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

Kentucky’s 1989 Supreme Court ruling *Rose v. Council for Better Education* marked a new era of state court-ordered finance reforms (Murray, Evans, and Schwab 1998; Berry 2007; Corcoran and Evans 2008; Sims 2011). The post-*Rose* court cases were increasingly argued on adequacy grounds, meaning that plaintiffs were no longer able to successfully argue that all children had a right to an equal education. Adequacy-based language required that students received sufficient resources to meet academic standards laid out by the state.

At least two factors distinguish these court rulings from rulings that took place in the decades prior. First, three years after the *Rose* case, the gap in per-pupil spending between low- and high-income districts narrowed noticeably and continued to close (Corcoran et al. 2004; Jackson, Johnson, and Persico 2016). How much this change in spending should be attributed to Adequacy cases or other secular changes is not known. Whether these rulings had an impact on the level and distribution of resources and academic outcomes is, therefore, an unresolved and open question.

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

From an economic standpoint, the causal relationship between spending and desirable outcomes is also of interest because the share of gross domestic product that the United States spends on public elementary and secondary education has remained substantial, ranging between 3.4 percent and 4.2 percent since the 1970s. Given the large share of spending on education, it would be useful to know if these resources are well spent. The goal of this paper is to provide a robust description of the causal relationship between fiscal shocks at the state level and student outcomes at the district level between academic years 1990–91 and 2009–10.

Our paper is therefore motivated by two overlapping research questions:

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

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

---

Using district-level, aggregate data from the Common Core of Data (CCD), we estimate the causal effects of court-ordered finance reforms on real per-pupil revenues and graduation rates. We specifically analyze court decisions that overturned states’ school finance systems based partly or fully on Adequacy grounds. The academic outcome of interest is the high school graduation rate, which is the most distal academic outcome that can be observed in our data for students living in states undergoing Adequacy-based, court-ordered finance reform.

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

We then subject the model to various sensitivity tests by permuting secular time trends, correlated random trends (unit-specific trends), and, to account for cross-sectional dependence, interactive fixed effects. In total, we estimate fifteen complementary models. Generally, the results are robust to model fit: Relative to the benchmark model, interactive fixed effects and alternative specifications of the secular time trend have modest effects on point estimates. When correlated random trends are excluded, the effect on graduation rates disappears; however, estimating unit-specific time trends at different levels of aggregation (the state and district) and using different functional forms (linear and quadratic) produces estimates that are similar to those of the preferred model. Thus, there is evidence that treatment and control substate units (especially high-poverty districts) have different secular trends, but consistent point estimates are estimable under the assumption that pretreatment trends can be approximated with a functional form.

Further exploration of the data reveals that, for states undergoing reform, (1) the marginal effect of an increase in real per-pupil revenues increased graduation rates in poorer districts but not wealthier districts; (2) spending and graduation rates increased faster in high-poverty districts than in low-poverty districts; and (3) the pattern of results identified in the aggregate model varies substantially for individual states undergoing court order, as some states received additional revenues without a corresponding change in graduation rates (and vice versa).

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

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

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

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

Recognizing this possibility, Sims (2011) and Corcoran and Evans (2015) compare the effect these cases had on spending in low- and high-income (or high- and low-poverty) districts. They find that these Adequacy cases increased spending in higher-poverty districts, and more resources were allocated to districts based on observable indicators of student need, such as free lunch eligibility status and the percent of students who are minority. Lafortune, Rothstein, and Schanzenbach (2018) extend this sample to year 2013 and find complementary results, although the authors include non-Adequacy cases as well as legislative reforms in their evaluation.

As school spending increased in predominantly low-income districts, it was natural to ask whether increased spending positively affected student outcomes. Numerous papers have attempted to identify at the national level causal effects of spending on achievement, graduation, and earnings (Card and Payne 2002; Hoxby 2001; Jackson, Johnson, and Persico 2016; Lafortune, Rothstein, and Schanzenbach 2018). While the specific econometric methods vary, each of the studies included here assume that court rulings result in exogenous shocks to the state’s education finance system. The results of Card 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.

Jackson, Johnson, and Persico (2016) construct a longer panel (with respect to time), using restricted-use individual-level data, reaching back to children born between 1955 and 1985, to test the effects of these cases on revenues, graduation rates, and adult earnings. Leveraging variation across cohorts in exposure to fiscal reform, that 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.

Using data from the National Assessment of Educational Progress, Lafortune, Rothstein, and Schanzenbach (2018) investigate the effects of sixty-eight post-1990 legislative and court-ordered reforms on student academic outcomes. The authors find that spending and achievement gaps narrowed between high-income and low-income school districts in states undergoing reform during this period, and that an extra $6,200 over ten years in per-pupil spending in districts with below-average income raised achievement by 0.1 standard deviations relative to average-income districts.

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

1. We examine the effects on revenues and graduation rates of Adequacy-based court cases taking place between 1989 and 2010. By including cases and data up to year 2010 and investigating the effects of these cases on graduation rates, this analysis updates and extends the work of Sims (2011). This analysis also complements the work of Lafortune, Rothstein, and Schanzenbach (2018). Those authors identify proximal effects of post-1990 finance reform on student achievement outcomes, whereas we examine the most distal outcomes available for students in this era by looking at graduation rates.

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

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

3. DATA
The dataset is the compilation of several public-use surveys that are administered by the National Center for Education Statistics (NCES) and the U.S. Census Bureau. The analytic sample is constructed using the following datasets: Local Education Agency (School District) Finance Survey (F-33); Local Education Agency (School District) Universe Survey; Local Education Agency Universe Survey Longitudinal Data File:
Court-Ordered Finance Reforms


Data are available beginning in the 1990–91 school year up to the 2009–10 school year. The dataset is a panel of aggregated data, containing U.S. 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, percentages of eighth-grade students receiving diplomas four years later (graduation rates), total enrollment, percentages 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 percentages 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 highest poverty quartile for the state.

Total revenues are the sum of federal, local, and state revenues in each district. The variable total revenues is then divided by the total number of students in the district to provide a per-pupil estimate; deflated by the U.S. Consumer Price Index, All Urban Consumers index to convert the figure to real terms; and then converted to the natural logarithm. Graduation rates are defined as the total number of diploma recipients in year \( t \) as a share of the number of eighth graders in year \( t - 4 \), a measure that Heckman and LaFontaine (2010) show is not susceptible to the downward bias caused by using lagged ninth-grade enrollment in the denominator. Graduation rates are top-coded so that they take a maximum value of 1. The demographic race variables come from the school-level file from the CCD and are aggregated to the district level; percentages are calculated by dividing by total enrollment. Special education counts come from the district-level CCD. Child poverty is a variable from the Small Area Income and Poverty Estimates.

To define the analytic sample, we apply the following restrictions. First, Hawaii and the District of Columbia are removed from the sample, as each place has only one school district. Montana is also removed because the state is missing a substantial amount of graduation rate data. Only unified districts are included so that it is possible to link changes in revenues to graduation rates at a later date. Unified districts are defined as those districts that serve students in either pre-kindergarten, kindergarten, or first grade through the twelfth grade. For the variables total enrollment, graduation rates, and FLE, New York City Public Schools (NYCPS) reports its data as thirty-three geographic districts in the nonfiscal surveys; for total revenues, NYCPS is treated as a consolidated district. For this reason, the nonfiscal data are combined into a single district. As suggested in the NCES documentation, NYCPS’s supervisory union number is used to aggregate the geographical districts into a single entity. Online Appendix A.3 outlines our approach to addressing data errors in the analytic sample.

---

2. Web links to each of the data sources are available in a separate online appendix that can be accessed on Education Finance and Policy’s Web site at www.mitpressjournals.org/doi/suppl/10.1162/edfp_a_00236. Please see online Appendix A.1.

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

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

5. In dropping all nonunified school districts, we exclude 3,561 districts.
Table 1. Adequacy Era Court-Ordered Finance Reform Years

<table>
<thead>
<tr>
<th>State Name</th>
<th>Year of 1st Overturn</th>
<th>Funding Formula Adopted</th>
<th>Case Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>2009</td>
<td>LE + MFP</td>
<td>Moore v. State</td>
</tr>
<tr>
<td>Kansas</td>
<td>2005</td>
<td>EP + LE + MFP + SL</td>
<td>Montoy v. State (Montoy II)</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>1993</td>
<td>MFP</td>
<td>McDuff v. Secretary of Executive Office of Education</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>1997</td>
<td>MFP</td>
<td>Claremont School District v. Governor</td>
</tr>
<tr>
<td>Ohio</td>
<td>1997</td>
<td>MFP</td>
<td>DeRolph v. Ohio (DeRolph I)</td>
</tr>
</tbody>
</table>

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

To the analytic sample, we add the first year a state’s funding regime was overturned in the Adequacy era. The base set of court cases comes from Corcoran and Evans (2008), and were updated using data from the National Education Access Network. The analytic file consists of 188,752 district-year observations. The total number of unified districts is 9,916. The average graduation rate is about 77 percent and average log per-pupil spending is 8.94 (total real per-pupil revenues are about $7,950.20). Nonweighted summary statistics are slightly larger in magnitude.

4. ECONOMETRIC SPECIFICATIONS AND MODEL SENSITIVITY

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

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

Finally, because treatment occurs at the state level and our outcomes are at the district level, there are several ways one may choose to specify the estimation 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., pretreatment trends) and unobserved factors, such as cross-sectional dependence. We outline these alternative modeling choices and discuss their implications relative to the benchmark model. We then estimate fifteen different models in which we permute these model specifications—the result of which is that our benchmark model is, in most cases, insensitive to these modeling choices. This suggests the exogeneity of the timing of court-ordered finance reform is defensible.

### Event Study: Testing the Exogeneity of Court Rulings

We first assess the exogeneity of the court rulings, as these will be treated as shocks to identify the causal effects of finance reform. Using an event study model, it can be shown whether it is appropriate to treat adequacy-based decisions overturning states’ school finance systems as exogenous shocks to both revenues and graduation rates. To model the heterogeneous impact of these cases across districts within states, treatment

---

**Table 2. Descriptive Statistics for Outcome Variables**

<table>
<thead>
<tr>
<th></th>
<th>Weighted</th>
<th></th>
<th>Unweighted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Graduation rates</td>
<td>77</td>
<td>15</td>
<td>82</td>
<td>14</td>
</tr>
<tr>
<td>Log revenues</td>
<td>8.94</td>
<td>0.26</td>
<td>8.97</td>
<td>0.29</td>
</tr>
<tr>
<td>Total revenues</td>
<td>7,950.2</td>
<td>2,352.12</td>
<td>8,224.31</td>
<td>2,782.13</td>
</tr>
<tr>
<td>Percent minority</td>
<td>32</td>
<td>30</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>Percent black</td>
<td>17</td>
<td>22</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>Percent Hispanic</td>
<td>15</td>
<td>22</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Percent special education</td>
<td>12</td>
<td>5</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Percent child poverty</td>
<td>16</td>
<td>10</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Log enrollment</td>
<td>9.43</td>
<td>1.62</td>
<td>7.4</td>
<td>1.25</td>
</tr>
</tbody>
</table>

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.
parameters are indexed by state-specific poverty quartiles, defined using free lunch eligibility status.9

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

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

\[
Y_{sqdt} = \theta_d + \delta_{qt} + \sum_{n=1}^{16} \gamma_{(q,-n)}[1(Q_{sd} = q) \times 1(t - t_s^n = -n) \times D_s] \\
+ \sum_{n=1}^{19} \gamma_{(q,+n)}[1(Q_{sd} = q) \times 1(t - t_s^n = n) \times D_s] + \epsilon_{sqdt}, \tag{1}
\]

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

The parameters of interest are the \( \gamma_{q,n} \), which are the estimates of the effect of school finance reform in quartile \( q \) in treatment year \( n \) on \( Y_{sqdt} \). In total, the model provides estimates for sixteen anticipatory effects and nineteen post-treatment effects for each poverty quartile.10 Although there are twenty-one post-treatment effects available, treatment effect years 19 through 21 are combined into a single binary indicator, as there are only two treatment states that contribute to causal identification in these later years.11 In

9. This is the approach taken by Sims (2011).
10. Because treatment is based on the first year of an Adequacy-era reform in state \( s \), estimated effects in the post-treatment period will reflect the first Adequacy-era court-ordered finance reform and possibly any subsequent reforms in later years. We privilege the first court-ordered reform because the timing of subsequent reforms is more likely to be endogenously timed with the first reform.
11. Although our data sample has twenty years of data, we have up to twenty-one potential treatment effects, as Kentucky (KY) had its reform in 1989 and our panel begins in the 1990–91 academic year. Therefore, KY does...
Court-Ordered Finance Reforms

Notes: Differences-in-differences with treatment effects estimated nonparametrically before and after reform. Model accounts for district fixed effects ($\theta_d$) and free lunch eligible (FLE)-by-year fixed effects ($\delta_{tq}$). Range caps show 90 percent confidence intervals. Standard errors are clustered at the state level, which is the level at which treatment occurs. Black lines with hollow circles reflect weighted least squares models, where the weights are district enrollment; gray lines with hollow triangles reflect ordinary least squares models.

Figure 1. Event Study of Log Per-Pupil Revenues

reporting our results, only estimates for years 1 through 7 before and after reform are reported, as estimates outside this range suffer from precision loss. For example, very few states were treated early enough to contribute information to the post-treatment effect estimates in later years (see table 1). Because we include $\delta_{tq}$ fixed effects, the estimates of $\gamma_{q, n}$ are identified using the variation within quartile $q$ by year $t$ cells; consequently, results are not directly comparable across the quartiles.

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

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.
Christopher A. Candelaria and Kenneth A. Shores

Differences-in-differences with treatment effects estimated nonparametrically before and after reform. Model accounts for district fixed effects ($\theta_d$) and free lunch eligible (FLE)-by-year fixed effects ($\delta_{tq}$). Range caps show 90 percent confidence intervals. Standard errors are clustered at the state level, which is the level at which treatment occurs. Black lines with hollow circles reflect weighted least squares models, where the weights are district enrollment; gray lines with hollow triangles reflect ordinary least squares models.

Figure 2. Event Study of Graduation Rates

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

The results of the event study suggest the court rulings were an exogenous shock to revenues, but the adoption of court order may have been preceded by downward trends in graduation rates among treated districts in poverty quartile 4. We proceed by estimating a traditional differences-in-differences specification that restricts the pretreatment period to zero. To account for the potentially heterogeneous secular trends between
treated and control units, linear correlated random trends are included for each state by poverty quartile group.\textsuperscript{12}

**Benchmark Differences-in-Differences Model**

The preferred benchmark model takes the following form:

\[
Y_{sqdt} = \theta_d + \delta_{qt} + \psi_{sqt} + \sum_{n=1}^{19} \gamma_{(q,n)}[1(Q_{sd} = q) \times 1(t - t^\ast = n)]\times D_s + \varepsilon_{sqdt},
\]

where the pretreatment period is restricted to be zero, and the term $\psi_{sqt}$, which is a state by quartile-specific linear time trend, is added. As before, serial correlation is accounted for by clustering standard errors at the state level (Bertrand, Duflo, and Mullainathan 2004). Together, this model absorbs approximately 9,800 district fixed effects, 76 year effects, and 192 continuous fixed effects (state-by-FLE quartile interacted with linear time), and estimates 76 treatment variables for years 1 to 19 and quartiles 1 through 4. Wolfers (2006) shows that including $\psi_{sqt}$ can introduce bias when only a singular post-treatment effect is estimated, but this bias can be eliminated if the entire treatment response is estimated nonparametrically. For this reason, we include 76 treatment variables for 19 treatment years and four poverty quartiles. The variables for equation 2 are included for different reasons, which are explained below.

**Secular Time Trends**

In equation 2, secular time trends are modeled with FLE quartile-by-year fixed effects, denoted as $\delta_{qt}$. These $\delta_{tq}$ fixed effects are included instead of standard year fixed effects $\delta_t$ to establish a more plausible and relevant counterfactual trend for treated districts. Quartile-by-year fixed effects indicate whether finance reform increased $Y_{sqdt}$ in high-poverty districts relative to high-poverty districts without reform, whereas the year fixed effect alone provides a relative comparison to trends occurring at the state level.

**Correlated Random Trends**

In equation 2, the parameter $\psi_{sqt}$ is included because there was some evidence that the timing of a state’s ruling was correlated with trends in graduation rates among high-poverty districts. Including $\psi_{sqt}$ adjusts for all state-by-quartile secular trends but assumes these trends follow a functional form. Equation 2 assumes the functional form is linear but this assumption can be tested, as described in section 5.

**Weighting**

In equation 2, treatment heterogeneity is explicitly modeled by disaggregating treatment effects into four poverty quartiles. Although treatment heterogeneity is modeled across poverty quartiles, other sources of treatment heterogeneity may be missed. Weighting a regression model by group size is traditionally used to correct for heteroskedasticity, but it also provides a way to test for the presence of additional unobserved heterogeneity. According to asymptotic theory, the probability limits of OLS and

\textsuperscript{12.} In section 5, quadratic correlated random trends are included as well, to allow for a more flexible functional form.
WLS should be consistent. Thus, regardless of how one weights the data, the point estimates between the two models should not dramatically differ. When OLS and WLS estimates do diverge substantially, there is concern that the model is not correctly specified, and it may be due in part to unobserved heterogeneity associated with the weighting variable (DuMouchel and Duncan 1983; Solon, Haider, and Wooldridge 2015).

5. RESULTS

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

Benchmark Differences-in-Differences Model Results

In this section, we report the causal effect estimates of court-ordered finance reform on real per-pupil revenues and graduation rates. In our tables and graphs, we only report treatment effect estimates for years 1 through 7, as the number of states in the treatment group changes substantially over time.13 Tables and graphs display both weighted and unweighted estimates, where the weight is time-varying district enrollment. FLE quartile 1 represents low-poverty districts, and FLE quartile 4 represents high-poverty districts.

Revenues

Results in table 3 reveal that court-ordered finance reforms increased revenues in all FLE quartiles in the years after treatment, though not every point estimate is significant at conventional levels. Because the model includes FLE quartile-by-year fixed effects, point estimates for a given quartile are interpreted relative to other FLE quartiles that are in the control group. For example, in year 7 after treatment, districts in FLE quartile 1 had revenues that were 12.8 percent higher than they would have been in the absence of treatment, with the counterfactual trend established by nontreated districts also in FLE quartile 1 (significant at the 5 percent level). In FLE quartile 4, we find that the revenues were 11.5 percent higher relative to what they would have been in the absence of reform, relative to nontreated districts in FLE quartile 4 (significant at the 1 percent level).14

---

13. See online Appendix B.2 for additional details about this decision, a table showing the treatment year in which a given treatment state exits the analytic sample, and estimated effects for revenues and graduation rates in high-poverty districts for the fifteen-year post-treatment period.

14. One point to note: It would be wrong to conclude that the 12.8 percent effect is larger than the 11.5 percent effect, because the 12.8 percent effect for quartile 1 may be relative to only a modest increase in nontreated low-poverty
Table 3. Change in Log Per-Pupil Revenues, by Free Lunch Eligible (FLE) Quartile and Quantile

<table>
<thead>
<tr>
<th>Treatment Year</th>
<th>Weighted</th>
<th>Unweighted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FLE 1</td>
<td>FLE 2</td>
</tr>
<tr>
<td>1</td>
<td>0.043*</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>2</td>
<td>0.061*</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>3</td>
<td>0.05*</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>4</td>
<td>0.072*</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>5</td>
<td>0.098*</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>6</td>
<td>0.129*</td>
<td>0.105+</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>7</td>
<td>0.128*</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>R²</td>
<td>0.909</td>
<td>0.909</td>
</tr>
</tbody>
</table>

Notes: This table shows point estimates and standard errors for nonparametric differences-in-differences estimator. Model accounts for district fixed effects ($θ_d$), FLE-by-year fixed effects ($δ_{tq}$), and state-by-FLE linear time trends ($ψ_{sq}$). FLE quartiles are indexed by $FLE \in 1, 2, 3, 4$. Column “FLE Cont.” corresponds to models in which the additional control variable year-by-FLE percentile ($δ_{tQ}$) is included. Point estimates for continuous model are interpreted as change in revenues for one-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.

$+ p < 0.10; * p < 0.05; ** p < 0.01; *** p < 0.001.$

The unweighted results in table 3 also suggest that revenues increased. Although the magnitudes differ slightly, the coefficient estimates are comparable to the corresponding weighted results. Overall, revenues increased across all FLE poverty quartiles in states with court-ordered finance reform, relative to equivalent poverty quartiles in nontreated states.

**Graduation Rates**

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

districts, whereas the 11.5 percent effect for quartile 4 may be relative to a steep increase in nontreated high-poverty districts. Later, we present models and results to test whether poorer districts received greater revenues and graduation rates following reform.
Table 4. Change in Graduation Rates, by Free-Lunch Eligible (FLE) Quartile and Quantile

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Weighted</th>
<th></th>
<th>Unweighted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FLE 1</td>
<td>FLE 2</td>
<td>FLE 3</td>
<td>FLE 4</td>
</tr>
<tr>
<td>Year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.01*</td>
<td>0.026***</td>
<td>0.016</td>
<td>0.02*</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.008)</td>
<td>(0.01)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>2</td>
<td>0.007</td>
<td>0.008</td>
<td>0.003</td>
<td>0.037**</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>3</td>
<td>0.008</td>
<td>0.02</td>
<td>0.013</td>
<td>0.077***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>4</td>
<td>0.021*</td>
<td>0.034*</td>
<td>0.026</td>
<td>0.074***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.015)</td>
<td>(0.019)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>5</td>
<td>0.029+</td>
<td>0.04*</td>
<td>0.035+</td>
<td>0.091***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.015)</td>
<td>(0.02)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>6</td>
<td>0.032*</td>
<td>0.046**</td>
<td>0.043*</td>
<td>0.11***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.017)</td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>7</td>
<td>0.015</td>
<td>0.043*</td>
<td>0.04+</td>
<td>0.115***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>R²</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Notes: This table shows point estimates and standard errors for nonparametric differences-in-differences estimator. Model accounts for district fixed effects ($\theta_d$), FLE-by-year fixed effects ($\delta_{tq}$), and state-by-FLE linear time trends ($\psi_{sq}$). FLE quartiles are indexed by FLE $\in 1, 2, 3, 4$. Column “FLE Cont.” Corresponds to models in which the additional control variable year-by-FLE percentile ($\delta_{tQ}$) 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.

$+p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001.$

The unweighted graduation results in table 4 are similar to the weighted results. One key difference is that the point estimates tend to be smaller in FLE quartile 4. For example, graduation rates are 6.8 percentage points higher in year 7, compared with 11.5 percentage points for weighted models. This reduction in effect size for unweighted results applies only to FLE quartile 4.\(^{15}\)

Model Sensitivity

Here we outline alternative model specifications. These alternatives provide checks on the exogeneity assumptions of the differences-in-differences model, as well as provide upper and lower bounds for how much point estimates depart from the preferred model. We describe the alternative model specifications below.

Alternative Model Specifications: Secular Time Trends

Replacing the secular trend $\delta_{tq}$ with $\delta_t$ models the counterfactual time trend as the average trend among all never-treated districts and those districts that are awaiting treatment in a given time period. Specifying the model with $\delta_t$ corresponds to a more traditional differences-in-differences specification of secular time trends in panel data models, allowing for comparisons of treatment effect estimates across the FLE poverty quartiles. The assumption of this model is that there is a common counterfactual trend for high- and low-poverty districts.

\(^{15}\) See online Appendix B.1 for an extended discussion of how to interpret differences between weighted and OLS estimates.
Alternative Model Specifications: Correlated Random Trend

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

Alternative Model Specifications: Cross-Sectional Dependence

We estimate models in which we assume our error term has a factor structure denoted by $\lambda'_s F_t$. Following Bai (2009), we define $\lambda_s$ 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 in the model. For our sensitivity tests, we estimate models in which the number of included factors $r$ is an element of the set $\{1, 2, 3\}$.

To estimate the $\lambda'_s F_t$ factor structure in equation 2, we use the method of principal components as described by Bai (2009) and implemented by Moon and Weidner (2015).

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

Model Sensitivity: Results

The preferred model, referred to as the benchmark model, indicates a meaningful positive and significant effect of court order on the outcomes of interest. Results from this model assume the benchmark model makes correct assumptions about secular trends, correlated random trends, and cross-sectional dependence. To examine model sensitivity, we focus attention on districts in the highest poverty quartile (i.e., FLE quartile 4) and assess the extent to which results are sensitive to these modeling choices.

Figures 3 and 4 plot the WLS causal effect estimates of finance reform on the logarithm of real per-pupil revenues and graduation rates, respectively, across a variety of model specifications. We display corresponding figures featuring OLS estimates and report regression tables with point estimates and standard errors in online Appendix B.3. Although not all possible combinations are plotted, the thirteen models shown in the figures do provide insight into sensitivity of the causal effect estimates. In both figures, our benchmark differences-in-differences model corresponds to the second model in

---

16. Moon and Weidner (2015) show that point estimates stabilize once the number of factors $r$ equals the true number of factors $r^*$ and that when $r > r^*$, 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 and Horenstein (2013).

17. We describe the basic steps of the procedure in online Appendix C.
Christopher A. Candelaria and Kenneth A. Shores

Notes:
For the benchmark model plot, range caps show 90 percent confidence intervals around the point estimates. Across both plots, the benchmark model is indicated by a thick, black dashed line. The legend corresponds to the sensitivity models plot and shows three perturbations of the benchmark model, delimited by “;”. ST denotes the type of nonparametric secular trend: (a) free lunch eligible (FLE) quartile by year fixed effects or (b) year fixed effects. CRT denotes the type of correlated random trends: (a) none, corresponding to no CRT; (b) state by FLE quartile fixed effects interacted with linear time; (c) state fixed effects interacted with linear time; (d) district fixed effects interacted with linear time. CSD denotes type of cross-sectional dependence adjustment: (a) 0, which implies no factor structure; (b) 1, which is a 1-factor model; (c) 2, which is a 2-factor model; or (d) 3, which is a 3-factor model. Bracketed numbers indicate the column location of point estimates for model m in online Appendix tables B.2–B.5.

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

the legend, which is denoted by a thick, dashed line with a hollow diamond marker symbol and the following triple:

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

ST refers to the type of secular trend, which can be either FLE-by-year or year fixed effects. CRT refers to the type of correlated random trend in the model, which can include no trend or one of the following trends: state, state by FLE quartile, or district-specific trends, along with the quadratic form of state by FLE quartile. CSD refers to the number of factors included to account for cross-sectional dependence. A model with factor number 0 does not account for cross-sectional dependence, whereas models with 1, 2, or 3 account for models with 1, 2, and 3 factors, respectively.18

Results from figure 3 reveal that, on average, the benchmark model closely resembles the causal effects on revenues in FLE quartile from other specifications. Among the WLS models, the benchmark results are lower, except for models that include quadratic forms of the correlated random trends or three factors. The largest effects are produced by a specification that includes year fixed effects for the secular trend, a correlated

18. Bracketed numbers indicate the column location for those model estimates available in online Appendix B.3.
random trend at the level of the state, and does not adjust for cross-sectional dependence (indicated in figure 3 by gray hollow diamonds with a short dash line).

It may be illuminating to compare gray and black hollow diamonds in figure 3, as these estimates correspond to models that ignore correlated random trends and CSD but differ in how they estimate the counterfactual trends. Models corresponding to hollow gray diamonds assume the counterfactual trend is homogeneous, and models corresponding to hollow black diamonds assume the counterfactual trend is heterogeneous, indexed by poverty quartile. Hollow gray diamonds are consistently larger than hollow black diamond point estimates. This difference indicates that the assumption of a homogeneous counterfactual trend ($\delta_t$) overestimates the treatment effect for high-poverty districts, since revenues were increasing faster in high-poverty districts in nontreated states relative to low-poverty districts in nontreated states.

Figure 4 reports the variability of effect sizes for graduation rate estimates. Among the WLS estimates, the benchmark model overstates the causal effects on graduation rates in FLE quartile 4 relative to all other graduation rate models. Of particular interest is the influence of correlated random trends. In general, the point estimates for graduation rates cluster together, within the range of 6 to 12 percentage points at treatment year 7, including models that account for up to three factors and allow for the unit-specific
time trend to be quadratic. Only when the correlated random trend is excluded entirely do point estimates drift downward so that they are small and insignificant (indicated by gray or black hollow diamonds and a dotted line).

By looking at the level of aggregation of the correlated random trend, it is possible to learn about the nature of the pretreatment trends in the dependent variable. Looking only at models that adjust for year by FLE effects (i.e., black markers and lines), point estimates for models that include a state-level time trend (indicated by the hollow diamond with short dashed line) are significant and nearer to our preferred model but still differ by about 5 percentage points. By including district-by-year effects (approximately 9,800 linear time trends, indicated by the hollow diamond and dashed line), the point estimates nearly coincide with the preferred model. This suggests that treatment and control groups do have different pretreatment trends, and these differences largely occur at the substate level. Finally, note that allowing the time trend to be quadratic (black diamond and solid line) has negligible effect on the model. Stability in these point estimates suggests that misSpecifying the functional form has not introduced bias.

There is little to no evidence of cross-sectional dependence in the timing of these Adequacy-era court cases. Recall that CSD can occur if there are heterogeneous responses to common shocks or if there is interdependence between units. When factor variables are included (with factors equal to 1, 2 or 3), point estimates are slightly attenuated relative to the benchmark model; however, it is important to emphasize that the true number of factors, \( r \), is not known. Unfortunately, due to finite sample bias, we cannot include additional factors because the results become highly unstable.\(^\text{19}\)

Overall, the benchmark model tends to understate effect sizes for real per-pupil revenues, while for discussion, estimates for graduation rates are sensitive to the exclusion of correlated random trends. Recall that a pretreatment trend was observed for graduation rates in FLE quartile 4, which motivated the inclusion of a correlated random trend. To assess whether the trend was properly modeled, we included trends at different levels of aggregation and with different functional forms. Point estimates are largely insensitive to changes in these domains, and the variation in effect size around our benchmark estimates is less than 5 percentage points. While there is evidence that treatment and control units (especially in high-poverty districts) have different secular trends, it is possible to estimate consistent point estimates once the pretreatment trends are modeled with the correct functional form.

### Equalizing Effects

Equation 2 estimates levels of change in revenues and graduation rates, comparing high- (low-) poverty districts in treated states to high- (low-) poverty districts in non-treated states. Comparing point estimates across poverty quartiles is problematic because the counterfactual trends are estimated for each poverty quartile. Assuming a homogeneous counterfactual trend (\( \delta_i \)) allows for within-state comparisons but, as shown in figure 3, this assumption belies differences in secular trends across the poverty quartile. To test whether revenues and graduation rates increased more in high-poverty districts relative to low-poverty districts following court order in a way that allows for heterogeneity among poverty groups, we construct a variable that ranks

---

\(^{\text{19}}\) See Moon and Weidner (2015) for discussion.
districts within a state based on the percentage of students qualifying for free lunch status in 1989 (prior to the first reform). This ranking is then converted into a percentile by dividing by the total number of districts in that state. Compared with percentages qualifying for free lunch, these rank-orderings put districts on a common metric, and are analogous to FLE quartiles, but with a continuous quantile rank-ordering.

The model that is estimated is analogous to equation 2 with two changes:

(i) The variable \(\delta_t Q\) is added, 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 100; and

(ii) The treatment indicators and corresponding coefficients \(\sum_{n=1}^{19} \gamma_{(q,n)}[1(Q_d = q) \times 1(t - t^*_{s,n} = n) \times D_s]\) are now set equal to \(\sum_{n=1}^{19} \gamma_{(q,n)}[Q \times 1(t - t^*_{s,n} = n) \times D_s]\).

Item 1 provides a counterfactual trend with respect to how much nontreated states are “equalizing” \(Y_{sqdt}\) with respect to \(Q\). The secular trend in these models now adjusts for the rate that revenues and graduation rates are changing across FLE quantiles among untreated districts as well as the FLE specific average annual trend \((\delta_{tq})\).

For item 2, the interpretation of the point estimates on the treatment year indicators is the marginal change in the outcome variable \(Y_{sqdt}\) given a one-unit change in FLE quantile within a state. For revenues, a point estimate of 0.0001 is equivalent to a 0.01 percent change in per-pupil total revenues for each one-unit rank-order increase in FLE status within a state. For graduation rates, a point estimate of 0.0001 is equivalent to a 0.01 percentage point increase for each one-unit rank-order increase. A positive coefficient indicates more revenues and graduation rates are going to poorer districts within a state. Equation 2 is estimated for both WLS and OLS, with the modifications just described. These results can be seen for both dependent variables in tables 3 and 4. The columns of interest are columns 5 and 10, which are labeled FLE Continuous.

After court-ordered reform, revenues increased across poverty quantiles, as indicated by the positive slope coefficients in table 3. Seven years after reform, a 10-unit increase in FLE percentile is associated with a 1.5 percent increase in per-pupil revenues for the weighted regression. For the unweighted regression, a 10-unit increase is associated with a 1.8 percent increase. As discussed in online Appendix B.1, neither the weighted nor unweighted model results dominate each other, so we can view the slope coefficient as having a lower bound of 1.5 percent and an upper bound of 1.8 percent. Assuming the treatment is linear, these results suggest that districts in the 90th percentile would have had per-pupil revenues that were between 12.32 to 14.56 percent higher than districts in the 10th percentile.

Table 4 also shows that court-ordered reform increased graduation rates across the FLE distribution. Seven years after reform, a 10-unit increase in FLE percentile is associated with a 1.05 percentage point increase in graduation rates for the weighted regression. For the unweighted model the corresponding point estimate is 0.77 percentage points, a difference in effect size commensurate with previously discussed results.

20. Models that include only \(\delta_t\) instead of \(\delta_{tq}\), not shown, produce nearly identical results.
Assuming linearity, districts in the 90th percentile would have had graduation rates that were between 6.16 and 8.4 percentage points higher than districts in the 10th percentile.

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

**The Effect of Revenues on Graduation Rates**

Whether money matters for academic outcomes is an old and highly debated question (Coleman et al. 1966). While a substantial number of past studies do not find strong evidence that money improves academic outcomes, on average (Hanushek 2003), the latest research suggests that a relationship exists when there is a financial shock to the school’s funding system and if that shock has a larger effect on subpopulations of individuals (Jackson, Johnson, and Persico 2016; Lafortune, Rothstein, and Schanzenbach 2018). We address whether money matters by using the shocks of court-ordered finance reform as instruments for real revenues per pupil, allowing the identification of the causal effect of money on graduation rates for specific poverty quartiles and across the poverty distribution. Although Jackson, Johnson, and Persico (2016) perform a similar analysis, we exclusively leverage reforms from the Adequacy era to provide exogenous variation in revenues, and we disaggregate the causal effects of these reforms into poverty quartiles and quantiles. Specifically, we interact the endogenous regressor per-pupil log revenues with each poverty quartile, which provides two-stage least squares estimates for the marginal effect of a log dollar change in revenues on graduation rates for each poverty quartile.21

Table 5 presents results from the two-stage least squares regressions. There are two panels: panel A displays weighted results and panel B displays unweighted results; for simplicity, we focus the narrative on the weighted results. Each panel has a total of six columns. Column 1 is a two-stage least squares regression in which log real revenues per pupil are instrumented by the quartile-specific indicators that reflect the number of years after a court-ordered finance reform has taken place. Columns 2 through 5 are separate FLE-quartile regressions, where quartile-specific log real revenues per pupil are instrumented by the corresponding quartile-specific treatment year indicators. Finally, column 6 is a model that instruments log real revenues per pupil with the years-after-court-order binary indicators interacted with a continuous within-state poverty quantile rank-ordering variable. The first-stage regression equations are reflected in equation 2, with the continuous model adaptations for column 6. Standard errors that are clustered at the state level are in parentheses, but errors that are clustered at the district level are provided in brackets for reference.22

In the specification estimating the average causal effect for all districts (column 1), a 10 percent increase in revenues per pupil causes a 1.97 percentage point increase in

---

21. Additional details about the two-stage least squares estimation process can be found in online Appendix B.4.
22. Jackson, Johnson, and Persico (2016) cluster at the district level for their two-stage least squares models.
### Table 5. Two-Stage Least Squares Results

<table>
<thead>
<tr>
<th></th>
<th>A: Weighted</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full</td>
<td>FLE 1</td>
<td>FLE 2</td>
<td>FLE 3</td>
<td>FLE 4</td>
<td>Cont.</td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>0.197</td>
<td>0.0725</td>
<td>0.127</td>
<td>0.123</td>
<td>0.506</td>
<td>0.359</td>
<td></td>
</tr>
<tr>
<td>(0.162)</td>
<td>(0.200)</td>
<td>(0.133)</td>
<td>(0.286)</td>
<td>(0.293)</td>
<td>(0.197)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st-stage F-statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster on state</td>
<td>–</td>
<td>156.42</td>
<td>281.69</td>
<td>249.03</td>
<td>219.91</td>
<td>70.48</td>
<td></td>
</tr>
<tr>
<td>Cluster on district</td>
<td>12.41</td>
<td>12.54</td>
<td>17.49</td>
<td>10.57</td>
<td>9.01</td>
<td>19.55</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>188752</td>
<td>45178</td>
<td>48675</td>
<td>49052</td>
<td>45847</td>
<td>188752</td>
<td></td>
</tr>
</tbody>
</table>

|                | B: Unweighted                |          |          |          |          |          |          |
|                | Full | FLE 1 | FLE 2 | FLE 3 | FLE 4 | Cont. |
| Full           | 0.278 | 0.148 | 0.265 | 0.306 | 0.379 | 0.347   |
| (0.112)        | (0.103) | (0.122) | (0.141) | (0.168) | (0.133) |          |
| 1st-stage F-statistics |          |          |          |          |          |          |
| Cluster on state | –    | 101.66 | 439.84 | 140.27 | 520.96 | 454.86 |
| Cluster on district | 17.69 | 16.85 | 21.76 | 16.67 | 15.44 | 39.80   |
| Observations   | 188752 | 45178 | 48675 | 49052 | 45847 | 188752 |

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

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

All models account for district fixed effects ($\theta_d$), free-lunch eligible (FLE)-by-year fixed effects ($\delta_{tq}$), and state-by-FLE linear time trends ($\psi_{stq}$). Model 6 adds a year-by-FLE percentile ($\delta_{tQ}$) control variable.

Standard errors in parentheses are clustered at the state level; standard errors in square brackets are clustered at the district level.

When clustering errors at the state level, FLE quartile coefficients are not statistically different from each other in either ordinary least squares (OLS) or weighted least squares (WLS) models. However, when clustering at the district level and estimating via WLS, FLE quartile 4 is statistically different from quartiles 1 and 2 at the 5 percent level and quartile 3 at the 10 percent level. When estimating via OLS, quartile 4 is statistically different from quartile 1 at the 1 percent level.

$^+p < 0.10; ^*p < 0.05; ^**p < 0.01; ^***p < 0.001$.

Although our results suggest that school resources affect graduation rates, the primary threat to validity is whether changes in graduation rates following court-ordered finance reform can be wholly attributable to changes in school spending. For the highest-poverty districts, our results suggest that a 10 percent increase in revenues per pupil causes a 5.06 percentage point increase in graduation rates, which is significant at the 10 percent level for errors clustered at the state and 1 percent level for errors clustered at the district. There is no evidence of an effect for FLE quartiles 1 through 3, even when clustering at the district level.
exclusion restriction to hold, we must assume that the court reforms affect graduation rates only through their effect on spending. This assumption is violated in cases where court-ordered finance affects other unobserved policy changes that also affect graduation rates. For example, in addition to increasing spending in high-poverty districts, the state may also adopt an incentive policy to bring higher-quality teachers to more impoverished districts. In such a scenario, graduation rates resulting from court order are not separable from the change in spending and the unobserved programmatic change.

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

**State-by-State Heterogeneity**

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

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

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

---

23. This design was first motivated by Hoxby (2001), who used a simulated instruments methodology.
To estimate state-specific treatment effects, we modify our benchmark model, equation \(2\) as follows:

\[
Y_{sqdt} = \theta_d + \delta_{qt} + \psi_{sq} t + \xi_{(s,q)} \left[ 1(\text{State} = s) \times 1(Q_{sd} = q) \times \text{YearsAfter}_{st} \times D_s \right] \\
+ \sum_{n=8}^{19} \gamma_{(s,q,n)} \left[ 1(\text{State} = s) \times 1(Q_{sd} = q) \times 1(t - t^*_s = n) \times D_s \right] + \varepsilon_{sqdt}, \tag{3}
\]

which estimates state-specific slopes for treatment years 1 through 7 and nonparametric treatment effects for the remainder of the treatment period. The variable \(\text{YearsAfter}_{st}\) takes value 0 in all years before a court-ordered finance reform in state \(s\) and values 1 through 7 in the first seven years after reform. Estimating state-specific slopes instead of nonparametric year effects simplifies presentation and reduces noise; limiting state-specific slopes to the first seven years following reform fixes each state-specific effect to a common response period. Note that for each state, the remainder of the treatment period is estimated nonparametrically. Because the number of variables estimated exceeds the number of states, we now cluster standard errors at the district level.

Previous research has been conducted for three states with an Adequacy ruling in this period. For Massachusetts, Guryan (2001) gains identification from discontinuities in the Massachusetts funding formula and finds that state aid resulting from the 1993 Massachusetts Education Reform Act did increase per-pupil spending, and this increase in spending improved fourth-grade test scores but not eighth-grade test scores. For Kentucky, Clark (2003) compares changes in the correlation between spending and district income after Kentucky’s 1989 Rose case against changes in the same correlation for districts in Tennessee. Relative to Tennessee, the correlation between spending and district income decreased in Kentucky following reform—meaning that spending became more equal—but the effects on student achievement were minimal. Finally, in New Jersey, Resch (2008) finds the effect of the 1997 Abbott case increased revenues in Abbott districts relative to other, high-poverty districts within New Jersey, but that these changes in spending improved test scores for only subgroups of students.

24. Because Montana was removed from our sample due to data quality issues, we obtain state-specific estimates for twelve states.

25. Linear slope models are analogous to comparative interrupted time series, in which the treatment slopes are estimated net of pretreatment trends (estimated via \(\psi_{sq} t\) and quartile-year fixed effects). A linear approximation is preferred in this case to improve precision and simplify presentation. In online Appendix B.5, we display results for models in which the full post-treatment period for each treated state-by-poverty quartile is estimated nonparametrically. Results from these nonparametric models are nearly identical to the linear specification, but with greater imprecision. Additional analysis of the state-specific responses to court order are provided to further validate these results.

26. The analytic strategy we use here makes comparisons to these studies difficult. Kentucky was treated so early that we lack pretreatment information about the state: for this reason, we do not include it here. Moreover, in Clark’s study of Kentucky, Tennessee serves as the counterfactual, but Tennessee had its Adequacy case in 1995, and we find evidence that Tennessee, following its reform, spent less revenue in high-poverty districts, net of its pre-reform trend, which would exaggerate a Kentucky-specific effect. For New Jersey, our treatment pool includes all poor districts within the state, whereas Resch compares high-poverty non-Abbott districts in New Jersey to the Abbott districts; therefore, any positive finding identified by Resch could be attenuated when we pool treated and control together. For Massachusetts, the identification strategy we pursue is not well equipped for a state that underwent treatment so early in the analytic sample. As shown in figure B.12 in online...
Notes: Bars correspond to point estimates from an augmented version of the benchmark model (i.e., differences-in-differences with treatment effects estimated nonparametrically after reform, accounting for district fixed effects \((\theta_d)\) FLE-by-year fixed effects \((\delta_{tq})\) and state-by-FLE linear time trends \((\psi_{stq})\) in which 12 treatment state indicator variables \((\Sigma_{i=1}^{12})\) are interacted with FLE quartiles and a linear treatment slope for years 1 to 7. This model provides estimates of treatment effects for 12 treated states, 4 quartiles, as a function of time following court order.

**Figure 5.** Treatment Heterogeneity Between States (Free-Lunch Eligible [FLE] Quartile 4 Estimates)

Figure 5 displays estimates of the slope parameters of interest, \(\xi(s,q)\), from WLS and OLS models for the dependent variables log per-pupil revenues and graduation rates for FLE quartile 4 districts in each of the twelve treated states.\(^{27}\) The top-left panel provides state-specific results for which the effect of court order on revenues and graduation rates is both positive; the top-right panel provides results for which the effect is both negative; the bottom-left panel provides results for which the effect on revenues is positive and graduation rates is negative; the bottom-right panel provides results for which the effect on revenues is negative and graduation rates is positive.\(^{28}\)

Note, first, that most states that had an Adequacy ruling during this period increased both revenues and graduation rates.\(^{29}\) The fact that most state-specific responses to court order correspond to the estimated average effect provides support for the benchmark model. Following court order, Tennessee lost both revenues and graduation rates. As discussed in online Appendix B.5, spending in Tennessee’s poorest districts was increasing rapidly prior to court order; net of that prior trend, spending dropped in Tennessee following court order.

The results for Massachusetts, New Hampshire, North Carolina, and Wyoming require further investigation. In Wyoming and New Hampshire, court order resulted in

---

**Appendix B.5.** Identification is strongly influenced by controlling for the pretreatment trend, and for Massachusetts, this trend is identified from two years of data.

\(^{27}\) In online Appendix B.5, we provide point estimates and standard errors for each of the values displayed. Most estimates are statistically significant, and we easily reject the null hypothesis that state-specific slopes are equal; however, estimates that are imprecisely estimated should be cautiously interpreted.

\(^{28}\) WLS and OLS results for each state are nearly identical, except for graduation rates in Tennessee (although the point estimate is not distinguishable from zero in both the WLS and OLS cases) and revenues in North Carolina (where the point estimate is slightly negative in the WLS case and slightly positive in the OLS case but not distinguishable from zero in both cases). For this reason, we focus on WLS in the discussion.

\(^{29}\) Revenues estimates for Arkansas and New Jersey are not statistically significant.
a positive change on revenues ($p < 0.001$) without affecting graduation rates. In Massachusetts and North Carolina, conversely, graduation rates increased (NC $p < 0.05$; MA $p > 0.1$) despite a loss (or nonincrease) in revenues. These results provide evidence that the relationship between spending and graduation rates is not moncausal or reducible to the adopted funding formula. Massachusetts, New Hampshire, and Wyoming all adopted minimum foundation plans, but so did every other state except New York and North Carolina. North Carolina’s adoption of an equalization plan and flat grant matches the plans adopted by New York, but spending increased in New York whereas spending did not in North Carolina. Therefore, while most states that received additional funds were able to increase high school graduation rates, there are some states for which money was not needed to produce academic gains and others for which money was not sufficient.

**Robustness Checks**

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

To test for selective migration, we estimate the benchmark model on four dependent variables: 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, and percent of students receiving special education. We describe results here for FLE quartile 4, and in online Appendix B.6 we present tables of results for FLE quartiles 1 through 4 and the continuous model.30 Overall, there is no evidence of selective migration to treated districts in FLE quartile 4. None of the point estimates for percent minority are statistically significant, nor are they large in magnitude. We also find no evidence of changes in the percentage of children in poverty and those in special education.31

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 and Solon 2011; Solon, Haider, and Wooldridge 2015). Although transformations of the dependent variable do affect the interpretation of the marginal effect, results should be similar in terms of significance and sign. We replicate results using total per-pupil revenues instead of log revenues, and results for WLS and OLS regressions are significant in all treatment years, which matches the pattern of significance of the main revenues results in table 3. Moreover, all point estimates are positive across both tables.

31. In online Appendix tables B.7, B.8, and B.9, we do find that the composition of minority students decreased and that the percentage of special education students increased in lower-poverty districts (i.e., FLE quartiles 1 to 3). However, when we look across the district poverty distribution to examine whether there are any changes in the minority composition, and percentage of special education for an increase in the poverty percentile rank, we find no statistically significant effects.
6. CONCLUSION

We highlight four contributions of this paper. First, this paper tests whether Adequacy-based cases increased revenues and graduation rates above and beyond the secular changes taking place in high-poverty districts across the country, as other states have made internal decisions to change funding formulas without the need of court intervention (Jackson, Johnson, and Persico 2016; Lafortune, Rothstein, and Schanzenbach 2018). We find consistent evidence that these court cases did have positive effects on these two outcomes. High-poverty districts in states undergoing reform increased revenues and graduation rates relative to high-poverty districts in nontreated states; in addition, these effects were relatively more equalizing compared to trends taking place in other states across the United States. The academic outcome we evaluate—student graduation rates—is the most distal outcome available for students experiencing changes brought about by court order during this period, and for this reason the evaluation is a timely appraisal of the effect of additional spending on student academic outcomes.

Second, the paper provides robustness analysis that is relevant to key recent findings in the school finance literature (Jackson, Johnson, and Persico 2016; Lafortune, Rothstein, and Schanzenbach 2018). We demonstrate that the effects of court order are quite robust to model specification. When we alter specifications of secular trends, correlated random trends, and cross-sectional dependence, estimates do not dramatically change.

Third, we can use our results to benchmark the studies by Jackson, Johnson, and Persico (2016) and Lafortune, Rothstein, and Schanzenbach (2018). Using court cases that predominantly predate the Adequacy era, Jackson, Johnson and Persico find that a 10 percent increase in per-pupil spending in all twelve school-age years increases high school graduation rates by 9.8 percentage points for low-income children. Our two-stage least squares results are smaller, at 5 percentage points for FLE quartile 4. This may be because high-poverty districts were already treated in some capacity prior to 1990, thereby attenuating subsequent effects. Alternatively, if nontreated states increased efforts to promote graduation in high-poverty districts in the absence of court order, results would likewise be attenuated.

Lafortune, Rothstein, and Schanzenbach (2018) include court cases during the Adequacy era (but expand the set to include legislative and non-Adequacy cases) and find that in districts with below-average income (relative to the state), ten years of school-age exposure increases achievement 0.1 standard deviations. The most comparable results we produce are from the reduced form versions of the continuous models. Seven years after court order, a 50-percentile point increase in district poverty equates to a 5-percentage point increase in graduation. Although it is difficult to benchmark these different outcomes, the results are complementary.

Fourth, this paper highlights two aspects of heterogeneity in response to Adequacy rulings. The first is that the marginal effect of an additional dollar of spending is more productive in poor districts than it is in nonpoor districts. Point estimates from table 5 for FLE quartile 4 are between 2.3 to 5.1 times larger than those for FLE quartile 1, and point estimates for FLE quartile 1 are small and imprecise. This suggests that there may be upper limits to spending, but these limits were not reached in high-poverty districts in the United States during this period. The second is that state responses
Court-Ordered Finance Reforms
to court order led to different results, and these differences are not solely attributable to funding formula. Specifically, future research should address how states with little to no observed change in resources, such as North Carolina and Massachusetts (and to a lesser extent Arkansas and New Jersey), were able to improve graduation rates, whereas states like New Hampshire and Wyoming increased revenues but failed to translate those resources into improvements in high school completion.

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

ACKNOWLEDGMENTS
Both authors acknowledge generous support from the Institute for Educational Sciences (grant #R305B090016) in funding this work. Candelaria thanks the Karr Family and Shores thanks the National Academy of Education for additional funding support. The authors especially thank Tom Dee, Matthieu Gomez, Susanna Loeb, C. Kirabo Jackson, Sean Reardon, Jeffrey Smith, Martin Weidner, Justin Wolfers, and two anonymous reviewers for helpful comments and suggestions. All errors are our own.

REFERENCES


Correia, Sergio. 2014. reghdfe: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Statistical Software Components (5457874). Boston College Department of Economics.


