COURT-ORDERED FINANCE REFORMS IN THE ADEQUACY ERA: HETEROGENEOUS CAUSAL EFFECTS AND SENSITIVITY[*]

Christopher A. Candelaria[†] and Kenneth A. Shores[‡]

[†]Vanderbilt University
[‡]University of Pennsylvania

March 21, 2017

Abstract

We provide new evidence about the effect of court-ordered finance reforms that took place between 1989 and 2010 on per-pupil revenues and graduation rates. We account for heterogeneity in the treated and counterfactual groups to estimate the effect of overturning a state's finance system. Seven years after reform, the highest poverty quartile in a treated state experienced a 11.5 to 12.1 percent increase in per-pupil spending and a 6.8 to 11.5 percentage point increase in graduation rates. We subject the model to various sensitivity tests, which provide upper and lower bounds on the estimates. Estimates range, in most cases, from 6 to 12 percentage points for graduation rates. *JEL* codes: C23, I26

Keywords: School Finance, Adequacy Era, Differences-in-Differences, Correlated Random Trends, Cross-Sectional Dependence

1 Introduction

Kentucky's 1989 Supreme Court ruling *Rose v. Council for Better Education* marked a new era of State court-ordered finance reforms (Murray, Evans, & Schwab, 1998; Berry, 2007; Corcoran & Evans, 2008; Sims, 2011). At least two factors distinguish these court rulings from rulings that took place in the decades prior. First, the post-*Rose* court cases were increasingly argued on adequacy grounds, meaning that plaintiffs were no longer able to successfully argue that all children had a right to an equal education. Adequacy-based language required that students received sufficient resources to meet academic standards laid out by the state. Moreover, three years after the *Rose* case, the gap in per pupil spending between low and high income districts narrowed noticeably and continued to close (Corcoran, Evans, Godwin, Murray, & Schwab, 2004; Jackson, Johnson, & Persico, 2016). How much this change in spending should be attributed to Adequacy cases or other secular changes is not known. Whether these rulings had an impact on the level and distribution of resources and academic outcomes is, therefore, an unresolved, open question.

Second, these cases took place in an era when many states had already undergone substantial changes to their funding systems. As Jackson et al. (2016) note, by 1990, 10 states had court-mandated reforms, 30 states had legislative reforms, and 39 states had changed their funding formula. Moreover, per pupil expenditures increased by about a factor of 1.6 between 1972 and 1992. While spending gaps between high and low income districts persisted during this period, spending increases were, for the most part, evenly split between high and low income districts. Against the backdrop of these changes in spending dynamics—increased spending for low income students and a greater emphasis on equity across most states—it is not known whether additional aid to low income districts would produce additional academic benefits.

From an economic standpoint, the causal relationship between spending and desirable outcomes is also of interest because the share of GDP that the United States spends on public elementary and secondary education has remained fairly large, ranging between 3.4 to 4.2 percent since the 1970s.[1] Given the large share of spending on education, it would be useful to know if these

---

[1]Estimates based on authors' calculations using data from Tables 106.10 and 106.20 from the Digest of Education

resources are well spent. The goal of this paper is to provide a robust description of the causal relationship between fiscal shocks at the state level and student outcomes at the district level between academic years 1990–91 and 2009–10.

Our paper is therefore motivated by two overlapping research questions:

1. Did Adequacy-based court rulings in favor of plaintiffs lead to changes in the level of spending for low-income students and the distribution of spending within a state, relative to changes taking place in the rest of the country?

2. To the extent these Adequacy-based rulings did change the level and distribution of spending, did these changes lead to improvements in academic outcomes for low-income students?

Using district-level, aggregate data from the Common Core (CCD), we estimate the causal effects of court-ordered finance reforms on the natural logarithm of real per-pupil revenues and graduation rates. We specifically analyze court decisions that overturned states' school finance systems based partly or fully on adequacy grounds. The academic outcome of interest is high school graduation rates, which are the most distal academic outcome that can be observed for students living in states undergoing Adequacy-based court order in this period. For this reason, high school graduation rates are one of the best indicators of the effects of court ordered finance reform during this period.

To identify causal effects, we estimate a heterogeneous differences-in-differences model that accounts for a poverty quartile-by-year secular trend and state-by-poverty quartile linear time trends, which serves as our benchmark specification. Using this specification, we find that high poverty districts in states that had their finance regimes overthrown by court order experienced an increase in real per pupil revenues by approximately 11.5 to 12.1 percent and graduation rates by approximately 6.8 to 11.5 percentage points seven years following reform.

We then subject the model to various sensitivity tests by permuting secular time trends, correlated random trends (unit-specific trends), and, to account for cross-sectional dependence, interactive fixed effects. In total, we estimate 15 complementary models. Generally, the results are robust

---

Statistics, 2014 edition.

to model fit: relative to the benchmark model, interactive fixed effects and alternative specifications of the secular time trend have modest effects on point estimates, while excluding correlated random trends eliminates the effect on graduation rates. Estimating the linear time trend using different levels of aggregation (the state and district) produces estimates that are similar to those of the preferred model, as does allowing the trend to be quadratic. There is evidence that treatment and control sub-state units have different secular trends, but conditionally exogenous point estimates are available under the assumption that sub-unit pre-treatment trends can be approximated with a functional form.

Further exploration of the data reveal that (a) the marginal effect of an increase in per-pupil log revenues increased graduation rates in poorer districts but not wealthier districts; (b) these reforms equalized spending and graduation rates within states; (c) and that the pattern of results identified in the aggregate model vary substantially for individual states undergoing court order, as some states received additional revenues without a corresponding change in graduation rates (and vice-versa).

In summary, this paper explores the heterogeneous causal impact of Adequacy-era court ordered finance reform. We find that Adequacy-based court cases overturning a state's financial system for the period 1990-91 to 2009-10 had an effect on revenues and graduation rates, that these results are robust to a wide variety of modeling choices, that the effect was equalizing, and that states responded differently to court order. The heterogeneity in how states responded to court order motivates further inquiry. While on average, we observe positive effects in both spending and graduation rates, we also observe states that were able to improve academic outcomes without increases in spending, as well as states that failed to make uses of increased spending to improve academic outcomes. A better understanding of these state-specific contexts produces a fruitful area for future research, especially if we wish to understand the circumstances in which fiscal changes can be more or less productive for student academic outcomes.

2   Background on Court-Ordered Finance Reforms

Beginning in the 1970s, unequal education funding between districts within states led plaintiffs to sue state governments in an attempt to equalize funding. These disparities were largely the result of differences in the local share of education funding sourced by a district's property tax base. In almost all cases, plaintiffs argued that the funding disparities were unconstitutional based on equity grounds. They believed that the state was responsible for providing equal spending per pupil across districts so that all children had an equal opportunity for success in the public education system. However, by the mid to late 1980s, the equity argument was not very successful in persuading the court justices (Heise, 1995; West & Peterson, 2007; Baker & Green, 2008; Koski, 2009; Springer, Liu, & Guthrie, 2009).

Starting with the 1989 *Rose v. Council for Better Education* ruling in Kentucky, plaintiffs began challenging the constitutionality of state school financing systems based on adequacy grounds. In order to benefit from public education, the plaintiffs believed it was the state's responsibility to provide an adequate level of funding so that all students could attain the state's minimum threshold for proficiency on various academic and non-academic outcomes. The *Rose* case marked the beginning of the so-called "Adequacy Era;" the majority of cases that followed were brought to state courthouses based on adequacy grounds.

The effects of Adequacy court cases on the level and distribution of spending, relative to so-called "Equity" cases that took place prior to 1989, have been somewhat mixed. One comparison of interest is whether Adequacy or Equity cases differentially affected the variation in spending across districts within states. Using data up to year 2002, a number of papers failed to find differences between these types of cases (Springer et al., 2009; Berry, 2007; Corcoran & Evans, 2008). However, variation in spending might be unaffected by these policies if spending between low and high income districts is already equal, and if the effect of these reforms is to channel additional resources to low income areas; in fact, variation may increase as a result of these reforms.

Recognizing this possibility, Sims (2011) and Corcoran & Evans (2015) compare the effect

these cases had on spending in low and high income (or high and low poverty) districts. They find that these Adequacy cases increased spending in higher poverty districts, and that more resources were allocated to districts based on observable indicators of student need, such as free lunch eligibility (FLE) status and the percents of students who are minority. Lafortune, Schanzenbach, & Rothstein (2016) extend this sample to year 2013 and find complementary results, although the authors include non-Adequacy cases as well as legislative reforms in their evaluation.

As school spending increased in predominantly low income districts, it was natural to ask whether that increased spending positively affected student outcomes. A number of papers have attempted to identify causal effects of spending on achievement, graduation and earnings (Card & Payne, 2002; Hoxby, 2001; Jackson et al., 2016; Lafortune et al., 2016). While the specific econometric methods vary, each of the studies included here assume that court rulings result in exogenous shocks to the state's education finance system. The results of Card & Payne (2002) and Hoxby (2001) were in conflict, but these were hampered by data limitations, as only a simple pre-post- contrast between treatment and control states was available to the researchers, thereby limiting their capacity to verify the identifying assumptions of the differences-in-differences model.

Jackson et al. (2016) construct a much longer panel (with respect to time), using restricted-use individual-level data reaching back to children born between 1955 and 1985, to test the effects of these cases on revenues, graduation rates and adult earnings. Leveraging variation across cohorts in exposure to fiscal reform, this study finds large effects from court order on school spending, graduation and adult outcomes, and these results are especially pronounced for individuals coming from low-income households and districts.

Lafortune et al. (2016) investigate the effects of 68 post-1990 legislative and court-ordered reforms on student academic outcomes using data from the National Assessment of Educational Progress (NAEP). The authors find that spending and achievement gaps narrowed between high income and low income school districts in states undergoing reform during this period, and that an extra $1,000 in per-pupil spending raises achievement scores 10 years later by 0.18 standard deviations.

In this paper, we make three contributions to the school finance literature.

1. We examine the effects of Adequacy-based court cases taking place between 1989 and 2010 on revenues and graduation rates. By including cases and data up to year 2010 and investigating the effects of these cases on graduation rates, this analysis updates and extends the work of (Sims, 2011). This analysis also complements the work of (Lafortune et al., 2016). LaFortune and colleagues identify proximal effects of post-1990 finance reform on student achievement outcomes, while we examine the most distal outcomes available for students in this era by looking at graduation rates.

2. We subject our preferred econometric model to a bevy of sensitivity analyses. We present results that are robust to modeling choices that account for secular trends, correlated random trends, and cross-sectional dependence. In so doing, we provide upper and lower bounds on effect sizes by permuting these variables and demonstrate the plausibility of the exogeneity of these court rulings.

3. We detail the heterogeneity of these effects between poverty quantiles and treated states. We compare the impact of these cases between high poverty districts in treated and untreated states, as well estimate the extent to which these reforms changed the distribution of spending between treated and non-treated states. We then detail the heterogeneous responses to court order across all of the treated states in our sample. Exploring the state-specific heterogeneity helps both to unpack the average treatment effect, as well as provide local context for each of these court cases.

3 Data

The data set is the compilation of several public-use surveys that are administered by the National Center for Education Statistics and the U.S. Census Bureau. The analytic sample is constructed using the following data sets: Local Education Agency (School District) Finance Survey (F-33); Local Education Agency (School District) Universe Survey; Local Education Agency Universe

Survey Longitudinal Data File: 1986-1998 (13-year); Local Education Agency (School District) Universe Survey Dropout and Completion Data; and Public Elementary/Secondary School Universe Survey.[2]

Data are available beginning in the 1990-91 school year up to the 2009-10 school year. The data set is a panel of aggregated data, containing United States district and state identifiers, indexed across time. The panel includes the following variables: counts of free lunch eligible (FLE) students, per pupil log and total revenues, percents of 8th grade students receiving diplomas 4 years later (graduation rates), total enrollment, percents of students that are black, Hispanic, minority (an aggregate of all non-white race groups), special education, and children in poverty. Counts of FLE students are turned into district-specific percentages, from which within state rankings of districts based on the percents of students qualifying for free lunch are made. Using FLE data from 1989-90, we divide states into FLE quartiles, where quartile four is the highest poverty quartile for the state.[3] Total revenues are the sum of federal, local, and state revenues in each district. The variable total revenues is then divided by the total number of students in the district to provide a per pupil estimate; deflated by the US CPI, All Urban Consumers Index to convert the figure to real terms; and then converted to the natural logarithm. Graduation rates are defined as the total number of diploma recipients in year t as a share of the number of 8th graders in year $t - 4$, a measure which Heckman & LaFontaine (2010) shows is not susceptible to the downward bias caused by using lagged 9th grade enrollment in the denominator. Graduation rates are top-coded, so that they take a maximum value of 1.[4] The demographic race variables come from the school-level file from the Common Core and are aggregated to the district level; percents are calculated by dividing by total enrollment. Special education counts come from the district level Common Core of Data. Child poverty is a variable we take from the Small Area Income and Poverty Estimates (SAIPE).

To define the analytic sample, we apply the following restrictions. First, Hawaii and the District

---

[2]Web links to each of the data sources are listed in Appendix A.1.

[3]Missing FLE data for that year were imputed by NCES interpolation methods. Data are found on Local Education Agency Universe Survey Longitudinal Data File (13-year). These quartiles are derived from the percentages of Free Lunch Eligible (FLE) students reported at the district level in each state in 1990. The year is fixed at the start of our sample because the poverty distribution could be affected by treatment over time.

[4]In Appendix A.2, we describe additional details about variable construction.

of Columbia are removed from the sample, as each place has only one school district. Montana is also removed because the state is missing a substantial amount of graduation rate data. Only unified districts are included so that it is possible to link changes in revenues to graduation rates at a later data. Unified districts are defined as those districts that serve students in either Pre-Kindergarten, Kindergarten, or 1st grade through the 12th grade. For the variables total enrollment, graduation rates, and FLE, NYCPS reports its data as 33 geographic districts in the nonfiscal surveys; for total revenues, NYCPS is treated as a consolidated district. For this reason, the non-fiscal data are combined into a single district. As suggested in the NCES documentation, NYCPS's supervisory union number is used to aggregate the geographical districts into a single entity. Appendix A.3 outlines our approach to addressing data errors in our analytic sample.

To the analytic sample, we add the first year a state's funding regime was overturned in the Adequacy Era. The base set of court cases comes from Corcoran & Evans (2008), and were updated using data from the National Education Access Network.[5] Table 1 lists the court cases we are considering. As shown, there are a total of thirteen states that had their school finance systems overturned during the Adequacy Era.[6] Kentucky was the first to have its system overturned in 1989 and Alaska was the most recent; its finance system was overturned in 2009.

[Insert Table 1 Here]

Table 2 provides summary statistics of the key variables in the analytic sample.[7] The analytic file consists of 188,752 district-year observations. The total number of unified districts is 9,916. The average graduation rate is about 77 percent and average log per pupil spending is 8.94 (total real per pupil revenues are about $7,590). Non-weighted summary statistic provide are slightly larger in magnitude.

[Insert Table 2 Here]

---

[5]The National Education Access Network provides up-to-date information about school finance reform and litigation. Source: `http://schoolfunding.info/`.

[6]As previously mentioned, we drop Montana from the analysis due to inadequate graduation rate data.

[7]We drop New York City Public Schools from the analytic sample because it is an outlier district in our data set. We provide a detailed explanation of why we do this in Appendix B.1.

4   Econometric specifications and model sensitivity

In this section, we describe our empirical strategy to estimate the causal effects of court-ordered finance reform at the state level on real log revenues per student and graduation rates at the district level. We first specify a fully non-parametric event study model, which allows us to test whether state level Supreme Court rulings constitute shocks that affect revenues and graduation rates for each poverty quartile across the states. The event study model, also known as a Granger-style model (Angrist & Pischke, 2008), allows us to examine the dynamic nature of a treatment effect, both in the preceding and subsequent years of treatment.

Based on the findings from the event study, we then posit our benchmark model, which is a differences-in-differences equation that adjusts for poverty-by-year fixed effects and unit-specific time trends, or correlated random trends (Wooldridge, 2005, 2010). The poverty-by-year effects allow us to test whether court order improved outcomes in high-poverty (and remaining poverty quartile) districts relative to other high-poverty (and remaining poverty quartile) districts in states without reform. Unit-specific trends account for downward pre-treatment trends in graduation rates among high-poverty districts that were observed in the event study.

Finally, because treatment occurs at the state level and our outcomes are at the district level, there are several ways one may choose to specify the estimating equation. For example, there are choices about whether and how to model the counterfactual time trend, and how to adjust for correlated random trends (i.e., pre-treatment trends) and unobserved factors, such as cross-sectional dependence. We outline these alternative modeling choices and discuss their implications relative to the benchmark model. We then estimate 15 different models in which we permute these model specifications, the result of which is that our benchmark model is, in most cases, insensitive to these modeling choices. This suggests that the exogeneity of the timing of court-ordered finance reform is defensible.

9

4.1 Event Study: Testing the Exogeneity of Court Rulings

We first assess the exogeneity of the court rulings, as these will be treated as shocks to identify the causal effects of finance reform. Using an event study model, it can be shown whether it is appropriate to treat adequacy-based decisions overturning states' school finance systems as exogenous shocks to both revenues and graduation rates. In order to model the heterogeneous impact of these cases across districts within states, treatment parameters are indexed by state-specific poverty quartiles, defined using free lunch eligibility status.[8]

In the event study, overturning a state's school finance system defines treatment, and the year of the court ruling defines the potential shock. Using binary indicator variables, the dynamic treatment response is estimated non-parametrically in the years before and after the court ruling decision date. This approach captures both anticipatory and dynamic treatment effects, respectively. Although some states had multiple court decisions favoring the plaintiff, only the first Adequacy ruling in favor of plaintiffs is counted as treatment, as the effects of these subsequent cases are not identified.

Because states were subject to reform in different years and not all states had reform, a set of both treatment and control districts identify the parameters in the event study model. If the coefficients on the indicators of the anticipatory effects do not exhibit any systematic patterns or they are indistinguishable from zero, then there is suggestive evidence that treatment is exogenous. Failure to find systematic patterns and statistically significant estimates suggests that treatment and control groups were following similar trends prior to treatment. In a regression framework, the event model takes the following form:

$$(1) \quad Y_{sqdt} = \theta_d + \delta_{qt} + \sum_{n=1}^{16} \gamma_{(q,-n)}[\mathbb{1}(Q_{sd} = q) \times \mathbb{1}(t - t_s^* = -n) \times D_s]$$

$$+ \sum_{n=1}^{19} \gamma_{(q,+n)}[\mathbb{1}(Q_{sd} = q) \times \mathbb{1}(t - t_s^* = n) \times D_s] + \varepsilon_{sqdt},$$

where Y_{sqdt} is an outcome of interest—real log revenues per student or graduation rates—in state

---

[8]This is the approach taken by Sims (2011).

s, in poverty quartile q, in district d, in year t; θ_d is a district-specific fixed effect; δ_{tq} is a time by poverty quartile-specific fixed effect; D_s is a binary indicator denoting whether state s had a reform; t_s^* is the first year of the reform in state s; indicator function $\mathbb{1}(Q_{sd} = q)$ takes value 1 when district d in state s is in quartile q, where $q \in \{1, 2, 3, 4\}$; and ε_{sdt} is assumed to be a mean zero, random error term. In the model, $Q_{sd} = 4$ represents the highest poverty districts within a state. To account for serial correlation, all point estimates have standard errors that are clustered by state, the level at which treatment occurs (Bertrand, Duflo, & Mullainathan, 2004). The model is estimated using Stata's reghdfe command (Correia, 2014).

The parameters of interest are the $\gamma_{q,n}$, which are the estimates of the effect of school finance reform in quartile q in treatment year n on Y_{sqdt}. In total, the model provides estimates for 16 anticipatory effects and 19 post-treatment effects for each poverty quartile. Although there are 21 post treatment effects available, treatment effect years 19 though 21 are combined into a single binary indicator, as there are only two treatment states that contribute to causal identification in these later years.[9] In reporting our results, only estimates for years 1 through 7 before and after reform are reported, as estimates outside this range suffer from precision loss. For example, very few states were treated early enough to contribute information to the post-treatment effect estimates in later years (see Table 1). Because we include δ_{tq} fixed effects, the estimates of $\gamma_{q,n}$ are identified using the variation within quartile q by year t cells; consequently, results are not comparable across the quartiles.[10]

Figure 1 plots the results of the event study model for the logarithm of real per pupil revenues for each of the quartiles. For comparison, ordinary least squares (OLS) (gray lines with hollow triangles) are plotted alongside weighted least squares (WLS) results (black lines with hollow circles), where the weights are time-varying district enrollment. The figure shows that there are no discernible trends in the log of real per pupil revenues in the years before a court decision, and almost all the anticipatory effects are statistically indistinguishable from zero at the 10 percent level.

---

[9]While our data sample has 20 years of data, we have up to 21 potential treatment effects, as KY had its reform in 1989 and our panel begins in the 1990-91 academic year. Therefore, KY does not contribute to a treatment effect estimate in the first year of reform, but it does contribute to effect estimates in all subsequent treatment years.

[10]We discuss the implications of this decision in the sections that follow.

In the years after court-ordered finance reform, real per pupil revenues increased in all quartiles. In summary, here there is evidence that the timing of the court decisions was unpredictable (i.e., there are no anticipatory effects) and that court order increased revenues.

[Insert Figure 1 Here]

Results for graduation rates are shown in Figure 2. With respect to the anticipatory effects, there are no statistically significant estimates at the 10 percent level across all the quartiles; however, there is an observable downward trend for the high poverty districts in quartile four, which flattens out three years prior to court order. This can indicate that the timing of treatment is correlated with the slope of the graduation rate in treated districts; for example, states may have been more likely to pass court ordered finance reform because they observed downward trending graduation rates in high poverty districts. Such an occurrence would be a violation of the common trends assumption for differences-in-differences estimation, which is that, in the absence of treatment, treated and control units have similar secular trends. In terms of treatment effects, there is evidence that there was an effect in quartiles 2 through 4, as each has an upward trend in the years after treatment, and results are significant in some years. There is no evidence of an effect in quartile 1, as the trend is flat and effects are not significant.

[Insert Figure 2 Here]

The results of the event study suggest that the court rulings were an exogenous shock to revenues, but that the adoption of court order may have been preceded by downward trends in graduation rates among treated districts in poverty quartile four. We proceed by estimating a traditional differences-in-differences specification that restricts the pre-treatment period to zero. To account for the potentially heterogeneous secular trends between treated and control units, linear correlated random trends are included for each state by poverty quartile group.[11]

---

[11]In Section 5.2 (Model Sensitivity), quadratic correlated random trends are included as well, to allow for a more flexible functional form.

4.2 Benchmark differences-in-differences model

The preferred benchmark model takes the following form:

$$(2) \quad Y_{sqdt} = \theta_d + \delta_{qt} + \psi_{sq}t + \sum_{n=1}^{19} \gamma_{(q,+n)}[\mathbb{1}(Q_{sd} = q) \times \mathbb{1}(t - t_s^* = n) \times D_s] + \varepsilon_{sqdt},$$

pre-treatment period is restricted to be zero, and the variables $\psi_{sq}t$ are added, which are a state by quartile-specific linear time trend. As before, serial correlation is accounted for by clustering standard errors at the state level. (Bertrand et al., 2004). All together, this model absorbs approximately 9,800 district fixed effects, 76 year effects, and 192 continuous fixed effects (state-by-FLE quartile interacted with linear time), and estimates 76 treatment variables for years 1-19 and quartiles one through four. Wolfers (2006) shows that including $\psi_{sq}t$ can introduce bias when only a singular post-treatment effect is estimated, but this bias can by eliminated if the entire treatment response is estimated non-parametrically. For this reason we include 76 treatment variables for 19 treatment years and four poverty quartiles. The variables for Equation (2) are included for different reasons, which are explained below.

4.2.1 Secular time trends

In Equation (2), secular time trends are modeled with FLE quartile-by-year fixed effects, denoted as δ_{tq}. These δ_{tq} fixed effects are included instead of standard year fixed effects δ_t to establish a more plausible counterfactual trend for treated districts. The main reason is that higher poverty districts in both treated and untreated states were increasing in revenues and graduation rates; therefore, by indexing the year fixed effects with FLE quartiles, treated districts in a given quartile are compared to non-treated districts in the same quartile.

4.2.2 Correlated random trends

In Equation (2), the parameter $\psi_{sq}t$ is included because there was some evidence that the timing of a state's ruling was correlated with trends in graduation rates among high poverty districts. Including $\psi_{sq}t$ adjusts for all state-by-quartile secular trends, but assumes that these trends follow

a functional form. Equation (2) assumes the functional form is linear, but this assumption can be tested, as described in Section 5.2: Model Sensitivity.

4.2.3 Weighting

In Equation (2), treatment heterogeneity is explicitly modeled by disaggregating treatment effects into four poverty quartiles. Although treatment heterogeneity is modeled across poverty quartiles, other sources of treatment heterogeneity may be missed. Weighting a regression model by group size is traditionally used to correct for heteroskedasticity, but it also provides a way to test for the presence of additional unobserved heterogeneity. According to asymptotic theory, the probability limits of OLS and weighted least squares should be consistent. Thus, regardless of how you weight the data, the point estimates between the two models should not dramatically differ. When OLS and WLS estimates do diverge substantially, there is concern that the model is not correctly specified, and it may be due in part to unobserved heterogeneity associated with the weighting variable (DuMouchel & Duncan, 1983; Solon, Haider, & Wooldridge, 2015).

5 Results

We present our results in six parts. First, we present the causal effect estimates of court-ordered finance reforms using our preferred differences-in-differences model. Second, we examine the extent to which the point estimates from the benchmark model are sensitive to assumptions about secular trends, correlated random trends, and cross-sectional dependence. Third, we assess whether reforms were equalizing across the FLE poverty distribution; we wish to test formally whether, within treated states, poorer districts benefited more relative to richer districts in terms of revenues and graduation rates. Fourth, we estimate two-stage least squares regressions to isolate the causal effect of school spending on graduation rates. Fifth, we disaggregate treatment effects for each state with a court order during this period. Finally, we conclude with a series of robustness checks that allow us to gauge the validity of our causal estimates.

5.1 Benchmark differences-in-differences model results

5.1.1 Revenues

The causal effect estimates of court-ordered finance reform on the logarithm of per pupil revenues and graduation rates are reported in Tables 3 and 4, respectively. FLE quartile 1 represents low-poverty districts, and FLE quartile four represents high-poverty districts. Only treatment effect years 1 though 7 are reported because the number of states in the treated group changes dramatically over time. Six years after treatment, Alaska and Kansas no longer contribute treatment information; in years 7 and 8, we lose North Carolina and New York.[12] Both weighted and unweighted estimates are displayed, where the weight is time-varying district enrollment.

Results posted in Table 3 reveal that court-ordered finance reforms increased revenues in all FLE quartiles in the years after treatment, though not every point estimate is significant at conventional levels. Because the model includes FLE quartile-by-year fixed effects, point estimates for a given quartile are interpreted relative to other FLE quartiles that are in the control group. For example, in year 7 after treatment, districts in FLE quartile 1 had revenues that were 12.8 percent higher than they would have been in the absence of treatment, with the counterfactual trend established by non-treated districts in FLE quartile 1 (significant at the 5 percent level). In FLE quartile four, we find that the revenues were 11.5 percent higher relative to what they would have been in the absence of reform, relative to non-treated districts in FLE quartile four (significant at the 1 percent level).[13]

The unweighted results in Table 3 also suggest that revenues increased. Although the magnitudes differ slightly, the coefficient estimates are quite similar to the corresponding weighted results. Overall, revenues increased across all FLE poverty quartiles in states with Court order,

---

[12]Throughout the rest of paper we restrict the description of our results for years less than or equal to 7, though estimation occurs for the entire panel of data. Appendix Table E.1 displays states with court order and the treatment year the state exits.

[13]One point to note: it would be wrong to conclude that the 12.8 percent effect is larger than the 11.5 percent effect, because the 12.8 percent effect for quartile 1 may be relative to only a modest increase in non-treated low poverty districts, whereas the 11.5 percent effect for quartile four may be relative to a steep increase in non-treated high poverty districts. Later, we present models and results to test whether poorer districts received greater revenues and graduation rates following reform.

relative to equivalent poverty quartiles in non-treated states.

[Insert Table 3 Here]

5.1.2 Graduation rates

With respect to graduation rates, the weighted results in Table 4 show that court-ordered finance reforms were consistently positive and significant among districts in FLE quartile four. In the first year after reform, graduation rates in quartile four increased modestly by 2.0 percentage points. By treatment year 7, however, graduation rates increased by 11.5 percentage points, which is significant at the 0.1 percent level. Each treatment year effect corresponds to a different cohort of students. Therefore, the dynamic treatment response pattern across all 7 years is consistent with the notion that graduation rates do not increase instantaneously; longer exposure to increased revenues catalyzes changes in academic outcomes. There is modest evidence that FLE quartiles 2 and 3 improved graduation rates following court order, though these point estimates are not consistently significant and are smaller in magnitude than those in FLE 4. The lowest-poverty districts in FLE quartile 1 have some significant effects, between one to three percentage points, but the point estimates show no evidence of a strong, upward trend over time.

The unweighted graduation results in Table 4 tell a similar story as the weighted results. One key difference is that the point estimates tend to be smaller in FLE quartile four. For example, graduation rates are 6.8 percentage points higher in year 7 than they would have been in the absence of treatment. This reduction in effect size for unweighted results applies only to FLE quartile four.[14]

[Insert Table 4 Here]

5.2 Model Sensitivity

Here we outline alternative model specifications. These alternatives provide checks on the exogeneity assumptions of the differences-in-differences model, as well as provide upper and lower

---

[14]See Appendix B.1 for an extended discussion of how to interpret differences between weighted and ordinary least squares estimates.

bounds for how much point estimates depart from the preferred model. We describe the alternative model specifications below.

5.2.1 Alternative model specifications: Secular time trends

Replacing the secular trend δ_{tq} with δ_t models the counterfactual time trend as the average trend among all never-treated districts and those districts that are awaiting treatment in a given time period. Specifying the model with δ_t corresponds to a more traditional differences-in-differences specification of secular time trends in panel data models, allowing for comparisons of treatment effect estimates across the FLE poverty quartiles.

5.2.2 Alternative model specifications: Correlated random trend

The correlated random trend parameter $\psi_{sq}t$ is replaced with one of the following parameters: $0, \psi_s t, \psi_d t, \psi_{sq}t^2, \psi_s t^2$, where 0 implies the absence of a correlated random trend. That is, we either do not estimate a pre-treatment trend, we allow the pre-treatment trend to be estimated at the state and district levels, or we allow the time element to have a quadratic functional form.

5.2.3 Alternative model specifications: Cross-sectional dependence

We estimate models in which we assume our error term has a factor structure denoted by $\lambda_s' F_t$. Following Bai (2009), we define λ_s as a vector of factor loadings and F_t as a vector of common factors. Each of these vectors is of size r, which is the number of factors included the model. For our sensitivity tests, we estimate models in which the number of included factors r is an element of the set $\{1, 2, 3\}$.[15] To estimate the $\lambda_s' F_t$ factor structure in Equation (2), we use the method of principal components as described by Bai (2009) and implemented by Moon & Weidner (2015) and Gomez (2015b).[16]

---

[15]Moon & Weidner (2015) show that point estimates stabilize once the number of factors r equals the true number of factors r° and that when $r > r^\circ$, there is no bias. However, this is only true when the time dimension t in the panel approaches infinity. When t is small, it is possible to increase bias by including too many factors. See Table IV in their paper, as well as Onatski (2010) and Ahn & Horenstein (2013).

[16]We describe the basic steps of the procedure in Appendix C.

In the differences-in-differences framework, the factor structure has a natural interpretation. Namely, the common factors F_t represent macroeconomic shocks that affect all the units (e.g., recessions, financial crises, and national policy changes), and the factor loadings λ_s capture how states are differentially affected by these shocks. Of particular concern is the presence of inter-dependence, which can result if one state's Supreme Court ruling affects the chances of another state's ruling. This would violate the identifying assumptions of the differences-in-differences model and result in bias, unless that interdependence is accounted for (Bai, 2009; Pesaran & Pick, 2007). Additional details about factor structure estimation are in Appendix C.

5.2.4 Model Sensitivity: Results

The preferred model indicates a meaningful positive and significant effect of court order on the outcomes of interest. To examine model sensitivity, we focus attention on districts in the highest poverty quartile (i.e., FLE quartile four). The evidence suggests there was an effect for graduation rates and revenues for the FLE quartile four districts, but these results assume that the benchmark model makes correct assumptions about secular trends, correlated random trends, and cross-sectional dependence. We now assess the extent to which results are sensitive to these modeling choices.

Figures 3 and 4 plot the causal effect estimates of finance reform on the logarithm of real per pupil revenues and graduation rates, respectively, across a variety of model specifications. We report regression tables with point estimates and standard errors in the Appendix.[17] While not all possible combinations are plotted, the 13 models shown here do provide insight into sensitivity of the causal effect estimates.

In both figures, our benchmark differences-in-differences model corresponds to the second model in the legend, which is denoted by a thick, dashed line with a hollow diamond marker symbol and the following triple:

---

[17]For log per pupil revenues, see Appendix Tables E.2 and E.3 report weighted results and unweighted results, respectively. For graduation rates, see Appendix Tables E.4 and E.5 report weighted and unweighted results, respectively. Point estimates and standard errors for these models are presented in the Appendix.

ST: FLE by year; CRT: State by FLE; CSD: 1 (semicolon is the delimiter).

ST refers to the type of secular trend, which can be either FLE-by-year or year fixed effects. CRT refers to the type of correlated random trend in the model, which can include state-, state by FLE quartile-, or district-specific trends, along with the quadratic form of state by FLE quartile. CSD refers to the number of factors included to account for cross-sectional dependence. A model with factor number 0 does not account for cross-sectional dependence, while models with 1, 2, or 3 account for models with 1, 2, and 3 factors, respectively.[18]

Results from Figure 3 reveal that, on average, the benchmark model closely resembles the causal effects on revenues in FLE quartile from other specifications. For WLS models, the benchmark results are lower, except for models that include quadratic forms of the correlated random trends or three factors. For OLS, the benchmark results are once again higher than models that include quadratic state-by-FLE time trends and factors. The largest effects in both the weighted and unweighted regressions are produced by a specification that includes year fixed effects for the secular trend, a correlated random trend at the level of the state, and does not adjust for cross-sectional dependence (indicated in Figure 3 by gray hollow diamonds with a short dash line).

It may be illuminating to compare gray and black hollow diamonds in Figure 3, as these reflect models that ignore correlated random trends and CSD but differ in how they estimate the counterfactual trends. Models corresponding to hollow gray diamonds assume the counterfactual trend is homogeneous, while models corresponding to hollow black diamonds assume the counterfactual trend is heterogeneous, indexed by poverty quartile. Hollow gray diamonds are consistently larger than hollow black diamond point estimates, for both weighted and unweighted regressions. As hypothesized, this indicates that the assumption of a homogeneous counterfactual trend (δ_t) overestimates the treatment effect for high poverty districts, since revenues were increasing faster in high poverty districts in non-treated states relative to low poverty districts in non-treated states.

[Insert Figure 3 Here]

Figure 4 reports the variability of effect sizes for graduation rate estimates. For WLS estimates,

---

[18]Bracketed numbers indicate the column location for those model estimates available in Appendix Tables E.2, E.3, E.4, E.5.

the benchmark model overstates the causal effects on graduation rates in FLE quartile four relative to all other graduation rate models. Of particular interest is the influence of correlated random trends. In general, the point estimates for graduation rates cluster together, within the range of 6 to 12 percentage points at treatment year 7, including models that account for up to three factors and allow for the unit-specific time trend to be quadratic. Only when the correlated random trend is excluded do point estimates drift downwards so that they are small and insignificant (indicated by gray hollow diamonds and a dotted line).

By looking at the level of aggregation of the correlated random trend, it is possible to learn something about the nature of the pre-treatment trends in the dependent variable. Looking only at models that adjust for year by FLE effects (i.e., black markers and lines), point estimates for models that include a state-level time trend (indicated by the hollow diamond with short dashed line) are significant and nearer to our preferred model but still differ by about 5 percentage points. Including district-by-year effects (approximately 9,800 linear time trends, indicated by the hollow diamond and dashed line) the point estimates nearly coincide with the preferred model. This suggests that treatment and control groups do have different pre-treatment trends, and that these differences largely occur at the sub-state level.

Finally, note that allowing the time trend to be quadratic has very little effect on the model, even though the pre-treatment effects for graduation rates flattened three years prior to treatment. This suggests that mis-specifying the functional form has not introduced bias.

There is little to no evidence of cross-sectional dependence in the timing of these Court cases. Recall that CSD can occur if there is heterogeneous responses to common shocks or if there is interdependence between units. When factor variables are included (with factors equal to one, two or three), point estimates are slightly attenuated relative to the benchmark model for weighted results.

For unweighted results, the pattern of results is largely consistent. Excluding correlated random trends drives results down, but including correlated random trends at different levels of aggregation, or by allowing the time trend to be quadratic, reproduces benchmark results within plus or minus

three percentage points. Including three factors does bring the point estimates down such that they are equivalent to models that do not include any correction for correlated random trends, but it is important to emphasize that the true number of factors r is not known. Unfortunately, due to finite sample bias, we cannot include as many factors as is necessary to make the errors i.i.d..[19]

[Insert Figure 4 Here]

Overall, the preferred model tends to understate effect sizes for real log revenues and that causal effect estimates for graduation rates are variable depending on whether we adjust for correlated random trends. When we adjust for pre-treatment trends (especially at a sub-state level) point estimates for graduation rates stabilize and the variation around our benchmark estimates is less than 5 percentage points.

5.3 Equalizing effects

Equation (2) estimates levels of change in log revenues and graduation rates, comparing high (low) poverty districts in treated states to high (low) poverty districts in non-treated states. Comparing point estimates across poverty quartiles is problematic because the counterfactual trends are estimated for each poverty quartile. Assuming a homogeneous counterfactual trend (δ_t) allows for within state comparisons, but this assumption belies differences in secular trends across poverty quartile. In order to test whether revenues and graduation rates increased more in high poverty districts relative to low poverty districts following court order in a way that allows for heterogeneity among poverty groups, we construct a variable that ranks districts within a state based on the percents of students qualifying for FLE status in 1989. This ranking is then converted into a percentile by dividing by the total number of districts in that state. Compared to percents qualifying for free lunch, these rank-orderings put districts on a common metric, and are analogous to FLE quartiles, but with a continuous quantile rank-ordering.

The model that is estimated is analogous to Equation (2) with two changes:

1. The parameter $\delta_t Q$ is added, which is a continuous fixed effect variable that controls for

---

[19]See Moon & Weidner (2015) for discussion.

year-specific linear changes in Y_{sqdt} with respect to Q, where Q is a continuous within-state poverty quantile rank-ordering variable bounded between 0 and 100.

2. The treatment indicators $\sum_{n=1}^{19} \gamma_{(q,n)}[\mathbb{1}(Q_{sd} = q) \times \mathbb{1}(t - t_s^* = n) \times D_s]$ are now set equal to

$$\sum_{n=1}^{19} \gamma_{(q,n)}[Q \times \mathbb{1}(t - t_s^* = n) \times D_s].$$

Item (1) provides a counterfactual trend with respect to how much non-treated states are "equalizing" Y_{sqdt} with respect to Q. The secular trend in these models now adjusts for the rate that revenues and graduation rates are changing across FLE quantiles among untreated districts as well as the FLE specific average annual trend (δ_{tq}).[20]

For Item (2), the interpretation of the point estimates on the treatment year indicators is the marginal change in the outcome variable Y_{sqdt} given a one-unit change in FLE quantile within a state. For revenues, a point estimate of 0.0001 is equivalent to a 0.01 percent change in per pupil total revenues for each one-unit rank-order increase in FLE status within a state. For graduation rates, a point estimate of 0.0001 is equivalent to a 0.01 percentage point increase for each one-unit rank-order increase. A positive coefficient indicates that more revenues and graduation rates are going to poorer districts within a state.

Equation (2) is estimated for both WLS and OLS, with the modifications described just above. These results can be seen for both dependent variables in Table 3 and 4. The columns of interest are columns 5 and 10, which are labeled FLE Continuous.

After court-ordered reform, revenues increased across poverty quantiles, as indicated by the positive slope coefficients in Table 3. Seven years after reform, a 10-unit increase in FLE percentile is associated with a 1.5 percent increase in per-pupil log revenues for the weighted regression. For the unweighted regression, a 10-unit increase is associated with a 1.8 percent increase. As discussed in Appendix Section B.1 (Understanding differences between weighted and unweighted results), neither the weighted nor unweighted model results dominate each other, so we can view the slope coefficient as having a lower bound of 1.5 percent and an upper bound of 1.8 percent.

---

[20]Models that include only δ_t instead of δ_{tq}, not shown, are nearly identical.

Assuming the treatment is linear, these results suggest that districts in the 90th percentile would have had per pupil revenues that were between 12.32 to 14.56 percent higher than districts in the 10th percentile.

Table 4 also shows that court-ordered reform increased graduation rates across the FLE distribution. Seven years after reform, a 10-unit increase in FLE percentile is associated with a 1.05 percentage point increase in graduation rates for the weighted regression. For the unweighted model the corresponding point estimate is 0.77 percentage points, a difference in effect size commensurate with previously discussed results. Assuming linearity, districts in the 90th percentile would have had graduation rates that were between 6.16 to 8.4 percentage points higher than districts in the 10th percentile.

Figures 1 and 2 showed that high poverty quartiles in states undergoing reform experienced an increase in both revenues and graduation rates, centered around the timing of reform. It was also evident that for graduation rates this increase was larger than the increase in the other FLE quartiles. The results of these models indicate that states undergoing reform shifted more revenues and graduation rates to higher poverty districts, relative to shifts taking place in non-treated states.

5.4 The Effect of Revenues on Graduation Rates

Whether money matters for academic outcomes is an old and highly debated question (Coleman et al., 1966). While a substantial number of past studies do not find strong evidence that money improves academic outcomes, on average (Hanushek, 2003), the latest research suggests that a relationship exists when examining when there is a financial shock to the school's funding system and if that shock has a larger effect on sub-populations of individuals (Jackson et al., 2016; Lafortune et al., 2016). We address whether money matters by using the shocks of court-ordered finance reform as instruments for real revenues per pupil, allowing the identification of the causal effect of money on graduation rates for particular poverty quartiles and across the poverty distribution. While Jackson et al. (2016) perform a similar analysis, we exclusively leverage reforms from the Adequacy era to provide exogenous variation in revenues, and we disaggregate the causal effects

23

of these reforms into poverty quartiles and quantiles. Specifically, we interact the endogenous regressor per pupil log revenues with each poverty quartile, which provides two-stage least squares estimates for the marginal effect of a log dollar change in revenues on graduation rates for each poverty quartile.

Table 5 contains results from the two-stage least squares regressions. There are two panels: panel (a) displays weighted results and panel (b) displays unweighted results; for simplicity we focus the narrative on the weighted results. Each panel has a total of six columns. Column (1) is a two-stage least squares regression in which log real revenues per pupil are instrumented by the quartile-specific indicators that reflect the number of years after a court-ordered finance reform has taken place. Columns (2) through (5) are separate FLE-quartile regressions, where quartile-specific log real revenues per pupil are instrumented by the corresponding quartile-specific treatment year indicators. Finally, column (6) is a model that instruments log real revenues per pupil with the years-after-court-order binary indicators interacted with a continuous within-state poverty quantile rank-ordering variable. The first stage regressions equations are reflected in Equation (2), with the continuous model adaptations described above for Column (6). Standard errors that are clustered at the state level are in parentheses, but errors that are clustered at the district level are provided in brackets for reference.

[Insert Table 5 Here]

In the specification that estimates the average causal effect for all districts (Column 1), a 10 percent increase in revenues per pupil causes a 1.97 percentage point increase in graduation rates, but the result is not significant at conventional levels when errors are clustered at the state level. When we disaggregate the causal effect for each poverty quartile, however, there is meaningful variation in the effectiveness of school resources. For the highest-poverty districts, our results suggest that a 10 percent increase in revenues per pupil causes a 5.06 percentage point increase in graduation rates, which is significant at the 10 percent level. There is no evidence of an effect for FLE quartiles 1 through 3.

While our results suggest that school resources affect graduation rates, the primary threat to va-

lidity is whether changes in graduation rates following court-ordered finance reform can be wholly attributable to changes in school spending. For the exclusion restriction to hold, we must assume that the court reforms affect graduation rates only through their effect on spending. This assumption is violated in cases where court ordered finance had an effect on other unobserved policy changes that also affect graduation rates. For example, in addition to increasing spending in high poverty districts, the state may also adopt an incentive policy to bring higher quality teachers to more impoverished districts. In such a scenario, graduation rates resulting from court order are not separable from the change in spending and the unobserved programmatic change.

In the following section, we attempt to characterize the extent to which reforms varied across the states by interacting the treatment indicators with state-specific indicators. We ultimately show that all court-ordered finance reforms are not created equal and that differences in response to court order cannot be attributed to funding formula alone.

5.5   State by State Heterogeneity

We wish to investigate the extent to which treatment effects vary among states undergoing reform in this period. Jackson et al. (2016), Card & Payne (2002), and others have looked at treatment heterogeneity by constructing funding formula indicator variables (e.g., Minimum Foundation Grants, Local Effort, etc.) and estimating treatment effects based on the type of funding formula resulting from court order.[21] Descriptions of these funding formula are available in Hightower, Mitani, & Swanson (2010) and Jackson, Johnson, & Persico (2014). Further exploring this heterogeneity by looking at state specific responses to Court order is well motivated for at least two reasons.

First, as can be seen in Table 1, many states (such as Massachusetts, New Hampshire, Ohio and Tennessee) continued to have or adopted identical funding formula following court order. Lumping these four states together under a common indicator variable might overlook important variation. Second, many states did not change their funding formula following court order. This suggests that the shock to a state's educational system may not have operated strictly through the

---

[21]This approach was first motivated by (Hoxby, 2001), who used a simulated instruments approach.

funding formula. Thus, aggregating treatment effects to funding formula indicator variables further overlooks treatment heterogeneity.

Finally, if one is willing to assume that the counterfactual can be approximated by a common pool of untreated and not yet treated states (i.e., by including year fixed effects), then we can estimate each state's response to Court order. Estimating the counterfactual in this way has been done before, as both Jackson et al. (2016) and Card & Payne (2002) model the counterfactual as a year effect, which effectively assumes that the effect of each funding formula can be identified from a common pool of untreated and not yet treated states. In the absence of better counterfactual groups, if we have reasons to investigate heterogeneity within funding formula type, it is no problem to disaggregate the treatment effect to each state.

In order to estimate state-specific treatment effects, we modify our benchmark model, Equation (2) as follows:

$$(3) \quad Y_{sqdt} = \theta_d + \delta_{qt} + \psi_{sq}t + \xi_{(s,q)}[\mathbb{1}(State = s) \times \mathbb{1}(Q_{sd} = q) \times YearsAfter_{st} \times D_s]$$

$$+ \sum_{n=8}^{19} \gamma_{(s,q,n)}[\mathbb{1}(State = s) \times \mathbb{1}(Q_{sd} = q) \times \mathbb{1}(t - t_s^* = n) \times D_s] + \varepsilon_{sqdt},$$

which estimates state-specific slopes for treatment years one through seven and non-parametric treatment effects for the remainder of the treatment period. The variable $YearsAfter_{st}$ takes value 0 in all years before a court-ordered finance reform in state s and values 1 through 7 in the first seven years after reform.[22] Figure 5 displays estimates of the the slope parameters of interest, $\xi_{(s,q)}$, from weighted and ordinary least squares models for the dependent variables log per pupil revenues and graduation rates for each of the 12 treated states.[23]

Previous research has been conducted for three states with an Adequacy ruling in this pe-

---

[22]Because Montana was removed from our sample due to data quality issues, we obtain state-specific estimates for 12 states.

[23]Linear slope models are analogous to comparative interrupted time series, in which the treatment slopes are estimated net of pre-treatment trends (estimated via $\psi_{sq}t$) and quartile-year fixed effects. A linear approximation is preferred in this case to improve precision. In the Appendix we display results for models in which the full post-treatment period for each treated state-by-poverty quartile is estimated non-parametrically. Results from these non-parametric models are nearly identical to the linear specification, but with greater imprecision. Additional analysis of the state-specific responses to court order are provided to further validate these results.

riod. For Massachusetts, Guryan (2001) gains identification from discontinuities in the the Massachusetts funding formula and finds that state aid resulting from the 1993 Massachusetts Education Reform Act did increase per pupil spending and that this increase in spending improved 4th grade test scores but not 8th grade test scores. For Kentucky, Clark et al. (2003) compares changes in the correlation between spending and district income after Kentucky's 1989 *Rose* case against changes in the same correlation for district in Tennessee. Relative to Tennessee, the correlation between spending and district income decreased in Kentucky–meaning that spending became more equal–following reform, but the effects on student achievement were minimal. Finally, in New Jersey, Resch (2008) finds that the effect of the 1997 *Abbott* case increased revenues in Abbott districts relative to other, high poverty districts within New Jersey, but that these changes in spending improved test scores for only subgroups of students.[24]

[Insert Figure 5 Here]

Figure 5 provides WLS and OLS results on the left- and right-hand sides, respectively. The top-left panel provides state-specific results for which the effect of court order on revenues and graduation rates is both positive; the top-right panel provides results for which the effect is both negative; the bottom-left panel provides results for which the effect on revenues is positive and graduation rates is negative; the bottom-right panel provides results for which the effect on revenues is negative and graduation rates is positive.[25]

Note, first, that the majority of states that had an Adequacy ruling during this period increased

---

[24]The analytic strategy we employ here makes comparisons to these studies difficult. Kentucky was treated so early that we lack pre-treatment information about the state; for this reason, we do not include it here. Moreover, in Clark's study of Kentucky, Tennessee serves as the counterfactual, but Tennessee had its Adequacy case in 1995, and we find evidence that Tennessee, following its reform, spent less revenues in high poverty districts, net of its pre-reform trend, which would exaggerate a Kentucky-specific effect. For New Jersey, our treatment pool includes all poor districts within the state, whereas Resch compares high poverty non-Abbott districts in New Jersey to the Abbott districts; therefore, any positive finding identified by Resch could be attenuated when we pool treated and control together. For Massachusetts, the identification strategy we pursue is not well equipped for a state that underwent treatment so early in the analytic sample. As shown in Figure D.8, identification is strongly influenced by controlling for the pre-treatment trend, and for Massachusetts, this trend is identified from two years of data.

[25]WLS and OLS results for each state are nearly identical, with the exception of graduation rates in Tennessee (although the point estimate is not distinguishable from zero in both the WLS and OLS cases) and revenues in North Carolina (where the point estimate is slightly negative in the WLS case and slightly positive in the OLS case but not distinguishable from zero in both cases). For this reason, we focus on weighted least squares in the discussion.

both revenues and graduation rates.[26] The fact that the majority of state-specific responses to court order correspond to the average effect provides support for the benchmark model. Following court order, Tennessee lost both revenues and graduation rates. As discussed in Appendix C.3, spending in Tennessee's poorest districts was increasing rapidly prior to court order; net of that prior trend, spending dropped in Tennessee following court order.

The results for Massachusetts, New Hampshire, North Carolina, and Wyoming require further investigation. In Wyoming and New Hampshire, court order resulted in a positive change on revenues without affecting graduation rates. In Massachusetts and North Carolina, conversely, graduation rates increased despite a loss (or non-increase) in revenues. These results provide evidence that the relationship between spending and graduation rates is not mono-causal or reducible to the adopted funding formula. Massachusetts, New Hampshire and Wyoming all adopted minimum foundation plans, but so did every other state except New York and North Carolina. North Carolina's adoption of an equalization plan and flat grant matches the plans adopted by New York, but spending increased in New York whereas spending did not in North Carolina. Therefore, while the majority of states that received additional funds were able to increase high school graduation rates, there are some states for which money was not needed to produce academic gains and others for which money was not sufficient.

5.6   Robustness checks

One threat to internal validity using aggregated data is selective migration. If treatment induces a change in population and this change in population affects graduation rates, then the results using aggregate graduation rates will be biased. Such a source of bias would occur if, for example, parents that value education were more likely to move to areas that experienced increases in school spending. If there is evidence of population changes resulting from treatment, and if these population characteristics are correlated with the outcome variable, there may be bias.

To test for selective migration, we estimate the benchmark model on four dependent variables:

---

[26]Revenues estimates for Arkansas and New Jersey are not statistically significant.

logarithm of total district enrollment, percent minority (sum of Hispanic and black population shares within a district), percent of children in poverty from the Census Bureau's Small Area Income and Poverty Estimates (SAIPE), and percent of students receiving special education. In Table 6 we present results for FLE quartile four.[27]

Table 6 shows that there is no strong evidence of selective migration to treated districts in FLE quartile four. None of the point estimates for percent minority are statistically significant, nor are they large in magnitude. We also find no evidence of changes in the percentage of children in poverty and those in special education.[28]

Prior research suggests that nonlinear transformations of the dependent variable (e.g., taking the natural logarithm of the dependent variable) might produce treatment effect estimates that are substantially different from the original variable (Lee & Solon, 2011; Solon et al., 2015). While transformations of the dependent variable do affect the interpretation of the marginal effect, we should see similar patterns in terms of significance and sign. Table 6, shows that WLS and OLS results for total per pupil revenues are significant in all treatment years, which matches the pattern of significance of the main revenues results in Table 3. Moreover, all point estimates are positive across both tables. Appendix Table E.9 shows estimates for all FLE quartiles and is qualitatively similar in terms of significance and sign to the results in Table 3.

[Insert Table 6 Here]

6 Conclusion

We highlight three contributions of this paper. First, this paper tests whether Adequacy-based cases increased revenues and graduation rates above and beyond the secular changes taking place in high poverty districts across the country, as other states have made internal decisions to change

---

[27]Although we only report the benchmark treatment effect estimates for FLE quartile four in Table 6, results for all other quartiles and the continuous model appear in the Appendix. See Tables E.6, E.7, and E.8 in Appendix E.

[28]In Appendix tables E.6, E.7, and E.8, we do find that the composition of minority students decreased and that the percentage of special education students increased in lower poverty districts (i.e. FLE quartiles one to three). However, when we look across the district poverty distribution to examine whether there are any changes in the minority composition and percentage of special education for a increase in the poverty percentile rank, we find no statistically significant effects.

funding formula without the need of court intervention (Jackson et al., 2016; Lafortune et al., 2016). We find consistent evidence that these court cases did have positive effects on these two outcomes. High poverty districts in states undergoing reform increased revenues and graduation rates relative to high poverty districts in non-treated states; in addition, these effects were relatively more equalizing compared to trends taking place in other states across the United States. The academic outcome we evaluate, student graduation rates, is the most distal outcome available for students experiencing changes brought about by court order, and for this reason the evaluation is a timely appraisal of the effect of additional spending on student academic outcomes.

Second, the paper provides robustness analysis that is relevant to key recent findings in the school finance literature (Jackson et al., 2016; Lafortune et al., 2016). We demonstrate that the effects of court order are quite robust to model specification. When we alter specifications of secular trends, correlated random trends, and cross sectional dependence, estimates do not dramatically change.

Third, this paper highlights two aspects of heterogeneity in response to Adequacy rulings. The first is that the marginal effect of an additional dollar of spending is more productive in poor districts than it is in non-poor districts. Point estimates From Table 5 for FLE four are between 2.3 to 5.1 times larger than those for FLE one, and point estimates for FLE one are small and imprecise. This suggests that there may be upper limits to spending, but that these limits have not been reached in high poverty districts in the United States. The second is that state responses to court order led to different results, and that these differences are not solely attributable to funding formula. Specifically, future research should address how states with little to no observed change in resources, such as North Carolina and Massachusetts (and to a lesser extent Arkansas and New Jersey), were able to improve graduation rates, while states like New Hampshire and Wyoming increased revenues but failed to translate those resources into improvements in high school completion.

Taken together, it is evident that there are mechanisms between monetary increases and student outcomes that are not well understood; further research can hopefully shed light on these matters.

References

AHN, S. C. & HORENSTEIN, A. R. (2013). "Eigenvalue ratio test for the number of factors." *Econometrica*, *81*(3), 1203–1227.

ANGRIST, J. D. & PISCHKE, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's companion.*: Princeton University Press.

BAI, J. (2009). "Panel data models with interactive fixed effects." *Econometrica*, *77*(4), 1229–1279.

BAKER, B. & GREEN, P. (2008). "Conceptions of equity and adequacy in school finance." *Handbook of research in education finance and policy*, 203–221.

BERRY, C. (2007). "The impact of school finance judgments on state fiscal policy." *School money trials: The legal pursuit of educational adequacy*, 213–242.

BERTRAND, M., DUFLO, E., & MULLAINATHAN, S. (2004). "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics*, *119*(1), 249–275. DOI: 10.1162/003355304772839588.

CARD, D. & PAYNE, A. A. (2002). "School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores." *Journal of Public Economics*, *83*(1), 49–82.

CLARK, M. A. ET AL. (2003). "Education reform, redistribution, and student achievement: Evidence from the Kentucky Education Reform Act." *PhD diss. Princeton University*.

COLEMAN, J. S., CAMPBELL, E. Q., HOBSON, C. J., MCPARTLAND, J., MOOD, A. M., WEINFELD, F. D., & YORK, R. L. (1966). "Equality of Educational Opportunity." Report OE-38001, National Center for Educational Statistics.

CORCORAN, S., EVANS, W. N., GODWIN, J., MURRAY, S. E., & SCHWAB, R. M. (2004). "The changing distribution of education finance, 1972–1997." *Social inequality*, 433–465.

CORCORAN, S. P. & EVANS, W. N. (2008). "Equity, Adequacy, and the Evolving State Role in Education Finance." In H. F. Ladd & E. B. Fiske (Eds.) *Handbook of Research in Education Finance and Policy*.: Routledge, 149–207.

CORCORAN, S. & EVANS, W. (2015). , New York, NY: Routledge.

CORREIA, S. (2014). "REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects." Statistical Software Components, Boston College Department of Economics, July.

DUMOUCHEL, W. H. & DUNCAN, G. J. (1983). "Using sample survey weights in multiple regression analyses of stratified samples." *Journal of the American Statistical Association*, *78*(383), 535–543.

FRISCH, R. & WAUGH, F. V. (1933). "Partial time regressions as compared with individual trends." *Econometrica: Journal of the Econometric Society*, 387–401.

GOMEZ, M. (2015a). "FixedEffectModels: Julia package for linear and iv models with high dimensional categorical variables.." `https://github.com/matthieugomez/FixedEffectModels.jl`.

——— (2015b). "SparseFactorModels: Julia package for unbalanced factor models and interactive fixed effects models." `https://github.com/matthieugomez/SparseFactorModels.jl`.

GURYAN, J. (2001). "Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts."Technical report, National Bureau of Economic Research.

HANUSHEK, E. A. (2003). "The Failure of Input-Based Schooling Policies." *The Economic Journal*, *113*, F64–F98.

HECKMAN, J. J. & LAFONTAINE, P. A. (2010). "The American high school graduation rate: Trends and levels." *The review of economics and statistics*, *92*(2), 244–262.

HEISE, M. (1995). "State constitutions, school finance litigation, and the third wave: From equity to adequacy." *Temple L. Rev.*, *68*, 1151.

HIGHTOWER, A. M., MITANI, H., & SWANSON, C. B. (2010). *State policies that pay: A survey of school finance policies and outcomes.*: Editorial Projects in Education.

HOXBY, C. M. (2001). "All School Finance Equalizations are Not Created Equal." *The Quar-*

terly Journal of Economics, *116*(4), 1189–1231. DOI: `http://dx.doi.org/10.1162/003355301753265552`.

ILIN, A. & RAIKO, T. (2010). "Practical approaches to principal component analysis in the presence of missing values." *The Journal of Machine Learning Research*, *11*, 1957–2000.

JACKSON, C. K., JOHNSON, R. C., & PERSICO, C. (2016). "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *Quarterly Journal of Economics*.

JACKSON, C. K., JOHNSON, R., & PERSICO, C. (2014). "The Effect of School Finance Reforms on the Distribution of Spending, Academic Achievement, and Adult Outcomes." Working Paper No. 20118, National Bureau of Economic Research. URL: `http://www.nber.org/papers/w20118`.

KOSKI, W. S. (2009). "Evolving Role of the Courts in School Reform Twenty Years after Rose, The." *Ky. LJ*, *98*, 789.

LAFORTUNE, J., SCHANZENBACH, D., & ROTHSTEIN, J. (2016). "School Finance Reform and the Distribution of Student Achievement." Working Paper No. 22011, National Bureau of Economic Research.

LEE, J. Y. & SOLON, G. (2011). "The fragility of estimated effects of unilateral divorce laws on divorce rates." *The BE Journal of Economic Analysis & Policy*, *11*(1).

LOVELL, M. C. (1963). "Seasonal adjustment of economic time series and multiple regression analysis." *Journal of the American Statistical Association*, *58*(304), 993–1010.

MOON, H. R. & WEIDNER, M. (2015). "Linear regression for panel with unknown number of factors as interactive fixed effects." *Econometrica*, *83*(4), 1543–1579.

MURRAY, S. E., EVANS, W. N., & SCHWAB, R. M. (1998). "Education-finance reform and the distribution of education resources." *American Economic Review*, 789–812.

ONATSKI, A. (2010). "Determining the number of factors from empirical distribution of eigenvalues." *The Review of Economics and Statistics*, *92*(4), 1004–1016.

PESARAN, M. H. & PICK, A. (2007). "Econometric issues in the analysis of contagion." *Journal of Economic Dynamics and Control*, *31*(4), 1245–1277.

RAIKO, T., ILIN, A., & KARHUNEN, J. (2008). "Principal component analysis for sparse high-dimensional data." In *Neural Information Processing*., 566–575, Springer.

RESCH, A. M. (2008). *Three essays on resources in education*.: ProQuest.

SIMS, D. P. (2011). "Lifting All Boats? Finance Litigation, Education Resources, and Student Needs in the Post-Rose Era." *Education Finance & Policy*, *6*(4), 455–485.

SOLON, G., HAIDER, S. J., & WOOLDRIDGE, J. M. (2015). "What are we weighting for?" *Journal of Human Resources*, *50*(2), 301–316.

SPRINGER, M. G., LIU, K., & GUTHRIE, J. W. (2009). "The Impact of School Finance Litigation on Resource Distribution: A Comparison of Court-Mandated Equity and Adequacy Reforms." *Education Economics*, *17*(4), 421–444.

WEST, M. R. & PETERSON, P. E. (2007). *School money trials: The legal pursuit of educational adequacy*.: Brookings Institution Press.

WOLFERS, J. (2006). "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results." *American Economic Review*, *96*(5), 1802–1820.

WOOLDRIDGE, J. M. (2005). "Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models." *Review of Economics and Statistics*, *87*(2), 385–390.

——— (2010). *Econometric analysis of cross section and panel data*.: MIT press.

WRIGHT, S. & HOLT, J. N. (1985). "An inexact levenberg-marquardt method for large sparse nonlinear least squres." *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics*, *26*(04), 387–403.

Figure 1: Event Study of Log Per Pupil Revenues
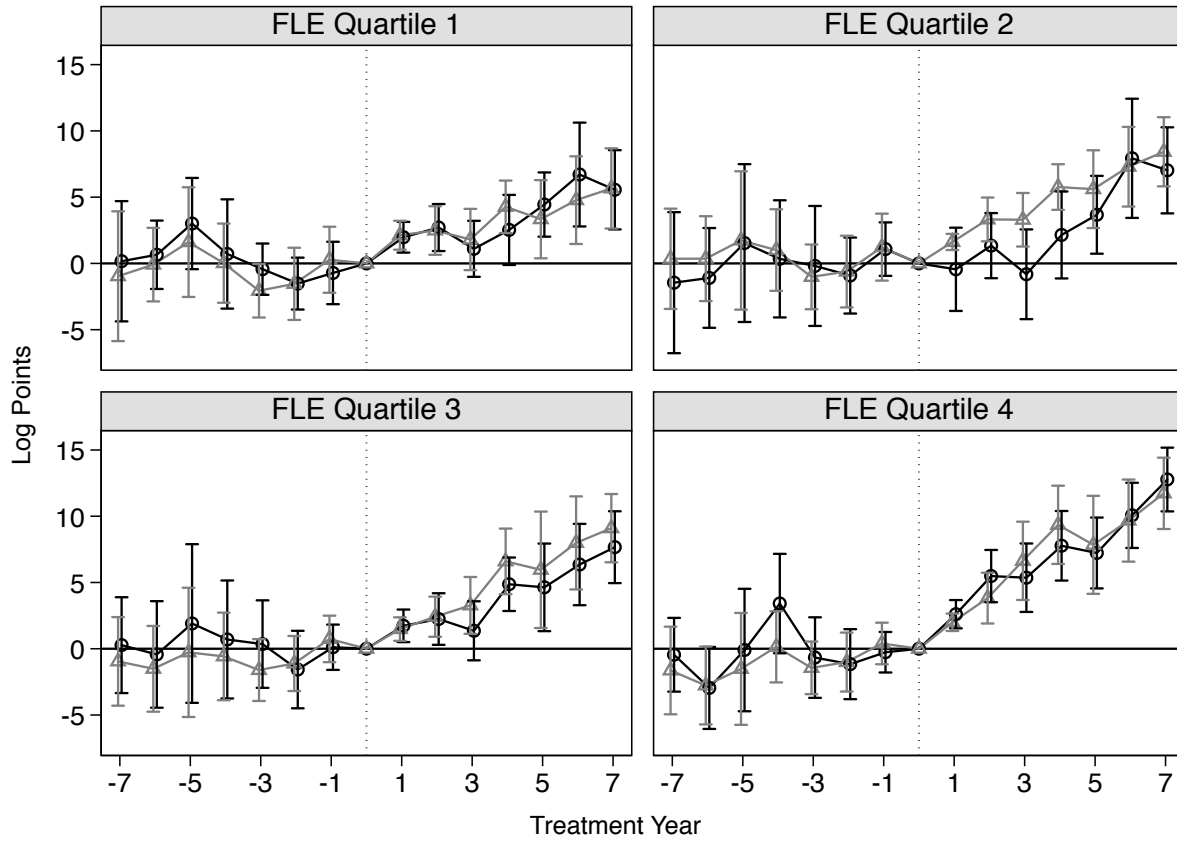
Notes: Differences-in-differences with treatment effects estimated non-parametrically before and after reform. Model accounts for district fixed effects (θ_d) and FLE-by-year fixed effects (δ_{tq}). Range caps show 90 percent confidence intervals. Standard errors are clustered at the state level, which is the level at which treatment occurs. Black lines with hollow circles reflect weighted least squares models, where the weights are district enrollment; gray lines with hollow triangles reflect OLS models.

Figure 2: Event Study of Graduation Rates
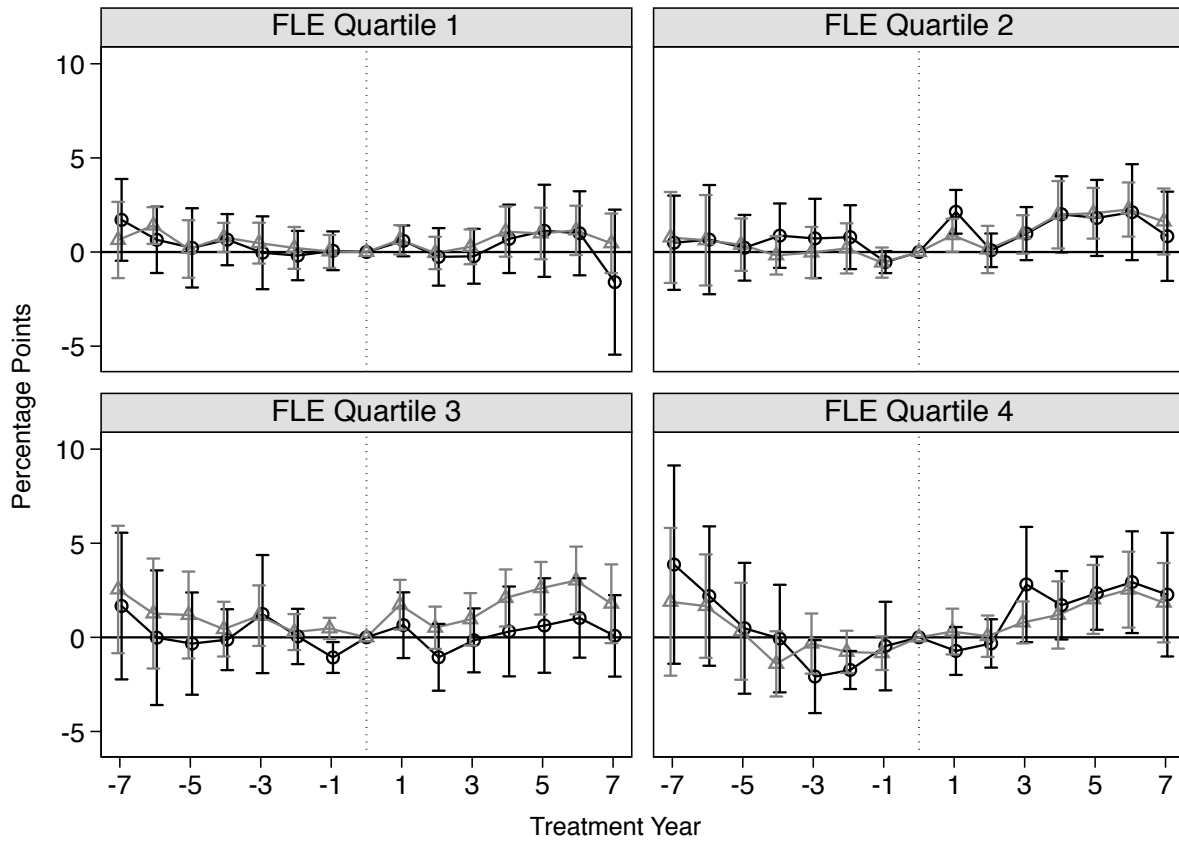
Notes: Differences-in-differences with treatment effects estimated non-parametrically before and after reform. Model accounts for district fixed effects (θ_d) and FLE-by-year fixed effects (δ_{tq}). Range caps show 90 percent confidence intervals. Standard errors are clustered at the state level, which is the level at which treatment occurs. Black lines with hollow circles reflect weighted least squares models, where the weights are district enrollment; gray lines with hollow triangles reflect OLS models.

Figure 3: Model Sensitivity: Changes in Estimates for Log Per Pupil Revenues across Models

Legend:

- CRT: District; ST: FLE by year; CSD: 0 [9]
- **CRT: State by FLE; ST: FLE by year; CSD: 0 [2]**
- ---CRT: State by FLE; ST: FLE by year; CSD: 1 [3]
- ---CRT: State by FLE; ST: FLE by year; CSD: 2 [4]
- ---CRT: State by FLE; ST: FLE by year; CSD: 3 [5]
- CRT: State by FLE^2; ST: FLE by year; CSD: 0 [6]
- CRT: State; ST: FLE by year; CSD: 0 [7]
- CRT: None; ST: FLE by year; CSD: 0 [1]
- CRT: District; ST: year; CSD: 0 [15]
- CRT: State by FLE; ST: year; CSD: 0 [11]
- CRT: State by FLE^2; ST: year; CSD: 0 [12]
- CRT: State; ST: year; CSD: 0 [13]
- CRT: None; ST: year; CSD: 0 [10]

Notes: Legend shows three parameter changes, delimited by ";". ST denotes the type of nonparametric *secular trend*: (a) FLE quartile by year fixed effects or (b) year fixed effects. CRT denotes the type of *correlated random trends*: (a) none, corresponding to no CRT; (b) state by FLE quartile fixed effects interacted with linear time; (c) state fixed effects interacted with linear time; (d) district fixed effects interacted with linear time. CSD denotes type of cross-sectional dependence adjustment: (a) 0, which implies no factor structure; (b) 1, which is a 1 factor model; (c) 2, which is a 2 factor model; or (d) 3, which is a 3 factor model. Bracketed numbers indicate the column location of point estimates for model m in Tables E.2 and E.3. Benchmark model is indicated by a thick, black dashed line.

Figure 4: Model Sensitivity: Changes in Estimates for Graduation Rates across Models

Notes: Legend shows three parameter changes, delimited by ";". ST denotes the type of nonparametric *secular trend*: (a) FLE quartile by year fixed effects or (b) year fixed effects. CRT denotes the type of *correlated random trends*: (a) none, corresponding to no CRT; (b) state by FLE quartile fixed effects interacted with linear time; (c) state fixed effects interacted with linear time; (d) district fixed effects interacted with linear time. CSD denotes type of cross-sectional dependence adjustment: (a) 0, which implies no factor structure; (b) 1, which is a 1 factor model; (c) 2, which is a 2 factor model; or (d) 3, which is a 3 factor model. Bracketed numbers indicate the column location of point estimates for model m in Tables E.2 and E.3. Benchmark model is indicated by a thick, black dashed line.

Figure 5: Treatment Heterogeneity between States

(a) Weighted Least Squares

(b) Ordinary Least Squares

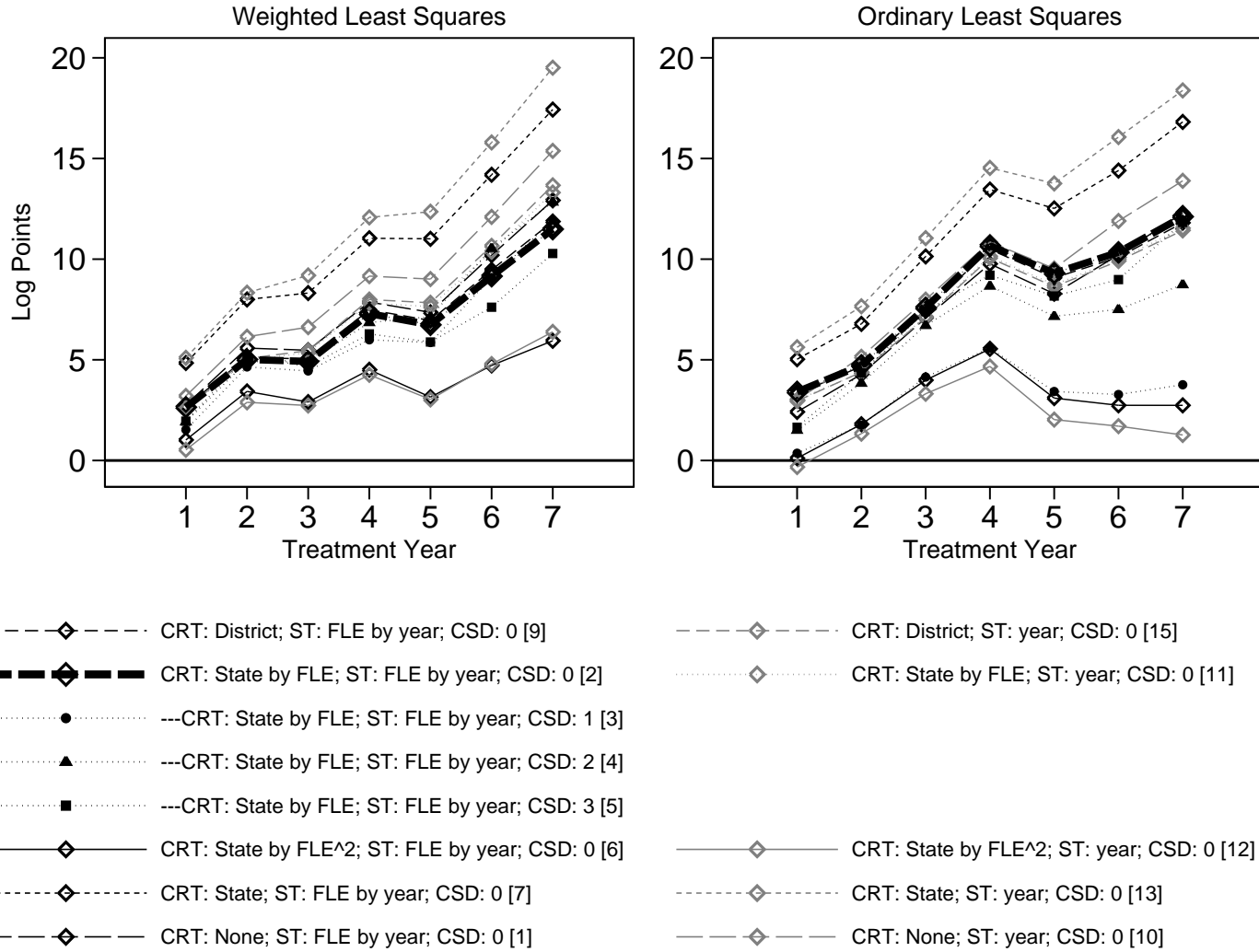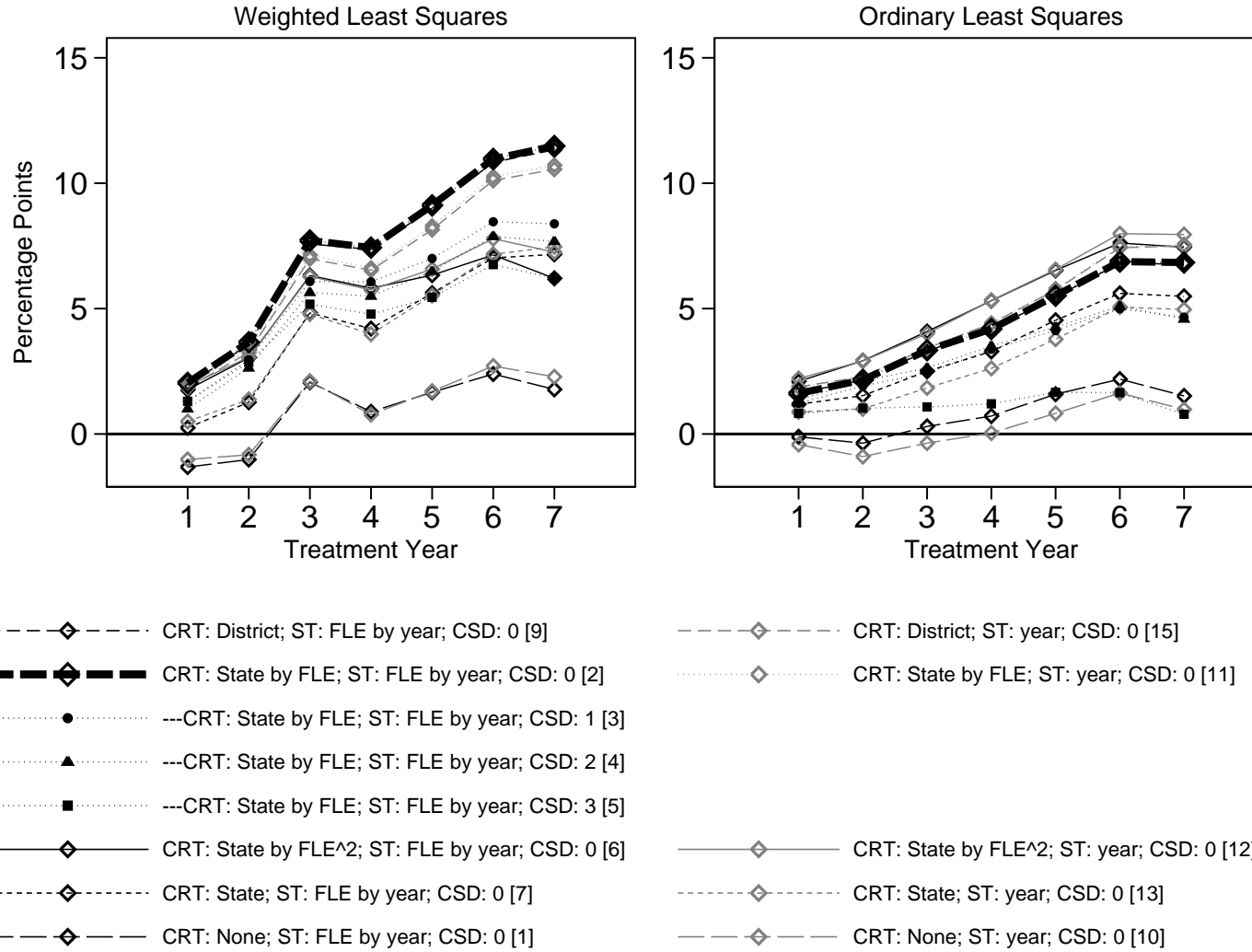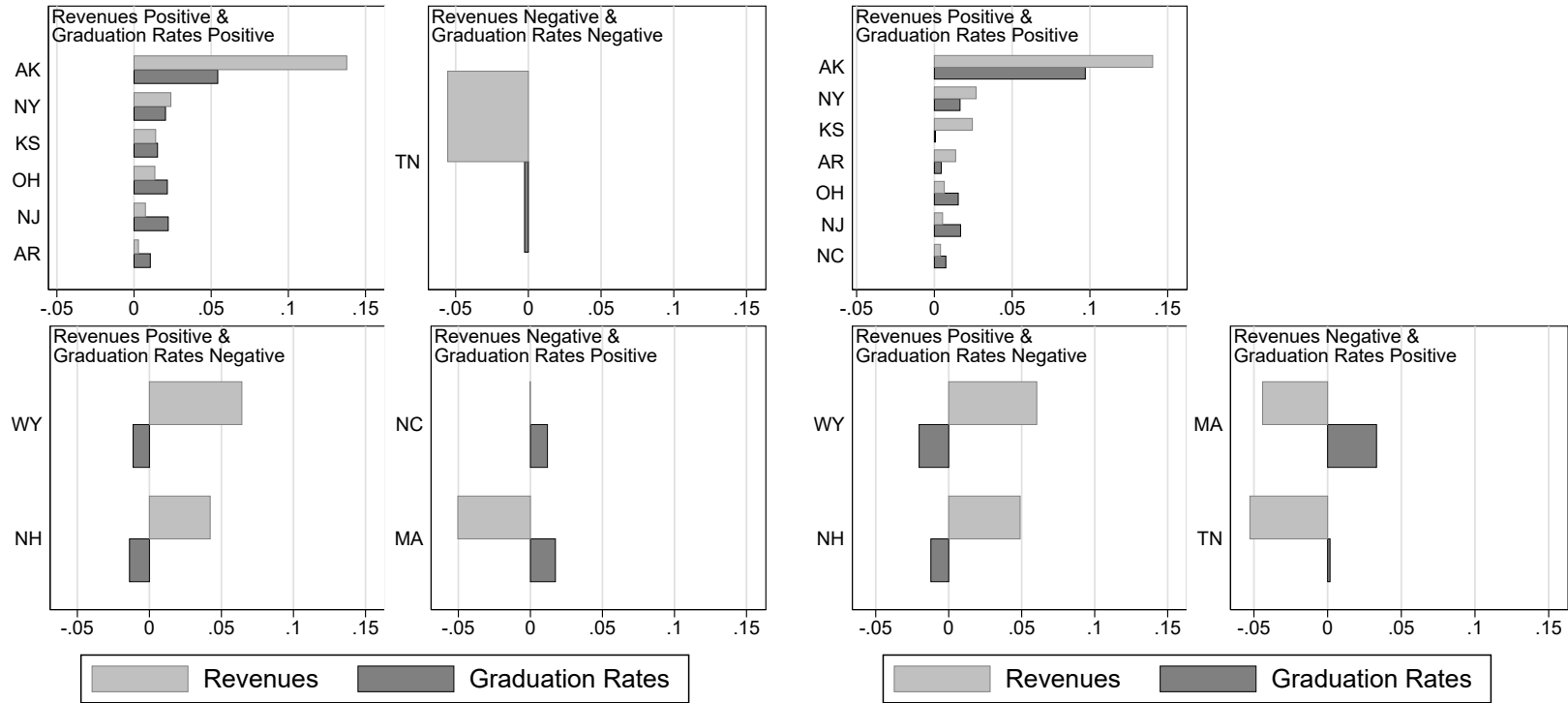Bars correspond to point estimates from an augmented version of the benchmark model (i.e. differences-in-differences with treatment effects estimated non-parametrically after reform, accounting for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq}), and state-by-FLE linear time trends ($\psi_{sq}t$)), in which 12 treatment state indicator variables ($\sum_{s=1}^{11}$) are interacted with FLE quartiles and a linear treatment slope for years 1 to 7. This model provides estimates of treatment effects for 12 treated states, 4 quartiles, as a function of time following court order.

Tables

Table 1: Adequacy Era Court-Ordered Finance Reform Years

State Name	Year of 1st Overturn	Funding Formula Adopted	Case Name
Alaska	2009	LE + MFP	*Moore v. State*
Arkansas	2002	EP + MFP + SL	*Lake View School District No. 25 v. Huckabee (Lakeview III)*
Kansas	2005	EP + LE + MFP + SL	*Montoy v. State (Montoy II)*
Kentucky	1989	EP + LE + MFP	*Rose v. Council for Better Education*
Massachusetts	1993	MFP	*McDuff v. Secretary of Executive Office of Education*
Montana	2005	EP + MFP + SL	*Columbia Falls Elementary School District No. 6 v. Montana*
New Hampshire	1997	MFP	*Claremont School District v. Governor*
New Jersey	1997	EP + MFP	*Abbott v. Burke (Abbott IV)*
New York	2003	EP + FG	*Campaign for Fiscal Equity, Inc. v. New York*
North Carolina	2004	EP + FG	*Hoke County Board of Education v. North Carolina*
Ohio	1997	MFP	*DeRolph v. Ohio (DeRolph I)*
Tennessee	1995	MFP	*Tennessee Small School Systems v. McWherter (II)*
Wyoming	1995	EP + MFP + SL	*Campbell County School District v. Wyoming (Campbell II)*

Notes: The table shows the first year in which a state's education finance system was overturned on adequacy grounds; we also provide the name of the case and the funding formula adopted following court order. The primary source of data from this table is Corcoran & Evans (2008). We have updated their table with information provided by ACCESS, Education Finance Litigation: `http://schoolfunding.info/`. The adopted funding formula is taken from Jackson et al. (2014) and updated for Alaska from Hightower et al. (2010). EP corresponds to an equalization plan; FG corresponds to to flat grant; LE to local effort equalization; MFP corresponds to minimum foundation plan; SL to spending limits. See Hightower et al. (2010) and Jackson et al. (2014) for description of funding formula.

Table 2: Descriptive Statistics for Outcome Variables

	Weighted		Unweighted	
	Mean	SD	Mean	SD
Graduation Rates	0.77	0.15	0.82	0.14
Log Revenues	8.94	0.26	8.97	0.29
Total Revenues	7950.2	2352.12	8224.31	2782.13
Percent Minority	0.32	0.3	0.16	0.23
Percent Black	0.17	0.22	0.08	0.17
Percent Hispanic	0.15	0.22	0.08	0.16
Percent Special Education	0.12	0.05	0.13	0.05
Percent Child Poverty	0.16	0.1	0.16	0.09
Log Enrollment	9.43	1.62	7.4	1.25

Notes: This table provides means and standard deviations for the outcome variables used in this paper. Summary statistics shown here that are weighted use the district enrollment across the sample.

Table 3: Change in Log Per Pupil Revenues, by Free Lunch Eligible Quartile and Quantile

Treatment	Weighted					Unweighted				
Year	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.
1	.043 *	.006	.032	.026 *	.00036	.048 *	.04 +	.04 *	.034 *	.00054 *
	(.018)	(.04)	(.022)	(.013)	(.00025)	(.024)	(.021)	(.018)	(.014)	(.00023)
2	.061 *	.027	.033	.05 *	.0006 +	.06 *	.059 *	.045 *	.047 **	.0007 *
	(.024)	(.036)	(.025)	(.019)	(.00031)	(.029)	(.025)	(.02)	(.018)	(.00028)
3	.05 +	.007	.03	.049 *	.00051	.061 +	.065 *	.058 *	.076 ***	.00097 **
	(.029)	(.05)	(.028)	(.022)	(.00036)	(.035)	(.03)	(.024)	(.019)	(.0003)
4	.072 *	.04	.068 *	.073 **	.00093 *	.093 *	.099 **	.098 ***	.107 ***	.00146 ***
	(.032)	(.055)	(.029)	(.024)	(.00036)	(.039)	(.034)	(.025)	(.018)	(.0003)
5	.098 *	.059	.069 *	.067 *	.00095 *	.091 +	.106 **	.096 **	.093 ***	.00138 **
	(.039)	(.05)	(.033)	(.03)	(.00043)	(.049)	(.04)	(.033)	(.026)	(.00043)
6	.129 *	.105 +	.088 *	.092 *	.00134 *	.113 +	.127 **	.116 **	.103 **	.0016 **
	(.052)	(.053)	(.041)	(.037)	(.00053)	(.061)	(.049)	(.041)	(.038)	(.00055)
7	.128 *	.104	.1 *	.115 **	.00154 **	.13 +	.145 **	.129 **	.121 **	.00182 **
	(.058)	(.069)	(.047)	(.036)	(.00057)	(.066)	(.053)	(.044)	(.039)	(.00056)
R^2	.909	.909	.909	.909	.909	.887	.887	.887	.887	.887

Notes: This table shows point estimates and standard errors for non-parametric differences-in-differences estimator. Model accounts for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq}), and state-by-FLE linear time trends ($\psi_{sq}t$). FLE quartiles are indexed by FLE $\in 1, 2, 3, 4$. Column "FLE Cont." corresponds to models in which the additional control variable year-by-FLE percentile ($\delta_t Q$) is included. Point estimates for continuous model are interpreted as change in revenues for 1-unit change in poverty percentile rank within a state, relative to change in percentile rank in states without court order. All standard errors are clustered at the state level. (Significance indicated $+ < .10, * < .05, ** < .01, *** < .001$)

Table 4: Change in Graduation Rates, by Free Lunch Eligible Quartile and Quantile

Treatment	Weighted					Unweighted				
Year	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.
1	.01 *	.026 ***	.016	.02 *	.00026 *	.009 *	.017 *	.018 *	.016 +	.00023 *
	(.005)	(.008)	(.01)	(.009)	(.0001)	(.004)	(.008)	(.008)	(.008)	(.0001)
2	.007	.008	.003	.037 **	.00028 *	.006	.015 *	.011	.022 *	.00022 +
	(.009)	(.007)	(.012)	(.012)	(.00013)	(.005)	(.008)	(.01)	(.01)	(.00012)
3	.008	.02	.013	.077 ***	.00066 ***	.011	.026 *	.018	.033 **	.00035 *
	(.006)	(.013)	(.012)	(.017)	(.00015)	(.007)	(.011)	(.014)	(.011)	(.00015)
4	.021 *	.034 *	.026	.074 ***	.00071 **	.022 **	.04 **	.033 *	.042 *	.00053 *
	(.011)	(.015)	(.019)	(.018)	(.00022)	(.008)	(.015)	(.015)	(.017)	(.00021)
5	.029 +	.04 *	.035 +	.091 ***	.00087 ***	.024 **	.046 **	.042 *	.055 **	.00066 **
	(.017)	(.015)	(.02)	(.017)	(.0002)	(.008)	(.016)	(.017)	(.019)	(.00022)
6	.032 *	.046 **	.043 *	.11 ***	.00105 ***	.029 **	.052 **	.051 *	.069 **	.00082 **
	(.016)	(.017)	(.019)	(.019)	(.0002)	(.01)	(.019)	(.021)	(.022)	(.00028)
7	.015	.043 *	.04 +	.115 ***	.00105 ***	.027 *	.052 *	.044 +	.068 *	.00077 *
	(.019)	(.021)	(.022)	(.02)	(.00021)	(.013)	(.024)	(.026)	(.027)	(.00034)
R^2	.78	.78	.78	.78	.78	.567	.567	.567	.567	.567

Notes: This table shows point estimates and standard errors for non-parametric differences-in-differences estimator. Model accounts for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq}), and state-by-FLE linear time trends ($\psi_{sq}t$). FLE quartiles are indexed by FLE $\in 1, 2, 3, 4$. Column "FLE Cont." corresponds to models in which the additional control variable year-by-FLE percentile ($\delta_t Q$) is included. Point estimates for continuous model are interpreted as change in revenues for 1-unit change in poverty percentile rank within a state, relative to change in percentile rank in states without court order. All standard errors are clustered at the state level. (Significance indicated $+ < .10, * < .05, ** < .01, *** < .001$)

Table 5: Two-Stage Least Squares Results

(a) Weighted Results

	(1) Full	(2) FLE 1	(3) FLE 2	(4) FLE 3	(5) FLE 4	(6) Cont.
$\log(Rev/Pupil)$	0.197	0.0725	0.127	0.123	0.506	0.359
	(0.162)	(0.200)	(0.133)	(0.286)	$(0.293)^{+}$	$(0.197)^{+}$
	[0.0514]***	[0.0817]	[0.0840]	[0.145]	[0.143]***	[0.0774]***
N	188752	45178	48675	49052	45847	188752

(b) Unweighted Results

	(1) Full	(2) FLE 1	(3) FLE 2	(4) FLE 3	(5) FLE 4	(6) Cont.
$\log(Rev/Pupil)$	0.278	0.148	0.265	0.306	0.379	0.347
	$(0.112)^{*}$	(0.103)	$(0.122)^{*}$	$(0.141)^{*}$	$(0.168)^{*}$	$(0.133)^{*}$
	[0.0279]***	[0.0535]**	[0.0465]***	[0.0590]***	[0.0657]***	[0.0373]***
N	188752	45178	48675	49052	45847	188752

Notes: Panels (a) and (b) reflect estimates from two-stage least squares models. Panel (a) displays coefficients for weighted models and panel (b) display unweighted. The variable of interest is $\log(Rev/Pupil)$, which is the natural logarithm of real per-pupil revenues at the district level.

In model (1), the excluded instruments from the second stage are quartile-specific indicators that reflect the number of years after a court-ordered finance reform has taken place; there are a total of 72 indicators. In models (2) to (5), the excluded instruments are the years-after-court-order-treatment indicators for the given quartile under investigation. For each quartile, the number of excluded instruments is 18. In model (6), the excluded instruments are the years-after-court-order-treatment indicators interacted with the continuous percentile variable.

All models account for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq}), and state-by-FLE linear time trends ($\psi_{sq}t$). Model (6) adds a year-by-FLE percentile ($\delta_t Q$) control variable.

All standard errors are clustered at the state level. (Significance: $^{+}\ p < 0.10$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table 6: Robustness Checks using Benchmark Model Specification - FLE Quartile 4 Coefficients Reported

Treatment	Pct. Minority		Pct. Poverty		Pct. Special Ed.		Total Revenues	
Year	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
1	-.005	-.002	.012	.002	.022	.017	294 *	381 *
	(.008)	(.002)	(.009)	(.009)	(.027)	(.019)	(161)	(179)
2	.005	.000	-.005	-.012	.028	.021	549 *	478 *
	(.004)	(.002)	(.016)	(.012)	(.03)	(.022)	(221)	(194)
3	.002	.000	-.001	-.011	.019	.013	561 *	765 **
	(.003)	(.002)	(.008)	(.010)	(.035)	(.026)	(231)	(219)
4	.001	.000	.010	-.006	.038	.024	839 **	1095 ***
	(.004)	(.003)	(.011)	(.011)	(.031)	(.023)	(252)	(226)
5	.001	.002	-.005	-.010	.045	.032	867 **	985 **
	(.004)	(.003)	(.008)	(.011)	(.034)	(.026)	(346)	(323)
6	.001	.001	.004	-.008	.049	.034	1230 **	1190 *
	(.005)	(.004)	(.012)	(.015)	(.04)	(.031)	(462)	(517)
7	-.003	.000	-.005	-.011	.049	.036	1530 ***	1370 **
	(.006)	(.005)	(.018)	(.017)	(.045)	(.034)	(412)	(473)
R^2	.956	.971	.949	.911	.432	.564	.901	.865

Notes: This table shows results for different demographic variables and total revenues for FLE Quartile 4 from our benchmark model, which accounts for district fixed effects (θ_d), year fixed effects (δ_t), and state-by-FLE linear time trends ($\psi_{sq}t$). All standard errors are clustered at the state level. Full results, estimated for each quartile and for the continuous model, are reported in the Appendix. (Significance indicated: $+ < .10, * < .05, ** < .01, *** < .001$)

Appendices

A Data Appendix

A.1 Data Sources

Our data set is the compilation of several public-use surveys that are administered by the National Center for Education Statistics and the U.S. Census Bureau. We construct our analytic sample using the following data sets: Local Education Agency (School District) Finance Survey (F-33); Local Education Agency (School District) Universe Survey; Local Education Agency Universe Survey Longitudinal Data File: 1986-1998 (13-year); Local Education Agency (School District) Universe Survey Dropout and Completion Data; and Public Elementary/Secondary School Universe Survey. Web links to each of the data sources are listed below:

Note: All links last accessed June 2015.

Local Education Agency Finance Survey (F-33)

Source: `https://nces.ed.gov/ccd/f33agency.asp`

Local Education Agency Universe Survey

Source: `https://nces.ed.gov/ccd/pubagency.asp`

Local Education Agency Universe Survey Longitudinal Data File (13-year)

Source: `https://nces.ed.gov/ccd/CCD13YR.ASP`

Local Education Agency Universe Survey Dropout and Completion Data

Source: `https://nces.ed.gov/ccd/drpagency.asp`

Public Elementary/Secondary School Universe Survey Data

Source: `https://nces.ed.gov/ccd/pubschuniv.asp`

A.2 Variable Construction

We construct total real revenues per student, our first outcome of interest, using the F-33 survey, where total revenues is the sum of federal, local, and state revenues in each district.[29] We divide this value by the total number of students in the district and deflate by the US CPI, All Urban Consumers Index to convert the figure to year 2000 dollars. Because of large outliers in the F-33 survey, we replace with missing values observations with real total revenues per student that are either 1.5 times larger than the 95th percentile or 0.5 times smaller than the 5th percentile of per-student total revenues within a state. We do this to prevent large outliers from driving our results.[30]

Our measure of graduation rates is a combination of data from the Local Education Agency (School District) Universe Survey, the Public Elementary/Secondary School Universe Survey, and the Local Education Agency (School District) Universe Survey Dropout and Completion Data. From the school-level file, we extract the number of 8th graders in each school, and we aggregate these school-level data to obtain district-level data in each year. From the Local Education Agency data files we construct a time series of total diploma recipients. Data on total diploma recipients was not part of the Local Education Agency universe files as of academic year 1997-98, so beginning with that year, we use the the Dropout and Completion public-use files to obtain the diploma data. We calculate the graduation rate as the total number of diploma recipients in year t as a share of the number of 8th graders in year $t-4$, a measure which Heckman (2010) shows is not susceptible to the downward biased caused by using lagged 9th grade enrollment in the denominator. We top-coded the graduation rates, so that they take a maximum value of 1.

---

[29]For years 1990-91, 1992-93, and 1993-94, we obtained district-level data from Kforce Government Solutions, Inc. These data are public-use files.

[30]The outlier adjustment described above has been used in other studies; for example, see Murray, Evans and Schwab (1998). We generally find that outliers are usually found in districts with very small enrollment.

A.3 Analytic Sample: Additional Details

We observed a series of errors for some state-year-dependent variable combinations. In some states, counts of minority students were mistakenly reported as 0, when in fact they were missing. This occurred in Tennessee, Indiana, and Nevada in years 2000-2005, 2000, and 2005, respectively. The special education variable had two distinct problems. For three states it was mistakenly coded as 0 when it should have been coded as missing. This occurred in Missouri, Colorado, and Vermont in years 2004, 2010 and 2010, respectively. We also observed state-wide 20 percentage point increases in special education enrollment for two states, which immediately returned to pre-spike levels in the year after. This occurred in Oregon and Mississippi in years 2004 and 2007, respectively. Finally, graduation rate data also spiked dramatically before returning to pre-spike levels in three state-years. This occurred in Wyoming, Kentucky and Tennessee in years 1992, 1994 and 1998, respectively. In each of these state-year instances where data were either inappropriately coded as zero or fluctuated due to data error, we coded the value as missing.

B Additional Results

B.1 Understanding differences between weighted and unweighted results

Although the discrepancy between weighted and unweighted results is an indication of model mis-specification, we present some evidence that the mis-specification is driven, in part, by unmodeled treatment effect heterogeneity that varies by district size (Solon et al., 2015). To examine this, we discuss the New York City public schools district (NYCPS) as a case study. Throughout all our analyses, we have excluded NYCPS, the largest school district in the United States, because of its strong influence on the point estimates in FLE quartile 4 when weighting by district enrollment.[31]

In Figure D.1, the causal effect estimates for treatment years 1 to 7 are shown for both the logarithm of per pupil revenues (left panel) and graduation rates (right panel) when NYCPS is included in the sample and when it is not. For the unweighted regressions, it does not matter

---

[31]To be clear, NYCPS was removed from our analytic sample before estimating the benchmark regressions above. When NYCPS is included, the point estimates for graduation rates is much larger, as shown below.

whether NYCPS is included, as all districts are weighted equally and removing one district has little effect on overall outcomes. The weighted regression results, however, show that the inclusion of the NYCPS district produces point estimates for revenues that are systematically higher than weighted results that exclude NYCPS. A similar story holds for graduation rates beginning in treatment year 4. After excluding NYCPS, the weighted model results are closer to the unweighted point estimates, though they still do not perfectly align, suggesting additional treatment effect heterogeneity among high poverty districts.

[Insert Figure D.1 Here]

Examining NYCPS provides just one example of how treatment heterogeneity might be related to district size. As shown in Appendix Figure D.2, NYCPS Graduation rates and teacher salaries increased after 2003, the year New York had its first court ruling. Enrollment, percent poverty, class size and percent minority decreased during this period. Each of these are potential mechanisms for improving graduation rates and likely contribute to the large treatment effect we observe in the weighted results. NYCPS is dropped for all analyses; dropping the next set of largest districts does not have such a dramatic effect on the results as dropping NYCPS does. For this reason, all other districts are retained.

The use of weighted least squares does not necessarily provide a particular estimand of interest, such as the population average *partial* effect (Solon et al., 2015); instead, OLS and WLS results identify different weighted averages of complex heterogenous effects that vary according to district size. In the presence of these heterogeneous effects neither set of results—weighted or unweighted—should be preferred. Trying to model the heterogeneity is also quite complex, as illustrated with the NYCPS case study. Overall, WLS and OLS point estimates are generally consistent in terms of sign; however, the magnitude of the effect size tends to differ. In light of this, we continue to show weighted and unweighted estimates in our tables.

B.2 State Heterogeneity

In this section, we provide further information about heterogeneity among states in their response to court order. To begin, we show estimates for a modified differences-in-differences equation, where the full vector of treatment indicator variables (four poverty quartiles for 19 treatment years) are interacted with the 12 states undergoing court order. These results are intended to provide assurance that the assumption of a linear response to court order we made previously defensible. Results for weighted and ordinary least squares estimates are shown in Figures D.3 and D.4.

[Insert Figure D.3 Here]

[Insert Figure D.4 Here]

We plot effect estimates for treatment years one through seven for both log per pupil revenues and graduation rates among quartile 4 treatment states. In Figure D.3, panel (a) displays results for two states in which revenues increased on average across the seven treatment years and graduation rates were either null or negative on average; panel (b) displays results for five states in which both revenues and graduation rates increased on average; panel (c) displays results for one state in which both revenues and graduation rates decreased on average; panel (d) displays results for two states in which revenues were null or negative and graduation rates were positive on average.

Note that the non-parametric results nearly perfectly coincide with the linear versions: Wyoming and New Hampshire are shown to have had positive revenues effects and negative graduation effects. Tennessee is shown to have a negative response in both revenues and graduation rates. Massachusetts and North Carolina are shown to have had a positive response in terms of graduation rates and a negative response in revenues. These state-specific results match what was shown when we assumed treatment response was linear. We identify positive revenues effects in New Jersey for the linear model and nothing in the non-parametric model, but the magnitudes in both models are very small and not distinguishable from zero.

In order to provide better understanding of both the state-specific response to court ordered finance reform and the influence that correlated random trends have on estimation, we plot aggregate trends of log per pupil revenues and graduation rates for each treated state. In order to make

50

these trends comparable to regression models, the data are generated in the following way:

1. The weighted average of log per pupil revenues and graduation rates are calculated for each state, FLE quartile and year, with the exception that all non-treated states are combined into a single "counterfactual" state. These collapsed data now include observations for years 1991 to 2010, quartiles one through four, and 12 treated states along with the aggregated "counterfactual" state.

2. These aggregate data are detrended and demeaned for all years prior to court order. The estimated pre-court order slope is then projected onto the rest of the treatment years. Predicted values based on the estimated state-quartile mean and pre-treatment slope are then used to construct residuals.

We plot residuals for log per pupil revenues and graduation rates for each treated state; to approximate the counterfactual trend, we plot residuals (generated from the same regression) for the "counterfactual" quartile four. The residuals are plotted as solid and dashed thin black lines, where solid corresponds to a treated state and dashed to the counterfactual trend. The average over the post-treatment period is shown as a thick black solid and dashed line, respectively. Finally, in order to demonstrate the consequence of de-trending, we also plot the non-detrended residuals (demeaned only) for treated state quartile four. We include 11 figures, excluding Kentucky which lacks pre-treatment data.

[Insert Figure D.5 Here]

[Insert Figure D.6 Here]

[Insert Figure D.7 Here]

[Insert Figure D.8 Here]

[Insert Figure D.9 Here]

[Insert Figure D.10 Here]

[Insert Figure D.11 Here]

[Insert Figure D.12 Here]

[Insert Figure D.13 Here]

[Insert Figure D.14 Here]

[Insert Figure D.15 Here]

These figures are informative for two reasons. First, in each case (except New Jersey), the average difference between quartile four treated and quartile four counterfactual coincides with the state-specific differences-in- differences estimates shown above in Section 5.5 (State by State Heterogeneity). In other words, these figures provide face validity for the state-specific estimates generated from the linear and non-parametric models. New Jersey is the only state for which the depiction of change in revenues shown here departs from the estimated change: the descriptive figures show a negative change in revenues following court order, whereas our regression equations identify a positive effect. It should be noted that the positive effect that is identified in Equation 3 is not distinguishable from zero. Moreover, New Jersey's pre-1997 revenues trend is highly unusual. There is an evident shift in revenues following the first 1994 Abbott case (which was not ruled on Adequacy grounds), but that shift disappears in two years. It is difficult to estimate the pre-treatment trend with such a large disruption.

The usefulness of estimating a pre-treatment trend for these treated quartiles is evident in many cases. Tennessee (Figure D.14) perhaps provides the most compelling example. Note the solid gray line indicating the demeaned revenues trend; it is increasing steeply prior to reform. Unless we account for that pre-treatment trend, we will mis-attribute the effect of court order to it. This is what is meant by a "unit-specific secular trend;" year and unit effects do not account for it. We can see from the same line that there is a slight downward shift in revenues that begins in 1995, right at the time of the state's first Adequacy ruling. The pre-trend is accounted for in these figures by regressing log revenues on time for that quartile for the period prior to 1995; we construct residuals by projecting the pre-trend through the rest of the data and plot those residuals as a solid black line. De-trending has a profound effect on the inference; we now see that Tennessee's change in revenues, conditional on its pre-trend, is decreasing relative to the de-trended counterfactual trend (indicated by the dashed black line). A similar effect of de-trending can be seen in Arkansas,

52

Kansas, New York, and Ohio (Figures D.6, D.7, D.11, and D.13).

In some cases, the influence of a pre-trend is more ambiguous. Massachusetts (Figure D.8) has very little pre-trend information, so projecting the pre-trend outwards throughout the rest of the treatment period can have a profound and not well identified effect. For this reason, state specific estimates provide noisy and, from our point of view, descriptive evidence about the effect of court order at the state level.

C   Technical Appendix: Interactive Fixed Effect Models

C.1   Estimation Procedure Overview

Here we briefly outline the procedure to estimate an interactive fixed effect model. For convenience, we refer all non-factor structure coefficients using the vector β.

The procedure begins by choosing starting values for the β vector, which we denote as $\tilde{\beta}$. Then, the following steps are carried out:

(1) Calculate the residuals of the OLS estimator *excluding* the factor structure using $\tilde{\beta}$.

(2) Estimate the factor loadings, λ_s, and the common factors, F_t, on the residual vector obtained in step [1] using the Levenberg–Marquardt algorithm (Wright & Holt, 1985).

(3) After estimating the factor structure, we remove it from the regressors using partitioned regression. Then, we re-estimate the model using a least squares objective function in order to obtain a new estimate of $\tilde{\beta}$.

(4) Steps [1] to [3] are repeated until the following stopping condition is achieved: After comparing each element of the vector $\tilde{\beta}$ obtained in step [3] with the previous estimate $\tilde{\beta}^{old}$, we stop if $|\tilde{\beta}_k - \tilde{\beta}_k^{old}| < 10^{-8}$ for all k. If this condition is not achieved, then we stop if the difference in the least squares objective function calculated in step [3] is greater than the Total Sum of Squares multiplied by 10^{-10}. This stopping condition takes effect when the estimator is not converging.

C.2   Sparse Factor Models

Traditional approaches to factor methods specify factor loadings at the lowest unit of analysis, which in this paper is the district level, d. However, we are interested in accounting for interdependence between states, the level at which treatment occurs. While principal components analysis generally requires one observation per i-by-t, sparse factor methods are available that allow for multiple i per t, as in our case, where we have multiple districts d within states s for every year

t. Moreover, sparse factor methods allow for missing data, thereby obviating the need to implement interpolation or multiple imputation to fill in missing data (Ilin & Raiko, 2010; Raiko, Ilin, & Karhunen, 2008; Wright & Holt, 1985).

The estimated factors and factor loadings $\lambda_s' F_t$ can be decomposed into the variables $\theta_s F_t + \delta_t \lambda_s$ and entered directly into the regression model, thus allowing us to cluster the standard errors at the state level; we describe the approach to estimating standard errors below in Appendix Section C.3. These non-treatment parameters are eliminated using high dimensional fixed effects according to the Frisch-Waugh-Lovell theorem (Frisch & Waugh, 1933; Lovell, 1963).[32]

C.3 Computing Standard Errors

In this section, we show how to obtain a variety of standard error structures using the estimates of the factor loadings and common factors that we obtain from the iterative procedure used in Moon and Weidner (2014). This may be necessary when, for example, the unit of observation is at a lower level of aggregation than the unit where serial correlation is expected. We thank Martin Weidner for suggesting this procedure and for providing details about it.

First, we note that the interactive fixed effects (IFE) procedure provides us with a vector of estimates for λ_i and F_t. We denote the estimated parameter vectors as $\hat{\lambda}_i$ and $\hat{F}_t$.

The goal is to use these estimates to estimate the model using OLS in a statistical package that allows for different error structures such as clustered standard errors. We want to estimate the following equation:

$$(4) \quad Y_{it} = \alpha_i + \theta_t + D_{it}' \beta + \lambda_i' F_t + \varepsilon_{it},$$

where we have $\hat{\lambda}_i$ and $\hat{F}_t$ from the IFE procedure. If we perform a first-order Taylor expansion about $\lambda_i' F_t$, then we can utilize our estimates and obtain the same coefficients on our treatment indicators using the IFE procedure.

In general, the first-order Taylor expansion for a function of two variables λ and F about the

---

[32]The model is estimated in Julia using the package FixedEffects.jl and SparseFactorModels.jl. (Gomez, 2015a,b)

points $\hat{\lambda}$ and $\hat{F}$ is as follows:

$$(5) \quad f(\lambda, F) \approx f(\hat{\lambda}, \hat{F}) + \frac{\partial f}{\partial \lambda}\Big|_{\lambda=\hat{\lambda}, F=\hat{F}}(\lambda - \hat{\lambda}) + \frac{\partial f}{\partial F}\Big|_{\lambda=\hat{\lambda}, F=\hat{F}}(F - \hat{F})$$

In our case, we have $f(\lambda, F) = \lambda_i' F_t$. After applying the Taylor expansion, we obtain:

$$(6) \quad \lambda_i' F_t \approx \lambda_i' \hat{F}_t + \hat{\lambda}_i' F_t - \hat{\lambda}_i' \hat{F}_t.$$

Next, we define $\alpha_i = \lambda - \hat{\lambda}$ and $F_t = \theta_t$. Therefore, we obtain:

$$(7) \quad \lambda_i' F_t \approx \alpha_i \hat{F}_t + \hat{\lambda}_i \theta_t.$$

The model we can estimate via OLS is

$$(8) \quad Y_{it} = \alpha_i + \theta_t + D_{it}'\beta + \alpha_i \hat{F}_t + \hat{\lambda}_i \theta_t + \varepsilon_{it},$$

where the α_i and the θ_t are parameters that are to be estimated. We can easily apply various standard error corrections to this model in a wide variety of statistics packages.

Proposition C.1. *OLS estimation of Equation (8) is identical to OLS estimation of Equation (4).*

Proof. For this proof, we show that the first order conditions (FOCs) are identical. Thus, the OLS estimates will also be identical.

To begin, we first simplify notation. We let $X_{it} = \alpha_i + \theta_t + D_{it}$. There is no loss of generality in performing this step, as we can think of estimating the unit-specific fixed effects and the time-specific fixed using the least-squares dummy variable (LSDV) method.[33]

Part 1: Obtain FOCs for Equation (4)

Step 1: Specify the objective function:

$$\underset{\hat{\beta}, \hat{\lambda}_i, \hat{F}_t}{\text{minimize}} \quad \Psi = \sum_i \sum_t (y_{it} - X_{it}'\hat{\beta} - \hat{\lambda}_i' \hat{F}_t)^2$$

---

[33]In this paper, we actually remove the fixed effects using the within transformation.

Step 2: Obtain partial derivatives for each parameter and set them equal to zero

$$\frac{\partial \Psi}{\partial \hat{\beta}} = -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \hat{\lambda}_i' \hat{F}_t)(X_{it}) \right] = 0$$

$$\frac{\partial \Psi}{\partial \hat{\lambda}_i} = -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \hat{\lambda}_i' \hat{F}_t)(\hat{F}_t) \right] = 0$$

$$\frac{\partial \Psi}{\partial \hat{F}_t} = -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \hat{\lambda}_i' \hat{F}_t)(\hat{\lambda}_i) \right] = 0$$

Part 2: Obtain FOCs for Equation (8)

Step 1: Specify the objective function:

$$\underset{\hat{\beta},\hat{\lambda}_i,\hat{F}_t}{\text{minimize}} \quad \Phi = \sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \alpha_i \hat{F}_t - \hat{\lambda}_i \theta_t)^2$$

Step 2: Obtain partial derivatives for each parameter and set them equal to zero

$$\frac{\partial \Phi}{\partial \hat{\beta}} = -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \alpha_i \hat{F}_t - \hat{\lambda}_i \theta_t)(X_{it}) \right]$$

$$= -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \hat{\lambda}_i' \hat{F}_t)(X_{it}) \right] = 0$$

$$\frac{\partial \Phi}{\partial \hat{\lambda}_i} = -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \alpha_i \hat{F}_t - \hat{\lambda}_i \theta_t)(\hat{F}_t) \right]$$

$$= -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \hat{\lambda}_i' \hat{F}_t)(\hat{F}_t) \right] = 0$$

$$\frac{\partial \Phi}{\partial \hat{F}_t} = -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \alpha_i \hat{F}_t - \hat{\lambda}_i \theta_t)(\hat{\lambda}_i) \right]$$

$$= -2 \left[\sum_i \sum_t (y_{it} - X'_{it}\hat{\beta} - \hat{\lambda}_i' \hat{F}_t)(\hat{\lambda}_i) \right] = 0$$

To simplify the FOCs, we use the fact that we had defined $\alpha_i = \lambda - \hat{\lambda}$ and $F_t = \theta_t$. Given that we evaluate the minimization problem at $\hat{\beta}$, $\hat{\lambda}_i$, and $\hat{F}_t$, α_i will be equal to zero. Finally, we can replace $\hat{\lambda}_i \theta_t$ with $\hat{\lambda}_i' \hat{F}_t$. Thus, the FOCs for Equations (4) and (8) are identical. $\qquad \square$

D Additional Figures

Figure D.1: Change in Log Per Pupil Revenues & Graduation Rates after Court Ruling, FLE Quartile 4, with and without NYCPS
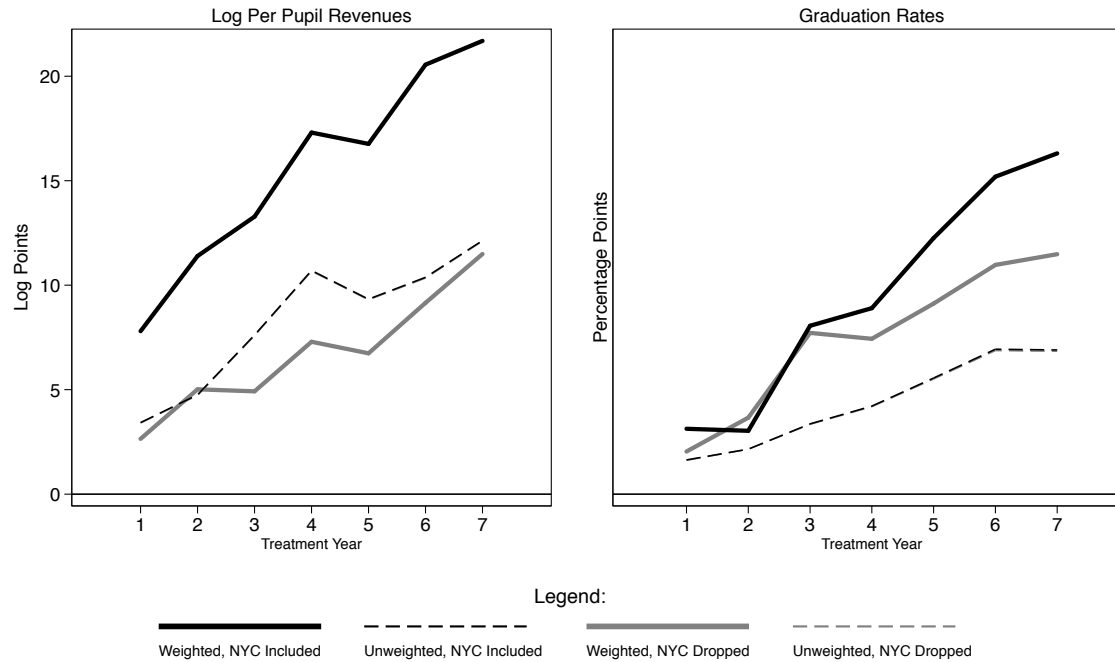
Notes: Differences-in-differences with treatment effects estimated non-parametrically after reform. Model accounts for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq}), and state-by-FLE linear time trends ($\psi_{sq}t$). Left panel corresponds to results for log revenues; right panel to results for graduation rates. Black lines are for models that include NCYPS; gray lines are for models that exclude NYCPS. Solid lines are for models that include enrollment as analytic weight; dashed lines are for unweighted models. Unweighted models completely overlap (black and gray dashed lines are not distinguishable). When NCYPS is removed, point estimates for weighted models are closer to unweighted models.

Figure D.2: New York City Public Schools Potential Mediators & Outcomes
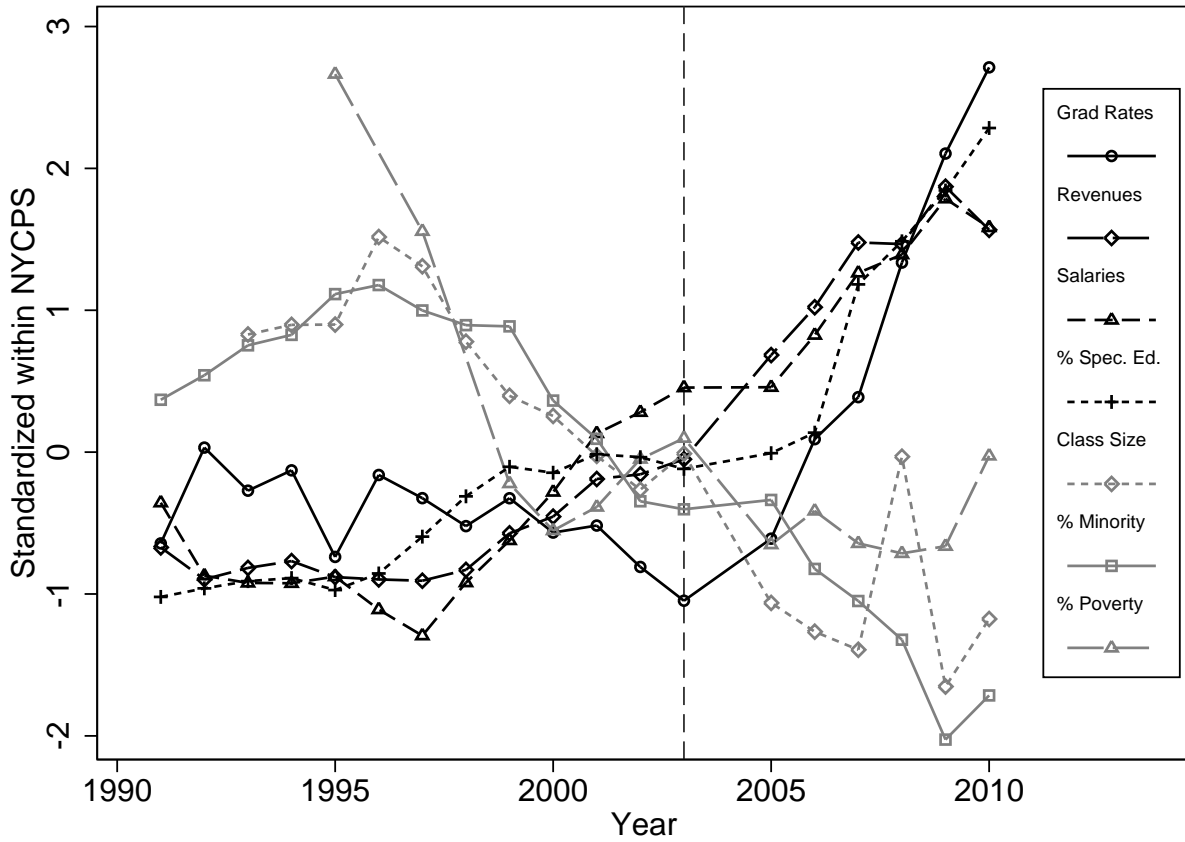
NYCPS is the largest district in the United States, and it was also demonstrably improving in many ways during this period. Here, we plot standardized beta coefficients for various outcomes of interest. These are standardized to be mean zero with standard deviation one for NYCPS across the entire time period. Graduation rates and teacher salaries increased after 2003, the year New York had its first court ruling. Enrollment, percent poverty, class size and percent minority decreased during this period. All of these potential mechanisms for improving graduation rates are intended to be controlled for in a differences-in-differences framework through the inclusion of year dummies.

Figure D.3: Treatment Heterogeneity between States (WLS)

Dashed lines reflect log per pupil revenues; solid lines reflect graduation rates.

Notes: Lines reflect point estimates from an augmented version of the benchmark model (i.e. differences-in-differences with treatment effects estimated non-parametrically after reform, accounting for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq}), and state-by-FLE linear time trends ($\psi_{sq}t$)), in which we include 12 treatment state indicator variables ($\sum_{s=1}^{11}$) and interact these with the FLE quartile and treatment year variables ($\sum_{q=1}^{4}\sum_{n=1}^{16}$). This provides estimates of treatment effects for 12 treated states (11 shown, as Alaska has only one year post-treatment), 4 quartiles, for each year following timing of Court order. Here we overlay log per pupil revenues and graduation rates for FLE quartile 4 (graduation rates solid; revenues dashed). Marker colors and symbols each uniquely identify a treatment state.

Figure D.4: Treatment Heterogeneity between States (OLS)

Dashed lines reflect log per pupil revenues; solid lines reflect graduation rates.

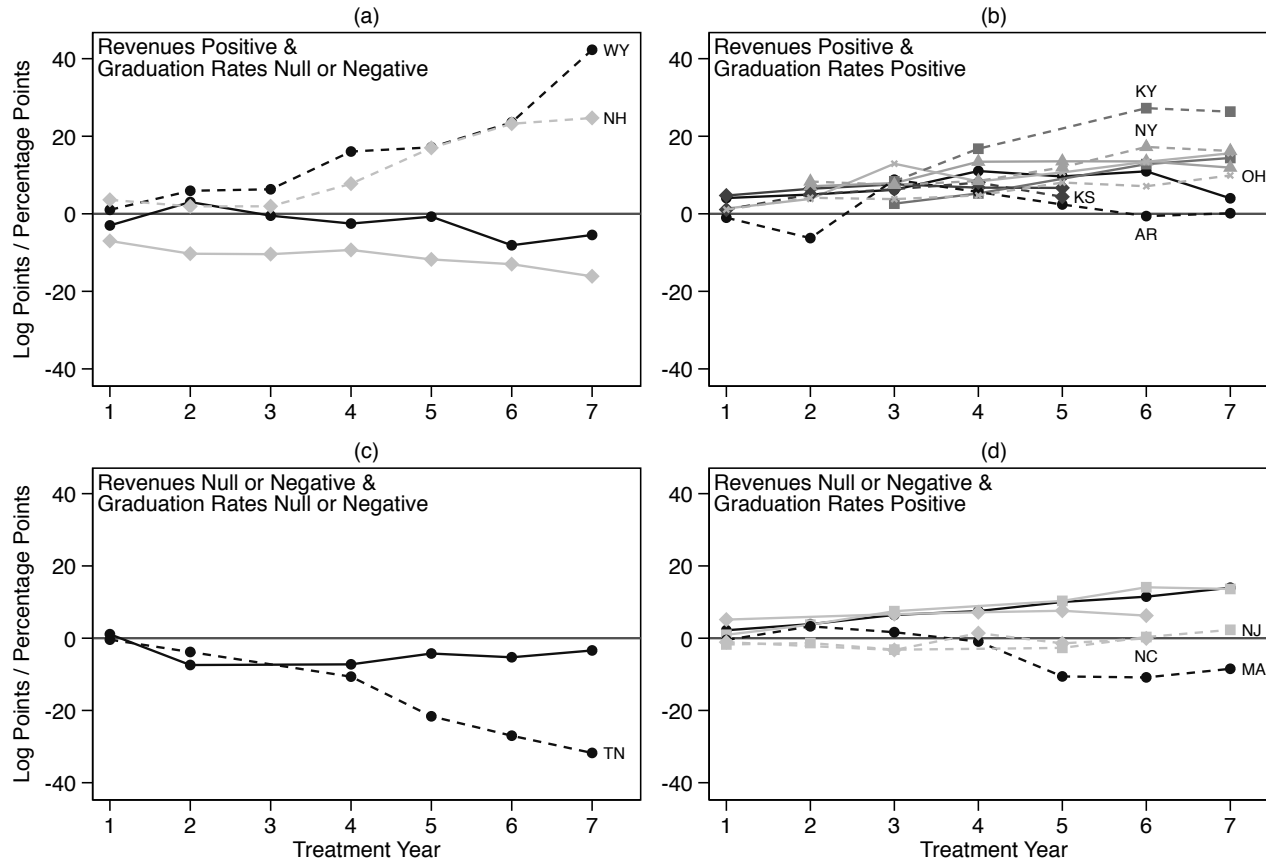Notes: Lines reflect point estimates from an augmented version of the benchmark model (i.e. differences-in-differences with treatment effects estimated non-parametrically after reform, accounting for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq}), and state-by-FLE linear time trends ($\psi_{sq}t$)), in which we include 12 treatment state indicator variables ($\sum_{s=1}^{11}$) and interact these with the FLE quartile and treatment year variables ($\sum_{q=1}^{4}\sum_{n=1}^{16}$). This provides estimates of treatment effects for 12 treated states (11 shown, as Alaska has only one year post-treatment), 4 quartiles, for each year following timing of Court order. Here we overlay log per pupil revenues and graduation rates for FLE quartile 4 (graduation rates solid; revenues dashed). Marker colors and symbols each uniquely identify a treatment state.

Figure D.5: Revenues and Graduation Changes by State (AK)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

Figure D.6: Revenues and Graduation Changes by State (AR)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

Figure D.7: Revenues and Graduation Changes by State (KS)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

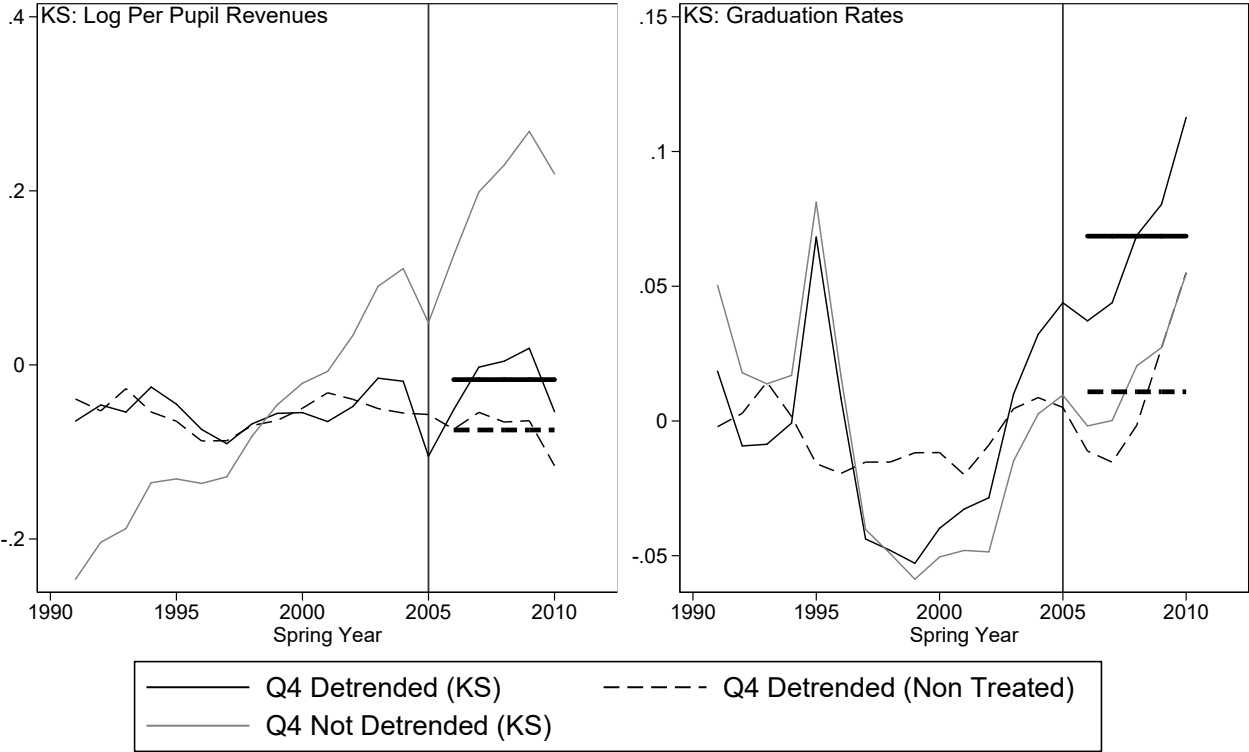Figure D.8: Revenues and Graduation Changes by State (MA)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

Figure D.9: Revenues and Graduation Changes by State (NH)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

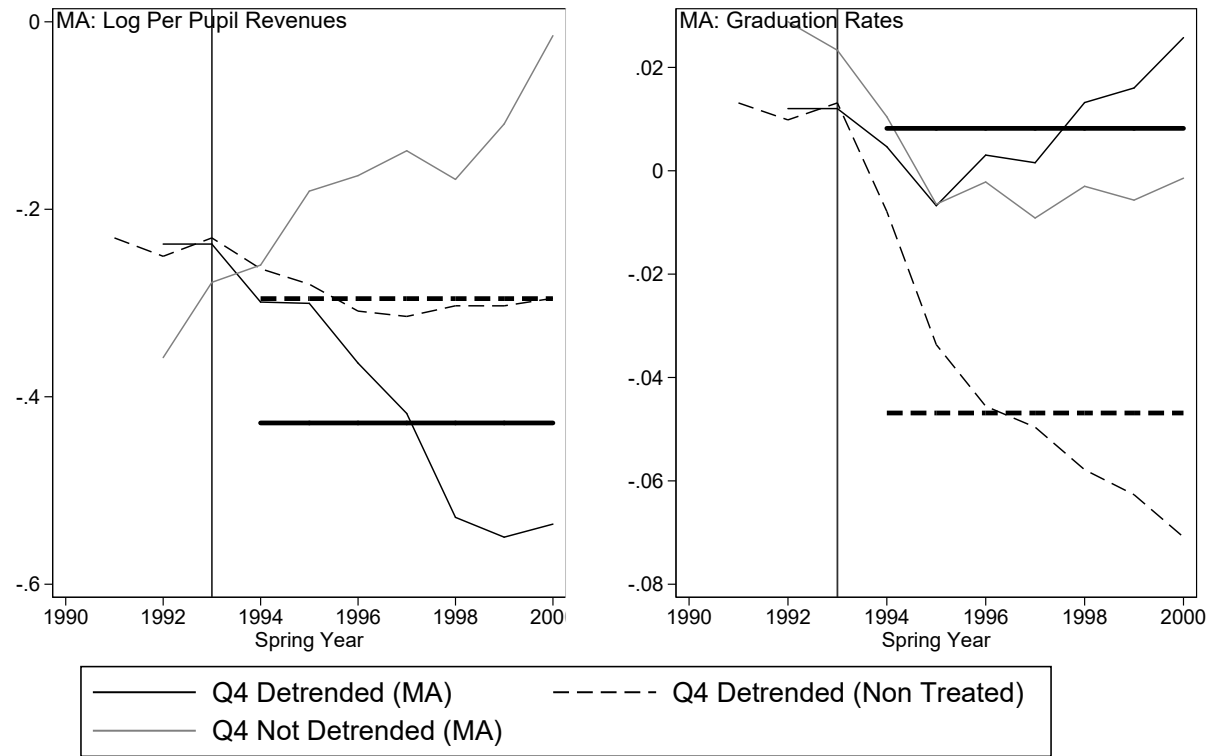Figure D.10: Revenues and Graduation Changes by State (NJ)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

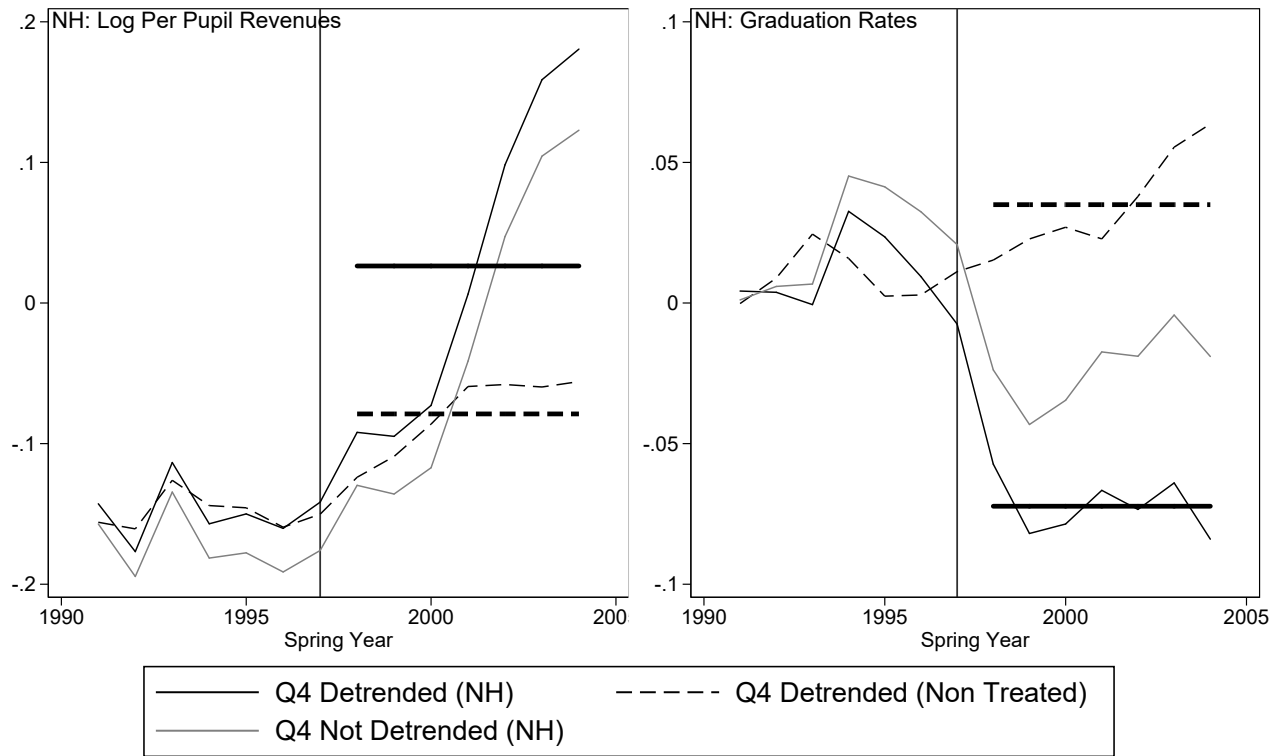Figure D.11: Revenues and Graduation Changes by State (NY)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

Figure D.12: Revenues and Graduation Changes by State (NC)
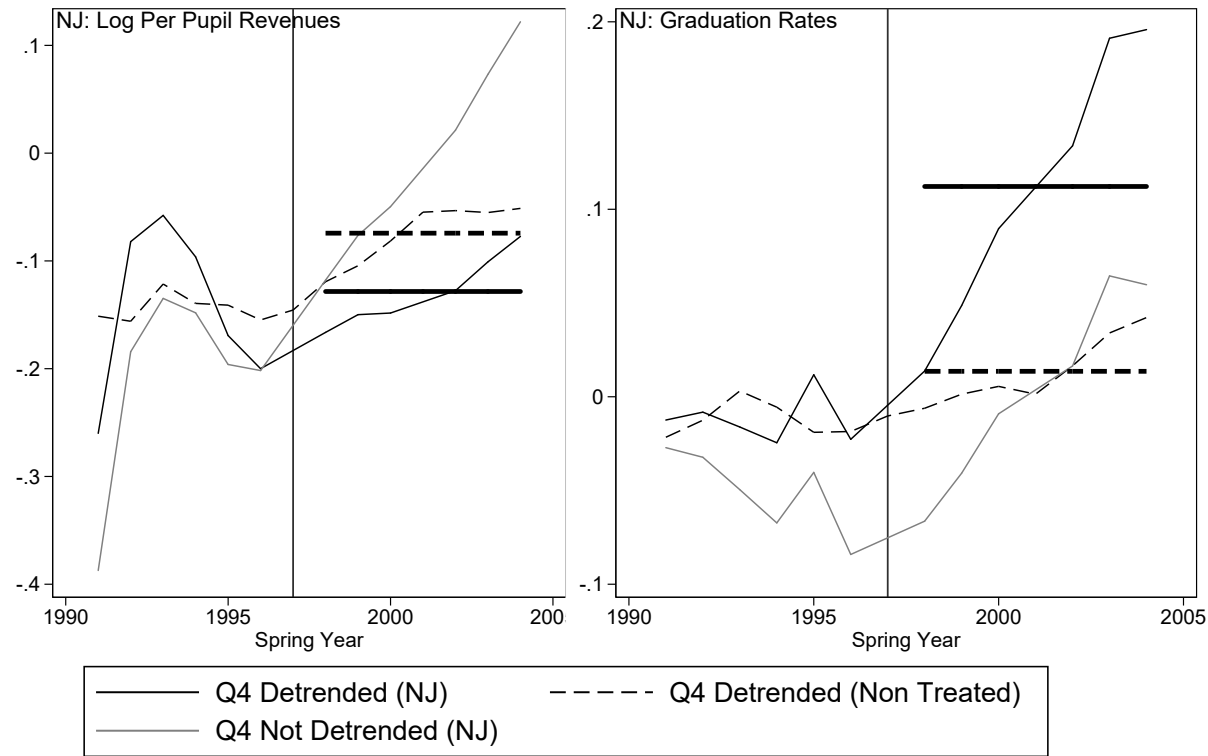
Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

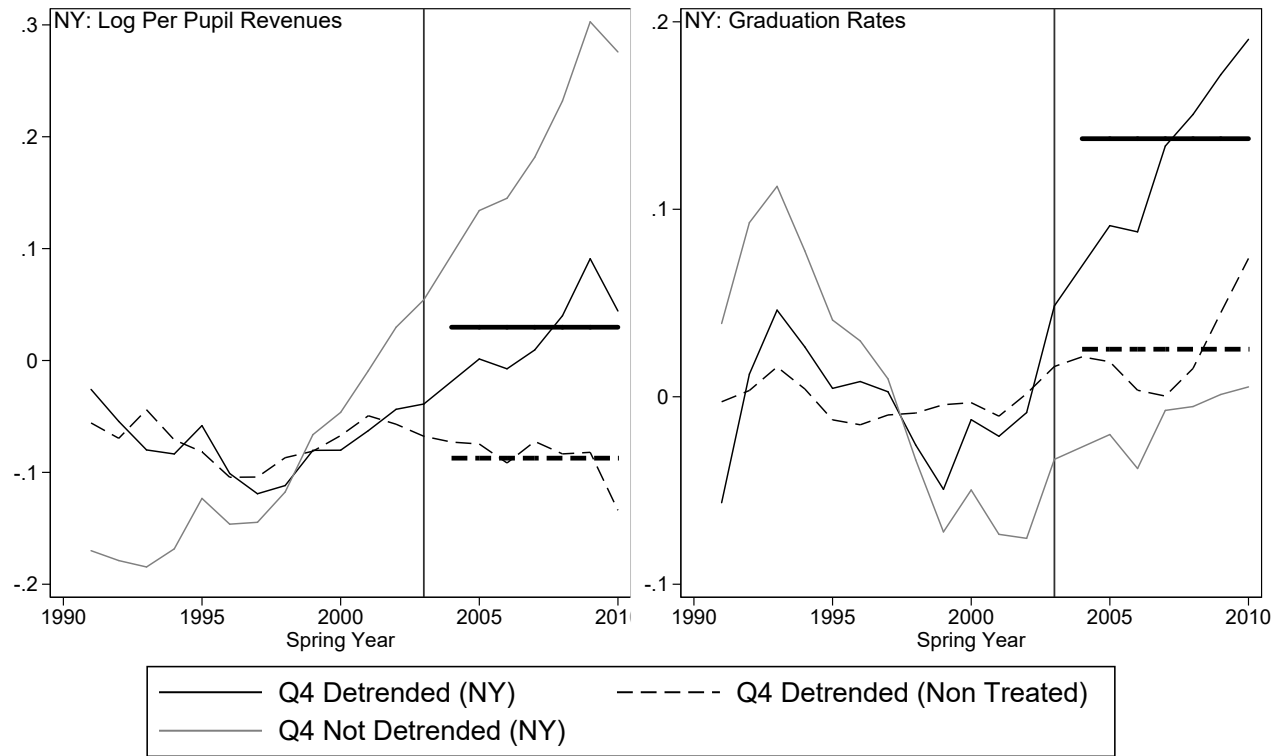Figure D.13: Revenues and Graduation Changes by State (OH)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

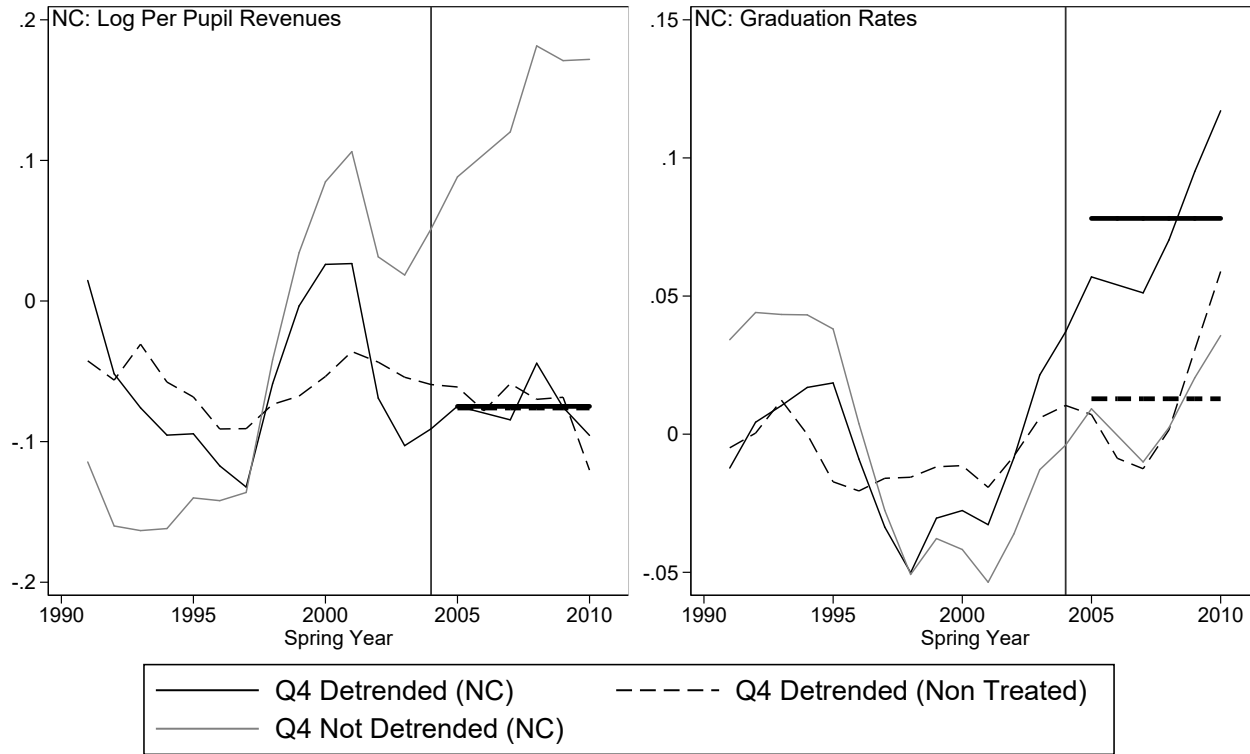Figure D.14: Revenues and Graduation Changes by State (TN)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

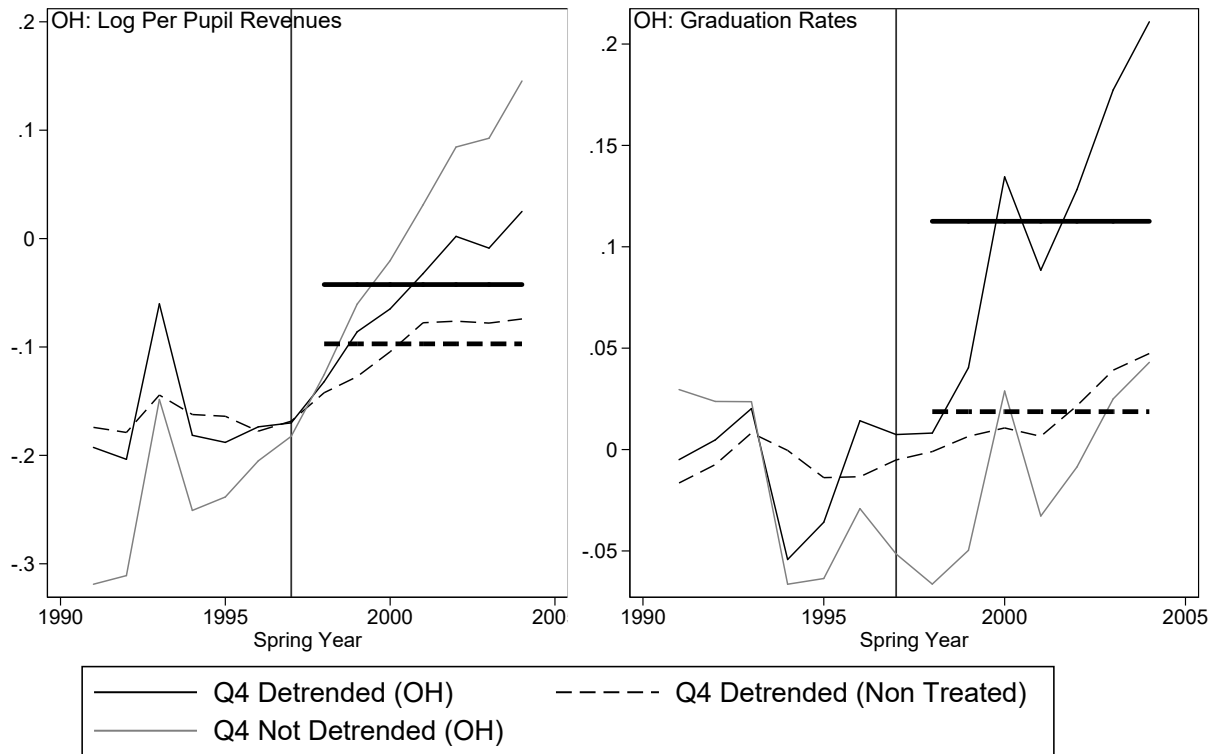Figure D.15: Revenues and Graduation Changes by State (WY)

Notes: Data are generated in the following way: The analytic sample is collapsed as a weighted average by state, poverty quartile and year. Non-treated states are pooled together into a common "counterfactual" treated state. These averages are then demeaned and detrended, by removing state-by-quartile time trends and fixed effects. The residuals from these models are shown in the figure. Solid black lines indicate demeaned and detrended revenues and graduation rates for treated state poverty quartile four. Dashed black lines indicate demeaned and detrended revenues and graduation rates averaged across non-treated states poverty quartile four. Solid gray lines indicate demeaned (not detrended) revenues and graduation rates for treated state poverty quartile four. Thick solid black lines correspond to the average of the residuals for treated states following reform; thick dashed lines correspond to the average of the residuals for non-treated states. Vertical line indicates timing of court order.

E   Additional Tables

Table E.1: States Exiting from Treatment Status

States	Treatment Year
N/A	1
AK	2
AK	3
AK	4
AK	5
AK, KS	6
AK, KS, NC	7
AK, KS, NC, NY	8
AK, AR, KS, NC, NY	9
AK, AR, KS, NC, NY	10
AK, AR, KS, NC, NY	11
AK, AR, KS, NC, NY	12
AK, AR, KS, NC, NY	13
AK, AR, KS, NC, NH, NJ, NY, OH	14
AK, AR, KS, NC, NH, NJ, NY, OH	15
AK, AR, KS, NC, NH, NJ, NY, OH, TN, WY	16
AK, AR, KS, NC, NH, NJ, NY, OH, TN, WY	17
AK, AR, KS, MA, NH, NJ. NC, NY, OH, TN, WY	18
AK, AR, KS, MA, NH, NJ, NC, NY, OH, TN, WY	19 on

Notes: This table lists states that no longer contribute treatment information in year t of the dynamic treatment response period. This occurs for states that were treated relatively late in the panel of data that is available. For example, our data set ends in year 2010, and Table 1 shows that Kansas had their system overturned in 2005. Thus, Kansas contributes to the identification of the treatment response in treatment years 1 through 5, corresponding to 2006 to 2010, inclusive. It does not contribute to treatment years 6 and beyond.

Table E.2: Log Per Pupil Revenues Results, Model Sensitivity Specifications, Weighted Least Squares

Treatment Year							Model Number								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.028	.026	.015	.019	.02	.01	.048	.02	.027	.032	.025	.005	.051	.021	.026
	(.01)	(.013)	(.012)	(.016)	(.016)	(.009)	(.017)	(.011)	(.014)	(.009)	(.013)	(.01)	(.018)	(.01)	(.014)
2	.056	.05	.046	.049	.052	.034	.08	.05	.051	.061	.05	.029	.083	.053	.051
	(.013)	(.019)	(.022)	(.024)	(.023)	(.02)	(.024)	(.019)	(.02)	(.013)	(.022)	(.02)	(.025)	(.019)	(.023)
3	.055	.049	.044	.049	.047	.029	.083	.046	.05	.066	.054	.027	.092	.055	.055
	(.017)	(.022)	(.028)	(.033)	(.031)	(.026)	(.026)	(.025)	(.024)	(.016)	(.022)	(.027)	(.026)	(.024)	(.023)
4	.079	.073	.06	.069	.063	.045	.11	.061	.075	.092	.078	.043	.121	.071	.08
	(.017)	(.024)	(.032)	(.04)	(.039)	(.03)	(.028)	(.031)	(.025)	(.015)	(.021)	(.032)	(.027)	(.029)	(.022)
5	.074	.067	.059	.071	.059	.031	.11	.052	.07	.09	.076	.03	.124	.066	.078
	(.021)	(.03)	(.04)	(.052)	(.047)	(.039)	(.035)	(.038)	(.032)	(.021)	(.029)	(.041)	(.035)	(.037)	(.031)
6	.102	.092	.091	.105	.076	.047	.142	.071	.095	.121	.104	.048	.158	.089	.107
	(.018)	(.037)	(.044)	(.061)	(.054)	(.045)	(.046)	(.045)	(.039)	(.02)	(.035)	(.046)	(.046)	(.044)	(.037)
7	.129	.115	.119	.128	.103	.059	.174	.09	.119	.154	.133	.064	.195	.114	.137
	(.015)	(.036)	(.055)	(.084)	(.076)	(.065)	(.047)	(.063)	(.038)	(.016)	(.034)	(.067)	(.047)	(.063)	(.036)
Weight															
Yes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
No															
Fixed Effect															
FLE-Year	X	X	X	X	X	X	X	X	X						
Year										X	X	X	X	X	X
CRT															
None	X									X					
State-FLE		X	X	X	X						X				
State-FLE²						X						X			
State							X						X		
State²								X						X	
District									X						X
Factor															
0	X	X				X	X	X	X	X	X	X	X	X	X
1			X												
2				X											
3					X										

Benchmark model is highlighted. Point estimates and standard errors in parentheses are shown for various model specifications. Model choice is indicated in bottom of panel. We permute parameters using weights (yes/no), fixed effects (FLE-by-year/year), correlated random trends (none/state-by-FLE/state-by-FLE squared/state/state squared/district), and factors (0/1/2/3). Additional combinations available upon request.

Table E.3: Log Per Pupil Revenues Results, Model Sensitivity Specifications, OLS

Treatment Year	Model Number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.024	.034	.004	.015	.017	.001	.05	.013	.033	.03	.03	-.003	.056	.019	.03
	(.012)	(.014)	(.011)	(.013)	(.012)	(.012)	(.018)	(.012)	(.015)	(.012)	(.014)	(.013)	(.018)	(.01)	(.014)
2	.043	.047	.018	.038	.043	.018	.068	.032	.047	.051	.044	.013	.077	.041	.044
	(.017)	(.018)	(.02)	(.021)	(.021)	(.022)	(.024)	(.02)	(.018)	(.017)	(.017)	(.022)	(.024)	(.019)	(.018)
3	.071	.076	.041	.067	.074	.04	.101	.055	.075	.08	.072	.033	.11	.065	.071
	(.024)	(.019)	(.029)	(.028)	(.024)	(.03)	(.023)	(.026)	(.02)	(.024)	(.018)	(.032)	(.022)	(.026)	(.019)
4	.098	.107	.056	.087	.092	.055	.135	.069	.105	.109	.101	.047	.145	.08	.101
	(.024)	(.018)	(.034)	(.033)	(.031)	(.037)	(.025)	(.033)	(.019)	(.025)	(.018)	(.04)	(.025)	(.034)	(.018)
5	.083	.093	.034	.072	.082	.031	.125	.045	.091	.095	.087	.02	.138	.057	.086
	(.029)	(.026)	(.037)	(.036)	(.034)	(.041)	(.034)	(.037)	(.027)	(.03)	(.026)	(.044)	(.034)	(.037)	(.027)
6	.102	.103	.033	.075	.09	.027	.144	.043	.101	.119	.1	.017	.161	.06	.099
	(.022)	(.038)	(.032)	(.034)	(.036)	(.041)	(.044)	(.036)	(.04)	(.021)	(.036)	(.043)	(.042)	(.037)	(.037)
7	.123	.121	.038	.087	.118	.027	.168	.043	.118	.139	.115	.013	.184	.059	.114
	(.019)	(.039)	(.043)	(.045)	(.05)	(.058)	(.045)	(.054)	(.04)	(.018)	(.036)	(.06)	(.043)	(.055)	(.038)
Weight															
Yes															
No	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Fixed Effect															
FLE-Year	X	X	X	X	X	X	X	X	X						
Year										X	X	X	X	X	X
CRT															
None	X									X					
State-FLE		X	X	X	X						X				
State-FLE^2						X						X			
State							X						X		
State^2								X						X	
District									X						X
Factor															
0	X	X				X	X	X	X	X	X	X	X	X	X
1			X												
2				X											
3					X										

Benchmark model is highlighted. Point estimates and standard errors in parentheses are shown for various model specifications. Model choice is indicated in bottom of panel. We permute parameters using weights (yes/no), fixed effects (FLE-by-year/year), correlated random trends (none/state-by-FLE/state-by-FLE squared/state/state squared/district), and factors (0/1/2/3). Additional combinations available upon request.

Table E.4: Graduation Rates Results, Model Sensitivity Specifications, Weighted Least Squares

Treatment Year	Model Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-.013	.02	.013	.01	.013	.018	.002	.001	.019	-.01	.02	.019	.005	.004	.018
	(.009)	(.009)	(.006)	(.007)	(.009)	(.012)	(.007)	(.008)	(.009)	(.008)	(.006)	(.01)	(.005)	(.005)	(.006)
2	-.01	.037	.03	.026	.029	.03	.013	.012	.036	-.008	.033	.031	.014	.012	.033
	(.01)	(.012)	(.009)	(.01)	(.012)	(.019)	(.007)	(.012)	(.012)	(.011)	(.009)	(.016)	(.006)	(.01)	(.009)
3	.021	.077	.061	.056	.052	.063	.048	.045	.076	.021	.071	.063	.048	.044	.07
	(.02)	(.017)	(.019)	(.021)	(.015)	(.021)	(.014)	(.017)	(.017)	(.022)	(.017)	(.018)	(.016)	(.018)	(.017)
4	.009	.074	.061	.055	.048	.058	.042	.042	.074	.008	.066	.057	.04	.039	.065
	(.01)	(.018)	(.013)	(.013)	(.018)	(.035)	(.01)	(.023)	(.019)	(.011)	(.014)	(.031)	(.008)	(.02)	(.014)
5	.017	.091	.07	.065	.054	.063	.056	.052	.09	.017	.083	.066	.055	.051	.082
	(.014)	(.017)	(.016)	(.017)	(.02)	(.04)	(.01)	(.025)	(.018)	(.014)	(.013)	(.035)	(.01)	(.022)	(.014)
6	.024	.11	.085	.079	.068	.071	.07	.065	.108	.027	.103	.078	.072	.066	.101
	(.02)	(.019)	(.019)	(.02)	(.024)	(.049)	(.013)	(.029)	(.019)	(.021)	(.015)	(.043)	(.013)	(.028)	(.015)
7	.018	.115	.084	.077	.062	.062	.072	.064	.113	.023	.107	.073	.075	.066	.106
	(.024)	(.02)	(.023)	(.025)	(.026)	(.057)	(.015)	(.03)	(.021)	(.023)	(.018)	(.05)	(.016)	(.032)	(.018)
Weight															
Yes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
No															
Fixed Effect															
FLE-Year	X	X	X	X	X	X	X	X	X						
Year										X	X	X	X	X	X
CRT															
None	X									X					
State-FLE		X	X	X	X						X				
State-FLE^2						X						X			
State							X						X		
State^2								X						X	
District									X						X
Factor															
0	X	X				X	X	X	X	X	X	X	X	X	X
1			X												
2				X											
3					X										

Benchmark model is highlighted. Point estimates and standard errors in parentheses are shown for various model specifications. Model choice is indicated in bottom of panel. We permute parameters using weights (yes/no), fixed effects (FLE-by-year/year), correlated random trends (none/state-by-FLE/state-by-FLE squared/state/state squared/district), and factors (0/1/2/3). Additional combinations available upon request.

Table E.5: Graduation Rates Results, Model Sensitivity Specifications, OLS

Treatment Year	Model Number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-.001	.016	.012	.014	.008	.021	.012	.016	.016	-.004	.019	.022	.009	.012	.019
	(.007)	(.008)	(.006)	(.007)	(.006)	(.01)	(.006)	(.007)	(.009)	(.007)	(.009)	(.011)	(.006)	(.007)	(.009)
2	-.004	.022	.019	.021	.01	.029	.015	.021	.021	-.009	.023	.029	.01	.015	.023
	(.011)	(.01)	(.007)	(.008)	(.01)	(.014)	(.007)	(.009)	(.01)	(.012)	(.009)	(.014)	(.006)	(.009)	(.009)
3	.003	.033	.024	.026	.011	.041	.025	.029	.033	-.004	.034	.04	.018	.023	.034
	(.012)	(.011)	(.009)	(.009)	(.01)	(.018)	(.008)	(.012)	(.011)	(.012)	(.01)	(.019)	(.007)	(.012)	(.01)
4	.007	.042	.034	.035	.012	.053	.033	.042	.042	0	.044	.053	.026	.035	.044
	(.012)	(.017)	(.013)	(.014)	(.016)	(.028)	(.013)	(.018)	(.018)	(.012)	(.017)	(.029)	(.012)	(.019)	(.017)
5	.016	.055	.041	.043	.017	.065	.045	.053	.055	.008	.058	.066	.038	.046	.057
	(.013)	(.019)	(.015)	(.015)	(.017)	(.034)	(.013)	(.022)	(.02)	(.012)	(.018)	(.035)	(.013)	(.022)	(.019)
6	.022	.069	.05	.051	.016	.076	.056	.063	.068	.016	.075	.08	.051	.058	.074
	(.013)	(.022)	(.018)	(.017)	(.016)	(.045)	(.016)	(.03)	(.023)	(.013)	(.021)	(.045)	(.016)	(.03)	(.022)
7	.015	.068	.047	.046	.008	.075	.055	.062	.068	.01	.076	.08	.05	.057	.075
	(.013)	(.027)	(.023)	(.021)	(.018)	(.056)	(.019)	(.037)	(.028)	(.013)	(.026)	(.056)	(.02)	(.037)	(.027)
Weight															
Yes															
No	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Fixed Effect															
FLE-Year	X	X	X	X	X	X	X	X	X						
Year										X	X	X	X	X	X
CRT															
None	X									X					
State-FLE		X	X	X	X						X				
State-FLE^2						X						X			
State							X						X		
State^2								X						X	
District									X						X
Factor															
0	X	X				X	X	X	X	X	X	X	X	X	X
1			X												
2				X											
3					X										

Benchmark model is highlighted. Point estimates and standard errors in parentheses are shown for various model specifications. Model choice is indicated in bottom of panel. We permute parameters using weights (yes/no), fixed effects (FLE-by-year/year), correlated random trends (none/state-by-FLE/state-by-FLE squared/state/state squared/district), and factors (0/1/2/3). Additional combinations available upon request.

Table E.6: Robustness: Percent Minority, Preferred Model

Treatment	Weighted					Unweighted				
Year	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.
1	-.003 +	.003	.003	-.005	.0000	-.003 *	.000	.000	-.002	-.00001
	(.001)	(.002)	(.002)	(.008)	(.00007)	(.001)	(.001)	(.001)	(.002)	(.00002)
2	-.007 *	.000	.003	.005	.00006	-.004 **	-.002 *	-.001	.000	-.00001
	(.003)	(.001)	(.003)	(.004)	(.00006)	(.001)	(.001)	(.002)	(.002)	(.00002)
3	-.011	-.003	.002	.002	.00003	-.005 *	-.003 +	-.002	.000	-.00002
	(.007)	(.004)	(.005)	(.003)	(.00007)	(.002)	(.002)	(.002)	(.002)	(.00003)
4	-.01	-.003	.005	.001	.00003	-.005 +	-.003	.000	.000	-.00001
	(.009)	(.006)	(.006)	(.004)	(.00008)	(.003)	(.002)	(.003)	(.003)	(.00003)
5	-.003	.005	.01 *	.001	.00009	-.004 +	-.001	.002	.002	.00002
	(.003)	(.004)	(.004)	(.004)	(.00007)	(.002)	(.002)	(.003)	(.003)	(.00003)
6	-.005	.005	.013 *	.001	.0001	-.005 *	-.002	.001	.001	.0000
	(.003)	(.004)	(.005)	(.005)	(.00008)	(.002)	(.002)	(.003)	(.004)	(.00004)
7	-.01 +	.004	.014 *	-.003	.00007	-.008 **	-.003	.000	.000	-.00002
	(.006)	(.005)	(.006)	(.006)	(.0001)	(.003)	(.002)	(.004)	(.005)	(.00004)
R^2	.956	.956	.956	.956	.956	.971	.971	.971	.971	.971

Notes: This table shows point estimates and standard errors for non-parametric heterogeneous differences-in-differences estimator. Model accounts for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq} (or $delta_t$ for continuous models), state-by-FLE linear time trends ($\psi_{sq}t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). All standard errors are clustered at the state level. (Significance indicated $+ < .10, * < .05, ** < .01, *** < .001$)

Table E.7: Robustness: Percent Child Poverty (SAIPE), Preferred Model

Treatment	Weighted					Unweighted				
Year	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.
1	.005	.005	.006	.012	.00015	.000	.002	.007 +	.002	.00005
	(.004)	(.007)	(.006)	(.009)	(.00011)	(.004)	(.004)	(.004)	(.009)	(.00008)
2	-.002	-.006	-.010	-.005	-.00008	-.005	-.005	-.004	-.012	-.00012
	(.005)	(.005)	(.011)	(.016)	(.00016)	(.005)	(.004)	(.006)	(.012)	(.00012)
3	-.002	-.004	.000	-.001	.0000	-.004	-.003	-.002	-.011	-.0001
	(.004)	(.004)	(.006)	(.008)	(.00008)	(.006)	(.005)	(.004)	(.010)	(.00009)
4	-.002	-.008 +	-.003	.010	.00006	-.005	-.004	.000	-.006	-.00006
	(.005)	(.004)	(.009)	(.011)	(.00011)	(.006)	(.006)	(.005)	(.011)	(.0001)
5	-.004	-.006	-.004	-.005	-.00005	-.005	-.004	-.001	-.010	-.00009
	(.005)	(.007)	(.008)	(.008)	(.0001)	(.007)	(.006)	(.005)	(.011)	(.00009)
6	-.003	-.009	-.008	.004	.0000	-.006	-.007	-.001	-.008	-.00008
	(.01)	(.009)	(.01)	(.012)	(.00013)	(.009)	(.008)	(.007)	(.015)	(.00013)
7	-.008	-.017	-.013	-.005	-.00009	-.009	-.009	-.001	-.011	-.00011
	(.01)	(.011)	(.013)	(.018)	(.00019)	(.012)	(.012)	(.009)	(.017)	(.00017)
R^2	.949	.949	.949	.949	.95	.911	.911	.911	.911	.911

Notes: This table shows point estimates and standard errors for non-parametric heterogeneous differences-in-differences estimator. Model accounts for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq} (or $delta_t$ for continuous models), state-by-FLE linear time trends ($\psi_{sq}t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). All standard errors are clustered at the state level. (Significance indicated $+ < .10, * < .05, ** < .01, *** < .001$)

Table E.8: Robustness: Special Education, Preferred Model

Treatment	Weighted					Unweighted				
Year	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.
1	.019	.013	.016	.022	.00028	.018	.014	.016	.017	.00023
	(.012)	(.013)	(.015)	(.027)	(.00028)	(.015)	(.016)	(.017)	(.019)	(.00025)
2	.021	.017	.019	.028	.00035	.02	.019	.02	.021	.00029
	(.016)	(.016)	(.018)	(.03)	(.00032)	(.019)	(.018)	(.019)	(.022)	(.00028)
3	.018	.009	.012	.019	.00024	.015	.011	.012	.013	.00017
	(.019)	(.019)	(.022)	(.035)	(.00037)	(.022)	(.023)	(.024)	(.026)	(.00033)
4	.029 +	.024	.031 +	.038	.00048	.027	.025	.027	.024	.00035
	(.015)	(.015)	(.018)	(.031)	(.0003)	(.018)	(.019)	(.021)	(.023)	(.00028)
5	.038 *	.034 *	.036 +	.045	.00058 +	.033	.03	.032	.032	.00044
	(.016)	(.016)	(.02)	(.034)	(.00032)	(.02)	(.02)	(.023)	(.026)	(.00031)
6	.042 *	.04 *	.039	.049	.00065 +	.038	.035	.035	.034	.00048
	(.019)	(.019)	(.024)	(.04)	(.00038)	(.024)	(.024)	(.028)	(.031)	(.00037)
7	.043 +	.04 +	.044	.049	.00066	.038	.038	.04	.036	.00051
	(.022)	(.022)	(.028)	(.045)	(.00043)	(.027)	(.027)	(.031)	(.034)	(.00041)
R^2	.432	.432	.432	.432	.431	.564	.564	.564	.564	.563

Notes: This table shows point estimates and standard errors for non-parametric heterogeneous differences-in-differences estimator. Model accounts for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq} (or $delta_t$ for continuous models), state-by-FLE linear time trends ($\psi_{sq}t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). All standard errors are clustered at the state level. (Significance indicated $+ < .10, * < .05, ** < .01, *** < .001$)

Table E.9: Robustness: Total Per Pupil Revenues, Preferred Model

Treatment	Weighted					Unweighted				
Year	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont.
1	456 *	204	360	294 *	4.71	617 *	480 *	464 *	381 *	6.31 *
	(233)	(386)	(255)	(161)	(2.85)	(328)	(262)	(234)	(179)	(3.01)
2	640 *	407	394	549 *	7.46 *	692 *	638 *	462 *	478 *	7.19 *
	(263)	(346)	(278)	(221)	(3.34)	(349)	(283)	(214)	(194)	(2.95)
3	578 *	265	390	561 *	7.14 +	754 *	718 *	598 *	765 **	9.97 **
	(351)	(514)	(347)	(231)	(4.04)	(460)	(368)	(267)	(219)	(3.49)
4	785 *	578	762 *	839 **	11.67 **	1085 *	1048 *	996 **	1095 ***	15.07 ***
	(427)	(585)	(396)	(252)	(4.51)	(541)	(434)	(303)	(226)	(3.92)
5	1060 *	798	815 *	867 **	12.88 *	1123 *	1155 *	997 **	985 **	14.58 **
	(461)	(558)	(458)	(346)	(5.32)	(662)	(511)	(398)	(323)	(5.23)
6	1429 *	1355 *	1079 *	1230 **	18.38 **	1457 *	1453 *	1293 *	1190 *	18.25 *
	(600)	(611)	(585)	(462)	(6.93)	(843)	(647)	(529)	(517)	(7.53)
7	1460 *	1347 *	1186 *	1530 ***	21.08 **	1607 *	1625 *	1391 **	1370 **	20.37 **
	(689)	(730)	(600)	(412)	(6.94)	(873)	(658)	(523)	(473)	(7.18)
R^2	0.901	0.901	0.901	0.901	.901	0.865	0.865	0.865	0.865	.864

Notes: This table shows point estimates and standard errors for non-parametric heterogeneous differences-in-differences estimator. Model accounts for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq} (or $delta_t$ for continuous models), state-by-FLE linear time trends ($\psi_{sq}t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). All standard errors are clustered at the state level. (Significance indicated $+ < .10, * < .05, ** < .01, *** < .001$)