresponding author) is a professor and chair of economics at Tulane and director of the REACH center (dharri5@tulane.edu) 6823 St. Charles Ave. 208 Tilton Hall, New Orleans, LA 70118.

I. Introduction

Research consistently shows that charter schools have positive effects on the achievement of students who attend them (e.g., Sass 2006, Booker et al. 2007, Hoxby and Rockoff 2005, Abdulkadiroğlu et al. 2011, Angrist et al. 2016) as well as perhaps high school graduation (Chen & Harris, 2023).[1] However, as in other contexts, the average can mask what might be equally important variation. Charter school participant effects vary across states (CREDO, 2023), across cities (Angrist et al., 2016; Harris & Larsen, 2023; Chin et al., 2017) and across schools within states (Baude et al., 2014). The overall variation of charter effects is wider than traditional public schools, i.e., a large share of charter schools perform worse than the lowest-performing traditional public schools—and, conversely, the best charter schools are better than the highest-performing traditional public schools (Baude et al., 2020).

This wider variation is partially predictable. First, a key objective of charter school movement, from the beginning, was to create greater variety, or product differentiation (Glomm, Harris, & Lo, 2004; Gilraine et al., 2021). Variation was therefore part of the design. Second, and related, charter schooling relies more on market accountability and parents vary in their preferences for typically measured academic outcomes (Harris & Larsen, 2023; Beuermann et al., 2023), which amplifies that potential variety. Third, charter schools are less tightly regulated than traditional public schools. The rules imposed on traditional public schools (e.g., over personnel) might keep them from reaching their potential performance ceilings but also maintain

[1] See Cohodes and Roy (2023) for a review of lottery studies that include both test scores and high school graduation. Our conclusion here is very similar to their review. However, Cohodes and Roy (2023) report four lottery studies that have examined effects on high school graduation. Three of these have negative point estimates (Angrist et al. 2016; Cohodes & Feigenbaum, 2023; Angrist et al., 2023) and one shows positive effects (Dobbie & Fryer, 2011). However, this is a selected sub-sample of over-subscribed schools, whereas Chen and Harris (2023) find positive market-level effects when looking at the entire country.

a floor below which performance cannot fall.[2] Weakening these rules allows charter performance to spread out in both directions.

As more than one thousand school districts have had at least one charter school, there is much more to learn about how and why charter effects vary so wildly. Since the first charter school opened in Minnesota in 1990, charter schooling has grown so that seven percent of all school-age children in the U.S. now attend one, nearing the number of students in private schools. More recently, Republican-leaning states have also seen the expansion of universal education savings accounts or "super-vouchers." While we know less about the performance of the private schools where these funds are used, there is good reason to expect even wider variation in market-level effects from vouchers.[3] Understanding the variation in charter school effects therefore provides a window into school choice policies generally.

In this study, we attempt to better understand the variation in charter school effectiveness in a nationwide analysis of the "market-level" charter school effects. While some prior work has focused on variation in participant effects (CREDO, 2023; Baude et al., 2020), we consider variation in the full array of charter mechanisms: not only *participant effects* but also four other mechanisms: *near competitive effects* on the closest public schools (Hoxby 2003, Bettinger 2005, Bifulco and Ladd 2006, Sass 2006, Zimmer and Buddin 2009, Imberman 2011, Linick 2014, Cordes 2018, Ridley and Terrier 2022, Griffith 2019, Slungaard Mumma 2022), *far competitive effects* on other traditional public schools located within geographic school district (Hess 2004,

[2] There is also some evidence that rules about charter student admissions are not enforced and/or that charter schools do not follow these rules. For example, in an audit study, Bergman and McFarlin (2020) find that charter schools are less likely to respond to emails from fictitious parents indicating their children have disabilities. More generally, charter schools have at least 20 different ways to select students, often in indirect ways (Harris, 2020, Mommandi & Welner, 2021). This could make the variation in charter school performance look wider than it really is.
[3] The variation in private school effects is likely to be even wider because they face even less regulation and are allowed to charge tuition, which varies from an average of less than $10,000 per student to upwards of $150,000, signaling widely varying resource and outcome levels.

Holley, Egalite, and Lueken 2013; Figlio, Hart, and Karbownik 2022),[4] *closure effects* on low-performing traditional public schools (Engberg et al. 2012, Bross, Harris, and Liu 2023, Carlson and Lavertu 2016), and *match effects* (Abdulkadiroğlu, Agarwal, and Pathak 2017; Campos and Kearns 2022). When a charter school opens in a given geographic district, all of these mechanisms may be at work, although the participant effects seem to comprise about two-thirds of the total effect (Chen and Harris, 2023).[5]

Following Chen and Harris's (2023) analysis of average market-level treatment effects, we start with a generalized difference-in-differences (GDD) method and add interaction terms to understand effect heterogeneity. We are also among the first to compare effect heterogeneity from quasi-experimental methods with a machine learning approach, specifically causal forests (CF). This method was originally devised for effect heterogeneity analysis with randomized control trials (RCTs). The CF method avoids issues such as the multiple comparisons problem, selective reporting of interactions (i.e., "fishing"), and overly stringent parametric assumptions. Our analysis directly compares the two methods in the context of an important policy question and national data.

With both methods, the data come mainly from the National Longitudinal School Database (NLSD), which includes nearly all districts in the U.S. from school years 1995 to 2018. With both the GDD and CF, estimate the market-level effects of charter entry on a weighted average of outcomes of all TPS and charter schools located within each local market, i.e., the geographic school district. Both methods account for all time-invariant differences across

[4] There is very little evidence on far competitive effects. These citations are mainly case studies of individual districts. Also, the Figlio et al. citation refers to district responses to private schools.

[5] In addition to Chen and Harris (2023), a few prior studies have examined the total or market-level effects of charter schools (Gilraine, Petronijevic, and Singleton 2021, Ridley and Terrier 2022, Harris and Larsen 2023). Mehta (2017) and Ferreyra and Kosenok (2018) have also developed and estimated structural models of charter school entry, school input choices, and student school choices.

districts and assumes that there are no time-varying unobserved factors that are correlated with both changes in charter market share *and* changes in student outcomes.

Using the GDD method and the same data, Chen and Harris (2023) found that when charter market share increases by 10 percentage points, geographic district test scores increase by 0.01 standard deviations and high school graduation rates increase by 1-2 percentage points, on average. These analyses pass various endogeneity tests (e.g., placebos, parallel trends, and changes in observable demographics) suggesting that they reflect causal effects. Since these average treatment effects are based on more than a thousand treated districts, these data and methods allow ample opportunity to study how the effects vary.

A general challenge with effect heterogeneity analysis is that there are so many possible dimensions that some are bound to show up as statistically significant by chance—the multiple comparisons problem. For this reason, theory places a key role in narrowing down the potential options and avoiding spurious relationships. The next section therefore outlines theories regarding why the variation in effects across units might be related to characteristics/traits of those units, such as demographics, teacher supply, and district/state policies.

In our preliminary results, we find suggestive evidence that charter effects are more positive for low-income students and perhaps racial/ethnic minorities. We also find more positive effects in districts with higher initial test score levels. Prior research has also found suggestive evidence that charter effects are especially large in cities that are attractive to young people and prospective teachers, e.g., Boston Angrist et al. (2016) and New Orleans (Harris & Larsen, 2023). When we look at this nationally, however, we see no evidence that charter effects are related to the general attractiveness of cities, at least not after we control for average local income. Also, we see no evidence that geographic districts with university-based schools of education, another source of potential educator supply, are related to charter market-level effects, ceteris paribus.

We also study how charter effects vary by district and state policies: school funding, charter accountability, and other policies. We examine school funding in multiple ways, starting with direct measures of spending by both charter schools and traditional public schools (TPS).[6] As explained in the theory section, we include TPS funding because we are estimating market-level effects that depend partially on TPS competitive responses, which in turn may depend on their funding. Specifically, we consider the TPS-charter funding gap and overall spending per-pupil, averaged across TPS and charter schools. The market-level effects are larger where the local TPS-charter funding gap is larger. This might seem surprising but could reflect the formulaic connection between the TPS-charter funding gap and the effect that charter entry has on TPS funding.

Our measures of school spending are subject to considerable error (e.g., some funds pass through school districts on their way to charter schools, but it is difficult to know how they are being accounted for). For that reason, we also study school funding using "charter funding equity" ratings from industry groups that advocate on behalf of charter schools. These groups are well acquainted with how funding actually flows, in ways that our data might not reflect. The results provide suggestive evidence that charter effects are larger when states give charter schools more funding, relative to TPS.

We use the same industry group policy ratings to study other aspects of charter school policies. The results also provide suggestive evidence that charter effects are larger when states have weaker rules around the transparency of charter applications and renewal decisions.

[6] We use the term "traditional public schools" (TPS) because charter schools, legally speaking are also public schools or at least have ambiguous public status.

However, in general, non-funding policy seems to play a small role in explaining market-level effects.

The above analyses have practical implications for policy, e.g., how all schools are funded and how charter schools are held accountable. They also have implications for authorizers who have the ability to direct charter schools to particular locations where they are likely to do the most good. Another important policy topic is whether there is some optimal charter market share. These shares range from 0 percent, in most districts to a national average of 7 percent, to an extreme of essentially 100 percent in New Orleans and elsewhere. However, we find almost no evidence of diminishing marginal returns to charter market share, except possibly in high school graduation rates. Also, we theorize that charter effects might be smaller in districts where enrollment is growing quickly versus those where it is declining, but we see no evidence of this either.

The paper proceeds as follows. Section II outlines a theory of effect heterogeneity for charter schools. Section III introduces the data, drawing heavily from the Chen and Harris (2020). Section IV discusses the methods and identification, for both the GDD and CF methods. The results are shown in section V. Section VI concludes.

II. Theory

In parametric econometric methods, theory is important in effect heterogeneity analysis because there are usually large number of potential dimensions that can be studied and the odds of finding some statistically significant patterns by chance are high (the multiple comparisons problem). That is, it is easy to "fish" for some statistically significant pattern. Theory makes this less likely as it directs analysis to dimensions where we have good reasons to expect patterns.

Most of the empirical literature on the effect heterogeneity of charter schools has focused on dimensions that are commonly used in social science, but which are not firmly grounded in theory. In this section, we draw on the economic theory of the consumer and the firm, modified to address unusual features of the schooling market.

We avoid some of these problems by using the CF method but theory is still important. Spurious findings are still possible with this non-parametric method. Also, theory is helpful for interpreting many of the results, especially as we propose several new theories about why charter effects might vary.

II.A. Basic Model

We begin with certain economic assumptions that apply to all households and schools. Regarding the consumer, suppose that households maximize utility, which includes consumption and the well-being of their children (e.g., schooling outcomes). All households prefer higher levels of academic quality q (i.e., schools with larger participant effects) but vary in their tastes for a single other output h (e.g., religious instruction or sports).[7] They also vary in their incomes, which are positively correlated with the baseline achievement of their children.

Regarding the producer, suppose that each school can be placed in the product space $\{q, h\}$. Each of these two outputs has a single known production function. We assume, consistent with empirical data, that schools are run by non-profit organizations and, further, that these non-profits maximize revenue, all of which comes from state governments.[8] This implies that the price of schooling to consumers, both TPS and charter schools, is zero.

[7] This approach and notation follows Glomm, Harris, and Lo (2004).
[8] We ignore the tax structure that generates this revenue and just assume for simplicity a lump sum tax.

In the first period, we have only traditional public schools (TPS-only) that are run by the local school board. Households sort themselves, in a Tiebout (1956) sense, into TPS-only districts based on how well their tastes match the schooling characteristics. TPSs vary on q because of: (a) variation in spending, which comes from taxation on local households; and (b) random shocks in the efficiency of resource use. Output h is constant and normalized to $h=0$ (e.g., TPS cannot teach religion or they all provide sports). Since all households prefer higher q, and it is a normal good, households sort by income, allowing higher-income households to obtain higher q.

In the second period, some states adopt charter school laws that allow charter schools to open. Unlike TPS, charter schools can locate anywhere in the product space, including $h>0$. Similar to Glomm, Harris, and Lo (2004), charter schools choose to locate in districts where q is low and where the relative preference for h is high. Charter schools vary on q for the same reasons at TPSs. We call the subset of local districts in which charter schools locate in this second period the TPS+C districts.[9] TPSs might respond to this competition by improving q; this could occur in any period depending on how TPSs perceive the threat of charter schools.

The above discussion implies that there are four relevant institutional units: schools nested within districts nested within states, as well as (non-nested) authorizers. Whether prospective charter schools open, and where, is generally determined by a matching process. Schools have desired (revenue-maximizing) districts and states and desired authorizers. Unlike states and districts, authorizers are two-sided matches—schools and authorizers are choosing one

[9] For simplicity, we assume there are no private schools.

another.[10] This will become important when we interpret the empirical variation in effectiveness across units.

We are ultimately interested in understanding the effects, and effect heterogeneity, on whole markets when charter schools enter local markets, though this process is dynamic in ways that make it difficult to determine where charter entry is most efficacious. The entry of charter schools may result in some resorting of households across districts. But resorting may be minimal, especially in the short run, because: (a) many charter schools do not require students to live in the local district the way TPSs do;[11] (b) households are risk-averse and have limited information about the characteristics of new schools when they first open; (c) most households with children already have their children in TPS and there is inertia with respect to school moves; and (d) frictions in this housing market make such moves costly.[12]

II.B. Effect Heterogeneity on Quality

In general, and in the present study, we can only measure charter school entry effects on q. We know much less about where charter schools sit on the h dimension of the product space. So, the main question this section is meant to inform is, how might the effects of charter school entry and expansion vary on q? In this section, we reframe the above discussion to distinguish the various levels of the charter school system into a generic hierarchical structure that leads to the econometric framework discussed later. We also develop the model to facilitate the design of our empirical framework.

II.B.1. General Framework

[10] Many CMOs decide to open new charter schools while others decide whether to absorb existing charter schools. Also, some authorizers are districts, which connects the location decision to the authorizer decision.
[11] Households might decide to move to a lower-cost housing area and avoid paying for higher TPS quality, if they do not plan to utilize TPS, but the fixed costs of moving would make this costly.
[12] Resorting across districts where h-preferring households who did not initially locate in the TPS-C districts in t=1 do so in t=2; and non-h-preferring households in TPS+C districts move to TPS-Only (in t=2).

Since the total effect of charter schools comes mainly from the participant effect (Chen & Harris, 2023), we start by asking, where do high-quality charter schools come from? First, we assume that (potential) charter schools have a locational choice set. As suggested above, these potential schools decide where they wish to locate based on district characteristics. These characteristics might include factors such as district demographics that make them most likely to be successful in maximizing revenue. We label these location-based factors as X_l^*.

These are not the only factors affecting the market-level effect. The above locational preferences become authorizers' choice sets. Also, the effects depend on how TPS respond to charter competition. We might expect high-quality charter schools to locate in districts with lower-quality TPS, where charters can more easily compete and attract enrollments from lower-performing schools.[13]

We denote the market-level effect of charter school entry as q_m^* and, based on the above discussion, this involves multiple mechanisms such that $q_m^* = q_p^* + q_c^* + q_f^*$. In other words, the charter effect is comprised of the participant effect (q_p^*), the various competitive responses (q_c^*), and the improved fit between student needs and school services (q_f^*). For simplicity, we initially assume that the components are additively separable.

The participant effect is shaped by three main factors: the ability of authorizers to select high-quality charter schools from that choice set (Bross & Harris, 2023), other supports that authorizers might provide to facilitate school improvement, and the resources available to charter schools (Jackson & Mackevicius, 2024; Harris & McKenzie, 2023). These, in turn, are shaped

[13] It is not clear whether low-performing TPS would be best able to respond to such competition. If not, this may create a trade-off. Shifting enrollment from low-performing schools is helpful, but less between the competitive effect and the closure effect.

by state policies, which we consider to be exogenous in our analysis and are therefore labeled X_p^*.

Recall that there are three types of competitive responses. near effects, far effects, and closures. The competitive effects may be shaped by a variety of factors: local interest group politics, enrollment trends, school and district leadership, and parent responses.

State policy is obviously important here and that is no more true than with statewide charter authorizer policies. In most situations, school districts have little incentive to authorize charter schools because this creates competition with the district's existing TPS. It is therefore no surprise that the presence of an independent charter authorizer (i.e., an authorizer other than the school district itself) increases the number of charter schools that open (Harris & McKenzie, 2023). There is also some, more correlational evidence that independent authorizers tend to select charter applications that turn out to have higher value-added (Bross, Harris, & Liu, 2023; Harris & McKenzie, 2023). The presence of an independent authorizer might also change the way that districts authorize. For example, if a district believes that the state will authorize a given charter school, then the district might be more likely to recruit and/or authorize that same school itself, so that it can maintain control over it. This is an example of a far competitive response.

We cannot observe each mechanism q_p^*, q_c^*, and q_f^* directly. Instead, we focus on observable factors that may affect these mechanisms. These include state education policies (X_p^*), e.g., school funding and independent authorizers) and another factor not yet discussed: the local market context (X_l^*). There is some evidence that charter schools might have a greater impact in districts where the performance of TPS is low to start with (Chen & Harris, 2023) and where entering charter schools have missions that align with TPS, so that there is more direct competition.

The above framework is intended to set up an empirical model that we can estimate with the available data. Later, we start our analysis by estimating the effect of charter entry in each market. We note that these estimates will contain measurement error such that $q_m = q_m^* + \varepsilon_m$. Substituting from above, this implies further that $q_m = X_p + X_l + \varepsilon_m$, which we can empirically estimate. This does not allow us to isolate the participant, competitive, and fit responses, which have been examined elsewhere using other methods (Chen & Harris, 2023; Chin et al. 2019; Harris & Liu, 2019). However, as we will see in the next two sections, this framework does lead to hypotheses that inform policy design. We discuss theory pertaining to two broad categories of traits. The location/household characteristics include TPS quality, household income, and taste/demographic diversity. The policy characteristics include school spending, state charter policies, teacher supply, and teacher unionization.

II.B.2. Effect Heterogeneity by Location (Non-Policy) Characteristics

School Quality. As noted above, households vary in their children's baseline achievement, household incomes, and tastes for *h*. By definition, baseline achievement refers to achievement prior to school entry (i.e., it is not an outcome of current *q*). It could be that "skill begets skill" so that otherwise identical high-achieving students benefit more from higher charter-induced *q*. On the other hand, when *q* is high to begin with, it is less likely that newly entering charter schools have higher *q* than the market (or just nearby schools), which would imply the opposite relationship. Also, we would expect that high-*q* locations are closer to the production frontier and have room for improvement.

Parent Income. We might expect that local markets with higher household income would see small improvements in *q* because they have already sorted into districts with high TPS *q*. On the other hand, charter entry might mean that the higher demand for *q* among higher-income households leads to larger competitive pressures that increase *q* more. Which of these effects dominates depends on the underlying causes of variation in *q* among TPSs (and charter schools). In practice, it seems that charter participant effects (Cohodes & Parham, 2021) and market-level effects (Chen & Harris, 2023) are larger in lower-income areas.

Demographic Diversity. It is not obvious that charter schools should have any particular effect on different racial groups except insofar as race is correlated with other income and baseline achievement. However, we do know that charter schools tend to locate in areas that are more diverse on income and race (Glomm, Harris, & Lo, 2004). That locational preference is likely driven by demand and potential revenue generation, especially if income/race diversity is a proxy for tastes and there is pent-up demand for alternatives, given TPS tendency toward uniformity.

II.B.3. Effect Heterogeneity by Teacher Variables

Teacher Supply. Prior research suggests that charter schools operate under different production functions than TPSs; in particular, they rely more heavily on management in general, and business-oriented management in particular (Harris, 2020); they also hire younger teachers who are less likely to have traditional teaching credentials. This means that, in any given location, TPS and charter schools may have unequal access to their core inputs. They are looking for educators with different backgrounds. This is a central issue given that education is a service where the core input is almost entirely labor and the apparent importance of teacher quality in the education production function (e.g., Chetty, Friedman & Rockoff, 2014a).

Teacher Union Power. Teacher unions have the power to block certain changes, including the entry of charter schools. They also seem to influence TPS quality, positively by increasing school spending and negatively in other ways (Marianno, forthcoming). But we are controlling for both of these factors—school spending and TPS quality—in the analysis, so our main hypothesis is that unions block charter schools that are likely to attract many students, meaning that the charter effects we do observe will be smaller and possibly negative. (This is still in progress.)

II.B.4. Effect Heterogeneity by School Spending

State government policy toward charter schools can vary on three main dimensions: funding, accountability, and regulation. Empirically, charter schools receive less government revenue per-pupil and are subject to less regulation (e.g., charter teachers are generally not organized under collective bargaining). However, the extent of these differences can vary across states (and across authorizers within states). This is important, given evidence that increased school funding usually improves student outcomes, at least for low-income students. We discuss each of these below in turn.

Charter School Funding. School funding typically improves school quality (Jackson & Mackevicius, 2024), so we (initially) expect the charter participant effects, and therefore the market-level effects, to be larger when charter school funding is higher. But the issue is more complex than that. The effect of charter schools on market-level outcomes depends mostly on the *relationship* between charter spending and TPS spending. Where there is a large charter spending gap, it will be more difficult for charter schools to be more effective than TPS and attract students.

TPS School Funding Sensitivity. Even supposing a TPS-charter funding gap of zero, there is a second issue: When charter schools enter and shift enrollment from TPS, what happens to TPS revenue and what does this mean for how TPS respond to charter entry, i.e., competitive effects? TPS likely have a greater ability to respond when there are hold harmless or other provisions that confine funding losses from enrollment declines; however, those same provisions also reduce the incentive for TPSs to respond. This trade-off of input capacity and pressure makes it unclear whether TPS enrollment/funding sensitivity should be positively or negatively related to student outcomes. We also note that TPS funding sensitivity to charter entry is likely positively correlated with the charter funding levels discussed above, which shape the participant effect. If TPS lose considerable funding, then charter schools probably gain more of that funding.[14]

Average spending across TPS and charter schools. Another way to look spending is in terms of total spending across all schools (per pupil). That is, we can measure total spending across all schools, charter plus TPS, and divide by the total number of students across those schools.[15] To the extent that TPS are funded by local property taxes that they are not required to turn over to charter schools based on enrollment shifts, TPS funding per-pupil will increase. The overall average funding per pupil (TPS plus charter) would also increase unless charter spending per pupil is very low (below what TPSs lose). At the other extreme, if all of the funding shifts from TPS to charter schools based on enrollment, then average per-pupil funding would be unchanged. A key implication of this is that funding per-pupil is likely to increase and unlikely to drop. This could be a partial explanation for the overall positive charter treatment effect (Chen & Harris, 2023).

[14] We expect that two main elements of state policy will affect these measures of TPS school spending. The first is the share of funding coming from states, which is generally on a formula-based per-pupil allotment. The second is hold harmless provisions that protect TPS from losses in the share of funding coming from the state. In states with 100-percent local property tax funding, TPS have little to lose in the short run and we can expect the competitive effects to be more limited. Again, the direction of this effect is unclear because of the possible trade-off between capacity and pressure. Also, note that increased charter school spending (see prior sub-section) will tend to go along with larger drops in TPS funding when students shift to charter schools. Our analysis will control for charter spending, but there could also be multicollinearity issues that make it difficult to separate the two.

[15] For reasons explained later, we can measure this more accurately.

For the reasons above, it is still not clear whether higher average spending should be associated with better outcomes because more spending in total could mean that TPS do not have to respond. However, we can learn something about the net effect of total spending by estimating the effects *on* total spending and spending by sector. If the effect of total spending is null, for example, then we might be able learn whether this is because TPS lost minimal funds and did not respond.

In the analysis, we estimate the effect heterogeneity by simple charter spending, charter funding gap and charter and TPS funding level. We also estimate the average effect of charter entry on total spending per pupil to understand the mechanisms at work.

II.B.5. Effect Heterogeneity by Other (Non-Funding) State Policies

Charter effects on schooling markets might depend not only on how they are funded but how they are regulated, especially through state law. It is often difficult to capture regulation, but, in this case, we have data from three industry/advocacy groups: The Center for Education Reform (CER), the National Alliance for Public Charter Schools (NAPCS), and the National Alliance of Charter School Authorizers (NACSA). These groups rate and rank states on the availability of authorizers other than school districts, whether charter schools are automatically non-renewed for low-performance, the degree to which charter schools are exempt from the rules facing TPS. Two of the groups also rate states based on their charter funding equity, which is similar to the TPS-charter funding gap described above. See the data section for more on these measures.

II.B.6. Other Forms of Effect Heterogeneity

Effect Heterogeneity by School Types. The above discussion pointed to two key differences in charter schools: whether they are managed by CMOs (versus standalone schools) and which type of charter authorizer they operate under. Based on prior evidence, we would expect that the entry of charter schools would be more positive when they are managed by CMOs.

We might also expect larger effects when the authorizer is not the district itself. School districts are unlikely to make a competitive response to schools under their own control. On the other hand, the participant effects of district-authorized schools are less well known and could potentially offset any reduced competitive response.

Diminishing Marginal Returns. We have written generally about the direction of charter school effects, which might imply a linear relationship between charter market share and q. However, there are reasons to expect larger effects with early charter entrants and diminishing effects and charter market share grows. First, if authorizers are picking from menu of potential charter entrants, and they have some accurate signals of future quality then the initial charter schools may be of higher quality than subsequent entrants (Bross & Harris, 2017).[16] Second, and similarly, the TPS competitive response seems more likely to arise with initial entrants as this is when competition first becomes an issue and is most salient. When the first charter school enters, this is likely to draw newspaper headlines and discussion about the emergence of charter schools; this type of reaction seems less likely with the second and third charter schools.

Effect Heterogeneity by Enrollment Growth. We noted above how competitive responses may vary depending on funding formulas and authorizer type. District enrollment growth is another possibility. TPSs have more difficulty expanding when enrollment growth is high. While they are required to serve all students within their boundaries, they are less flexible, e.g., their

[16] A potential counterargument here is that, as the charter sector gains traction, more supporting structures may emerge to incubate better charter schools. This occurred in New Orleans for example, where a non-profit began incubating charter schools once it became clear that the city was going to be opening more such schools.

buildings have to meet fairly specific standards and they are inclined to open in shopping malls the way that some charter schools do. Anecdotally, California school districts did not generally oppose the entry of charter schools in the 1990s and 2000s when enrollments were growing rapidly, but they have more recently as charter shares have grown and enrollments began declining. This might also imply that districts with more enrollment growth experience a smaller competitive response. In addition, authorizers might be less stringent in accepting applications from charter operators, out of desperation to provide additional capacity. Both factors suggest that we should expect smaller effects of charter entry in districts where enrollment growth is high (though the very fact of enrollment growth also leads to questions about changes in demographics that could bias our estimates).

However, our approach also has more direct implications for policy. We will learn whether various state policies are associated with improved charter market effects and whether new policies might be able to better target charter schools to those contexts where they do the most good. Our analysis also evaluates the performance of charter schools in particular states and of particular charter school authorizers. In the latter case, this is only possible with independent authorizers (in the states that have them) as the analysis requires isolating the effects of charter authorization and charter school improvement efforts from other district behaviors.

III. Data and Descriptive Statistics

III.A. Geographic Districts and their Outcomes

This paper uses data from the National Longitudinal School Database (NLSD), which contains data for a near-census of all TPSs, charter schools, and private schools in the U.S. from 1991 to the most recent academic year.[17] The NLSD merges school and district-level data from Common Core of Data (CCD), Stanford Education Data Archive (SEDA), Census Small Area Income and Poverty Estimates (SAIPE), and other sources. We are most interested in the student outcomes, test scores and graduation rates, but also include a vector of district-level covariates in some models. The street addresses of charter school also allow us to place the location of each charter school within its geographic school district.[18] These districts define the scope of the market, i.e., which schools are assumed to compete with one another.

The NLSD implicitly sets the "market" boundaries as equal to those of the traditional public school district; that is, in terms of the geography of the market, the subscript $m = d$. (This is partly to align the estimated market-level effect with the district competitive effects.) However, $m \neq d$ in the sense that the geographic district does not make decisions in the way that the institutional district does.

The specific dependent variables are the average freshmen graduation rate (AFGR) (hereafter, graduation rate) and student math and English Language Arts (ELA) test scores. The AFGR uses aggregate student enrollment data to estimate the size of an incoming freshman class and aggregate counts of the number of diplomas awarded four years later. For example, the AFGR for a school year in 2006 is the total number of diploma recipients in 2006 is divided by the average enrollment of the 8th-grade class in 2002, the 9th-grade class in 2003, and the 10th-

[17] All the school years mentioned in this paper are spring school years unless specifically stated otherwise.
[18] The SEDA team uses the 2019 Elementary and Unified School District Boundaries https://nces.ed.gov/programs/edge/Geographic/DistrictBoundaries) to define the Geographic School District for test score sample (except for special education or virtual schools). We use the same boundaries to define the Geographic School District for the graduation sample.

grade class in 2004. The graduation rate sample includes schools covering grades 9-12 annually from 1995 to 2010.[19]

The test scores are available in 3rd through 8th grade in math and ELA over the 2009-2018 school years from the Stanford Education Data Archive (SEDA). The SEDA provides nationally comparable, publicly available test score data for U.S. public school districts (Ho 2020).[20] For interpretation purposes, we normalize the test scores to have means of zero and standard deviations of unity within the grade, year, and subject.

Treatment status is defined as the charter enrollment share, or the percentage of students attending publicly funded schools who are enrolled in charter schools. These market share measures are created separately by grade level under the theory that charter schools only affect the geographic district outcomes in the specific grades being served (Jinnai 2014, Slungaard Mumma 2022).

For both outcomes, data are aggregated to the geographic district level by averaging TPS outcomes with those of charter schools located within their boundaries[21] (weighted by enrollment share[22]). This is crucial to the analysis as it allows us to capture charter effects operating through

[19] We noticed some clear errors in the AFGR data, e.g., wild swings from year to year. In these cases, we imputed the errant observations using linear interpolation. (When the errant observation was in the first or last year of the panel, we imputed using the adjacent year.)

[20] To make estimates are comparable across states, grades, and years. The SEDA research team took the following steps: (1) estimate the location of each state's proficiency "thresholds" in the distribution of scores; (2) place the proficiency thresholds on the same scale using the National Assessment of Educational Progress (NAEP), a test taken by a representative sample of students in each state; (3) estimate the mean test scores in each school, district, county, metropolitan statistical area, commuting zone, and state from the raw data and the threshold estimates, and (4) create estimates of average scale scores and achievement growth measures. See details in SEDA website https://edopportunity.org/methods/. While this method is of course based on assumptions (e.g., that the NAEP sample is truly state representative and the distribution of state scores is the same as the NAEP distribution despite the differing content of these various tests), this could only result in bias if test error in charter-entering districts differs from never-charter districts (within grades and states and conditional on other included covariates) and changes over time in conjunction with charter school entry. We could not come up with potential examples where this might be the case.

[21] SEDA provides test score data at the geographic district level. We converted the graduation rate data to the geographic district level ourselves. Please refer to Appendix B for the description that how the sample was created.

[22] Weighting is necessary to obtain a nationally representative effect (e.g., to avoid counting New York City the same as small rural districts). However, as Solon, Haider, and Wooldridge (2015) note, weighting does not

all of the various market mechanisms. In contrast, prior studies of charter schools have focused on the outcomes of individual schools in the examination of school participant effects or the competitive effects of individual charter schools on nearby TPS. Here, we are interested in the total effect of all the mechanisms, which means we need to account for the outcomes of all traditional *and* charter schools. Aggregating outcomes to the geographic district level also allows us to sidestep the main problem with prior studies, i.e., selection bias from students moving across schools (within districts).

The test score (graduation) data are available for 11,399 (10,658) districts, including 1,597 (1,073) districts that have at least one charter school. The final samples for the test scores (graduation) analyses include 97 (98) percent of the nation's publicly funded schools and 99 (97) percent of nation's charter schools.[23] Table 1 presents the summary statistics for the outcome and control variables. Compared with TPSs, charter schools nationwide tend to enroll a larger proportion of African American and Hispanic students. Charter schools also more likely to be in urban districts and where achievement is relatively low.

We are identifying total charter effects leveraging the timing of charter entry. One key factor affecting this is the timing of the passage of charter laws that allow them to entry. The first law allowing the establishment of public charter schools was passed in Minnesota in 1991. As of fall 2022, charter school legislation had been passed in 45 states and the District of Columbia.[24] As a result of the growing number of states allowing charter schools, the percentage of publicly

necessarily yield a consistent estimate of the population average treatment effect if the covariate variance depends on the weight. They therefore advise reporting results weighted and unweighted; and testing for effect heterogeneity by weight. We provide these additional results and discussion later as well.

[23] Since geographic districts are the unit of analysis, an observation is only missing if all the schools in the unit-by-school-type are missing. There are likely some missing schools within districts. Appendix B describes in detail how we compose the samples.

[24] The states in which public charter school legislation had not been passed by that time were Montana, Nebraska, North Dakota, South Dakota, and Vermont.

funded schools that are charter schools increased from 0 to 7.3 percent, and their percentage of enrollment increased from 0 to 6.2 percent.[25]

To highlight the variation in charter market share we are exploiting, Figure 2 lists the trajectory of charter school growth from the point of the first charter entrant. We report separate lines based on the maximum charter (enrollment) market share that appears in our data. As expected, growth rates are faster in geographic districts that have higher eventual market shares and diminish over time.[26]

III.B. State Policy Data

We gathered state ratings of charter school policies published in 2014 by the Center for Education Reform (CER), the National Alliance for Public Charter School (NAPCS), and the National Association of Charter School Authorizers (NACSA). These rankings are useful for understanding regulation because these organizations are well-informed about charter laws and are likely to focus their rankings on policies that are genuinely important, at least to their charter constituents. While we do provide evidence about the relationship between each group's overall ranking and the various market outcomes, our main interest is in the information the rankings provide about the components of charter regulation. The proceeding sections provide a short description of each composite index and the index components that we included for our study.[27]

Center for Education Reform (CER). CER ranks states based on "laws that have a strong, permanent authorizing structures, equitable funding codified in law, and autonomy across state, district, and teacher rules and regulations, giving charters the freedom to do what they do best."

[25] Appendix Figure A1 presents the trends in charter school share and charter enrollment share from spring 1991 to spring 2018.
[26] This non-linear pattern partially reflects the opening of the first charter school followed by enrollment growth in those schools, not just opening more schools.
[27] We selected the highest weighted components from each sub-index that are likely to be relevant for school performance, openings, and closures.

The organization's three most heavily weighted criteria are: the presence of independent (alternate) authorizers, number of schools allowed, and 100% charter funding (similar to the above "equitable funding"). The number of schools refers to enrollment caps and whether states are approving charter schools on a regular basis.[28]

National Alliance for Public Charter School (NAPCS). NAPCS evaluates state charter laws based on charter school laws that promote the creation of "high-quality charter schools while holding underperforming schools and authorizers accountable."[29] The five most heavily weighted components used for our study include state ratings by performance-based contract required, clear processes for renewal/nonrenewal, automatic exemption of charter schools from laws applied to traditional public schools (TPS), no caps on the growth of public charter schools, and equitable operational funding. The performance-based contract requirement component provides higher ratings to state laws that have requirements for contracts that provide academic performance expectations, operational performance expectations, and school and authorizer rights and duties. The process for renewal component examines whether there are "clear processes" for renewal, nonrenewal, and revocation decisions, including school closure and dissolution procedures. The no charter caps category determines whether there are caps on growth for public charter schools in a state. Equitable operational funding compares general operational, transportation, and other categorical funding of charter schools to TPS.

National Alliance for Public Charter School (NACSA). NACSA's policy ratings emphasizes policies that "facilitate the development of successful charter schools and enhance accountability for schools and authorizers alike."[30] The three most heavily weighted components

[28] We note that this almost guarantees a relationship between this criterion and charter market share in the empirical analysis that follows.
[29] https://www.publiccharters.org/publications/model-law-supporting-high-quality-charter-public-schools
[30] http://www.qualitycharters.org/wp-content/uploads/2015/08/State-Policy-Analysis.pdf

include a rating for alternate authorizer, renewal standard, and default closure. The alternate authorizer ranks states based on whether they include authorizers other than school districts and whether the state can sanction poor authorizers. The renewal standard component ranks states by whether they require a strong standard that hold schools accountable for performance. A default closure policy standard means that there are minimum requirements for renewal and/or provisions for closure in extreme conditions prior to contract expiration.

Although all three industry group rankings are clearly interested in expanding charter schooling, they differ in important ways. First, NACSA, as an organization of charter authorizers, is more focused on state policies that directly pertain to the authorizer roles, whereas NAPCS and CER rankings are based on a broader range of factors (e.g., funding). Second, NACSA is more focused on government accountability, while CER focuses on market accountability—“autonomy” and “freedom” for charter schools. (The NAPCS ranking is somewhere in between the two.) The NACSA approach is noteworthy given that the group represents the government-designated organizations—authorizers—who are responsible for holding schools accountable.

IV. Identification Strategies

We use two methods to study effect heterogeneity, the GDD and CF. Each of these is described below, followed by a brief discussion of the differences between them.

IV.A. Generalized Difference-in-Differences (GDD)

The GDD identifies effects from within-district variation in dosage over time (for those ever treated) as well as differences between treated and untreated units (Mouganie, Ajeeb, and Hoekstra 2020, Dow et al. 2020). This is preferable to a simple DD, in cases such as this with continuous, time-varying treatment. A simple DD would ignore most of the variation in charter market share and focus only on the point where the first charter school enters and treatment begins.

The dependent variables in our GDD analysis are test scores ($Test_{igt}$) and high school graduation (GR_{it}) in district *i* and grade *g* during year *t* (test scores are grade-specific; graduation is not). We include geographic unit fixed effects μ_i (usually geographic school districts) as well as state-grade-year fixed effects λ_{sgt} and district-specific linear trends $Year_i$. This implies the following models:

$$GR_{it} = \alpha + \beta_1(ChartShare_{it} \cdot H_{is}) + X_{it}\gamma + \mu_i + \lambda_{st} + Year_i + \varepsilon_{it} \tag{1}$$

$$Test_{igt} = \alpha + \beta_2\left(ChartShare_{igt} \cdot H_{is}\right) + X_{it}\gamma + \mu_i + \lambda_{sgt} + Year_i + \varepsilon_{igt} \tag{2}$$

Treatment is the continuous, time-varying charter market share, $ChartShare_{it}$. This is interacted with effect heterogeneity dimension H_{is}. None of these dimensions is time-varying. For this reason, and because of the district fixed effect, we do not enter H_{is} as a separate variable. The coefficients of interest are represented by $\boldsymbol{\beta}$.

A potential challenge with so many different dimensions of effect heterogeneity (see the theory in section II) is that the above factors may be correlated with each other. For example, state policies may be correlated with household demographics. Also, states with more accountability might also have distinctive funding formulas. An advantage of this analysis is that we can measure and include essentially all of these factors at the same time to isolate the role of each.

The term X_{it} is a vector of time-varying student covariates (log enrollment, share of black, white, Hispanic, FRL, and special education students) and other district-level covariates (district revenue per student, district expenditure per student, student-teacher ratio, teacher salary, number of magnet schools, and an indicator of whether a charter law has gone into effect). As these covariates are potentially endogenous, they are only included in certain specifications and as partial tests for charter effects on the student population.

The error term ε_{igt} is clustered at the district level in the main specification. (We report results clustered at the state level in the Appendix, which only slightly reduces precision.) The estimates are weighted by the district size, so that the estimates are nationally representative; specifically, we weight by graduation rate denominator in the high school sample and by grade-level enrollment for math and ELA. In our baseline analyses, we estimate the GDD model for both the current year charter enrollment share and with a one-year lag to allow delayed effects.

Our method accounts for time-invariant unobserved differences across districts and our main identifying assumption is that there are no unobserved time-varying factors that are correlated with changes charter market share *and* changes in student outcomes. We note that, even if charter market share changed based on time-invariant unobserved district characteristics that affected student outcomes, this would not introduce bias if the timing of charter entry were conditionally random, which is possible given all the steps involved in opening a charter school (e.g., putting in an application, getting it approved, gaining access to a building, recruiting students, and hiring staff).

Still, it is worth considering specific scenarios under which this might be violated. First, charter schools are more likely to locate in districts with low contemporaneous student outcomes (Glomm, Harris, and Lo 2005), signaling that we also might expect charter schools to also locate where *expected* future TPS performance is lower, though this is unobservable in our analysis.[31]

[31] Relatedly, charter schools might open where districts experience idiosyncratic shocks, in which case future outcomes regress toward the mean. If charter schools prefer to locate near low-performing school, as noted above,

This would lead to a downward bias in our estimates because TPS outcomes comprise the majority of the weighted average outcomes.

Alternatively, local governments might change non-charter education policies at the same time they introduce charter schools, as part of a package of reforms. Since school districts are often charter authorizers, it may be that a change in school board politics leads both to an increase in charter schools and, for example, new reading and math curricula in TPS. In this case, we might falsely attribute outcome changes to charter schools that were actually caused by other policy changes. We discuss various tests for these and related types of violations in the next section, after describing our main results.

IV.B. Random Causal Forest

Below, we provide a brief description of the random causal forest (CF) method applied to its originally designed setting of randomized controlled trials. Next, we describe recent advances that extend this method to difference-in-differences settings and specifically, settings where identification relies on the estimation of many fixed effects (CF-DD). Finally, we discuss the relative advantages and disadvantages of CF-DD and Interacted GDD, as discussed above.

IV.B.1. Brief Description of Causal Forests

The CF method is designed to be a data-driven and nonparametric approach to estimate treatment effect heterogeneity. It begins with the training of many causal trees (Athey and Imbens, 2016). Each causal tree acts as an adapted matching estimator, where rather than

---

then this would yield an upward bias (i.e., charter schools enter because of the negative shock in existing schools, but then those outcomes bounce back in the next period). We do not see this scenario as very likely because it takes several years to create an organization that can assemble a charter application, submit the application and gain approval, hire personnel, purchase necessary capital, and recruit students. Also, with such a long-term investment, charter organizations are likely to consider multiple years of information in making their entry decisions, which reduces the chance of regression to the mean.

matching on covariates for the purpose of constructing valid control groups, we match on covariates to segment units into neighborhoods of similar treatment effects. Each causal tree begins with a random subset of the full population and a random subset of the covariate vector, **X**.[32] It then proceeds to recursively partition the data by grid-searching across all covariates and across all thresholds within covariates to select the split that maximizes the squared-difference in average treatment effects between the resulting partitions. The subset where a split occurs is known as a "parent node," and the resulting partitions are known as "child nodes." Further splits are then made on the child nodes, creating further child nodes, until stopping criteria are met[33], and we are left with a set of "terminal nodes" (a.k.a. "leaves").

Terminal nodes are essentially a roadmap directing a given out-of-bag observation to its destination, where this destination is a conditional average treatment effect (CATE) prediction. That is, each new observation enters with a particular covariate value combination, $\mathbf{x}_i$, and follows the splitting rules until it reaches the terminal node which corresponds to that value combination. The CATE prediction each observation receives is the average treatment effect of the training units which originally fell into that leaf. A clear limitation of singular causal tree is that CATE predictions will be lumpy, as the set of terminal nodes is finite. Additionally, the structure of a single tree is subject to the randomness that generated the initial training sample (Wager and Athey, 2018). Therefore, in a causal forest, this out-of-bag unit will be assigned a

[32] Readers more familiar with traditional random forests will note that this differs slightly from the bootstrapped sub-samples used to train regression/classification trees in pure prediction settings.
[33] The stopping rules are some of the "hyperparameters" (i.e. configuration parameters used to manage machine learning training processes) of decision trees which prevent overfitting. These stopping rules include (1) the minimum number of treat/control units in a child node and (2) the minimum share of the parent node that must be contained in a child node. Other relevant hyperparameters include: (1) the share of the training sample used as hold-out for honest estimation and/or (2) an optional imbalance penalty parameter

CATE prediction from each of the many trees, and the final CATE prediction it receives will be an average of these individual predictions across all trees.

Causal trees follow a unique process for internal cross-validation known as "honest estimation" (Athey and Imbens, 2016). That is, for each tree, the random training sub-sample is again split in two before training occurs. The first split is used to actually define the tree structure by following the recursive partitioning procedure defined above. The second split is used to populate the nodes. That is, units in the first split create the roadmap, and units in the second split follow the roadmap into their respective neighborhoods, and the CATE predictions of that causal tree are the resulting treatment effect estimates from each neighborhood among units in the second split.

We study the main patterns within the CATE predictions by examining three artifacts of the model: (a) the differences in average covariate values for units with positive CATEs versus those with negative CATEs; (b) the variable importance factors (VIFs), which are the share of splits that are made on the basis of that covariate across all trees (weighted by the relative depth at which they occurred so that earlier splits are weighted more heavily); and (c) the best linear projection (BLP), which is the linear relationship between each covariate and the CATE prediction, conducted via a regression analysis of doubly-robust (i.e. propensity score debiased) CATE predictions on covariate values. To interpret the results, we start by examining the differences in means between district-years with statistically significant positive CATE estimates vs. those with statistically significant negative CATE estimates.[34] These differences provide a profile of the "average positively impacted district" vs. the "average negatively impacted

[34] CATE variance is computed as the variance of CATE estimates for a particular unit across trees. Part of the innovation of Athey and Wager (2018) is that these variance estimates are asymptotically normal when trees are trained via "honesty", and therefore allow for the construction of valid confidence intervals.

district." However, this alone does not provide much insight into *which* covariates are driving the heterogeneity. Therefore, we look to the VIF to help determine which covariates from the subset are most important. With the VIF scores, we can examine the covariate means with an idea of the "important" differences that drive treatment effects. However, it is important to note that the relationship between a covariate and the treatment effect may be non-linear and/or interactive with other covariates. The BLP helps us understand which, if any, relationships are approximately linear. Non-linear relationships are signaled when the difference-in-means suggests one direction (or no direction) but the BLP suggests another (especially when the VIF is large).

IV.B.2. Modification for Difference-in-Differences Setting

The causal forest is designed for experimental settings where identification of average causal effects (and therefore heterogeneous causal effects) comes from conditional or unconditional random assignment. Specifically, the causal forest relies on conditional independence: controlling for the set of confounders **X**, treatment is as good as random. However, in difference-in-differences settings, the identification assumption is not conditional independence, but parallel trends.

In the case of fixed effects estimators, the parallel trends assumption can be expressed equivalently as the conditional independence assumption when the fixed effects are included among the set of confounders (Roth et al., 2023). However, one cannot identify heterogeneity in treatment effects when fixed effects are included in **X** because the impact of any time-invariant covariates on the CATE will be absorbed by the fixed effects themselves. To avoid this issue, we perform two-way within-transformations on outcomes and treatment assignment indicators in order to obtain numerically equivalent estimates of average treatment effects to the GDD, and which can now vary according to the influences of time-invariant factors. We follow the "causal forest with fixed effects (CFFE)" approach defined in Kattenberg et al. (2023), which conducts "local" recentering (i.e. repeated within-transformations before splits occur at each tree node) and assumes a Borusyak et al. (2024) version of the staggered parallel trends assumption for identification.

IV.B.3. Comparison with the Interacted GDD

The CF method has several key advantages over the interacted GDD. The CF handles non-linearities (e.g. covariate interactions and/or polynomial relationships) naturally via a nonparametric mapping of covariates to CATE estimates. In the interacted GDD, the researcher must manually specify potential non-linearities, which comes at the cost of statistical power in high-dimensional regressions or the risk of spurious estimates from multiple hypothesis testing. In this way, the CF may identify heterogeneity in settings with many covariates more reliably than interacted regression.

Less well understood is the way that the CF handles missing data. The GDD and other parametric econometric methods implicitly carry out complete case analysis; when any variable is missing it is dropped. In the CF, missing observations are treated as having a different value for the variable. For example, when a parent node is being split into two child nodes, it might split based on missing versus non-missing observations rather than based on the magnitudes of the non-missing observations. (The equivalent of this in the GDD is adding a vector of missing indicators.) This becomes important later as some of the effect heterogeneity dimensions have considerable missingness.

The main potential advantage of the GDD over the CF is that, under certain conditions, we get a more plausible estimate of the average causal effect of treatment. The CFFE is meant to produce an estimate identical to the GDD. The above advantages in effect heterogeneity analyses are largely irrelevant if they are based in misleading heterogeneity. In ongoing work, we are trying to bring the CF more in line with the GDD, so that provides the best of both worlds.

V. Results

V.A. Replicating the Average Treatment Effects

We start by replicating the average treatment effects from Chen and Harris (2023) without the interaction terms. The first column (Sample 1) of Table 2 shows the results for the overall sample. The second column (Sample 2) drops the districts whose effects initially appear in the 5th-95th percentile (see reasoning below). Finally, the third column (Sample 3) starts with the whole sample and then shows the results dropping districts that had charter schools at the beginning of the panel, so that the estimates focus on the effects of the first charter entrant(s) due to the possibility of diminishing returns. All of the specifications assume a one-year lag in the effect from charter entry to student outcomes.

The results generally reinforce Chen and Harris's (2023) finding that the average effects are positive for each outcome.[35] Fourteen of 18 estimates are positive and eight are precisely estimated. Four of the estimates for math are negative and highly imprecise. There is no clear pattern as to which samples/columns show the largest effects. The results are also uniformly positive and more precise when dropping the district-specific linear trends. Since the results with

[35] Some slight differences eist between Table 2A and Chen and Harris (2023). The latter paper did not use linear interpolation to clean the AFGR.

district trends are more likely to pass the parallel trends, we opt for this as our preferred method going forward while reporting. These results are also mirrored in the CF histograms (Figure 1b).

V.B. Distribution of Effects

V.B.1. Overall Distribution Across Geographic Districts

We show histograms of the distribution of effects using both the GDD (Figure 1a) and CF (Figure 1b) methods. Charter effects may vary across geographic districts and this is the first (and smallest) institutional unit we consider. Figure 1(a) shows the wide distribution of market-level effects on high school graduation, centered near zero. Some of these are extremely large, in both positive and negative directions. For example, more than 50 districts have coefficients larger than +2.0; a coefficient of that size implies that going from 0 to 10 percent charter market share (just above the national average) would increase the high school graduation rate by 20 percentage points. The CF results have a much narrower distribution.

The test score effect distributions also seem wider than we would expect. For example, a larger-than-average coefficient of +1.0 means that increasing the charter market share by 10 percentage points increases annual student test scores by 0.10 s.d.. Some cities with very large market shares have generated market-level effects of this size (Harris & Larsen, 2023; Baxter, 2024). On the other hand, Chen and Harris (2023) estimate that the charter participant effect is two-thirds of the total market effect, with the remainder pertaining to the various competitive effects and math effects. In that case, a market-level effect of +0.10 s.d., for a 10 percent increase in market share, would require an annual participant effect, averaged across charter schools, of 0.67 s.d..[36]

[36] On the other hand, Chen and Harris (2023) estimate that the charter participant effect is two-thirds of the total market effect, with the remainder pertaining to the various competitive effects and math effects. In that case, a market-level effect of +0.10 s.d. would require an annual participant effect, averaged across charter schools, of 0.67 s.d..

We are also interested in the more unusual and surprising shape of the effect distribution for both math and ELA. They show a large concentration of effects in the +0.1 to +0.6 range. It turns out that almost all of these are in California. When we remove this state (see Appendix Figures A1 and A2), the distribution is much closer to normally distributed, although more bimodal (ELA) or even trimodal (math).

V.C. Effect Heterogeneity by Unit Traits

Most of the rest of the study focuses on explaining the variation across districts and states, using both the GDD and CF methods. The factors are broken into four categories: student demographics (e.g., race and family income), market context (e.g., TPS baseline outcomes and teacher supply metrics), school finance (e.g., charter spending, TPS-charter funding gaps, and total average spending), and other charter school policies. We organize the analysis topic by topic, so that we can directly compare the results across methods and summarize the results for each set.

The GDD and CF results are somewhat difficult to compare. In both cases, we can and do consider statistical significance. In the GDD, this pertains to the usual significance testing of the treatment coefficient, shown in Table 2. In the CF, this pertains to the precision of the difference in mean covariates between leaves that show positive versus negative treatment effects (using bootstrap standard errors), shown in Tables 3a-3c. However, the CF method also yields the variable important factor (VIF), which also signals which effect magnitudes, regardless of statistical significance, are driving heterogeneity; these are shown in Figure 2. Given the advantages and disadvantages of each method described earlier, we conclude that effect heterogeneity exists when estimates that are robust across the two methods and precise in at least one of them. Then, we also consider the VIF to further refine which are most important overall.

Table 6 summarizes all of the information across methods. In addition to indicating statistical significance with asterisks, we indicate the ten most important factors according to the VIF (V1 indicates highest VIF and so on down to V10). We also note that, by design, we chose a select number of covariates for the GDD to avoid collinearity issues. This is not necessary in, and is indeed an advantage of, the CF. For this reason, we report results for more covariates in the CF results.

V.C.1. Effect Heterogeneity by Demographics

Prior research consistently found that charter schools are more effective for disadvantaged students, broadly defined (Cohodes & Parham, 2021). The theories around this have been based partly on the lower test scores these groups typically have or, relatedly, that these groups are served by schools that are more poorly performing.

We also find that demographics are the most important predictor of charter school effects. This is most easily seen in the summary in Table 6. Seven of the eight estimates for percent Black and Hispanic are positive and statistically significant and have a relatively high VIF. The results are more mixed for high school graduation, however. They are negative for Blacks. All the estimates are positive, significant, and important for Hispanics, across outcomes and methods. A similar pattern emerges for FRL. All but one of the estimates is positive, significant, and important, except for one of the two on high school graduation.

We included two demographic measures in the CF analysis that were not in the CF. A higher percentage of students with disabilities is associated with a reduction in test scores but an increase in graduation. This is consistent with prior evidence that charter schools are less able to serve students with disabilities (e.g., they are smaller, more specialized, and less likely to have teachers trained in serving these groups), which would explain the lower test scores. Students identified as special education are also more likely to shed that designation when they move to charter schools. Families might prefer this for reasons of social stigma, even if it might mean slower academic progress. This could lead to higher graduation rates.

The situation with total enrollment is more complex. This has a high VIF but the differences in means are only statistically significant in one case. This is a sign of non-linearities in the role of enrollment. To be precise, when a large number of splits in the CF are based on any factor, it could be that some of the splits are associated with positive effects and some negative effects, especially if splits occur early in the tree (which mechanically increases VIF), so that the covariate differences-in-means end up being similar and not statistically significant. We are still exploring these non-linearities in ongoing work. We do note that prior research has found that school choice policies are more effective in urban areas, which tend to have larger enrollment.

V.C.2. Effect Heterogeneity by Market Context

We included three types of market context variables: region type (e.g., urban and suburban), attractiveness to young teachers, and baseline academic performance in traditional public schools. Baseline performance has consistently high VIFs and, with test scores, they are in the expected direction; charter effects are larger when baseline TPS performance is low and it is easier for charter schools to improve upon TPS. However, continuing the prior theme, the results are the opposite with graduation. The results are not robust for the size of local university-based schools of education, which affects educator supply. These effects also seem unimportant based on the VIF. Finally, the region type variables are often statistically significant but never have high VIFs.

V.C.3.Effect Heterogeneity by School Funding

We include many different measures of school finance. School funding generally leads to improved student outcomes, including both test scores and high school graduation (though we have less evidence on the latter outcome). However, the theory on the relationship between school funding and charter effects is more complicated. Other things equal, we would expect charter schools to be more effective when they have more funding themselves. However, market-level effects depend on the funding to both TPS and charter schools and those are interconnected.

The school funding variables have high VIFs in the CF analysis. The TPS-charter funding gap is actually positively related to student outcomes. This is noteworthy because it means that charter schools have a more positive impact when they receive less, but this could be because TPS see less of an impact. [MORE LATER]

V.C.4. Effect Heterogeneity by State Policy

We considered five policy dimensions from the charter industry groups. These results are quite mixed between the two methods and, probably not coincidentally, have very low VIFs.

In Table 4 we estimate equation (1) and (2) where H_{is} is a vector of traits, mostly household characteristics (included at the district level) and state policies (necessarily at the state level), using data taken from the baseline year to avoid endogeneity concerns. We include all of these characteristics simultaneously (each column in a single regression) because they are all correlated with one another. (the appendix shows the results where the covariates are included one at a time). Table 5 shows the placebo test for these variables. In what follows, we draw only tentative conclusions about any dimension of effect heterogeneity that fails the placebo test (especially if the placebo effect is in the same direction and magnitude as the effect estimate in Table 4). We are also more cautious about estimates that are not robust across the three samples.

V.C.1. Effect Heterogeneity by Household Demographics

The top rows of Table 4 show the results for the interaction between charter market share and percent of the school enrollment that is Black, Hispanic, or eligible for free or reduced-price lunch (FRL). The coefficients are uniformly positive for FRL and often precise. They are also quite large. If we increase the charter share by 10 percentage points, in a district with no FRL students versus one that is 100% FRL, then the latter effect will be 0.03 to 0.04 s.d. larger than the former. Those results also generally pass the placebo test in Table 5.

The results for Black and Hispanic students are more ambiguous. They are inconsistent across outcomes (negative for graduation and positive for test scores) and often fail the placebo test. We also included a measure of baseline student performance. The results are also ambiguous in this case. While the coefficients in Table 4 are uniformly positive, most of the placebo coefficients are as well (and are of the same magnitude).

V.C.2. Effect Heterogeneity by State Charter Policy

The results in Table 4 simultaneously control for state policy ratings. We focus here on the NAPCS ratings, which provide the most policy categories. Again, these are: transparent charter application process (*Transpar*), performance-based contract required (*Perf* in Table 4), clear processes for renewal/nonrenewal (*Non_renew*), automatic exemption of charter schools from laws applied to traditional public schools (TPS) (*Rule_exempt*), no caps on the growth of public charter schools (*Nocaps*), and equitable operational funding, i.e., higher funding for charter schools relative to TPS (*Eq_fiunding*).

It initially appears, in Table 4, that having a transparent application process is associated with lower charter effects, while states providing more equitable funding have better charter performance, at least in test scores. For the other policy variables, the coefficients are less consistent across samples and outcomes and/or the equivalent placebo coefficients are of similar signs/magnitudes and sometimes precise.

VI. Conclusion

A common theme of the economics of education is that different educational units (teachers, schools, and so on) vary widely in their quality. A second common theme is that measurable traits in those units (teacher certification, school size, and so on) do not explain much if any of this variation. With more than one thousand school districts having at least one charter school, it is now possible to test whether that pattern holds for charter schools.

In some respects, we find the same pattern. First, do charter effects vary? Do some "bloom" and others "wilt"? Yes, as with teachers and schools, some of the effects, especially with respect to high school graduation are very large. We also see some, albeit less, variation in effects across states.

Second, can we explain that variation? Here, we see several noteworthy patterns. Districts with more low-income students benefit more from having charter schools. These results are not driven by the fact that we are controlling for so many factors at the same time in the main results (Table 4). They also hold when we estimate equation (1) with smaller blocks of variables.

Somewhat surprisingly, charter schools seem to improve geographic district outcomes when those districts were higher performing to begin with. This may be less surprising, however, when we note that some of the states with larger effects, especially Massachusetts (see Table 3), also have high baseline achievement levels; conversely, some of the states with smaller effects (Arkansas and South Carolina) have very low performance levels.

State policies may explain some of the variation in results across states. Charter policies that require less transparency in charter applications and renewal decisions are associated with more positive charter effects. One theory that could explain this is that authorizers, in states with stronger transparency, may be pressed to make their decisions more on the basis of public information, as opposed to private information that might have that is also indicative of future charter performance.

Less surprising is that charter schools seem more effective, as least with respect to test scores, when state policies provide more funding to charter schools. However, as noted in section II, the role of charter funding is more complicated than the state-level policy ratings account for. We certainly expect the charter participant effects to be more positive when charter schools have more funding but, for purposes of understanding the market-level effects, we also have to be concerned with what charter entry means for TPS in the same districts, which may influence the competitive responses. We are in the process of collecting data, at the district level, which will allow us to better understand these issues.

We see only limited evidence of diminishing marginal returns. The effects are larger for test scores when we focus only on the districts that have a first charter entry during the panel, which is suggestive of DMR. But this pattern no longer holds when we consider the full distribution of baseline charter entry. We also see little clear evidence that charter effects vary based on changes in total geographic district enrollment.

Another pattern throughout the paper is that the results differ between the high school outcomes and test outcomes. This is unsurprising for three reasons: (a) our test scores results pertain to elementary/middle schools while high school graduation pertains to a different set of schools—high schools; (b) the time span of the test score data differs from that of high school graduation, so the schools are also of different vintages; and (c) test scores and high school graduation likely have different education production functions—we do not expect schools to be equally effective in producing all outcomes. When we limit only to test scores, we see suggestive evidence that districts with declining total enrollment see more negative charter effects. We are also conducting analysis of the possibility that charet effects might display diminishing marginal returns, i.e., that the effect of charter schools is larger for initial entrants.

This work is still preliminary. In ongoing work, we are also: (a) applying shrinkage estimation techniques to understand how much of the variation is driven by measurement error; (b) collecting additional data about school types (e.g., CMO status) so we can test whether the effects are larger when certain types of schools enter; (c) collecting better data, at the district level, on charter school spending and teacher supply, so we can estimate the effects of these additional traits; and (d) carrying out additional analysis of the roles played by diminishing.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A Pathak. 2017. "The welfare effects of coordinated assignment: Evidence from the New York City high school match." *American Economic Review* 107 (12):3635-3689.

Abdulkadiroğlu, Atila, Joshua D Angrist, Susan M Dynarski, Thomas J Kane, and Parag A Pathak. 2011. "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots." *The Quarterly Journal of Economics* 126 (2):699-748.

Angrist, Joshua D, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher D Walters. 2013. "Charter Schools and the Road to College Readiness: The Effects on College Preparation, Attendance and Choice. Understanding Boston." *Boston Foundation*.

Angrist, Joshua D, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher R Walters. 2016. "Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice." *Journal of Labor Economics* 34 (2):275-318.

Angrist, Joshua D, Parag A Pathak, and Christopher R Walters. 2013. "Explaining charter school effectiveness." *American Economic Journal: Applied Economics* 5 (4):1-27.

Bergman, Peter, and Jr McFarlin, Isaac. 2020. Education for all? A nationwide audit study of school choice. National Bureau of Economic Research.

Bettinger, Eric P. 2005. "The effect of charter schools on charter students and public schools." *Economics of Education Review* 24 (2):133-147.

Bifulco, Robert, and Christian Buerger. 2015. "The influence of finance and accountability policies on location of New York state charter schools." *journal of education finance*:193-221.

Bifulco, Robert, and Helen F Ladd. 2006. "The impacts of charter schools on student achievement: Evidence from North Carolina." *Education Finance and Policy* 1 (1):50-90.

Booker, Kevin, Brian Gill, Tim Sass, and Ron Zimmer. 2014. Charter high schools' effects on educational attainment and earnings. Mathematica Policy Research.

Booker, Kevin, Scott M Gilpatric, Timothy Gronberg, and Dennis Jansen. 2007. "The impact of charter school attendance on student performance." *Journal of Public Economics* 91 (5-6):849-876.

Booker, Kevin, Scott M Gilpatric, Timothy Gronberg, and Dennis Jansen. 2008. "The effect of charter schools on traditional public school students in Texas: Are children who stay behind left behind?" *Journal of Urban Economics* 64 (1):123-145.

Booker, Kevin, Tim R Sass, Brian Gill, and Ron Zimmer. 2011. "The effects of charter high schools on educational attainment." *Journal of Labor Economics* 29 (2):377-415.

Branch, Gregory F, Eric A Hanushek, and Steven G Rivkin. 2012. Estimating the effect of leaders on public sector productivity: The case of school principals. National Bureau of Economic Research.

Brewer, Dominic, Ron Zimmer, Richard Buddin, Derrick Chau, and Brian Gill. 2003. "Charter school operations and performance: Evidence from California."

Bross, Whitney, Douglas N Harris, and Lihan Liu. 2023. "The effects of performance-based school closure and charter takeover on student performance." *Economics of Education Review* 94.

Buddin, Richard. 2012. "The impact of charter schools on public and private school enrollments." *Cato Institute Policy Analysis* (707).

Buerger, Christian, and Douglas N Harris. 2021. "The impact of government contracting out on spending: The case of public education in new orleans." *The American Review of Public Administration* 51 (2):139-154.

Campos, Christopher, and Caitlin Kearns. 2022. "The Impact of Neighborhood School Choice: Evidence from Los Angeles' Zones of Choice." *Available at SSRN 3830628*.

Carlson, Deven, and Stéphane Lavertu. 2016. "Charter school closure and student achievement: Evidence from Ohio." *Journal of Urban Economics* 95:31-48.

Chabrier, Julia, Sarah Cohodes, and Philip Oreopoulos. 2016. "What can we learn from charter school lotteries?" *Journal of Economic Perspectives* 30 (3):57-84.

Chakrabarti, Rajashri, and Joydeep Roy. 2016. "Do charter schools crowd out private school enrollment? Evidence from Michigan." *Journal of Urban Economics* 91:88-103.

Chin, M., Kane, T. J., Kozakowski, W., Schueler, B. E., & Staiger, D. O. (2019). School District Reform in Newark: Within- and Between-School Changes in Achievement Growth. ILR Review, 72(2), 323-354.

Cohodes, Sarah R, and Katharine S Parham. 2021. "Charter schools' effectiveness, mechanisms, and competitive influence."

Cordes, Sarah A. 2018. "In pursuit of the common good: The spillover effects of charter schools on public school students in New York City." *Education Finance and Policy* 13 (4):484-512.

CREDO, Center for Research on Education Outcomes. 2023. As a Matter of Fact: The National Charter School Study III Center for Research on Education Outcomes, Stanford University.

Curto, Vilsa E, and Jr Fryer, Roland G. 2014. "The potential of urban boarding schools for the poor: Evidence from SEED." *Journal of Labor Economics* 32 (1):65-93.

Demers, Alicia, Ira Nichols-Barrer, Elisa Steele, Maria Bartlett, and Philip Gleason. 2023. "Long-Term Impacts of KIPP Middle and High Schools on College Enrollment, Persistence, and Attainment." *Mathematica*.

Dinerstein, Michael, and Troy D Smith. 2021. "Quantifying the supply response of private schools to public policies." *American Economic Review* 111 (10):3376-3417.

Dobbie, Will, and Roland G Fryer. 2013. The medium-term impacts of high-achieving charter schools on non-test score outcomes. National Bureau of Economic Research.

Dow, William H, Anna Godøy, Christopher Lowenstein, and Michael Reich. 2020. "Can labor market policies reduce deaths of despair?" *Journal of health economics* 74:102372.

Engberg, John, Brian Gill, Gema Zamarro, and Ron Zimmer. 2012. "Closing schools in a shrinking district: Do student outcomes depend on which schools are closed?" *Journal of Urban Economics* 71 (2):189-203.

Ferreyra, Maria Marta, and Grigory Kosenok. 2018. "Charter school entry and school choice: The case of Washington, DC." *Journal of Public Economics* 159:160-182.

Figlio, David, and Cassandra Hart. 2014. "Competitive effects of means-tested school vouchers." *American Economic Journal: Applied Economics* 6 (1):133-56.

Figlio, David N, Cassandra Hart, and Krzysztof Karbownik. 2022. "Effects of Maturing Private School Choice Programs on Public School Students." *American Economic Journal: Economic Policy*.

Figlio, David N, Cassandra MD Hart, and Krzysztof Karbownik. 2021. "Competitive Effects of Charter Schools."

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Natalya Verbitsky-Savitz, Bing-ru Teh, Melissa Bowen, and Allison Demeritt. 2012. "Charter-school management organizations: Diverse strategies and diverse student impacts." *Mathematica Policy Research, Inc.*

Gilraine, Michael, Uros Petronijevic, and John D Singleton. 2021. "Horizontal differentiation and the policy effect of charter schools." *American Economic Journal: Economic Policy* 13 (3):239-76.

Glomm, Gerhard, Douglas N Harris, and Te-Fen Lo. 2005. "Charter school location." *Economics of Education Review* 24 (4):451-457.

Goldhaber, Dan D, and Eric R Eide. 2003. "Methodological thoughts on measuring the impact of private sector competition on the educational marketplace." *Educational Evaluation and Policy Analysis* 25 (2):217-232.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*.

Griffith, David. 2019. Rising Tide: Charter School Market Share and Student Achievement. Washington, DC: Thomas B. Fordham Institute.

Hanushek, Eric A, John F Kain, Steven G Rivkin, and Gregory F Branch. 2007. "Charter school quality and parental decision making with school choice." *Journal of public economics* 91 (5-6):823-848.

Harris, Douglas N. 2020. *Charter school city: What the end of traditional public schools in New Orleans Means for American education*: University of Chicago Press.
Harris, Douglas N, and Matthew Larsen. 2023. "The Effects of the New Orleans Post-Katrina Market-Based School Reforms on Medium-Term Student Outcomes." *Journal of Human Resources*
Harris, Douglas N, and Valentina Martinez-Pabon. 2022. "Extreme Measures: A National Descriptive Analysis of Closure and Restructuring of Traditional Public, Charter, and Private Schools." *Education Finance and Policy*:1-50.
Hess, Frederick M. 2004. *Revolution at the margins: The impact of competition on urban school systems*: Brookings Institution Press.
Ho, Andrew D. 2020. What Is the Stanford Education Data Archive Teaching Us About National Educational Achievement? : SAGE Publications Sage CA: Los Angeles, CA.
Holley, Marc J, Anna J Egalite, and Martin F Lueken. 2013. "Competition with charters motivates districts: New political circumstances, growing popularity." *Education Next* 13 (4):28-36.
Hoxby, Caroline M. 2000. "Does competition among public schools benefit students and taxpayers?" *American Economic Review* 90 (5):1209-1238.
Hoxby, Caroline Minter. 2003. "School choice and school productivity. Could school choice be a tide that lifts all boats?" In *The economics of school choice*, 287-342. University of Chicago Press.
Hoxby, Caroline Minter, and Jonah E Rockoff. 2004. *The impact of charter schools on student achievement*: Department of Economics, Harvard University Cambridge, MA.
Imberman, Scott A. 2011. "The effect of charter schools on achievement and behavior of public school students." *Journal of Public Economics* 95 (7-8):850-863.
Jabbar, Huriya. 2015. "“Every kid is money” market-like competition and school leader strategies in New Orleans." *Educational Evaluation and Policy Analysis* 37 (4):638-659.
Jackson, C. Kirabo and Claire L. Mackevicius (2024). What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the United States. *American Economic Journal: Applied Economics* 16(1): 412–46.
Jinnai, Yusuke. 2014. "Direct and indirect impact of charter schools’ entry on traditional public schools: New evidence from North Carolina." *Economics Letters* 124 (3):452-456.
Ladd, Helen F, and John D Singleton. 2020. "The fiscal externalities of charter schools: Evidence from North Carolina." *Education Finance and Policy* 15 (1):191-208.
Lincove, Jane Arnold, Deven Carlson, and Nathan Barrett. 2019. The effects of school closure on the teacher labor market: Evidence from portfolio management in New Orleans. Working Paper.
Linick, Matthew Allen. 2014. "Measuring Competition: Inconsistent definitions, inconsistent results." *education policy analysis archives* 22 (16):n16.
Loeb, Susanna, Jon Valant, and Matt Kasman. 2011. "Increasing choice in the market for schools: Recent reforms and their effects on student achievement." *National Tax Journal* 64 (1):141.
Lubienski, Christopher. 2007. "Marketing schools: Consumer goods and competitive incentives for consumer information." *Education and Urban Society* 40 (1):118-141.
Mehta, Nirav. 2017. "Competition in public school districts: charter school entry, student sorting, and school input determination." *International Economic Review* 58 (4):1089-1116.
Mouganie, Pierre, Ruba Ajeeb, and Mark Hoekstra. 2020. The effect of open-air waste burning on infant health: evidence from government failure in Lebanon. National Bureau of Economic Research.
Ni, Yongmei. 2009. "The impact of charter schools on the efficiency of traditional public schools: Evidence from Michigan." *Economics of Education Review* 28 (5):571-584.
Reardon, Sean F, John P Papay, Tara Kilbride, Katherine O Strunk, Joshua Cowen, Lily An, and Kate Donohue. 2019. "Can Repeated Aggregate Cross-Sectional Data Be Used to Measure Average Student Learning Rates? A Validation Study of Learning Rate Measures in the Stanford Education Data Archive. CEPA Working Paper No. 19-08." *Stanford Center for Education Policy Analysis*.
Ridley, Matthew, and Camille Terrier. 2022. "Fiscal and Education Spillovers from Charter School Expansion." *Conditionally accepted at Journal of Human Resources.*

Rothstein, Jesse. 2007. "Does competition among public schools benefit students and taxpayers? Comment." *American Economic Review* 97 (5):2026-2037.

Sass, Tim R. 2006. "Charter schools and student achievement in Florida." *Education Finance and Policy* 1 (1):91-122.

Slungaard Mumma, Kirsten. 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. doi: 10.1257/pol.20190457.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225 (2):175-199.

Toma, Eugenia F, Ron Zimmer, and John T Jones. 2006. "Beyond achievement: Enrollment consequences of charter schools in Michigan." *Advances in Applied Microeconomics* 14 (2):241-255.

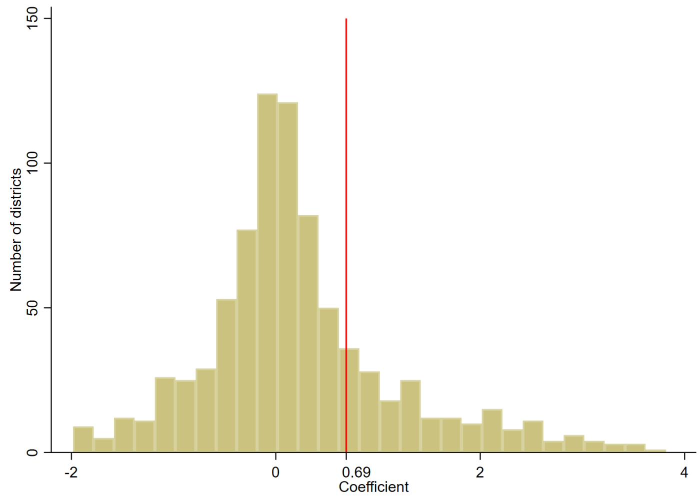
Wong, Mitchell D, Karen M Coller, Rebecca N Dudovitz, David P Kennedy, Richard Buddin, Martin F Shapiro, Sheryl H Kataoka, Arleen F Brown, Chi-Hong Tseng, and Peter Bergman. 2014. "Successful schools and risky behaviors among low-income adolescents." *Pediatrics* 134 (2):e389-e396.

Ziebarth, Todd. 2020. "Measuring up to the model: A ranking of state public charter school laws." *National Alliance for Public Charter Schools*.

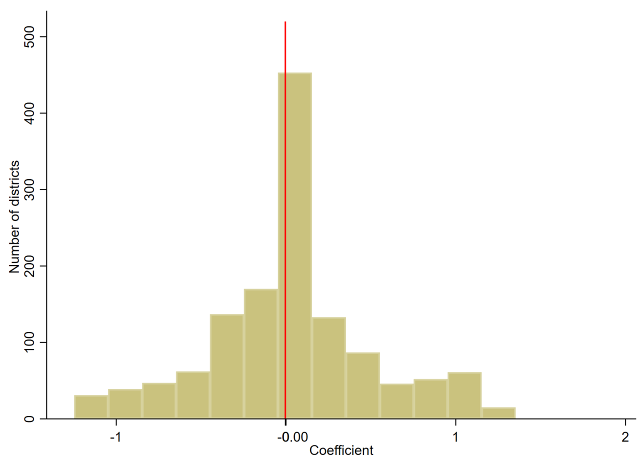
Zimmer, Ron, and Richard Buddin. 2009. "Is charter school competition in California improving the performance of traditional public schools?" *Public Administration Review* 69 (5):831-845.

Figure 1a Histogram plot of coefficient distribution

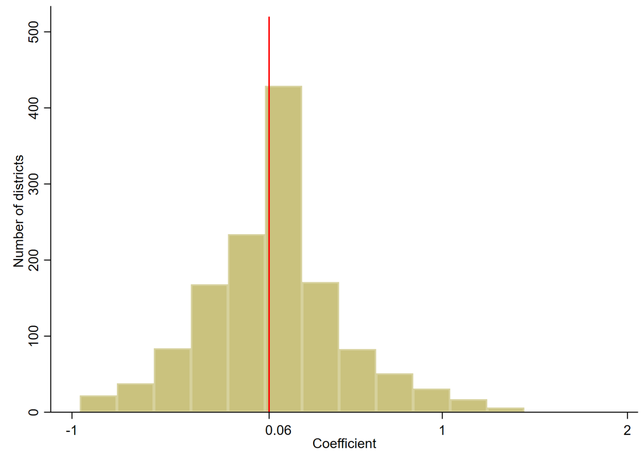
(a) Graduation rate



(b) Math

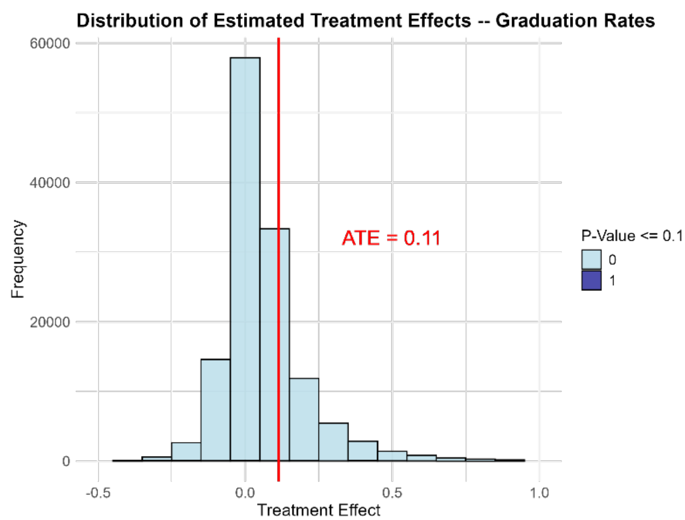


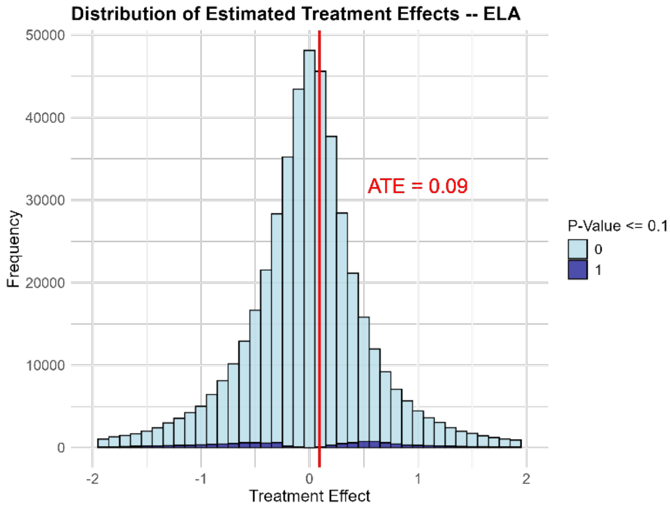
(c) ELA

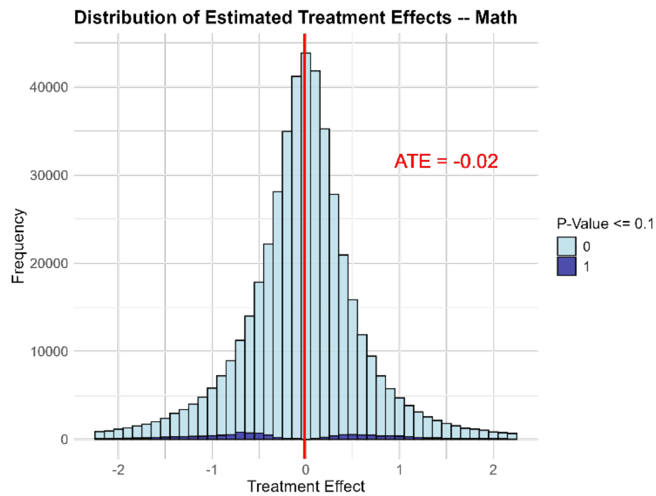


Note: This figure presents the histogram plots of estimates for the charter effects for each treated district. Specifically, for each treated district, we use the non-charter districts in the same state as the comparison group to do the GDD analysis. The coefficients below 5-th percentile and above 95-th percentile are not plotted here. The vertical red line is the mean of coefficients weighted by charter enrollment size.

Figure 1b: Causal Forest Histograms (district x year estimates)

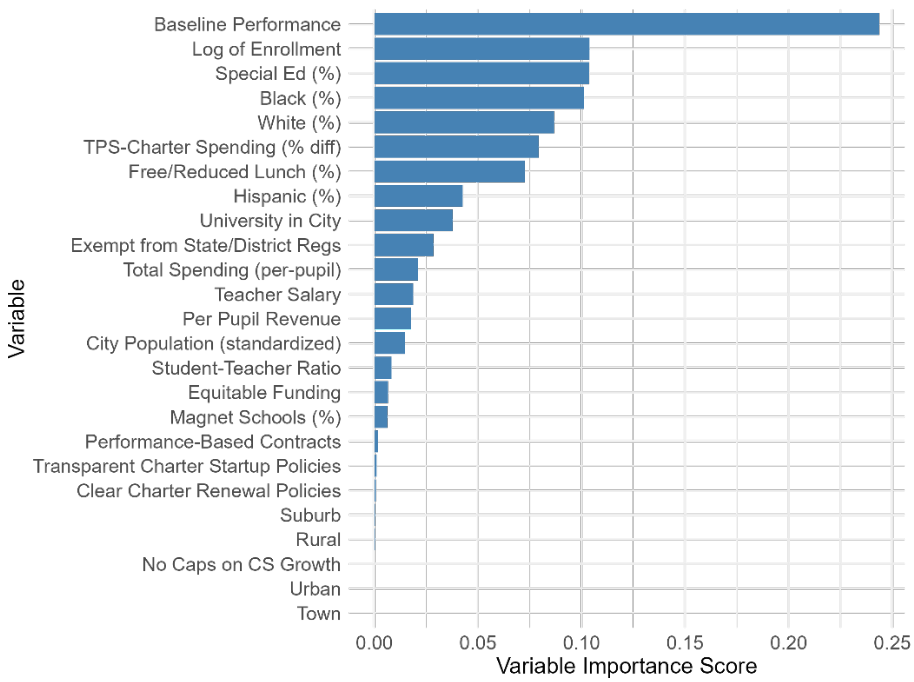




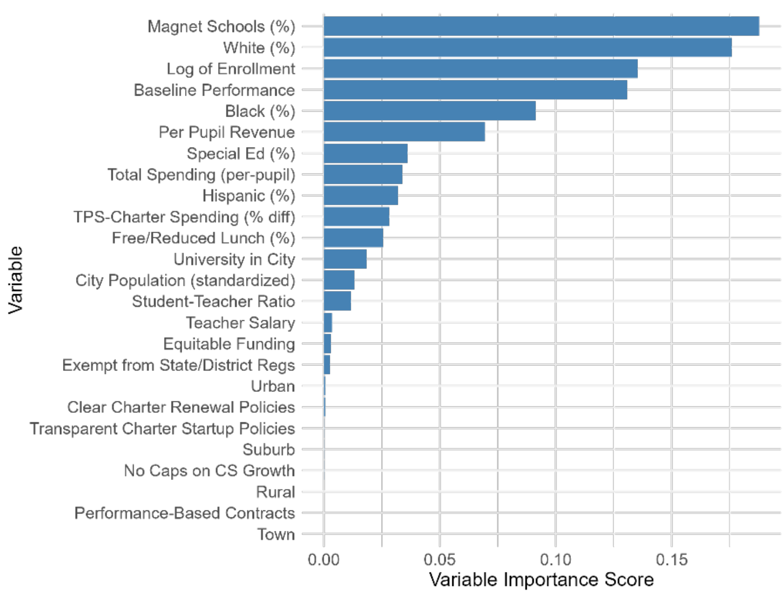


Notes: These figures are the distributions of individualized (district $\times$ year)-level treatment effects for each of the three main outcomes from the causal forests. We estimate standard errors for each ITE, and shade darker those districts with statistically significant (at the $p = 0.1$ level) ITE estimates. We also display the GRF-estimated ATE which comes from the aggregation of doubly-robust scores. These are interpreted as average partial effects for a given district $\times$ year. That is, each point represents $E\left[\frac{\partial \tau(\boldsymbol{X}_{it})}{\partial W_{it}}\right] = \frac{Cov[Y_{it}, W_{it} | \boldsymbol{X}_{it} = \boldsymbol{x}]}{Var[W_{it} | \boldsymbol{X}_{it} = \boldsymbol{x}]}$, the predicted treatment effect from increasing the charter share in district d in year t by 1 percentage point.

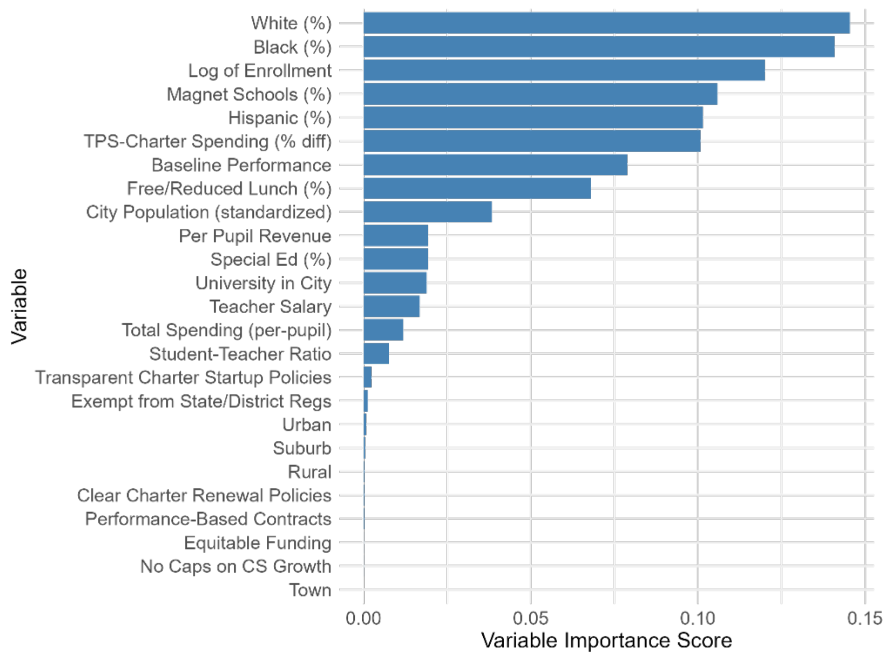
Figure 2: Causal Forest Variable Importance Factors (VIF)



(a) Graduation Rates



(b) Math Scores

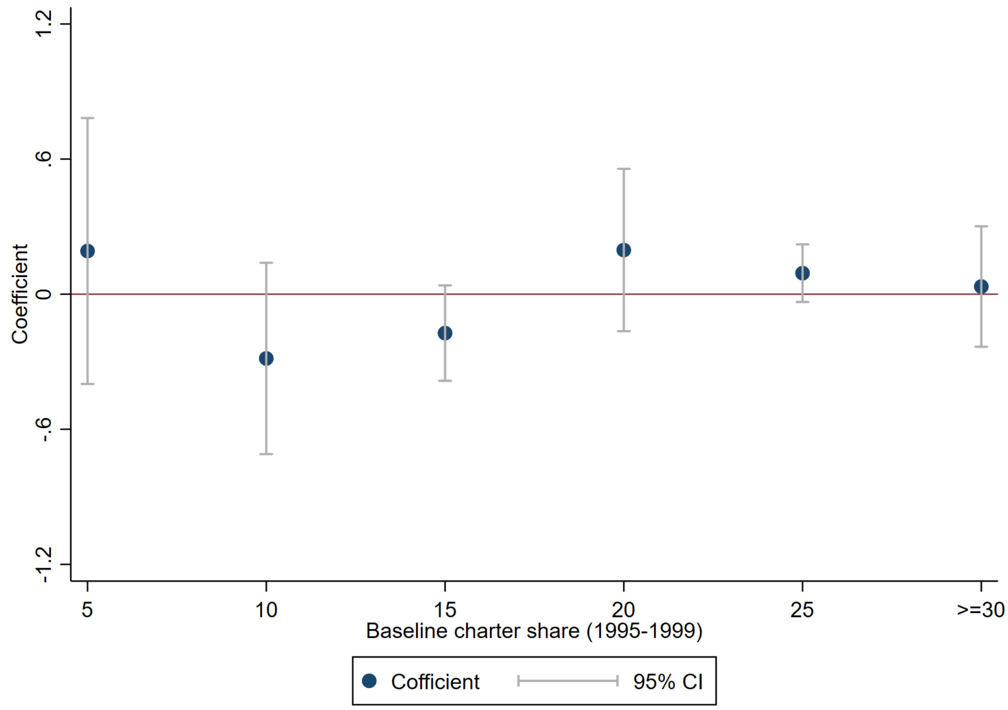


(c) ELA Scores

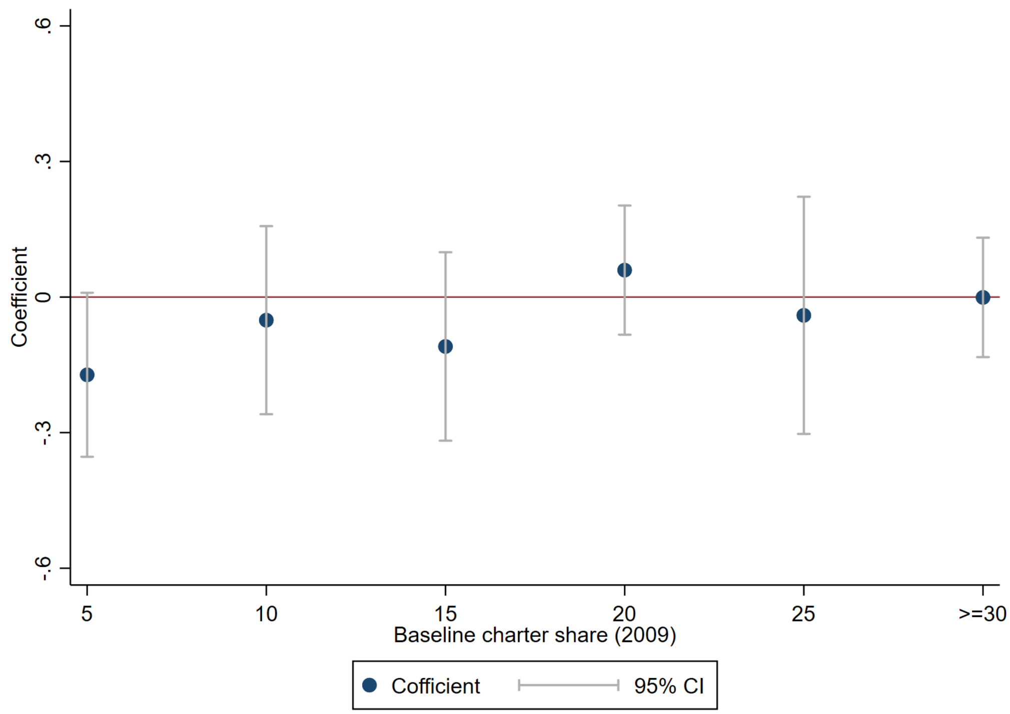
Notes: These figures display variable importance factors (VIFs), which amount to the share of trees which split along a given covariate, weighted by the depth at which the split occured.

Figure 3a Diminishing Returns

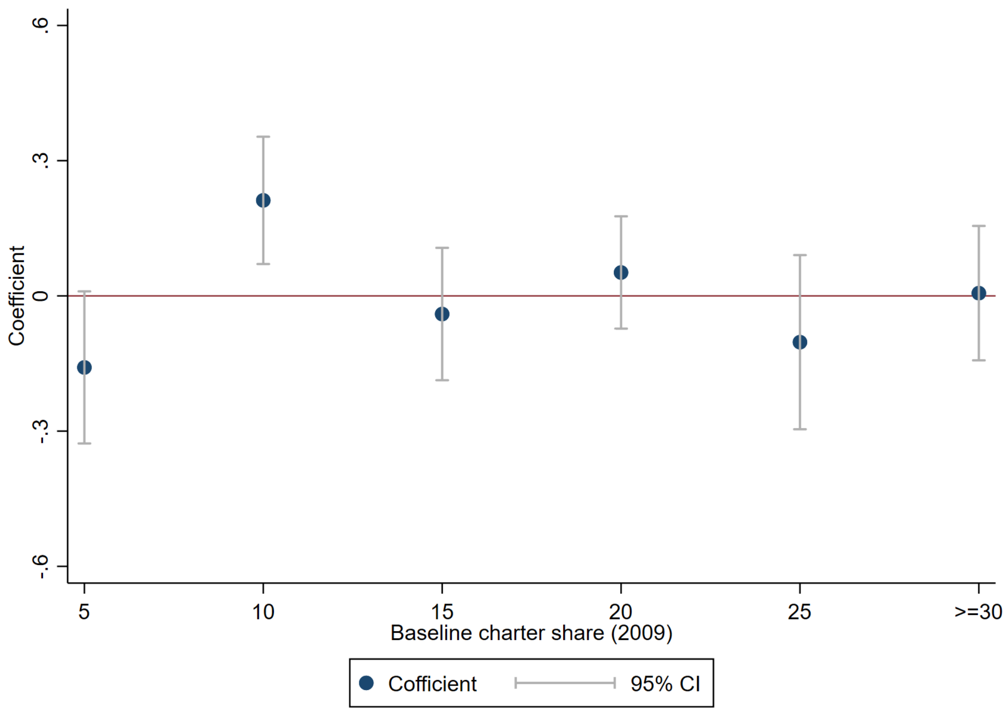
(a) Graduation rate



(b) Math

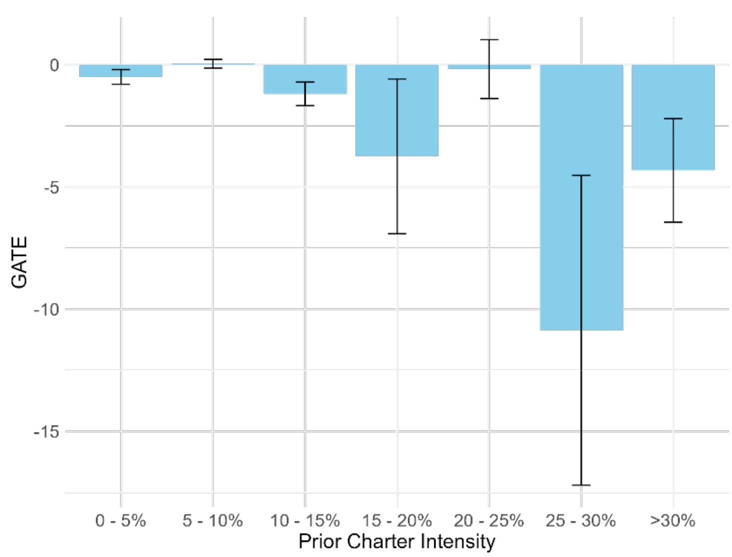


(c) ELA

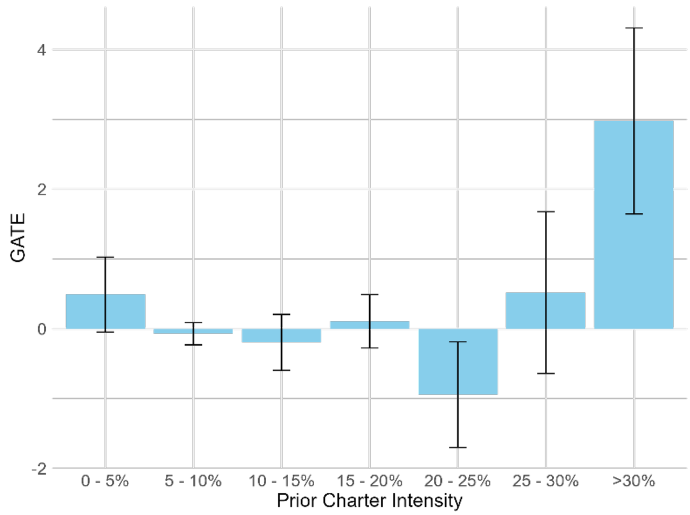


Note: This figure presents the coefficient of the interaction of charter share with baseline charter share groups. The baseline charter share is the charter share in the first period of the sample, but for graduation sample, we use the max charter share of 1995-1999 as the baseline charter share as only a few districts have charter schools in 1995.

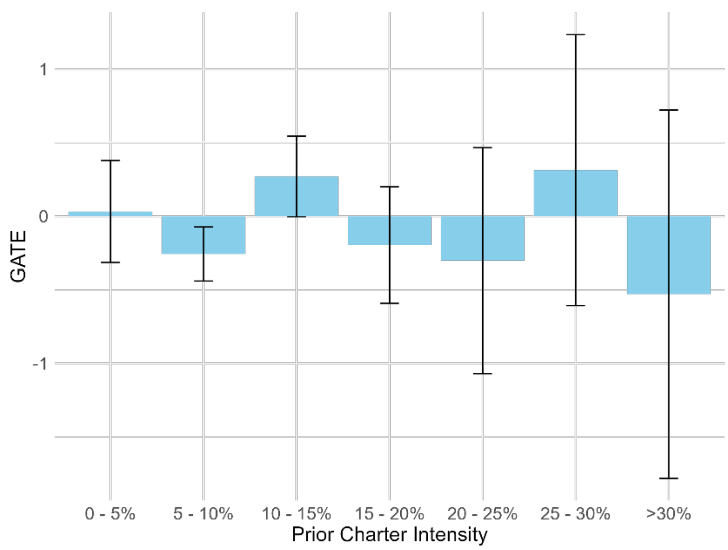
Figure 3b: Causal Forest diminishing returns



(a) Graduation Rates



(b) Math Scores



(c) ELA Scores

Table 1 Summary Statistics

Variable\Sample	Graduation rate			Test Score		
	Sample 1	Sample 2	Sample3	Sample 1	Sample 2	Sample3
Graduation rate	0.74	0.75	0.75			
Math				0.00	0.00	0.18
ELA				-0.01	-0.01	0.19
Black	0.17	0.16	0.17	0.17	0.16	0.12
Hispanic	0.15	0.14	0.14	0.24	0.24	0.18
FRL	0.29	0.29	0.28	0.53	0.53	0.47
Special education	0.13	0.13	0.13	0.13	0.13	0.14
Urban	0.32	0.31	0.30	0.30	0.30	0.14
Total spending	9,073	9,148	9,056	12,884	12,932	12,980
Charter spending	4,034	3,963	3,771	4,368	4,086	970
TPS - Charter spending %	11,583	11,297	11,784	11,214	11,170	11,496
Observations	133,418	131,077	132,943	476,725	461,947	420,636
N (district)	9,233	9,038	9,201	9,795	9,512	8,721

Notes: This table presents weighted means of outcome variables (graduation rate, math, and ELA) and control variables. graduation rate sample is weighted by the first-period enrollment. Sample (1) is the full sample, Sample (2) drops the noncredible-charter districts, Sample (3) drops the always-charter districts. Data source: National Longitudinal School Database.

Table 2 Effects of charter entry on student outcomes

	Sample 1		Sample 2		Sample 3	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Graduation rate						
Charter share	0.089	0.093*	0.089	0.093*	0.067	0.070
	[0.057]	[0.057]	[0.056]	[0.056]	[0.058]	[0.057]
R-squared	0.929	0.929	0.925	0.926	0.931	0.932
Observations	132,876	131,435	130,547	129,123	132,402	130,961
Panel B: Math						
Charter share	-0.029	-0.03	-0.028	-0.028	0.077**	0.073**
	[0.027]	[0.026]	[0.028]	[0.027]	[0.037]	[0.037]
R-squared	0.891	0.891	0.891	0.891	0.863	0.864
Observations	476,721	475,572	461,943	460,844	420,636	419,542
Panel C: ELA						
Charter share	0.056*	0.056*	0.055*	0.055*	0.049	0.045
	[0.030]	[0.031]	[0.029]	[0.030]	[0.039]	[0.039]
R-squared	0.911	0.911	0.911	0.911	0.888	0.888
Observations	476,721	475,572	461,943	460,844	420,636	419,542
District FE	Yes	Yes	Yes	Yes	Yes	Yes
State-by-year	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes
District trend	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows estimates of the effects of charter entry on student outcomes using different samples. Sample 1is the full sample; Sample 1 Censored is the specification that the outcome is censored to the 10th and 90th percentiles; Sample 2 drops the always-charter districts. Charter share in the data set range from 0 to 1. District control includes, share of black, Hispanic, FRL, special education, expenditure per student, urban, charter spending per student, TPS - Charter spending %. Regressions are weighted by the first-period high school enrollment for graduation rate and grade-level enrollment for Math and ELA. Standard errors are clustered at the district level.

Table 3 Effect heterogeneity analysis (include district time trend)

	Graduation rate			Math			ELA		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
Charter*Black	-0.258	-0.245	-0.550**	0.172	0.224**	0.258	0.286***	0.301***	0.140
	[0.276]	[0.278]	[0.229]	[0.110]	[0.106]	[0.190]	[0.104]	[0.102]	[0.191]
Charter*Hispanic	0.469	0.471	-0.125	0.032	0.072	0.403*	0.312**	0.310**	0.478**
	[0.598]	[0.585]	[0.268]	[0.146]	[0.147]	[0.216]	[0.127]	[0.126]	[0.222]
Charter*FRL	0.512	0.495	0.789***	0.513***	0.383**	-0.177	0.045	-0.010	-0.400**
	[0.327]	[0.331]	[0.283]	[0.161]	[0.155]	[0.196]	[0.143]	[0.143]	[0.204]
Charter*Performance	-0.607*	-0.588	-0.783**	0.640***	0.599***	0.236***	0.287***	0.275***	0.120*
	[0.353]	[0.368]	[0.365]	[0.071]	[0.068]	[0.062]	[0.041]	[0.042]	[0.065]
Charter* Nocaps	0.081	0.093	0.226	0.681***	0.685***	0.155	-0.143	-0.162	0.188
	[0.326]	[0.325]	[0.303]	[0.141]	[0.143]	[0.201]	[0.116]	[0.114]	[0.204]
Charter* Transpar	-0.290	-0.294	-0.502*	-0.269**	-0.263**	-0.115	-0.114	-0.125	0.012
	[0.283]	[0.280]	[0.266]	[0.133]	[0.129]	[0.271]	[0.124]	[0.121]	[0.253]
Charter* Perf	-0.330	-0.340	0.273	-0.099	-0.101	0.159	0.096	0.100	0.594**
	[0.322]	[0.313]	[0.280]	[0.134]	[0.129]	[0.272]	[0.105]	[0.099]	[0.262]
Charter* Non_renew	0.566*	0.539	0.562*	0.392**	0.355**	0.542*	0.503***	0.483***	0.264
	[0.343]	[0.339]	[0.317]	[0.167]	[0.163]	[0.279]	[0.143]	[0.138]	[0.278]
Charter*Exam	-0.078	-0.065	-0.124	-0.180	-0.162	-0.366*	-0.226**	-0.195**	-0.188
	[0.176]	[0.174]	[0.184]	[0.115]	[0.113]	[0.204]	[0.096]	[0.093]	[0.201]
Charter* Eq_funding	-0.082	-0.068	-0.106	0.118	0.098	-0.038	-0.065	-0.069	0.004
	[0.177]	[0.177]	[0.177]	[0.088]	[0.089]	[0.167]	[0.076]	[0.076]	[0.177]
Charter	0.323	0.311	0.206	-1.052***	-0.968***	-0.290	-0.226	-0.181	-0.411
	[0.432]	[0.433]	[0.392]	[0.175]	[0.171]	[0.256]	[0.139]	[0.132]	[0.259]
R-squared	0.930	0.927	0.932	0.886	0.885	0.858	0.906	0.906	0.881
Observations	153,012	150,622	152,522	551,231	535,650	494,430	551,231	535,650	494,430
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State by year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents the effect heterogeneity analysis. *Nocaps* refers to No Caps on the growth of charter schools in a state; *Transpar* refers to Transparent Charter Application, Review, and Decision-making Processes; *Perf* refers to Performance-based Charter School Contracts Required; *Non_renew* refers to Clear Processes for Renewal, Nonrenewal, and Revocation Decisions; *Exam* refers to Automatic Exemptions from Many State and District Laws and Regulations; *Eq_funding* refers to Equitable Access to Capital Funding and Facilities. These NAPCS sub categories rangs from 0 to 1 and higher values indicate a better policy. Sample (1) is the full sample, Sample (2) drops the noncredible-charter districts, Sample (3) drops the always-charter districts.

Table 4a: Causal Forest Differences-in-Means – Graduation Rates

Covariate	Significantly Positive	Significantly Negative	Difference (Positive - Negative)
Log of Enrollment	7.23	6.98	0.26**
White (%)	0.82	0.48	0.35***
Black (%)	0.05	0.39	-0.34***
Hispanic (%)	0.09	0.05	0.04***
Free/Reduced Lunch (%)	0.25	0.48	-0.23***
Special Ed (%)	0.26	0.13	0.13***
Baseline Performance	0.83	0.57	0.26***
Urban	0.07	0.07	0
Suburb	0.22	0.14	0.08***
Town	0.22	0.17	0.06*
Rural	0.48	0.62	-0.14***
Magnet Schools (%)	0.18	0.01	0.17
City Population (standardized)	0.00	0.00	0
University in City	-0.01	0.01	-0.01**
Per Pupil Revenue	11411.08	8940.67	2470.41***
Student-Teacher Ratio	14.20	14.93	-0.73***
Teacher Salary	81975.53	65372.43	16603.09***
TPS-Charter Spending (% diff)	0.03	0.01	0.02***
Total Spending (per-pupil)	11480.57	9178.68	2301.89***
Equitable Funding	6.21	5.06	1.15***
No Caps on CS Growth	8.26	9.12	-0.86***
Performance-Based Contracts	8.31	9.75	-1.44***
Transparent Charter Startup Policies	7.98	10.16	-2.18***
Clear Charter Renewal Policies	10.54	11.91	-1.37***
Exempt from State/District Regs	7.51	8.04	-0.53***
Number of Observations	438.00	277.00	715

Table 4b: Causal Forest Differences-in-Means – Math Scores

Covariate	Significantly Positive	Significantly Negative	Difference (Positive - Negative)
Log of Enrollment	7.29	7.26	0.03
White (%)	0.74	0.78	-0.04***
Black (%)	0.07	0.06	0**
Hispanic (%)	0.12	0.09	0.03***
Free/Reduced Lunch (%)	0.45	0.42	0.04***
Special Ed (%)	0.14	0.15	-0.01***
Baseline Performance	0.28	0.40	-0.12***
Urban	0.03	0.04	0
Suburb	0.29	0.31	-0.02*
Town	0.17	0.16	0.02***
Rural	0.50	0.50	0
Magnet Schools (%)	0.09	0.34	-0.25***
City Population (standardized)	0.00	0.00	0
University in City	0.00	0.00	0***
Per Pupil Revenue	13857.58	14541.93	-684.35***
Student-Teacher Ratio	14.77	14.23	0.54***
Teacher Salary	95478.39	96551.46	-1073.08***
TPS-Charter Spending (% diff)	0.01	0.00	0.01***
Total Spending (per-pupil)	13697.81	14348.39	-650.59***
Equitable Funding	5.54	5.44	0.1*
No Caps on CS Growth	8.59	8.44	0.15***
Performance-Based Contracts	8.44	8.55	-0.11**
Transparent Charter Startup Policies	8.39	8.68	-0.29***
Clear Charter Renewal Policies	10.51	10.37	0.13***
Exempt from State/District Regs	7.21	6.93	0.28***
Number of Observations	6002.00	7786.00	13788

Table 4c: Causal Forest Differences-in-Means – ELA Scores

Covariate	Significantly Positive	Significantly Negative	Difference (Positive - Negative)
Log of Enrollment	7.30	7.32	-0.02
White (%)	0.77	0.74	0.03***
Black (%)	0.06	0.06	0
Hispanic (%)	0.10	0.12	-0.02***
Free/Reduced Lunch (%)	0.45	0.42	0.03***
Special Ed (%)	0.15	0.15	-0.01***
Baseline Performance	0.31	0.42	-0.11***
Urban	0.03	0.06	-0.03***
Suburb	0.27	0.35	-0.09***
Town	0.18	0.14	0.04***
Rural	0.52	0.45	0.07***
Magnet Schools (%)	0.07	0.29	-0.22***
City Population (standardized)	0.00	0.01	-0.01***
University in City	0.00	-0.01	0.01***
Per Pupil Revenue	13411.55	15221.81	-1810.26***
Student-Teacher Ratio	14.88	14.23	0.66***
Teacher Salary	92433.94	100392.11	-7958.17***
TPS-Charter Spending (% diff)	0.01	0.01	0
Total Spending (per-pupil)	13194.43	15038.02	-1843.59***
Equitable Funding	5.41	5.57	-0.16***
No Caps on CS Growth	8.55	8.54	0.02
Performance-Based Contracts	8.79	8.43	0.35***
Transparent Charter Startup Policies	8.39	8.60	-0.21***
Clear Charter Renewal Policies	10.60	10.45	0.15***
Exempt from State/District Regs	7.33	7.01	0.32***
Number of Observations	6454.00	7172.00	13626

Table 5a: CF Best Linear Projection – Graduation Rates

Term	Estimate	Std. Error	t-stat	p-value
(Intercept)	-1.847**	0.721	-2.563	0.010
Baseline Performance	0.932**	0.380	2.453	0.014
Log of Enrollment	0.105**	0.047	2.247	0.025
Special Ed (%)	4.19	3.087	1.358	0.175
Black (%)	-0.898	0.667	-1.347	0.178
White (%)	-0.178	0.384	-0.463	0.643
TPS-Charter Spending (% diff)	-0.033	0.282	-0.116	0.908
Free/Reduced Lunch (%)	0.296	0.306	0.966	0.334
Hispanic (%)	0.157	0.423	0.371	0.711
University in City	-0.53*	0.317	-1.673	0.094
Exempt from State/District Regs	-0.019	0.034	-0.561	0.575

Table 5b: CF Best Linear Projection – Math Scores

Term	Estimate	Std. Error	t-stat	p-value
(Intercept)	1.389	0.911	1.525	0.127
White (%)	-0.186	0.485	-0.384	0.701
Black (%)	0.097	0.568	0.170	0.865
Log of Enrollment	-0.115*	0.065	-1.775	0.076
Magnet Schools (%)	0.007	0.005	1.433	0.152
Hispanic (%)	0.082	0.450	0.182	0.856
TPS-Charter Spending (% diff)	0.211	0.193	1.093	0.274
Baseline Performance	0.079	0.084	0.939	0.348
Free/Reduced Lunch (%)	0.024	0.428	0.056	0.955
City Population (standardized)	0.177*	0.096	1.843	0.065
Per Pupil Revenue	0	0.000	-1.396	0.163

Table 5c: CF Best Linear Projection – ELA Scores

Term	Estimate	Std. Error	t-stat	p-value
(Intercept)	0.969**	0.395	2.450	0.014
Magnet Schools (%)	0.004	0.007	0.630	0.528
White (%)	-0.58**	0.229	-2.532	0.011
Log of Enrollment	-0.008	0.032	-0.247	0.805
Baseline Performance	-0.017	0.047	-0.359	0.720
Black (%)	-0.576**	0.246	-2.342	0.019
Per Pupil Revenue	0	0.000	-1.112	0.266
Special Ed (%)	-2.143	1.328	-1.614	0.107
Total Spending (per-pupil)	0	0.000	0.627	0.530
Hispanic (%)	-0.745***	0.237	-3.148	0.002
TPS-Charter Spending (% diff)	0.301***	0.109	2.772	0.006