THE SENSITIVITY OF CAUSAL ESTIMATES FROM COURT-ORDERED FINANCE REFORM ON SPENDING AND GRADUATION RATES*

Christopher A. Candelaria and Kenneth A. Shores

Stanford University

November 20, 2015

Abstract

We provide new evidence about the effect of court-ordered finance reform on per-pupil revenues and graduation rates. We account for cross-sectional dependence and 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 4 to 12 percent increase in per-pupil spending and a 5 to 8 percentage point increase in graduation rates. We subject the model to various sensitivity tests. In most cases, point estimates for graduation rates are within 2 percentage points of our preferred model. *JEL* codes: C23, I26

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

I. Introduction

Whether school spending has an effect on student outcomes has been an open question in the economics literature, dating back to at least Coleman et al. (1966). The causal relationship between spending and desirable outcomes is of obvious interest, as the share of GDP that the United States spends on public elementary and secondary education has remained fairly large throughout the past five decades, ranging from 3.7 to 4.5 percent.[1] Given the large share of spending on education, it would be useful to know if these resources are well spent. Despite this interest, we lack stylized facts about the effects of spending on changes in student academic and adult outcomes. The goal of this paper is to provide a robust description of the causal relationship between fiscal shocks and student outcomes at the district level for US states undergoing financial reform for the period 1990-91 to 2009-10.

The opportunity for more robust descriptions of causal relationships with panel data emerges from two sources. The first is that data collection efforts have extended the time dimension of panel data, allowing for more sophisticated tests of the identifying assumptions of quasi-experimental methods such as differences-in-differences estimators. Previous research efforts on the effects of school spending were hampered by data limitations such as this.[2] The second is that recent econometric methods have been developed to better model unobserved treatment heterogeneity, counterfactual trends, pre-treatment trends (correlated random trends), and cross-sectional dependence (CSD). If any of these unobserved components are correlated with regressors, then the econometric model will be biased. Fortunately, it is possible to test for a wide variety of model specifications to determine the sensitivity of causal estimates to modeling choice.

Using district-level aggregate data from the Common Core (CCD), we estimate the effects of

---

[1]These estimates come from the Digest of Education Statistics, 2013 edition. As of the writing of this draft, the 2013 version is the most recent publication available.

[2]See for example, Hoxby (2001) and Card & Payne (2002).

fiscal shocks, where fiscal shocks are defined as a state's first Supreme Court ruling that overturns a given state's finance system, on the natural logarithm of per-pupil spending and graduation rates. Our estimation approach is designed to handle two aspects of the identification problem: first, treatment occurs at the state level and, second, there is treatment effect heterogeneity within states. At the state level, we are interested in identifying a plausibly exogenous shock to the state's finance system. Here, we wish to control for the presence of cross-sectional dependence (CSD), which can arise if there is cross-sectional heterogeneity in response to common shocks, spatial dependence or interdependence. We control for CSD by including interactive fixed effects, as suggested by Bai (2009). These interactive fixed effects are estimated at the state level and control for unobserved correlations between states in the panel.

It is known that unmodeled treatment heterogeneity can lead to bias if (a) the probability of treatment assignment varies with a variable X and (b) the treatment effect varies with a variable X (Angrist & Imbens, 1995; Angrist, 2004; Elwert & Winship, 2010). Here, we account for treatment heterogeneity by constructing a time-invariant poverty quartile indicator variable equal to 1 if a state's district is in one of four poverty quartiles for year 1989. Each of these poverty quartile variables is interacted with a treatment year variable, for a total of 76 (19 x 4) treatment interactions. In order to provide a counterfactual for each of these poverty quartiles, we estimate a poverty quartile-by-year secular trend indexed by poverty quartile-by-year fixed effects. Estimating treatment heterogeneity in this way can provide an unbiased estimate of the treatment effect if we have identified the correct heterogeneous variable.

A second concern is that the variable X (in this case a state's poverty quartile) is not ignorably assigned with respect to treatment. That is, a state's Court order may be plausibly exogenous with respect to the state but not with respect to sub-units within the state. To account for this, we estimate state-by-poverty quartile linear time trends (referred to as correlated random trends).

2

These provide pre-treatment balance with respect to state-poverty quartile trends in the dependent variable.

All together, we estimate a heterogeneous differences-in-differences model that accounts for (a) cross-sectional dependence at the state level, (b) a poverty quartile-by-year secular trend, and (c) state-by-poverty quartile linear time trends. In this preferred specification, we find that high poverty districts in states that had their finance regimes overthrown by Court order experienced an increase in log spending by 4 to 12 percent and graduation rates by 5 to 8 percentage points seven years following reform.

We then subject the model to various sensitivity tests by permuting the interactive fixed effects, secular time trends, and correlated random trends. In total we estimate 15 complementary models. Generally, the results are robust to model fit: relative to the preferred model, interactive fixed effects and the specification of the secular time trend have modest effects on point estimates. The model is quite sensitive to the presence of correlated random trends, however. When state-by-poverty quartile time trends are ignored or estimated at a higher level of aggregation (the state), the effects of reform on graduation rates are zero and precisely estimated. When we estimate the linear time trend using a lower level of aggregation (the district), point estimates are similar to those of the preferred model. We conclude that the timing of treatment is not exogenous with respect to sub-units within the state, but conditionally exogenous point estimates are available if we are willing to assume that the sub-unit pre-treatment trends can be approximated with a functional form.

To test the extent to which results are redistributive, we estimate slightly different models, allowing the effect of reform to be continuous across the poverty quantiles. That is, we interact treatment year variables with a continuous poverty quantile variable, while controlling for secular

changes in redistribution in untreated states. This provides an estimate of the marginal change in graduation rate for a one-unit increase in poverty percentile rank within a state. Here we see that the effect of reform was redistributive: for a 10 percentile increase in poverty within a treated state, per-pupil log revenues increased by 0.9 to 1.8 percent and graduation rates increased by 0.5 to 0.85 percentage points in year seven.

Because we have aggregate data, one threat to identification would occur if treatment induced demographic change and demographic variables correlate with outcomes. For instance, if state spending increased school quality but kept property taxes down, high income parents (with children who are presumably more likely to graduate) to schools housed in historically high poverty districts. To test for this possibility, we estimate our "redistributive" models substituting the original outcome variables for district level demographic variables that could indicate propensity to graduate: percent poor, percent minority, and percent special education. We find no evidence that the minority composition of high poverty districts changed after reform, but we do find that these districts experienced an *increase* in poverty and percent of students qualifying for special education. We would have to assume that increases in poverty and special education rates positively effect graduation rates for our results to be biased in a meaningful way.

Jackson, Johnson, & Persico (2015) find that a $1000 increase in spending resulting from court-ordered reform increased graduation rates by 19 percentage points. These results are not directly comparable to ours, as our data do not permit us to leverage variation resulting from differential exposure to spending. When we use total per pupil revenues as our outcome variable, we find that the highest poverty decile within a state would have increased spending by about $990 to $1,800, resulting in an increase in graduation by 5 to 8 percentage points. These results, albeit smaller in magnitude, are largely in keeping with Jackson and colleagues.

This paper makes substantive and methodological contributions. Substantively, we find that court cases overturning a state's financial system for the period 1991-2010 had an effect on revenues and graduation rates, that these results are robust to a wide variety of modeling choices, and that this effect was redistributive. Methodologically, we demonstrate the importance of transparency in research practice, especially as it applies to panel estimators. Modeling strategies for panel data sets encompass a wide selection of reasonable modeling choices, including specification of the secular trend, correlated random trends, and cross-sectional dependence. Unless there are strong priors for favoring one model over another, researchers can effectively and efficiently demonstrate the sensitivity of point estimates to modeling choice.

II. Background

State-level court-ordered finance reform beginning in 1989 has come to be known as the "Adequacy Era." These court rulings are often treated as fiscal shocks to state school funding systems. A number of papers have attempted to link these plausibly exogenous changes in spending to changes in other desired outcomes, like achievement, graduation and earnings (Card & Payne, 2002; Hoxby, 2001; Jackson et al., 2015). The results of Card and 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.

Most recently, Jackson et al. (2015) have constructed a much longer panel (with respect to time), with 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. 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. For this study, the sample of students are taken from the Panel Study of Income Dynamics (PSID), which is representative at the national level. One concern with these data is that identification is leveraged from variation between states over time, and results are disaggregated using within state variation around income. If individuals are sampled randomly, estimates may be imprecise but unbiased; if, however, individuals are sampled in a way that correlates with the timing of treatment, there may be bias. It is not obvious how to test whether this bias exists, but it does cast a shadow of doubt over their results.

The purpose of this paper is two-fold. First, it is important to show that results by Jackson et al. (2015) can be replicated across other data sets. Here, we use data from the Common Core (CCD), which provides aggregate spending and graduation rates for the universe of school districts in the United States. Both the PSID and CCD have a kind of Anna Karenina problem, in which each data set is unsatisfactory in its own way. The PSID follows individual students over time but has unobserved sampling issues that may correlate with treatment; the CCD contains the universe of districts but does not follow students over time and may not reveal sorting within districts. If results are qualitatively similar across different data, we can feel more confident that estimates do not reveal spurious correlations resulting from the sample generating process. Second, it is important to show that results are insensitive to similarly compelling modeling choices. Jackson and colleagues use graduation data from the Common Core of Data (CCD), as we do here, to corroborate results from the PSID sampleJackson et al. (2015). While they find a pattern of results from the CCD that is largely consistent with those results from the PSID, they do not test to see whether those results are sensitive to model specification. If we are to believe that results from the CCD sample largely corroborate results from the PSID, we must show that the identifying assumptions using the CCD sample are met. Our purpose is to 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 parameters.

III. Data

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.[3]

Our sample begins in the 1990-91 school year and ends in 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 4 is the top poverty quartile for the state. Total revenues are the sum of federal, local, and state revenues in each district. We divide this value

---

[3]Web links to each of the data sources are listed in the appendix.

by the total number of students in the district and deflate by the US CPI, All Urban Consumers Index to convert the figure to real terms. We then take the natural logarithm of this variable. Our graduation rates variable is 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) show is not susceptible to the downward bias caused by using lagged 9th grade enrollment in the denominator. We top-code graduation rates, 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. Child poverty is a variable we take from the Small Area Income and Poverty Estimates (SAIPE).

To define our analytic sample, we place some restrictions on the data, and we address an issue with New York City Public Schools (NYCPS). First, we drop Hawaii and the District of Columbia from our sample, as each place has only one school district. We also dropped Montana from our analysis because they were missing a substantial amount of graduation rate data. We keep only unified districts to exclude non-traditional districts and to remove charter-only districts. We define unified districts 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, we needed to combine the non-fiscal data into a single district. As suggested in the NCES documentation, we use NYCPS's supervisory union number to aggregate the geographical districts into a single entity.

We noticed 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

---

[4]In Appendix A, we describe where the data was gathered and cleaned, including URL information for where data can be found.

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.

To our 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 we updated the list using data from the National Education Access Network.[5] Table I lists the court cases we are considering. As shown, there are a total of twelve states that had their school finance systems overturned during the Adequacy Era. 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 I Here]

Table II provides summary statistics of the key variables in our interpolated data set, excluding New York City Public Schools (NYCPS).[6] We have a total of 188,752 district-year observations. The total number of unified districts in our sample 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). When we do not weight our data by district enrollment, we obtain similar figures, but

---

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

[6]We drop NYCPS because it is an outlier district in our data set. We provide a detailed explanation of why we do this in our results section.

they are slightly larger.

[Insert Table II Here]

IV. Econometric specifications and model sensitivity

In this section, we describe our empirical strategy to estimate the causal effects of school finance reform at the state level on real log revenues per student and graduation rates at the district level. We begin by positing our benchmark model, which is a differences-in-differences equation that models treatment heterogeneity across FLE poverty quartiles. We then explain what each of the parameter choices are designed to control for and why they are selected. Because treatment occurs at the state level and our outcomes are at the district level, there are several ways 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 unobservable factors such as cross-sectional dependence. We outline these alternative modeling choices and discuss their implications relative to the benchmark model.

IV.A. Benchmark differences-in-differences model

To identify the causal effects of finance reform, we leverage the plausibly exogenous variation resulting from state Supreme Court rulings overturning a given state's fiscal regime. Prior education finance studies have also relied on the exogenous nature of court rulings to estimate causal effects on fiscal and academic outcomes (see, for example, Jackson et al., 2015; Sims, 2011). After a lawsuit is filed against a state's education funding system, we assume that the timing of the Court's ruling is unpredictable. Under this assumption, the decision to overturn a funding system defines treatment and the date of the decision constitutes a random shock.

With panel data, the exogenous timing of court decisions generates a natural experiment that can be modeled using a differences-in-differences framework. States were subject to reform in different years and not all states had reform, which provides treatment and control groups over time. Our benchmark differences-in-differences model takes the following form:

$$(1) \quad Y_{sqdt} = \theta_d + \delta_{tq} + \psi_{sq}t + P'_{st}\beta_q + \lambda'_s F_t + \varepsilon_{sdt},$$

where Y_{sqdt} is an outcome of interest—real log revenues per student or graduation rates—in state 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, $\psi_{sq}t$ is a state by quartile-specific linear time trend; P_{it} is a policy variable that takes value 1 in the year state s has its first reform and remains value 1 for all subsequent years following reform; $\lambda'_s F_t$ is a factor structure that accounts for cross-sectional dependence at the state level; and ε_{sdt} is assumed to be a mean zero, random error term. 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).

Our parameters of interest are the β_q, which are the causal estimates of school finance reform in quartile q on Y_{sqdt}. We define q such that $q \in \{1, 2, 3, 4\}$, and the highest level of poverty is represented by quartile 4.[7] Throughout our paper, we parameterize the policy variable P'_{it} such that each poverty quartile's effect is estimated, so we do not have an omitted quartile. Moreover, we allow treatment effects to have a dynamic treatment response pattern (Wolfers, 2006). Each β_q, therefore, is a vector of average effect estimates in the years after reform in quartile q.

Overall, we estimate 19 treatment effects for each poverty quartile's vector of effects. Although there are 21 effects we can potentially estimate, we combine treatment effect years 19 though 21

---

[7]Using FLE data from 1989-90, we divide states into FLE quartiles, where quartile 4 is the top poverty quartile for the state.

into a single binary indicator, as there are only two treatment states that contribute to causal identification in these later years.[8] In reporting our results, we only report effect estimates for years 1 through 7 after reform. We do this because estimating treatment effects several years after treatment occurs results in precision loss and because because very few states were treated early enough to contribute information to treatment effect estimates in later years (see Table I). All together, this model absorbs approximately 9,800 district fixed effects, 76 year effects, 192 continuous fixed effects (state-by-FLE quartile interacted with linear time), as well as the factor variables $\lambda'_s F_t$. We estimate standard errors clustered at the state level by including the estimated factors and factor loadings as the covariates $\theta_s F_t + \delta_t \lambda_s$.[9] These non-treatment parameters are eliminated using high dimensional fixed effects according to the Frisch-Waugh-Lovell theorem (Frisch & Waugh, 1933; Lovell, 1963).[10]

IV.B.   Explaining model specifications

We now wish to articulate what the parameters from Equation (1) are controlling for and why they are included in the model. Researchers are presented with a number of modeling choices, and our goal here is to be transparent about what reasons we have for parameterizing the model the way we do. This also opens the possibility for subjecting the model to sensitivity analysis, in order to determine upper and lower bounds of point estimates, in relation to our preferred model. In particular, we examine choices related to cross-sectional dependence, secular trends, and correlated random trends. We also consider the difference between OLS regression and weighted least squares (WLS) regression, where the weights are measures of time varying district enrollment. As will be

---

[8]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.

[9]Appendix D shows how an estimated factor structure can be included in an OLS regression framework to obtain a variety of standard error structures.

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

shown in the results section, certain parameterizations of the differences-in-differences model have substantial impacts on point estimates relative to our benchmark model.

IV.B.1.   Cross-sectional dependence

In our benchmark model, the terms $\lambda'_s F_t$ specify that the error term has a factor structure that affects Y_{sqdt} and may be correlated with P'_{st}, the treatment indicators. 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; in our model, we set $r = 1$. 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 policy changes), and the factor loadings λ_s capture how states are differentially affected by these shocks. Of particular concern is the presence of interdependence, which can result if one state's Supreme Court ruling affects the chances of another state's ruling. This, of course, violates the identifying assumptions of the differences-in-differences model and results in bias, unless that interdependence is accounted for (Pesaran & Pick, 2007; Bai, 2009).

To estimate the $\lambda'_s F_t$ factor structure in Equation (1), we use the method of principal components as described by Bai (2009) and implemented by Moon & Weidner (2015); Gomez (2015b). The procedure begins by choosing starting values for the β_q vector, which we denote as $\tilde{\beta}_q$. 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.

Traditional approaches to factor methods specify factor loadings at the lowest unit of analysis, in this case 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 (Wright & Holt, 1985; Raiko, Ilin, & Karhunen, 2008; Ilin & Raiko, 2010).

IV.B.2. Secular time trends

Secular time trends, often specified non-parametrically using binary indicator variables, adjust for unobservable factors affecting outcome variables over time. The usual assumption is that these factors affect all units—in our case, districts—in the same way in a given year. Examples of unobservable factors include national policy changes and macroeconomic events such as recessions.

In our differences-in-differences model, having the correct specification of the secular trend is important because the treatment effect parameters are estimated relative to the trend. By OLS algebra, each year of the trend is estimated using variation from districts in non-treated states and

districts in treatment states awaiting reform; therefore, we must believe it represents the average counterfactual trend that treated districts would have had in the absence of treatment. If it is not modeled properly, the common trends assumption of the differences-in-differences model is violated and results would be biased.

In our benchmark specification, we control for secular time trends by including year by FLE quartile-by-year fixed effects, denoted as δ_{tq}. We include these δ_{tq} fixed effects, instead of standard year fixed effects, to establish a more plausible counterfactual trend for treated districts. We assume that there are unobserved characteristics common to districts of each FLE quartile that vary over time. In terms of the differences-in-differences framework, we are comparing treated districts in a given quartile before and after treatment relative to control districts before and after treatment in the *same* quartile.

IV.B.3. Correlated random trends

If the timing of a state's court ruling decision is correlated with unobserved trends at the state, district, or other group level (e.g., state by FLE quartile), we must control for these trends to obtain unbiased estimates of causal parameters. In the literature, these trends are formally called correlated random trends (Wooldridge, 2005, 2010), but they are often informally referred to as pre-treatment trends. Correlated random trends serve a distinct purpose from secular, non-parametric trends. While the secular trends help to establish a plausible counterfactual trend for the common trends assumption to hold, correlated random trends guard against omitted variable bias caused by an endogenous policy shock.

In Equation (1), the parameter $\psi_{sq}t$ is included because we believe the timing of the state ruling is correlated with pre-treatment trends within the state, approximated by a state-by-quartile-specific slope. The inclusion of this parameter aligns with the notion that the date on which a

court-ordered finance system is deemed unconstitutional is correlated with the slope of the FLE quartile trend within the state. For example, if graduation rates are steeply declining among the most impoverished districts within state s, we might expect a reform decision sooner than if the graduation rate had a mild, decreasing trend. Evidence for variation in pre-treatment trends within states can be seen in In Figure I. This Figure shows weighted mean log spending and graduation rates for states that experienced a Court ruling over time, where time is centered around the year of Court ruling. With respect to graduation rates, there is an obvious downward trend prior to a Court's ruling.

[Insert Figure I Here]

IV.B.4. Treatment heterogeneity

In Equation (1), we explicitly model treatment heterogeneity by disaggregating treatment effects into poverty quartiles, q (Meyer, 1995). These quartiles are derived from the percentages of Free Lunch Eligible (FLE) students reported at the district level in each state in 1989. We fix the year at the start of our sample because the poverty distribution could be affected by treatment over time. The quartiles are defined *within* each state.

While we capture treatment heterogeneity across poverty quartiles, we may fail to capture other sources of treatment heterogeneity. Weighting a regression model by group size is traditionally used to correct for heteroskedasticity, but it also provides a way to check whether unobserved heterogeneity is properly modeled. According to asymptotic theory, the probability limits of OLS and weighted least squares should be consistent. Regardless of how you weight the data, the point estimates between the two models should not dramatically differ. However, when OLS and WLS estimates diverge substantially, there is concern that the model is not correctly specified, and it may be due in part to unobserved heterogeneity (Solon, Haider, & Wooldridge, 2015). In our analyses,

16

the heterogeneity is associated with district size.

For our differences-in-differences specification, we examine the sensitivity of our point estimates to the inclusion of district-level, time-varying enrollment weights. We provide a detailed discussion about discrepancies between weighted and unweighted point estimates in the next section.

IV.C.   Alternative model specifications

Here we quickly outline alternative model specifications. In the Results section, we explore the sensitivity of our preferred model to alternative parameterizations. In general, we attempt to include alternative models that are typical in the panel methods and econometric literature. In total, we estimate 15 alternative models that broadly fall within the bounds of typical modeling choices. In the Results section, we provide upper and lower bounds for how much point estimates depart from our preferred model. Here, we outline the alternatives we estimate:

1. *Cross-sectional dependence*: We estimate models in which we assume $\lambda_s' F_t = 0$, as well as models in which the number of included factors r is $\in 1, 2, 3$.[11]

2. *Secular time trend*: We estimate models in which we set $\delta_{tq} = \delta_t$, thereby modeling the counterfactual time trend as constant across cross-sectional units.

3. *Correlated random trend*: We estimate models in which we set $\psi_{sq}t = \in \{\psi_s t, \psi_d t, \psi_{sq}t^2, \psi_s t^2\}$. That is, we allow the pre-treatment trend to be estimated at the state and district levels, as well as allowing the time trend to have a quadratic functional form.

---

[11]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).

V. Results

We present our results in four parts. We first show and discuss the causal effect estimates of court-ordered finance reforms using our benchmark differences-in-differences model. Second, we examine the extent to which our benchmark model point estimates are sensitive to assumptions about secular trends, correlated random effects, and cross-sectional dependence. Third, we assess whether reforms were redistributive 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. Finally, we conclude with a series of robustness checks that allow us to gauge the validity of our causal estimates.

V.A. Benchmark differences-in-differences model results

V.A.1. Revenues

We report our causal effect estimates of court-ordered finance reform on the logarithm of per pupil revenues and graduation rates in Tables IV and V, respectively. We obtain results by estimating our benchmark differences-in-differences model in Equation (1). FLE quartile 1 represents low-poverty districts, and FLE quartile 4 represents high-poverty districts. We only report treatment effect years 1 though 7 because the number of states in the treated group changes dramatically over time. Some states were treated very late (or very early), and we do not have enough years of data to follow them past 2010. As shown in Table III, 6 years after treatment, Alaska and Kansas no longer contribute treatment information; in years 7 and 8, we lose North Carolina and New York.[12] We display both weighted and unweighted estimates, where the weight is time-varying district enrollment.

---

[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.

[Insert Table III Here]

Examining the weighted results in Table IV, we find 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 our models include FLE quartile by year fixed effects, we cannot compare results across the quartiles; rather, point estimates for a given quartile should be interpreted as 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 12.7 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. This point estimate is significant at the 1 percent level. In FLE quartile 4, we find that the revenues were 11.9 percent higher relative to what they would have been in the absence of reform; this point estimate is significant at the 5 percent level.

Compared to weighted results, the unweighted results in Table IV suggest that revenues increased, but many of the point estimates are not significant. In FLE quartile 4, for example, all point estimates are positive, but none are significant. The magnitude of point estimates is also substantially smaller than those from the weighted regression. In year 7, point estimates across the quartiles are at least half as a small as the corresponding point estimates from the weighted model. When weighted and unweighted regression estimates diverge, there is evidence of unmodeled heterogeneity, which we will discuss momentarily. Overall, revenues increased across all FLE poverty quartiles in states with Court order, relative to equivalent poverty quartiles in non-treated states.

[Insert Table IV Here]

V.A.2. Graduation rates

With respect to graduation rates, the weighted results in Table V show that court-ordered finance reforms were consistently positive and significant among districts in FLE quartile 4. In the first

19

year after reform, graduation rates in quartile 4 increased modestly by 1.3 percentage points. By treatment year 7, however, graduation rates increased by 8.4 percentage points, which is significant at the 0.1 percent level. It is worth emphasizing that 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. We find 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 no significant effects, and the point estimates show no evidence of an upward trend over time.

The unweighted graduation results in Table V tell a similar story as the weighted results; one key difference is that the point estimates tend to be smaller. In FLE quartile 4, for example, graduation rates are 4.7 percentage points higher in year 7 that they would have been in the absence of treatment. This point estimate is almost half the size of the corresponding estimate when using weights. Although districts in the lowest-poverty quartile exhibit some marginally significant effects, these effects are small, and do not suggest a substantial increase from their levels before reform, which corresponds to the weighted regression results.

[Insert Table V Here]

V.A.3.   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

20

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.[13]

In Figure II, we plot the causal effect estimates for treatment years 1 to 7 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 whether NYCPS is included, as all districts are weighted equally. In the figure, the point estimates are identical and are represented by the dashed lines. 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.

[Insert Figure II Here]

Examining NYCPS provides just one example of how treatment heterogeneity might be related to district size. As shown in Appendix Figure B.1, 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. For all analyses, we drop NYCPS from our sample because its district enrollment weight is near 1 million throughout our sample period, which is an outlier in the distribution of district enrollment. Dropping the next set of largest districts does not have such a dramatic effect on the results as dropping NYCPS does. For this reason, we retain all other districts.

In line with Solon et al. (2015), we acknowledge that we are not necessarily estimating popula-tion average *partial* effects when using weighted least squares; instead, our OLS and WLS results

---

[13]To be clear, NYCPS was removed from our analytic sample before estimating the benchmark regressions above.

are identifying 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, our weighted and unweighted 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. As the unweighted results tend to produce smaller effect sizes than the the weighted results, the unweighted results may be considered lower bound estimates of the (heterogeneous) treatment effect, and the weighted results may be considered as upper bound estimates.

V.B. Model sensitivity

Our benchmark 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 4). We place emphasis on these districts because the effect of the policy is concentrated in high poverty district. The evidence suggests there was an effect for graduation rates and revenues for the FLE quartile 4 districts, but these results assume that our benchmark model makes correct assumptions about secular trends, correlated random trends, and cross-sectional dependence.[14]. We now assess the extent to which results are sensitive to these modeling choices.

Figures III and IV graphically show the variability of causal effect estimates of finance reform on the logarithm of per pupil revenues and graduation rates, respectively, across a variety of model specifications. Each marker symbol represents the difference between two point estimates, one from an alternative model specification and the other from our benchmark model; we calculate this dif-

---

[14]Although the unweighted revenue results are not statistically significant, they point estimates show a consistent positive effect

ference for treatment years 1 through 7. In the figures, we normalize our benchmark effect estimates to zero in each year. If a point estimate is greater than zero, then our model understates the effect from the alternative model; if it is less than zero, our model overstates the effect. We report more traditional regression tables with point estimates and standard errors in the Appendix.[15] We also report models with high order correlated random trends in the Appendix tables.[16]

While we do not report all possible combinations of different secular trend, correlated random trend, and cross-sectional dependence models, the 11 models we do show provide insight into sensitivity of the causal effect estimates. In both figures, our benchmark differences-in-differences model corresponds to the third model in the legend, which is denoted by an "x" marker symbol and the following triple:

- ST: FLE by year; CRT: State by FLE; CSD: 0 (delimiter is the semicolon).

ST refers to the type of secular trend, which we model as either FLE by year fixed effects or year fixed effects. CRT refers to the type of correlated random trend in the model, which we specify as state-specific, state by FLE quartile-specific, or district-specific trends.[17] Each of these trends is formed by interacting the appropriate fixed effects with a function of time, whether linear or quadratic. Finally, CSD refers to the number of factors we include 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 factor, 2 factors, and 3 factors, respectively.

In Figure III, we find that, on average, our benchmark model understates causal effects on revenues in FLE quartile 4 relative to all other graduation rate models. The largest effects in both

---

[15]For log per pupil revenues, Appendix Tables C.1 and C.2 report weighted results and unweighted results, respectively. For graduation rates, Appendix Tables C.3 and C.4 report weighted and unweighted results, respectively.

[16]We exclude correlated random trend estimates that allow the time trend to have a quadratic function. Quadratic correlated trends, at the state and state-FLE quartile levels, are very noisy, with standard errors at times greater than twice the magnitude of point estimates. These terms are shown in the Appendix.

[17]Bracketed numbers indicate the column location for those model estimates available in Appendix Tables C.1, C.2, C.3, C.4.

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 state level, and no adjustment for cross-sectional dependence. The point estimates from this model are large, and they are precisely estimated. Although these results suggest there were large revenue effects, we worry that this specification ignores omitted variables, such as CSD, as well as mis-specifies the counterfactual trend, all of which could result in upward bias.

Indeed, when we limit the identifying variation within a FLE poverty quartile and year, point estimates shrink in magnitude. Correlated random trends also matter in terms of the level at which they are specified. Using only state specific trends inflates the effect size. While bias generated from cross-sectional dependence might also be an issue, the revenue results suggest that point estimates are not greatly influenced when we account for it. This is encouraging as interdependence between states that affected the timing of reform is a serious threat to identification in the differences-in-differences framework and is not generally tested.

Finally, recall that unweighted log revenues results were smaller in magnitude than weighted results and generally noisier. When we subject the model to alternative specifications (in particular, ignoring cross-sectional dependence), we see that there is a great deal of heterogeneity in the magnitudes of these point estimates. We believe that our model is correct, but it is worth emphasizing that it is a lower bound relative to other modeling choices.

[Insert Figure III Here]

We consider the variability of graduation rate estimates in Figure IV. Unlike the revenues results, we have cases where our benchmark model both over- and under-states effect estimates relative to other models.

Of particular interest is the influence of correlated random trends. In the absence of specifying a

24

correlated random trend, point estimates are between 2 to 6 percentage points smaller than point estimates from our benchmark model. By including a state-level time trend (indicated by the larger solid diamond and hollow square), point estimates are nearer to our preferred model, but are still lower by about 2 percent. Modeling the time trends was motivated by the fact that FLE 4 districts were trending differently prior to reform than FLE districts 1 through 3, and this is reinforced here. Overall, we find evidence of omitted variable bias when correlated random trends are excluded from the model. By including the trend, the assumption is that we now have a correctly specified model *conditional* on the trend.

When our benchmark model understates the causal effects of other models, we find that the primary difference is whether an adjustment for cross-sectional dependence has been made. Our benchmark model accounts for a cross-sectional dependence using a 1-factor model. The models with the largest point estimates, relative to our benchmark model, do not account for cross-sectional dependence. It appears that treatment might be correlated with macroeconomic shocks affecting graduation rates, or that treatment might be induced by another state's pattern of graduation rates. After correcting for cross-sectional dependence, point estimates are not as large. It is important to emphasize that we do not know the true number of factors r to include in the model. Unfortunately, due to finite sample bias, we cannot include as many factors as is necessary to make the errors i.i.d.

[Insert Figure IV Here]

Overall, we find that our preferred model tends to *understate* effect sizes for real log revenues and that causal effect estimates for graduation rates are variable depending on how we model secular trends, correlated random trends, and cross-sectional dependence. When we adjust for pre-treatment trends (especially pre-treatment trends for state-by-FLE quartiles) point estimates for graduation rates stabilize and the variation around our benchmark estimates is less than 2 percentage points.

Changing parameters in a model is not trivial because some of these changes affect the identification strategy while others affect assumptions about omitted variable bias. To the extent possible, researchers should aim for transparency in reporting causal estimates where different model parameterizations may be applied. One of the aims of this paper is that transparency can be accomplished with relative ease, as depicted in Figures III and IV.

V.C.   Redistributive effects

To test whether revenues and graduation rates increased more in high poverty districts following court order, 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 we estimate is analogous to Equation (1) with three changes:

1. We replace δ_{tq} to equal δ_t, so that the secular trend in Y_{sqdt} is modeled as the average across all districts.

2. We add a parameter $\delta_t Q$, which is a continuous fixed effect variable that controls for 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 1.

3. The treatment variable $P'_{st}\beta_q$ is set to equal $P'_{st}Q$.

Item [1] now adjusts for the average annual trend in revenues and graduation rates among untreated districts. Item [2] is done to provide a counterfactual secular trend with respect to

how much non-treated states are "redistributing" 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 average annual trend.[18]

The interpretation of the point estimates on the treatment year indicators, indicated in Item [3], is the marginal change in our 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.

We perform OLS and WLS, for Equation (1), with the modifications described just above. These results can be seen for both dependent variables in Table IV and V. 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 IV. Seven years after reform, a 10-unit increase in FLE percentile is associated with a 0.9 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 previously discussed in Section V.A.3., neither the weighted nor unweighted model results dominate each other, so we can view the slope coefficient as having a lower bound of 0.9 percent and an upper bound of 1.8 percent. Assuming the treatment is linear, these results suggest that districts in the 90th percentile would have had per pupil revenues that were between 7.2 to 14.2 percent higher than districts in the 10th percentile.

Table V also shows that court-ordered reform increased graduation rates across the FLE distri-

---

[18]Models that include the FLE poverty quartile-by-year fixed effect, not shown, are nearly identical.

bution. Seven years after reform, a 10-unit increase in FLE percentile is associated with a 0.85 percentage point increase in graduation rates for the weighted regression. For the unweighted model the corresponding point estimate is 0.50 percentage points. Assuming linearity, districts in the 90th percentile would have had graduation rates that were between 4.0 to 6.8 percentage points higher than districts in the 10th percentile.

V.D.   Robustness Checks

The largest 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. To test for selective migration, we estimate the continuous version of our benchmark model on four dependent variables: 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. 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. Ex hypothesi, we would assume that our results would be upwardly biased if treatment decreased the percents of students who are minority, poor, or receiving special education, as these populations of students have been historically less likely to graduate (Stetser & Stillwell, 2014). Of course, we cannot test for within demographic sorting, which would occur if students more likely to graduate within the poor, minority and disabled populations we observe move into high poverty districts as a result of reform. The inability to test for within composition sorting is a limitation of our data. Although we only report the continuous treatment effect estimates of reform in Table VI,

results from our main benchmark model appear in the Appendix.[19]

Table VI shows that there is no strong evidence of selective migration to treated districts. None of the point estimates for percent minority are statistically significant, nor are they large in magnitude. There are some cases in which we obtain significance in terms of children in proverty (SAIPE) and the percentages of special education students, but these point estimates are positive and not consistently significant across the treatment years. If anything, the evidence from these models suggests our point estimates on the effect of reform are downwardly biased, as the demographic changes indicate an increase, relative to the change across poverty quantile in non-treated states, in the population of students that have been historically less likely to graduate.

In addition to considering selective migration, we also examine the robustness of our revenues dependent variable. 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. In Table VI, we find that weighted results are marginally significant in treatment years 6 and 7. We also find that the unweighted estimates are all statistically significant at the 5 percent level. This is the exact same pattern of significance that appears Table IV, the table of main revenues results. Moreover, all point estimates are positive across both tables. Appendix Table C.9, which shows estimates for all FLE quartiles is qualitatively similar in terms of significance and sign to the results in Table IV as well. Overall, we feel confident that our logarithmic transformation does not jeopardize the validity of our revenues results.

[Insert Table VI Here]

---

[19]Please see Tables C.5, C.6, C.7, and C.8 in Appendix C.

VI.   Conclusion

In this paper, we make both substantive and methodological contributions. Substantively, we demonstrate that states undergoing Court ordered finance reform in the period 1990-2010 experienced a sizable fiscal shock that primarily affected high poverty districts in the state. This fiscal shock led to a subsequent change in graduation rates in those states that likewise primarily benefited high poverty districts. The estimation of these effects is largely immune to a variety of model specifications that vary in how they account for cross-sectional dependence, pre-treatment trends and a heterogeneous secular trend. These effect sizes are, in turn, robust to changes in demographic composition, as we observe population composition variables that are fairly stable after reform.

Methodologically, we subject the differences-in-differences estimator to a wide range of specification checks, including cross-sectional dependence, correlated random trends, and secular trends. We efficiently present upper and lower bounds on point estimates for a range of model choices. Overall, we find that ignoring cross-sectional dependence and assuming a homogenous counterfactual trend do not meaningfully bias results in this application. However, we do observe substantial bias if we do not model pre-treatment trends at the state-by-poverty quartile. Without accounting for pre-treatment trends, effects of reform on graduation rates are insignificant and precisely estimated.

The provocative results from Jackson et al. (2015) should not be undervalued. They find that spending shocks resulting from Court order had major effects on a variety of student outcomes, including adult earnings. The question of whether school spending—a public investment of $700 billion dollar per year—can be causally linked to desirable outcomes has been a foundational question in public economics for the past 50 years. We have argued here that it is necessary to replicate these results using other data sets with better attributes or that have non-overlapping problems.

Moreover, given the richness of modern panel data sets, researchers are presented with a variety of plausibly equivalent modeling choices. In the absence of strong priors regarding model specification, the challenge for applied microeconomists is to efficiently convey the upper and lower bounds of estimates resulting from model choice. Here, we have shown that the effects of Court order are consistent and robust on a data set that contains the universe of school districts in the United States. Moreover, while results are sensitive to model fit, in most cases, point estimates for graduation rates from the preferred model never diverge by more than 2 percentage points.

References

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

ANGRIST, J. D. (2004). "Treatment effect heterogeneity in theory and practice*." *The Economic Journal*, *114*(494), C52–C83.

ANGRIST, J. & IMBENS, G. (1995). "Identification and estimation of local average treatment effects."

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

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.

COLEMAN, J. S., CAMPBELL, E. Q., HOBSON, C. J., MCPARTLAND, J., MOOD, A. M., WEINFELD, F. D., & YORK, R. (1966). "Equality of educational opportunity." *Washington, dc*, 1066–5684.

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.

ELWERT, F. & WINSHIP, C. (2010). "Effect heterogeneity and bias in main-effects-only regression models." *Heuristics, probability and causality: A tribute to Judea Pearl*, 327–36.

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.*. sha: 368df32285d72db2220e3a7e02671ebdff54613e edition. https://github.com/matthieugomez/FixedEffectModels.jl.

——— (2015b). *SparseFactorModels: Julia package for unbalanced factor models and interactive fixed effects models.*. sha: 368df32285d72db2220e3a7e02671ebdff54613e edition. https://github.com/matthieugomez/SparseFactorModels.jl.

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.

HOXBY, C. M. (2001). "All School Finance Equalizations are Not Created Equal." *The Quarterly 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. (2015). "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms."Technical report, 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.

MEYER, B. D. (1995). "Natural and Quasi-Experiments in Economics." *Journal of Business & Economic Statistics*, *13*(2), 151–161.

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

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.

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.

STETSER, M. C. & STILLWELL, R. (2014). "Public High School Four-Year On-Time Graduation Rates and Event Dropout Rates: School Years 2010-11 and 2011-12. First Look. NCES 2014-391.." *National Center for Education Statistics*.

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.

Figures

Figure I: Mean Log Per Pupil Revenues & Graduation Rates, Centered around Timing of Reform
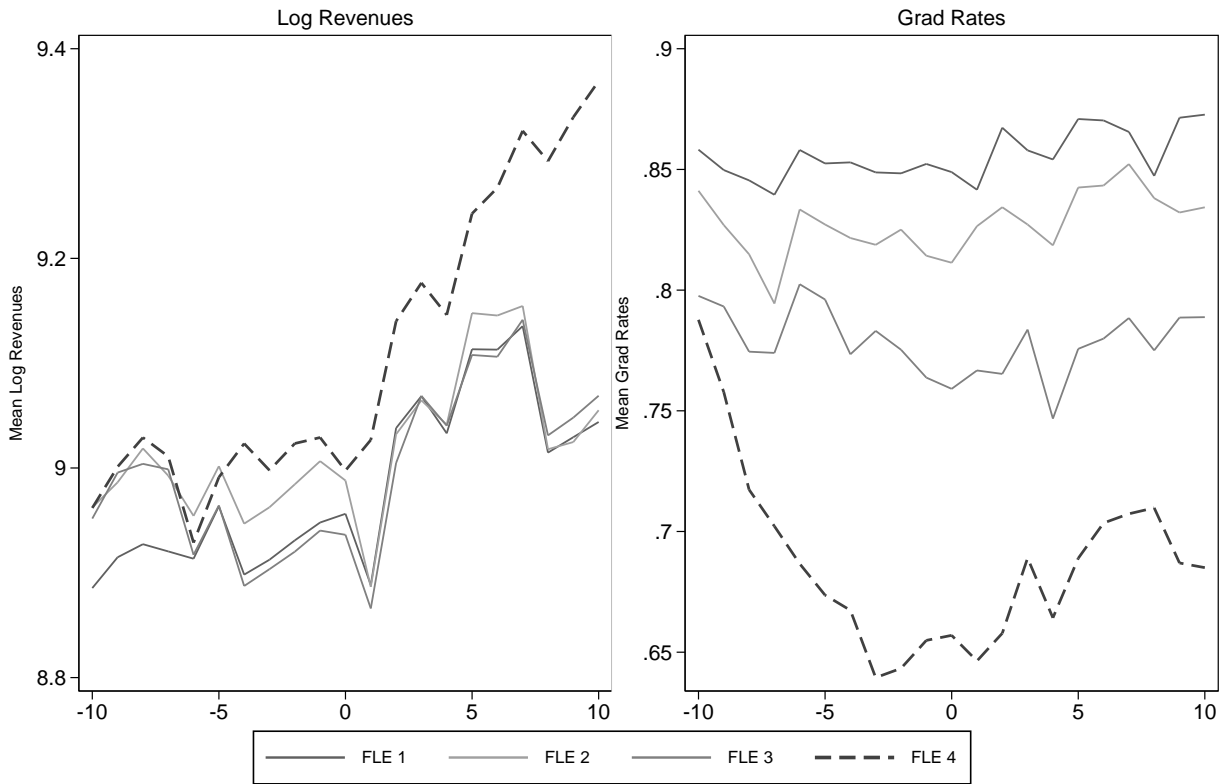
Population weighted mean log revenues and graduation rates for states undergoing court-ordered finance reform, by FLE quartile. Averages are for years 1990-2010, centered around first year of court-order. Note that common trends assumptions are not reflected in this figure, as FLE quartile 4 graduation rates are not estimated relative to FLE quartile 4 graduation rates in non-treated states. NYCPS excluded.

Figure II: Change in Log Per Pupil Revenues & Graduation Rates after Court Ruling, FLE Quartile 4, with and without NYCPS
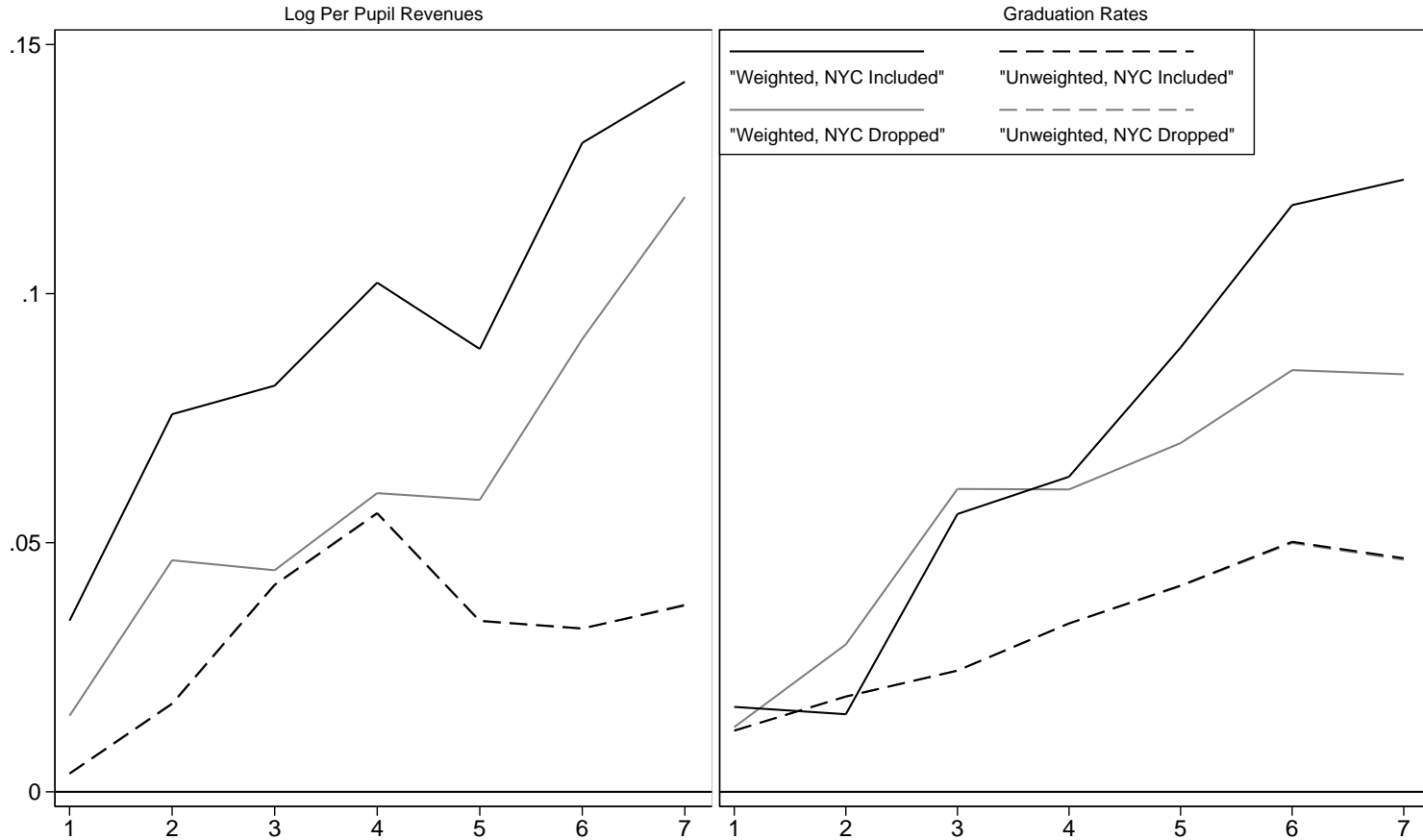
Notes: Differences-in-differences with treatment effects estimated non-parametrically after reform. Model accounts for district fixed effects (θ_d), FLE-by-year fixed effects (δ_{tq}), state-by-FLE linear time trends ($\psi_{sq}t$), and a state-level factor ($\lambda'_s F_t$). Left panel corresponds 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 III: Model Sensitivity: Changes in Estimates for Log Per Pupil Revenues across Models

Notes: Point estimate in treatment year t is shown as the difference between preferred model and model m, where model m variables are indicated in the legend. Point estimates along the x axis greater than zero indicate our preferred model *underestimates* the effect; greater than zero indicates our preferred model *overstates* the effect. Legend shows three parameter changes, delimited by ";". ST denotes the type of nonparametric *secular trend* under consideration: (a) FLE quartile by year fixed effects or (b) year fixed effects. CRT denotes the type of *correlated random trends* under consideration: (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 C.1 and C.2

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

Notes: Point estimate in treatment year t is shown as the difference between preferred model and model m, where model m variables are indicated in the legend. Point estimates along the x axis greater than zero indicate our preferred model *underestimates* the effect; greater than zero indicates our preferred model *overstates* the effect. Legend shows three parameter changes, delimited by ";". ST denotes the type of nonparametric *secular trend* under consideration: (a) FLE quartile by year fixed effects or (b) year fixed effects. CRT denotes the type of *correlated random trends* under consideration: (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 C.3 and C.4

Tables

Table I: Adequacy Era Court-Ordered Finance Reform Years

State Name	Year of 1st Overturn	Case Name
Alaska	2009	*Moore v. State*
Kansas	2005	*Montoy v. State (Montoy II)*
Kentucky	1989	*Rose v. Council for Better Education*
Massachusetts	1993	*McDuff v. Secretary of Executive Office of Education*
Montana	2005	*Columbia Falls Elementary School District No. 6 v. Montana*
North Carolina	2004	*Hoke County Board of Education v. North Carolina*
New Hampshire	1997	*Claremont School District v. Governor*
New Jersey	1997	*Abbott v. Burke (Abbott IV)*
New York	2003	*Campaign for Fiscal Equity, Inc. v. New York*
Ohio	1997	*DeRolph v. Ohio (DeRolph I)*
Tennessee	1995	*Tennessee Small School Systems v. McWherter (II)*
Wyoming	1995	*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. 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/`.

Table II: Descriptive Statistics for Outcome Variables

	Weighted		Unweighted	
	Mean	**SD**	**Mean**	**SD**
Graduation Rates	.77	[.15]	.82	[.14]
Log Revenues	8.94	[.26]	8.97	[.29]
Total Revenues	7950.2	[2352.12]	8224.31	[2782.13]
Percent Minority	.32	[.3]	.16	[.23]
Percent Black	.17	[.22]	.08	[.17]
Percent Hispanic	.15	[.22]	.08	[.16]
Percent Special Education	.12	[.05]	.13	[.05]
Percent Child Poverty	.16	[.1]	.16	[.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 III: 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, NY	14

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 2009, and Table I 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 2005 to 2009, inclusive. It does not contribute to treatment years 6 through 14.

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

Treatment Year	Weighted					Unweighted				
	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont
1	.029 **	-.004	.019	.015	.00001	.018	.008	.008	.004	.00054 *
	[.011]	[.034]	[.017]	[.012]	[.00019]	[.011]	[.009]	[.01]	[.011]	[.00023]
2	.054 **	.032	.03	.046 *	.00029	.032 +	.027	.014	.018	.0007 *
	[.017]	[.027]	[.021]	[.022]	[.00025]	[.017]	[.017]	[.016]	[.02]	[.00028]
3	.042 *	.013	.031	.044	.00014	.029 +	.028 +	.022	.041	.00097 **
	[.019]	[.041]	[.026]	[.028]	[.00032]	[.015]	[.015]	[.019]	[.029]	[.0003]
4	.058 *	.034	.059 *	.06 +	.00036	.046 **	.045 **	.045 +	.056	.00146 ***
	[.023]	[.046]	[.03]	[.032]	[.00035]	[.017]	[.017]	[.026]	[.034]	[.0003]
5	.081 **	.057	.063 +	.059	.00033	.036	.044	.035	.034	.00137 **
	[.025]	[.042]	[.035]	[.04]	[.0004]	[.026]	[.026]	[.036]	[.037]	[.00043]
6	.119 ***	.106 **	.089 *	.091 *	.0007 *	.049 +	.053 *	.041	.033	.0016 **
	[.035]	[.04]	[.038]	[.044]	[.00035]	[.027]	[.026]	[.031]	[.032]	[.00054]
7	.127 **	.116 *	.108 *	.119 *	.00092 *	.054	.058 +	.041	.038	.00183 **
	[.041]	[.057]	[.048]	[.055]	[.00046]	[.033]	[.032]	[.038]	[.043]	[.00056]
r-squared	.999	.999	.999	.999	.999	.894	.894	.894	.894	.887

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}), state-by-FLE linear time trends ($\psi_{sq}t$), and a state-level factor ($\lambda'_s F_t$). FLE quartiles are indexed by FLE $\in 1, 2, 3, 4$. Column "FLE Cont" corresponds to models in which FLE-by-year fixed effects are substituted for year fixed effects and 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 V: Change in Graduation Rates, by Free Lunch Eligible Quartile and Quantile

	Weighted					Unweighted				
Treatment Year	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont
1	.003	.015 ***	.008	.013 *	.00018 **	.006	.013 ***	.014 **	.012 *	.00016 **
	[.006]	[.004]	[.008]	[.006]	[.00006]	[.004]	[.003]	[.005]	[.006]	[.00005]
2	.005	.002	.001	.03 **	.00025 **	.005	.013 +	.009	.019 **	.00018 *
	[.009]	[.006]	[.009]	[.009]	[.00008]	[.005]	[.007]	[.006]	[.007]	[.00007]
3	-.003	.003	.001	.061 **	.00051 **	.003	.016 *	.008	.024 **	.00023 *
	[.008]	[.01]	[.007]	[.019]	[.00017]	[.005]	[.007]	[.01]	[.009]	[.00009]
4	.013	.019 *	.021 *	.061 ***	.00065 ***	.015 *	.032 **	.025 *	.034 **	.00043 **
	[.01]	[.008]	[.01]	[.013]	[.00011]	[.008]	[.011]	[.01]	[.013]	[.00014]
5	.013	.017 *	.02 +	.07 ***	.00072 ***	.012	.031 **	.027 *	.041 **	.00049 ***
	[.014]	[.009]	[.011]	[.016]	[.00013]	[.009]	[.01]	[.012]	[.015]	[.00015]
6	.012	.019	.024 *	.085 ***	.00088 ***	.013 +	.033 **	.032 *	.05 **	.00059 ***
	[.013]	[.012]	[.011]	[.019]	[.00017]	[.007]	[.013]	[.016]	[.018]	[.00017]
7	-.009	.01	.016	.084 ***	.00085 ***	.008	.03	.021	.047 *	.00052 *
	[.015]	[.016]	[.015]	[.023]	[.00021]	[.01]	[.018]	[.021]	[.023]	[.00021]
r-squared	.999	.999	.999	.999	.999	.573	.573	.573	.573	.573

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}), state-by-FLE linear time trends ($\psi_{sq}t$), and a state-level factor ($\lambda'_s F_t$). FLE quartiles are indexed by FLE $\in 1, 2, 3, 4$. Column "FLE Cont" corresponds to models in which FLE-by-year fixed effects are substituted for year fixed effects and the additional control variable year-by-FLE percentile ($\delta_t Q$) is included. Point estimates for continuous model are interpreted as change in graduation rates 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 $+ < .10, * < .05, ** < .01, *** < .001$)

42

Table VI: Robustness Check: Preferred Model Specification - Change in Demographic Characteristics Resulting from Court Order

	Log Enroll	Percent Minority		Percent SAIPE		Percent Sped		Total Revenues	
Treatment Year	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
1	-.00012	.00001	0	.00022 **	.00012 *	.0004	.00038	.18717	6.27757 *
	[.00007]	[.00005]	[.00002]	[.00008]	[.00006]	[.00027]	[.00023]	[2.00267]	[3.00265]
2	0	.00005	.00001	.00006	-.00003	.00046	.00044 +	3.00714	7.1655 *
	[.00007]	[.00006]	[.00003]	[.00011]	[.0001]	[.00029]	[.00025]	[2.5773]	[2.93854]
3	.00008	.00002	.00001	.00009	-.00002	.00031	.00026	1.75417	9.93939 **
	[.00011]	[.00007]	[.00004]	[.00006]	[.00005]	[.00033]	[.00029]	[3.41087]	[3.47845]
4	.00013	.00003	.00002	.00016 *	.00005	.00039	.00027	4.32547	15.0462 ***
	[.00013]	[.00009]	[.00004]	[.00007]	[.00008]	[.00023]	[.00017]	[4.26495]	[3.92201]
5	.00014	.00009	.00006	.00007	.00003	.00041 +	.00028	4.2002	14.5479 **
	[.00017]	[.00008]	[.00005]	[.00008]	[.00005]	[.00023]	[.00017]	[5.00702]	[5.23225]
6	.00014	.0001	.00005	.00012	.00002	.00049 +	.00033	8.91194 +	18.25232 *
	[.00019]	[.00008]	[.00005]	[.00008]	[.00007]	[.00028]	[.00021]	[4.75711]	[7.54563]
7	.0001	.00008	.00004	.00008	.00004	.00046	.00032	11.43805 +	20.3912 **
	[.00021]	[.0001]	[.00006]	[.00014]	[.00012]	[.0003]	[.00022]	[6.48416]	[7.19182]
r2	.992	.999	.971	.999	.914	.999	.6	.999	.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), year fixed effects (δ_t), state-by-FLE linear time trends ($\psi_{sq}t$), a state-level factor ($\lambda'_s F_t$), and a FLE percentile-by-year fixed effect ($\delta_t Q$). Point estimates are interpreted as change in dependent variable per 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 $+ < .10, * < .05, ** < .01, *** < .001$)

Appendices

A Data Appendix

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`

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.[20] 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.[21]

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 & LaFontaine (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.

---

[20]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.

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

B   Additional Figures

Figure B.1: New York City Public Schools Potential Mediators & Outcomes
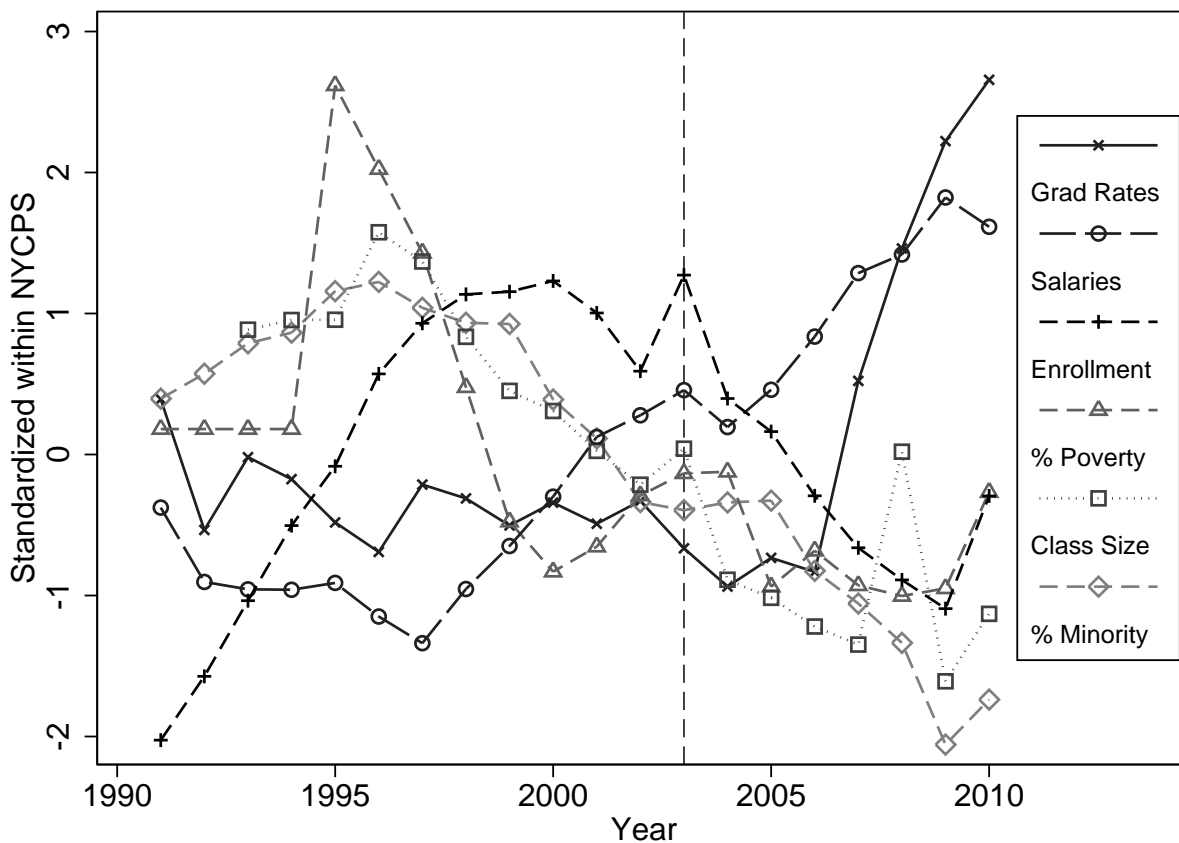
NYCPS is the largest district in the United States, and it was also demonstrably improving in many ways during this period. Here, we plot standardized beta coefficients for various outcomes of interest. These are standardized to be mean zero with standard deviation one for NYCPS across the entire time 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.

C Additional Tables

Table C.1: Log Per Pupil Revenues Results, Model Sensitivity Specifications, Weighted Least Squares

	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										

Preferred model highlighted. Point estimates and standard errors, indicated by brackets, 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). Not all combinations are available.

Table C.2: Log Per Pupil Revenues Results, Model Sensitivity Specifications, OLS

	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-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										

Preferred model highlighted. Point estimates and standard errors, indicated by brackets, 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). Not all combinations are available.

Table C.3: Graduation Rates Results, Model Sensitivity Specifications, Weighted Least Squares

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-0.013	0.02	0.013	0.01	0.013	0.018	0.002	0.001	0.019	-0.01	0.02	0.019	0.005	0.004	0.018
	[.009]	[.009]	[.006]	[.007]	[.009]	[.012]	[.007]	[.008]	[.009]	[.008]	[.006]	[.01]	[.005]	[.005]	[.006]
2	-0.01	0.037	0.03	0.026	0.029	0.03	0.013	0.012	0.036	-0.008	0.033	0.031	0.014	0.012	0.033
	[.01]	[.012]	[.009]	[.01]	[.012]	[.019]	[.007]	[.012]	[.012]	[.011]	[.009]	[.016]	[.006]	[.01]	[.009]
3	0.021	0.077	0.061	0.056	0.052	0.063	0.048	0.045	0.076	0.021	0.071	0.063	0.048	0.044	0.07
	[.02]	[.017]	[.019]	[.021]	[.015]	[.021]	[.014]	[.017]	[.017]	[.022]	[.017]	[.018]	[.016]	[.018]	[.017]
4	0.009	0.074	0.061	0.055	0.048	0.058	0.042	0.042	0.074	0.008	0.066	0.057	0.04	0.039	0.065
	[.01]	[.018]	[.013]	[.013]	[.018]	[.035]	[.01]	[.023]	[.019]	[.011]	[.014]	[.031]	[.008]	[.02]	[.014]
5	0.017	0.091	0.07	0.065	0.054	0.063	0.056	0.052	0.09	0.017	0.083	0.066	0.055	0.051	0.082
	[.014]	[.017]	[.016]	[.017]	[.02]	[.04]	[.01]	[.025]	[.018]	[.014]	[.013]	[.035]	[.01]	[.022]	[.014]
6	0.024	0.11	0.085	0.079	0.068	0.071	0.07	0.065	0.108	0.027	0.103	0.078	0.072	0.066	0.101
	[.02]	[.019]	[.019]	[.02]	[.024]	[.049]	[.013]	[.029]	[.019]	[.021]	[.015]	[.043]	[.013]	[.028]	[.015]
7	0.018	0.115	0.084	0.077	0.062	0.062	0.072	0.064	0.113	0.023	0.107	0.073	0.075	0.066	0.106
	[.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										

Preferred model highlighted. Point estimates and standard errors, indicated by brackets, 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). Not all combinations are available.

Table C.4: Graduation Rates Results, Model Sensitivity Specifications, OLS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-0.001	0.016	0.012	0.014	0.008	0.021	0.012	0.016	0.016	-0.004	0.019	0.022	0.009	0.012	0.019
	[.007]	[.008]	[.006]	[.007]	[.006]	[.01]	[.006]	[.007]	[.009]	[.007]	[.009]	[.011]	[.006]	[.007]	[.009]
2	-0.004	0.022	0.019	0.021	0.01	0.029	0.015	0.021	0.021	-0.009	0.023	0.029	0.01	0.015	0.023
	[.011]	[.01]	[.007]	[.008]	[.01]	[.014]	[.007]	[.009]	[.01]	[.012]	[.009]	[.014]	[.006]	[.009]	[.009]
3	0.003	0.033	0.024	0.026	0.011	0.041	0.025	0.029	0.033	-0.004	0.034	0.04	0.018	0.023	0.034
	[.012]	[.011]	[.009]	[.009]	[.01]	[.018]	[.008]	[.012]	[.011]	[.012]	[.01]	[.019]	[.007]	[.012]	[.01]
4	0.007	0.042	0.034	0.035	0.012	0.053	0.033	0.042	0.042	0	0.044	0.053	0.026	0.035	0.044
	[.012]	[.017]	[.013]	[.014]	[.016]	[.028]	[.013]	[.018]	[.018]	[.012]	[.017]	[.029]	[.012]	[.019]	[.017]
5	0.016	0.055	0.041	0.043	0.017	0.065	0.045	0.053	0.055	0.008	0.058	0.066	0.038	0.046	0.057
	[.013]	[.019]	[.015]	[.015]	[.017]	[.034]	[.013]	[.022]	[.02]	[.012]	[.018]	[.035]	[.013]	[.022]	[.019]
6	0.022	0.069	0.05	0.051	0.016	0.076	0.056	0.063	0.068	0.016	0.075	0.08	0.051	0.058	0.074
	[.013]	[.022]	[.018]	[.017]	[.016]	[.045]	[.016]	[.03]	[.023]	[.013]	[.021]	[.045]	[.016]	[.03]	[.022]
7	0.015	0.068	0.047	0.046	0.008	0.075	0.055	0.062	0.068	0.01	0.076	0.08	0.05	0.057	0.075
	[.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										

52

Preferred model highlighted. Point estimates and standard errors, indicated by brackets, 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). Not all combinations are available.

Table C.5: Robustness: Log Enrollment, Preferred Model

Treatment Year	Weighted					Unweighted				
	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont
1						-.012 **	-.01	-.015 +	-.008	-.00012
						[.004]	[.008]	[.009]	[.005]	[.00007]
2						-.011 *	-.01	-.011	.001	0
						[.006]	[.009]	[.007]	[.005]	[.00007]
3						-.016 +	-.007	-.012	.006	.00008
						[.008]	[.016]	[.013]	[.014]	[.00011]
4						-.021	-.012	-.013	.009	.00013
						[.013]	[.02]	[.018]	[.017]	[.00013]
5						-.025	-.015	-.017	.008	.00014
						[.017]	[.025]	[.022]	[.02]	[.00017]
6						-.034	-.024	-.025	.01	.00014
						[.021]	[.028]	[.025]	[.022]	[.00019]
7						-.044 +	-.033	-.035	.005	.0001
						[.024]	[.032]	[.028]	[.026]	[.00021]
r-squared						.992	.992	.992	.992	.992

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 ($\psi_{sq}t$), a state-level factor ($\lambda'_s F_t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). Point estimates are interpreted as change in dependent variable per 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 C.6: Robustness: Percent Minority, Preferred Model

Treatment Year	Weighted					Unweighted				
	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont
1	-.004 +	.003	.003	-.004	.00001	-.002 *	.001	0	-.001	0
	[.002]	[.002]	[.003]	[.005]	[.00005]	[.001]	[.001]	[.001]	[.001]	[.00002]
2	-.009 *	-.001	.002	.004	.00005	-.002 +	-.001	0	.002	.00001
	[.005]	[.002]	[.003]	[.004]	[.00006]	[.001]	[.001]	[.001]	[.002]	[.00003]
3	-.013	-.004	.002	.001	.00002	-.002	-.001	0	.002	.00001
	[.009]	[.005]	[.005]	[.003]	[.00007]	[.002]	[.002]	[.002]	[.002]	[.00004]
4	-.011	-.004	.006	0	.00003	-.002	-.001	.002	.003	.00002
	[.01]	[.006]	[.006]	[.004]	[.00009]	[.003]	[.003]	[.003]	[.002]	[.00004]
5	-.005	.004	.01 *	0	.00009	0	.002	.005 +	.005 *	.00006
	[.005]	[.003]	[.004]	[.004]	[.00008]	[.002]	[.002]	[.003]	[.002]	[.00005]
6	-.008	.004	.013 *	0	.0001	-.001	.001	.004	.005	.00005
	[.006]	[.004]	[.005]	[.005]	[.00008]	[.002]	[.003]	[.003]	[.003]	[.00005]
7	-.012	.003	.014 *	-.004	.00008	-.003	.001	.004	.004	.00004
	[.008]	[.005]	[.007]	[.005]	[.0001]	[.003]	[.003]	[.004]	[.003]	[.00006]
r-squared	1	1	1	1	1	.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 ($\psi_{sq}t$), a state-level factor ($\lambda_s' F_t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). Point estimates are interpreted as change in dependent variable per 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 C.7: Robustness: Percent Child Poverty (SAIPE), Preferred Model

Treatment Year	Weighted					Unweighted				
	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont
1	.012 ***	.015 **	.016 ***	.02 *	.00022 **	.008 *	.01 ***	.015 ***	.009	.00012 *
	[.003]	[.005]	[.004]	[.009]	[.00008]	[.004]	[.003]	[.003]	[.007]	[.00006]
2	.005	.002	-.003	.004	.00006	.004	.003	.004	-.004	-.00003
	[.004]	[.004]	[.009]	[.011]	[.00011]	[.005]	[.005]	[.006]	[.01]	[.0001]
3	.009 +	.008 *	.012 *	.011 *	.00009	.006	.007	.008	-.001	-.00002
	[.005]	[.004]	[.005]	[.005]	[.00006]	[.006]	[.006]	[.006]	[.005]	[.00005]
4	.013 *	.009	.015 +	.028 **	.00016 *	.008	.01	.013	.007	.00005
	[.006]	[.006]	[.008]	[.008]	[.00007]	[.008]	[.008]	[.009]	[.008]	[.00008]
5	.017 **	.017 **	.02 **	.019 **	.00007	.011	.012	.014	.006	.00003
	[.006]	[.005]	[.007]	[.007]	[.00008]	[.007]	[.007]	[.01]	[.004]	[.00005]
6	.024 **	.019 **	.023 **	.032 ***	.00012	.01	.011	.016	.009	.00002
	[.007]	[.007]	[.008]	[.009]	[.00008]	[.01]	[.01]	[.012]	[.009]	[.00007]
7	.027 **	.022 **	.027 **	.032 +	.00008	.013	.014	.021	.011	.00004
	[.009]	[.008]	[.01]	[.017]	[.00014]	[.012]	[.012]	[.016]	[.013]	[.00012]
r-squared	1	1	1	1	1	.914	.914	.914	.914	.914

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 ($\psi_{sq}t$), a state-level factor ($\lambda'_s F_t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). Point estimates are interpreted as change in dependent variable per 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 C.8: Robustness: Special Education, Preferred Model

Treatment Year	Weighted					Unweighted				
	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont
1	.024	.02	.022	.033	.0004	.029 +	.026	.027	.029	.00038
	[.015]	[.014]	[.016]	[.025]	[.00027]	[.015]	[.016]	[.016]	[.019]	[.00023]
2	.027	.027 +	.025	.038	.00046	.031 +	.031 +	.031 +	.033	.00044 +
	[.018]	[.016]	[.018]	[.027]	[.00029]	[.018]	[.017]	[.018]	[.02]	[.00025]
3	.022	.015	.016	.025	.00031	.022	.019	.019	.02	.00026
	[.018]	[.018]	[.021]	[.03]	[.00033]	[.019]	[.019]	[.021]	[.022]	[.00029]
4	.027 *	.02	.025 +	.029	.00039	.023 *	.021 *	.023 +	.02	.00027
	[.013]	[.013]	[.015]	[.023]	[.00023]	[.01]	[.01]	[.012]	[.015]	[.00017]
5	.032 **	.024 *	.026	.03	.00041 +	.024 *	.021 *	.022 +	.022	.00028
	[.012]	[.012]	[.016]	[.024]	[.00023]	[.01]	[.01]	[.012]	[.015]	[.00017]
6	.038 **	.032 *	.03	.034	.00049 +	.029 *	.026 *	.026 +	.023	.00033
	[.014]	[.014]	[.018]	[.029]	[.00028]	[.012]	[.013]	[.015]	[.018]	[.00021]
7	.036 *	.03 *	.031	.031	.00046	.026 *	.025 +	.027	.022	.00032
	[.015]	[.015]	[.019]	[.031]	[.0003]	[.012]	[.014]	[.016]	[.019]	[.00022]
r-squared	1	1	1	1	1	.601	.601	.601	.601	.6

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 ($\psi_{sq}t$), a state-level factor ($\lambda'_s F_t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). Point estimates are interpreted as change in dependent variable per 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 C.9: Robustness: Total Per Pupil Revenues, Preferred Model

Treatment Year	Weighted					Unweighted				
	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont	FLE 1	FLE 2	FLE 3	FLE 4	FLE Cont
1	452 *	219	373	310 *	0.187	211 *	47	47	-30	6.278 *
	[242]	[382]	[265]	[169]	[2.00267]	[94]	[70]	[101]	[122]	[3.00265]
2	631 *	393	392	563 *	3.007	355 **	258 *	100	123	7.166 *
	[282]	[359]	[282]	[232]	[2.5773]	[131]	[119]	[148]	[177]	[2.93854]
3	547	260	385	573 *	1.754	354 *	263 *	159	338	9.934 **
	[371]	[519]	[353]	[241]	[3.41087]	[144]	[125]	[182]	[277]	[3.47845]
4	782 *	584	760 *	853 **	4.325	538 ***	425 **	394 *	503	15.046 ***
	[443]	[586]	[402]	[263]	[4.26495]	[141]	[136]	[261]	[359]	[3.92201]
5	1005 *	770	803 *	893 *	4.200	445 *	381 *	246	255	14.548 **
	[485]	[582]	[475]	[364]	[5.00702]	[214]	[231]	[354]	[404]	[5.23225]
6	1369 *	1276 *	1061 *	1268 **	8.91194 +	653 **	531 **	388	316	18.252 *
	[630]	[648]	[607]	[486]	[4.75711]	[201]	[208]	[305]	[322]	[7.54563]
7	1428 *	1309 *	1167 *	1567 ***	11.43805 +	688 **	576 *	358	373	20.39 **
	[722]	[758]	[614]	[429]	[6.48416]	[248]	[297]	[409]	[474]	[7.19182]
r-squared	1	1	1	1	1	1	1	1	1	0.864

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 ($\psi_{sq}t$), a state-level factor ($\lambda_s' F_t$), and, in the case of continuous models, a FLE percentile-by-year fixed effect ($\delta_t Q$). Point estimates are interpreted as change in dependent variable per 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 $+ < .10, * < .05, ** < .01, *** < .001$)

D Technical Appendix: 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 & Weidner (2015). 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:

(D.1) $Y_{it} = \alpha_i + \theta_t + D'_{it}\beta + \lambda'_i F_t + \varepsilon_{it}$,

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

(D.2) $f(\lambda, F) \approx f(\hat{\lambda}, \hat{F}) + \frac{\partial f}{\partial \lambda}\Big|_{\lambda=\hat{\lambda}, F=\hat{F}}(\lambda - \hat{\lambda}) + \frac{\partial f}{\partial F}\Big|_{\lambda=\hat{\lambda}, F=\hat{F}}(F - \hat{F})$

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

$$\text{(D.3)} \quad \lambda_i' F_t \approx \lambda_i' \hat{F}_t + \hat{\lambda_i}' F_t - \hat{\lambda_i}' \hat{F}_t.$$

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

$$\text{(D.4)} \quad \lambda_i' F_t \approx \alpha_i \hat{F}_t + \hat{\lambda}_i \theta_t.$$

The model we can estimate via OLS is

$$\text{(D.5)} \quad Y_{it} = \alpha_i + \theta_t + D_{it}' \beta + \alpha_i \hat{F}_t + \hat{\lambda}_i \theta_t + \varepsilon_{it},$$

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 D.1. *OLS estimation of Equation* (D.5) *is identical to OLS estimation of Equation* (D.1).

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.[22]

Part 1: Obtain FOCs for Equation (D.1)

---

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

Step 1: Specify the objective function:

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

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 (D.5)

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 (D.1) and (D.5) are identical. $\square$