Spending More on the Poor? A Comprehensive Summary of State-Specific Responses to School Finance Reforms from 1990–2014

Kenneth A. Shores* Christopher A. Candelaria† Sarah E. Kabourek‡

* University of Delaware

† Vanderbilt University

‡ NORC at the University of Chicago

Draft Date: July 1, 2021

Author Note

For correspondence, please contact co-first authors Shores and Candelaria at kshores@udel.edu and chris.candelaria@vanderbilt.edu. We thank Dale Ballou, Eric Brunner, Elizabeth Cascio, Sean Corcoran, Thomas Downes, Nora Gordon, Gary Henry, Jesse Rothstein, and Diane Schanzenbach for valuable feedback and suggestions. We also thank Eli Ben-Michael for assistance with the `augsynth` program, which implements the ridge augmented synthetic control method.

Abstract

Sixty-seven school finance reforms (SFRs), a combination of court-ordered and legislative reforms, have taken place since 1990; however, there is little empirical evidence on the heterogeneity of SFR effects. In this study, we estimate the effects of SFRs on revenues and expenditures between 1990 and 2014 for 26 states. We find that, on average, per pupil spending increased, especially in low-income districts relative to high-income districts. However, underlying these average effect estimates, the distribution of state-level effect sizes ranges from negative to positive—there is substantial heterogeneity. When predicting SFR impacts, we find that multiple state-level SFRs, union strength, and some funding formula components are positively associated with SFR effect sizes in low-income districts. We also show that, on average, states without SFRs adopted funding formula components and increased K-12 state revenues similarly to states with SFRs.

Keywords: School Finance Reforms, Augmented Synthetic Controls, Treatment Effect Heterogeneity

Spending More on the Poor? A Comprehensive Summary of State-Specific Responses to School Finance Reforms from 1990–2014

Since Kentucky's Supreme Court ruling in 1989 (*Rose v. Council for Better Education*), attempts to reform educational financing have focused on providing students with enough resources to reach an adequate or sufficient level of achievement necessary for labor market and democratic participation (Koski & Reich, 2006). Many of these reforms have been "litigation-prompted" (Joondeph, 1994), meaning that the reform was initiated by a lawsuit and subsequent court ruling. In other states, legislation was passed without a lawsuit or court ruling, and in some cases, statutes were passed to forestall a ruling against the state.[1] We define a school finance reform (SFR) as a state-level reform that resulted from a court ruling (Jackson, Johnson, & Persico, 2016; Lafortune, Rothstein, & Schanzenbach, 2018) or a documented legislative statute as identified by Lafortune et al. (2018). Defining an SFR to incorporate both court rulings and statutes allows us to cast as wide a net as possible, but, as we discuss below, this list is unlikely to be an exhaustive description of the ways in which states increase educational investments.

There has been a spate of these state-level SFRs since 1990: sixty-seven documented reforms, resulting from a combination of court-ordered rulings and legislative statutes, occurred in 27 states between 1990 and 2014. Recent studies have shown that, on average, SFRs increase spending in poorer districts (Candelaria & Shores, 2019; Jackson et al., 2016; Lafortune et al., 2018; Sims, 2011) and improve student outcomes, including graduation rates (Candelaria & Shores, 2019; Jackson et al., 2016), test scores (Lafortune et al., 2018) and adult earnings (Jackson et al., 2016). These analyses of SFRs, however, are likely to overlook important heterogeneity among states.

We expect heterogeneity in the impacts of SFRs across states because the history of SFRs is one of diversity. Limiting our focus to SFRs that took place in the "adequacy era," after 1989–90, we observe some states changing their finance systems through court order,

[1] Pennsylvania's Act 61 of 2000 is an example of this.

others via legislative statute, and still others in response to both court and legislative activity. Some states had a single SFR during this period; others had multiple. Some states responded to SFRs by changing their funding formula; others retained key structural features of the formula but changed its sub-components or weights. Finally, some states were sued because facilities were deemed inadequate, while others were sued because aggregate spending was inadequate.[2]

Understanding the variability of SFR effects across different contexts can be useful for policymakers. If SFRs exhibit substantial variability in terms of their effects on spending and resource allocation, then pursuing a reform can be a risky option, even if SFRs, on average, have positive effects. For instance, to accommodate spending increases in public education required by SFRs, state lawmakers may need to disrupt financial budgets by reallocating funds among public expenditure categories or raise taxes to accommodate spending increases in public education (Baicker & Gordon, 2006; Liscow, 2018). Importantly, knowing which factors predict SFR effect sizes could mitigate some of the uncertainty associated with reform outcomes and can guide policymakers if some of these factors are levers over which the state has some control.

In this paper, we examine state-level variation in the effect sizes of SFRs on school revenues and spending among low- and high-income districts within states, variation in the types of resources districts purchased, and whether characteristics of the state's policy environment (i.e., its adopted funding formula, whether it was sued multiple times, the number of times the funding formula was changed following the SFR, and teacher union strength) are predictive of SFR impacts. Of 27 states with an SFR during this period, only five—Kansas, Kentucky, Maryland, Massachusetts, and Vermont—have been evaluated. This small sample of state-level case studies provides an incomplete picture of the SFR heterogeneity landscape.

[2] Many papers provide an overview of SFRs in this period (see, for example, Corcoran & Evans, 2015; Jackson, 2018; Roch & Howard, 2008a; West & Peterson, 2007).

In terms of scope, an SFR is a state-level event resulting from either court-order or a prominent legislative statute.[3] Consequently, control states will be those that those that did not have a court order or documented legislative statute during this period. Although states without documented SFRs serve as counterfactuals in this analysis, we should not assume that their educational investments were static. Indeed, as we document below, states without SFRs behaved similarly to states with SFRs in terms of funding formula adoption and state contribution to K-12 education.

To determine whether an SFR changed the level and distribution of spending in a state, our analytic strategy involves three steps. First, we account for SFR effect-size variability across states by separately examining the bottom and top income terciles defined by district-level household income from the 1990 Decennial Census. To do this, we assign the typical level of resources across districts in the bottom tercile of the income distribution and across districts in the top tercile within each state using the weighted median of the resource. This approach accounts for heterogeneity of SFR impacts within states while also providing a way to assess the progressivity of reforms between terciles.[4]

Second, we estimate state income tercile impacts by using a de-biased synthetic controls approach, the "augmented" synthetic control method, developed by Ben-Michael, Feller, and Rothstein (2020). We use this method to match each state that had a documented SFR to a pool of states with no documented SFRs, matching solely on pre-SFR levels of either revenues or spending. Given that we are interested in assessing the progressivity of reforms within states, we set the matching procedure to generate a unique

[3] As we discuss in the Data section, we base our lists on Jackson et al. (2016) and Lafortune et al. (2018). Prominent legislative statutes appear in Lafortune et al. (2018). The full list of SFR events is contained in Appendix Table A1: Court-Ordered and Legislative School Finance Reforms.

[4] An alternative approach is to compare effect sizes between districts with small and large shares of free-lunch eligible (FLE) students, as was done by Candelaria and Shores (2019). FLE rates are a useful measure for tracking policy, as many states allocate categorical aid based on FLE counts of students, but these rates are less useful for measuring whether poor and non-poor districts received more revenues or spending following an SFR, since FLE rates only weakly correlate with Census-based measures of income (Domina et al., 2018). Because our interest is in estimating the progressivity of these reforms and not the direct channels through which aid is allocated, we prefer the Census-based estimates and use a district's median household income.

comparison group, known formally as a synthetic control group, for each state income tercile with an SFR.[5] For each of the 26 states that had documented SFRs between 1990 and 2014, we then compare changes in K-12 revenues or spending, before and after reform, to matched changes in K-12 revenues or spending in the synthetic control group.[6]

Third, after obtaining estimated effects for states with SFRs, we then conduct descriptive analyses by linking these effects to variables describing the SFR policy environment. Based on the school finance literature, we examine the following predictor variables: state's funding formula, whether the state changed its funding formula multiple times, whether the state experienced multiple SFRs, whether legislation was passed following a ruling from the state supreme court, and the presence and strength of the state's teacher unions.

The short version of our findings is as follows: summarizing all estimated responses to SFRs across states, revenues increased by 5 percent ($p = 0.01$) and expenditures increased by 7 percent ($p < 0.01$) in low-income districts and by 2 ($p = 0.07$) and 0 ($p = 0.93$) percent, respectively, in high-income districts. In general, low-income districts increased spending in greater amounts relative to high-income districts after reform, meaning that SFRs tend to have progressive effects on school resource allocation, as has been documented elsewhere. Across low-income districts, a 10 percent increase in total spending corresponds to a 4.8 percent increase in current instructional spending ($p = 0.05$), a 3.6 percent increase in total salary spending ($p = 0.09$), and a 28 percent increase in capital spending ($p = 0.01$).

At the same time, the heterogeneity of responses to SFRs may temper enthusiasm for them. To illustrate this point, we summarize three findings where state-level SFRs did not

[5] The synthetic control group is a weighted average of non-SFR comparison state outcomes—per pupil revenues or expenditures—over time. The weights are based on an optimization procedure and are time invariant. Specific details appear in Appendix F: Methodological Details.

[6] Although 27 states had SFRs during our period of study, we can only estimate the effects of 26 reforms as Kentucky's reform occurred at the beginning of our analysis sample with no pre-treatment data, making it impossible to identify a comparison group before and after reform. See details in the Research Methods section.

necessarily guarantee a positive outcome for poorer districts. Specifically, we find that (1) two of 26 SFRs during the period of study decreased revenues and spending ($p < 0.10$); (2) revenues and expenditures did not increase in 15 and 9 states (fail to reject at $p < 0.10$), respectively; and (3) the estimated effects for revenues and expenditures is less than or equal to zero in 10 and 5 states, respectively. Thus, while on average these outcomes had positive effects among low-income districts, a close examination of the distribution reveals variability—mostly positive, but also some negative and imprecisely estimated null effects—among the states.

To explain these null results, we show that trends in funding formula adoption and per pupil state investment in education were changing similarly in states with and without SFRs. Mechanisms explaining changes to state education finance systems outside of documented SFRs are incomplete. Prominent examples occurred in states like Michigan and Florida, where the K-12 funding system was overturned by referendum. In other states, subtler changes occurred but were nevertheless meaningful. While fully accounting for all non-traditional SFRs is beyond the scope of this paper, we provide descriptive evidence that states can change K-12 spending in ways that did not result in an SFR classification but nevertheless had SFR-like effects. We consider this as an area for future research.

Examining revenues and expenditures sub-categories also reveals potential limitations from SFRs. First, we find that SFRs primarily work by increasing state revenues with some potential substitution away from local revenues. For tercile 1 districts, state revenues increased by 11 percent ($p = 0.02$), and local revenues decreased by 7 percent ($p = 0.15$). Though state revenues were "sticky"—a 10 percent increase in state revenues is associated with an 3.5 percent increase in total revenues ($p < 0.01$)—we see some substitution away from local revenues, with local revenues decreasing 4 percent for every 10 percent increase in state revenues ($p = 0.15$). Second, our results show that low-income districts emphasized capital spending versus total district salaries and instructional spending. To the extent that students are less likely to benefit from new construction versus other priorities (e.g., as supported by Jackson, 2018), the ability of SFRs to close achievement and educational attainment gaps may be impaired.

With respect to predictors of SFR efficacy, having multiple SFRs and, consistent with Brunner, Hyman, and Ju (2020), teacher union strength positively predict SFR progressivity. Subsequent to the SFR, states that incorporated equalization plans, or included categorical aid or spending limits after the SFR tended to have more progressive outcomes as well. Many factors fail to explain effect size heterogeneity. For example, variation is not explained by whether whether the reform was induced by the courts or legislature, nor by the year in which the SFR took place.

Despite the prevalence of SFR activity, expectations that SFRs will have heterogeneous effects, and significance of this variability for disadvantaged students, a comprehensive evaluation of these reforms has not been conducted. Our study fills this gap and illustrates how contemporary methodological approaches can be used to evaluate treatment effect variation in settings where randomization is impossible. Thus, our study complements and informs recent discussions about the heterogeneity of ostensibly common policy shocks and contextual correlates that explain this heterogeneity (see, for example, McEachin, Domina, & Penner, 2020; Orr et al., 2019; Yeager et al., 2019).

The paper proceeds as follows: (1) Previous Literature; (2) Data; (3) Research Methods; (4) Results; and (5) Conclusion.

Previous Literature

To date, studies of the effects of SFRs on revenues, expenditures, and student outcomes have either yielded an average effect of SFRs, leveraging the differential timing of reforms across states (i.e., across-state average SFR effect studies) or case-study SFR effects, based on a single reform in a given state. Overall, recent across-state studies that leverage the differential timing of an SFR across states have consistently found positive relationships between spending increases and student outcomes (Candelaria & Shores, 2019; Hyman, 2017; Jackson et al., 2016; Lafortune et al., 2018). These findings contrast from earlier studies, which did not provide consistent causal evidence that education spending increases improved student outcomes (e.g., Burtless, 1997; Greenwald, Hedges, & Laine, 1996; Hanushek, 1997)

Among the across-state average effect studies, there has been limited attention to the mechanisms through which school resource shocks improve student outcomes. Jackson et al. (2016) find that states undergoing SFRs increased the number of new teachers hired per student, suggesting that smaller class sizes are driving results, a mechanism supported by prior literature (Chetty et al., 2011; Fredriksson, Öckert, & Oosterbeek, 2012; Krueger, 1999). One challenge to this interpretation of mechanisms is that many SFRs target capital expenditures, resulting in capital expenditure increases (Jackson et al., 2016).

While capital expenditures have no direct impact on class sizes, they may improve student outcomes by increasing the time students spend in schools, for example by encouraging greater attendance, a result supported by evidence from facilities investments in California and Connecticut that boosted student achievement while increasing attendance (Lafortune & Schönholzer, 2017; Neilson & Zimmerman, 2014). At the same time, the evidence of a causal relationship between capital spending, time in school, and achievement is mixed. Of the seven studies reporting causal effects of capital spending (as summarized by Jackson, 2018), three report null findings on achievement and two of those three also find no direct effect on student attendance.[7]

In addition to considering mechanisms, across-state average SFR effect studies have evaluated the extent to which SFRs affect local spending. At the county and district levels, most evidence suggests that increased intergovernmental revenues from the state also increase total local K-12 revenues and expenditures (i.e., the "flypaper effect"), but localities benefiting from state aid also modestly reduce local revenues contributions.

[7] Cellini, Ferreira, and Rothstein (2010) do not include attendance as an outcome measure; Goncalves (2015) and Martorell, Stange, and McFarlin Jr (2016) directly test for attendance effects from increases to capital spending and find no relationship.

Hoxby (2001) provides one of the first results in the context of SFRs and shows that SFRs reduced total local effort in some states, depending on their funding formulae characteristics, which resulted in a leveling down of total expenditures for education. More recently, both Baicker and Gordon (2006) and Liscow (2018) find that intergovernmental revenue transfers for education from the state to county increased after finance reforms and that county-level education expenditures also increased; however, the expenditure result is not statistically significant in Baicker and Gordon (2006). Similarly, but at the district level, Lafortune et al. (2018) find that SFRs increased state education revenues directed to districts, resulting in increased total education revenues and expenditures, but localities also reduced local education revenues (though this last result is not statistically significant). Card and Payne (2002) calculate elasticities directly and find that increases to state education revenues also increase total education spending, with elasticities ranging between 0.53 to 0.66, indicating that each additional dollar of state funding increases total education spending between 53 to 66 cents.

Among case-study SFR effect studies, only 5 (of 27) states have been evaluated during this period, and among these studies, there has been little attention to which resources were purchased and how those resources might differentially affect student outcomes. Researchers have evaluated Kansas' 1992 School District Finance and Quality Performance Act (Duncombe & Johnston, 2004; Johnston & Duncombe, 1998), Kentucky's 1990 Kentucky Education Reform Act (Clark, 2003), Maryland's 2002 Bridge to Excellence in Public Schools Act (Chung, 2015), Massachusetts' 1993 Education Reform Act (Dee & Levine, 2004; Guryan, 2001), and Vermont's 1997 Equal Educational Opportunity Act (Downes, 2004).[8] Among these, results range from small and short-term increases to

[8] Michigan's 1994 Proposal A has been studied by multiple authors (Chaudhary, 2009; Cullen & Loeb, 2004; Hyman, 2017; Papke, 2008; Roy, 2011). Following, (Lafortune et al., 2018), we exclude this case because it was not an SFR, but instead came to a vote at the state level and was approved by voters as an amendment to the state constitution. Evaluations of New Jersey's 1997 *Abbott* and New York's 2003 *Campaign for Fiscal Equity* rulings are available as unpublished conference proceedings and dissertations (see Resch (2008) and Atchison (2017), respectively).

spending with no improvement in student outcomes (Kansas), moderate increases in spending with little to modest improvements in student outcomes (Kentucky and Maryland), to substantial increases in both spending and academic outcomes (Massachusetts and Vermont).

The limited study of these reforms and the heterogeneity of results provide impetus for a comprehensive study across multiple states. Further, given the variability in linkages between resource gains and academic improvements (e.g., Jackson, 2018), it suggests that variation in the type of resources states pursue resulting from SFRs is important as well. Therefore, we evaluate the impacts of SFRs in multiple domains, including per pupil total revenues, state revenues, local revenues, total expenditures, total current instructional expenditures, total salaries, and total capital outlays.

Data

To understand how SFRs impacted resources and to understand how these impacts varied by state income terciles, our analysis requires a tabulation of SFRs and a time series of dependent variables (i.e., total expenditures and expenditure categories such as instructional and capital). To assess which variables are then predictive of SFR effect size variation, we compile a time series, when possible, of state-level variables theorized to be predictive of SFR progressivity.

Tabulation of School Finance Reforms

We compile a list of all major school finance reforms beginning in 1990 by leveraging recent lists compiled by Jackson et al. (2016) and Lafortune et al. (2018). In cases where there was a disagreement between our two sources, we privileged Lafortune and colleagues because they provide supplemental research on case histories, have a more recent list, and include legislative SFRs. We made two substantive changes to the cases provided by Lafortune and colleagues. First, resolutions of court cases and legislative enactments were recorded in calendar years, but these calendar years need not align with academic years; for example, an event occurring in December of 2012 would be recorded as 2012, but would likely apply to the Fall and Spring of academic year 2012–13. We gathered the months and years in which cases were resolved or bills signed into law, and converted these events into academic calendar time. Second, in a few instances, a state had a court ruling and legislative bill passed in the same fiscal year but, based on the month, the ruling and bill occurred in adjacent academic years. In these cases, we separated the combined events into two events occurring in subsequent years.

Appendix Table A1 lists the school finance reform events under consideration. While we tabulate all court cases and legislative bills in Table A1, we require at least four year of pre-SFR outcome data to employ the ridge augmented synthetic control method we describe in the Research Methods section. Thus, all cases that occurred before academic year 1992–93 are excluded (and appear in bold typeface in Table A1), but if a state had multiple SFRs, then we use the first SFR with at least four years of pre-treatment data. Effectively, because Kentucky had one SFR in 1989–90, it is excluded from our data. New Jersey, Tennessee, and Texas had multiple SFRs, and we use their first SFR beginning in 1992–93 to assign treatment time.

Dependent Variables

Our primary data source is the Local Education Agency Finance Survey (F-33), which is collected by the U.S. Census Bureau and is audited and distributed by the National Center for Education Statistics (NCES). From the F-33, we extract total revenues and total expenditures. We also obtain the following revenues and expenditure subcategories: state revenues, local revenues, instructional staff support services expenditures, capital outlays, and total salary expenditures. The panel data set of fiscal outcomes we assemble spans academic years 1989–90 to 2013–14.[9] In our analyses, we scale

[9] During schools years 1990–91, 1992–93, and 1993–94, the full universe of school districts were not surveyed and are not included in the NCES release of data; however, we were able to obtain district-level data from sampled districts directly from the U.S. Census Bureau.

these data by total district enrollment and all dollar values are in 2014 USD using the Fall to Spring academic calendar of the Consumer Price Index (Shores & Candelaria, 2019).[10]

To assess heterogeneity across the income distribution, we place districts into income terciles based on their state's 1989 median income levels, which are reported at the district level and comes from the 1990 U.S. Decennial Census. These income data precede all reforms under consideration in this study. Districts in the bottom tercile are the poorest in the state; districts in the top tercile, the richest. The state-specific terciles remain fixed throughout all analyses to help mitigate bias from potential Tiebout sorting induced by school finance reforms. For each state income tercile and year, we then compute the weighted median of our outcome variables of interest, where the weights are based on annual district enrollment.[11]

Using these tercile measures, we can examine the extent to which school finance reforms improved outcomes, on average, in the poorest districts in a state. Moreover, we can examine the extent to which reforms were progressive by seeing whether bottom-tercile districts benefited more from school finance reform relative to top-tercile districts in the same state for a given outcome.

Table 1 shows means of all financial variables (i.e., revenues and expenditures), which is the first year of our sample period, for control states that never had an SFR and for treated states that did have a reform. We also include demographic variables to further characterize our analytic sample.[12] We provide statistics separately for terciles 1 and 3, corresponding to lower-income and higher-income districts, respectively. Reported means

[10] For some districts, there is volatility in the per pupil financial data, which is induced by large fluctuations in enrollment. To address this volatility, we apply sample restrictions and adjustments that are similar to Lafortune et al. (2018). We discuss our approach in Appendix B: Additional Details about Dependent Variables.

[11] We use the weighted median as a measure of central tendency because it is not sensitive to outlier observations nor skew within each tercile; thus, it better characterizes the typical values of each within-tercile outcome.

[12] Information about the construction of these demographic variables appears in Appendix B: Additional Details about Dependent Variables.

are simple arithmetic averages of weighted medians for financial variables and weighted means for demographic enrollment variables; sample standard deviations are reported in parentheses. In columns 3 and 6, we test whether treated and control state means are statistically equal to each other. Overall, baseline revenues and expenditures are indistinguishable between SFR and non-SFR states in both tercile 1 and 3 districts. Tercile 1 and 3 districts in SFR states have fewer Black students than non-SFR states, and in tercile 1 districts, SFR states have more Hispanic students.

[Table 1 about here.]

Predictors of SFR Efficacy

To assesses whether SFR-related policy variables influence SFR progressivity, we identify key factors that characterize the school finance context in the reform states based on the school finance literature. Variability in funding formulae will determine how much aid is allocated to low-income districts; some formulae, for example, provide targeted aid based on student characteristics while others place limits on local revenues contributions (Card & Payne, 2002; Hoxby, 2001). We also examine whether the SFRs were induced by the courts or the legislature (Langer & Brace, 2005), whether the state was subjected to multiple court rulings, which would indicate the state's compliance with court mandates (Roch & Howard, 2008b), whether the state changed its funding formula multiple times, and the prevalence and strength of teacher labor unions (Brunner et al., 2020). To this end, we generate variables to indicate whether the courts or legislature induced the SFR and whether the SFR was the first in the state. Further, we generate a panel dataset of funding formula for each state and year for the period 1990 to 2014. Details for the predictors are in Appendix C: Generating Panel Dataset of Funding Formula and Appendix E: Additional Predictors.

Research Methods

To estimate effect sizes for each state that underwent an SFR, we use the ridge augmented synthetic control method (ASCM) developed by Ben-Michael et al. (2020). In what follows, we discuss how to construct counterfactuals in a case-study environment such as ours using differences in means, synthetic control methods (SCM), and ridge ASCM. Intuition for the approach we take is included in the main text; technical details are put in Appendix F: Methodological Details.

Building Intuition: Four Approaches to Estimating Unit-Specific Treatment Effects

Our goal is to estimate individual effect sizes—the average treatment effect on the treated (ATT)—for each state-by-district income tercile that experienced an SFR.[13] To build intuition, we consider four options: (1) pre- and post- contrast in spending for a given state; (2) pre- and post- contrast in spending for a given state relative to the mean of non-treated states; (3) pre- and post- contrast in spending for a given state relative to a synthetic control mean; and (4) pre- and post- contrast in spending for a given state relative to a de-biased synthetic control mean. We discuss Alaska and New Mexico's SFRs, which both occurred in 1999-2000, as examples.

In the left panel of Figure 1, in blue (solid) and maroon (dashed), tercile 1 log per pupil total expenditures are shown for Alaska and New Mexico, respectively, and, in gray, the same data are shown for all other states not experiencing an SFR. The states' 1999-00 court rulings are displayed as a vertical line. If we only compared Alaska and New Mexico's spending spending before and after 2000 (i.e., option 1), ignoring the non-SFR comparison states, it would appear as if Alaska experienced a modest increase in revenues

[13] In practice, we estimate dynamic ATT effects (i.e., time-varying average treatment on the treated effects in each year after an SFR) for each state-by-district income tercile; however, we use the average of these dynamic ATTs—again, at the state-by-district income tercile level—as a summary measure of the impact of an SFR. For simplicity, we refer to the average of the dynamic ATT effects as the ATT effect instead of the average of the dynamic ATT effects. The dynamic effects are available in Appendix Figure G1.

four years after its reform, and New Mexico's revenues did not increase relative to its pre-SFR trend. However, we worry that these changes (or non-changes) may reflect experiences common to other states, such as economic recessions or bubbles. Thus, it would be useful to have a comparison group.

[Figure 1 about here.]

Options (2) through (4) represent different approaches to constructing comparison groups and are shown in the right panel of Figure 1. The top-right panel shows the role of these three approaches for Alaska, and the bottom-right panel includes New Mexico. The short-dashed red line allows each of the non-treated states to serve equally as counterfactual, or control, states to Alaska and New Mexico (option 2). Here, we simply take the difference between Alaska and New Mexico's tercile 1 log per pupil total revenues from the mean tercile 1 log per pupil revenues of the non-treated states at each point in time and plot these differences over time. As is evident, these non-treated states do not mimic either state's pre-SFR revenues levels or trajectories: Alaska's log revenues are much higher than the revenues among all non-SFR states but trending down, whereas New Mexico's are lower with no obvious trends. These patterns suggest that the simple average of non-SFR states do not uniformly make for good counterfactuals for Alaska or New Mexico.

Options (3) and (4) use matching methods to identify states that most resemble Alaska's and New Mexico's pre-SFR levels of tercile 1 log per pupil total expenditures. These matching methods assign time-invariant weights of differing values to control states based on their pre-treatment match. Option (3), shown in long-dashed yellow, is the traditional SCM, and option (4), shown in solid green, is the ridge ASCM. Comparing Alaska and New Mexico illustrates the benefit of ASCM in some cases.

For Alaska, there are no non-SFR states matching Alaska's level of spending, so applying positive weights to those states cannot generate a plausible set of controls. Thus, as is shown in the top-right panel, the SCM (illustrated as the dashed yellow line) generates poor pre-treatment match quality. ASCM, on the other hand, does much better, as illustrated by the solid green line in the top-right panel. Intuitively, the ASCM can be framed as a bias-correcting approach that uses regression adjustment to reduce the deviation between the pre-treatment outcomes for a treated state and the pre-treatment outcomes for the comparison states that form the synthetic control group (Abadie, in press; Ben-Michael et al., 2020). Allowing the weights to be negative gives more flexibility for this adjustment, though the weights must still sum to one (Ben-Michael et al., 2020).

In contrast, the SCM and ASCM methods yield nearly identical results for New Mexico. This result is illustrated in the bottom-right panel, where the solid green line representing the ASCM effect sizes perfectly overlays the SCM effect sizes. In this case, New Mexico is surrounded by many non-treated comparison units that can be combined into a synthetic control unit to match its pre-treatment outcomes. Thus, in general, if there are available controls for the treated unit, the augmentation procedure of ASCM is less emphasized; it is when the available pool of control states is sparse (i.e., outside the convex hull) that augmentation becomes more important.

The consequences for these different approaches can be seen in the post-period beginning in 2000-01. Option (4) shows a comparable positive effect of the SFR for both Alaska and New Mexico at about 2 log points, whereas option (1) shows a much larger effect for Alaska and a much smaller effect for New Mexico. Technical details for the ASCM and how it is distinguished from SCM can be found in Appendix F: Methodological Details.

Results

Our results proceed as follows: (1) We first compute average treatment on the treated effects (ATTs) for terciles 1 and 3 in each state and use these averages to describe heterogeneity for multiple resource types. Then, (2) to better understand the SFR environment in which SFRs are productive, we leverage the tercile 1 and tercile 3 ATTs as outcome variables in prediction models with covariates derived from the SFR political landscape.

Effect Size Heterogeneity

To distinguish within study variance (i.e., sampling variability for a given state income tercile) from between study variance (i.e., how much true variation there is among state-specific ATTs), we take a meta-analysis approach to describing effect size heterogeneity. This meta-analytic approach is feasible since the within study variance is estimated with the row-based jackknife, meaning we can calculate the between study variation as:

$$\hat{\tau_j}^2 = \frac{Q - (K-1)}{\sum w_k - \frac{\sum w_k^2}{\sum w_k}},$$

where j indexes expenditure outcomes for an income tercile; k indexes the number of SFR states, and K is the total number of treated states; $Q = \sum w_k \left(\bar{\hat{\gamma}}_{1k}^{\text{aug}} - \sum w_k \bar{\hat{\gamma}}_{1k}^{\text{aug}} / \sum w_k \right)^2$; and $w = 1/\widehat{\text{var}}(\bar{\hat{\gamma}}_{1k}^{\text{aug}})$. In other words, $\hat{\tau}^2$ is the effect size variance weighted by the inverse of the sampling variability of each effect (DerSimonian & Laird, 1986; Petropoulou & Mavridis, 2017).

Because we generate synthetic controls from a common (and repeating) pool of states without documented SFRs, the effect sizes we generate for individual states undergoing SFRs may be correlated. The extent of these correlations is not known *ex ante*, since the weights that are generated for the control states are generated separately for each year an SFR occurs; thus, control states that are weighted heavily for one SFR state may not be weighted heavily for other SFR states. Nevertheless, we model the potential correlation in estimated effect sizes among SFR states with the robust variance estimation technique of Hedges, Tipton, and Johnson (2010). Robust variance estimation complements the inverse-variance weight of the meta-analysis with an estimate of the between-state effect size correlation, i.e., $w_{ij} = 1/\left\{ \left(v_{\cdot j} + \widehat{\text{var}}(\bar{\hat{\gamma}}_{1k}^{\text{aug}}) \right) \left[1 + (k_j - 1)\, \rho \right] \right\}$. The specific value of ρ is chosen by the researcher; we used a value of 0.8 but tried other values ranging from 0.5 to

0.9 in increments of 0.05 and our results were nearly unchanged.

State income tercile effect sizes for per pupil total revenues, and state and local revenues, total expenditures, and capital and salary expenditures are shown in Figures 2, 3, 4, and 5 below. We report results for terciles 1 and 3 for total revenues and total expenditures, in the first and second panels, respectively. For the sub-categories of revenues and expenditures, we emphasize tercile 1 outcomes and relegate tercile 3 outcomes to Appendix H: Additional Heterogeneity Results for Terciles 1 and 3.[14] For each outcome variable, states are sorted alphabetically. Next to each state abbreviation we include the pre-treatment logged value of the revenues or expenditures variable. The vertical dashed line and diamond shows the meta-analytic average of the ATTs from each SFR state, and the magnitude of this meta-analytic average is shown in the last row of the right column; each state's ATT and 90 percent confidence interval is shown in the right column. The estimated standard deviation of the true effect sizes τ and the p-value for the meta-analytic average of the ATTs are shown as figure notes. The displayed error bars indicate 90 percent confidence intervals; occasionally these intervals are long enough so as to distort the axis and are therefore displayed with a $\longrightarrow$ or $\longleftarrow$ to indicate the confidence interval exceeds the axis range.

Per Pupil Total Revenues. In Figure 2, the meta-analytic average of the ATTs for log total revenues in terciles 1 and 3 are 5 percent ($p = 0.02$) and 2 percent ($p = 0.07$), respectively, meaning that, on average, SFRs increased spending in bottom-income districts more than they did for top-income districts (results are nearly identical whether we include SFRs that took place during the Great Recession or not, i.e., those in Indiana, Pennsylvania, and Washington). However, this average belies considerable heterogeneity. For tercile 1 districts, the estimated standard deviation of the true effect sizes τ is 12 percent. Assuming that effect sizes are normally distributed, this means that 90 percent of

[14] See Appendix Figure H1 for state and local revenue effects for tercile 3 and Appendix Figure H3 for instructional, salary, and capital expenditures for tercile 3. Additionally, model performance and sensitivity results are available in Appendix G: Assessing Ridge ASCM Estimates.

the estimated effect sizes are expected to fall within a range of -15 and 25 percent (i.e., the meta-analytic average of the ATTs $\pm 1.645 \times \tau$). Estimated effect sizes range between a 12 percent reduction in total revenues (Idaho) and a 23 percent increase in total revenues (Wyoming). For tercile 3 districts, 90 percent of the estimated effect sizes are expected to fall within a range of -3 and 7 percent, with estimated effect sizes ranging from -6 percent (Arizona) to 15 percent (Wyoming). In total, at 10 percent significance, SFRs increased total revenues in 9 of 26 state-district bottom income terciles. Therefore, revenues among low-income districts in nearly two-thirds of states undergoing SFRs did not increase relative to synthetic counterfactuals, though we cannot rule out effect sizes that are either very large or negative, due to sampling variability. We discuss interpretation of noisy estimates in the Discussion section.

To characterize these estimates, we can benchmark them against the linear secular trend in log revenues among tercile 1 districts from 1990 to 2008 (before the Great Recession) and the decline in log revenues from 2009 to 2013. We estimate these trends directly using our district-by-year data that have not been collapsed to the tercile-by-year level. The linear secular trend in log revenues among tercile 1 districts is 0.03, and the estimated linear decline in log revenues among tercile 1 districts during the Great Recession period is 0.02. Thus, our estimates of the average effect of an SFR are between 1.67 and 2.5 times larger than the linear change over time and the change induced by the Great Recession.

[Figure 2 about here.]

Comparison to Earlier Case Studies. The results for total revenues described above align with the limited number of prior case studies conducted on state-specific SFRs. For Massachusetts, Maryland, and Vermont, the positive impact of their states' SFRs identified here (both in terms of log total revenues and log total expenditures) match work by Chung (2015, Maryland), Dee and Levine (2004) and Guryan (2001) (Massachusetts) and Downes (2004, Vermont). We find no impact from Kansas' SFR, which matches the two studies of Kansas from Johnston and Duncombe (1998) and Duncombe and Johnston (2004), who find a brief increase in relative equity among high- and low-income districts in Kansas followed by a return to baseline levels of equity. Our longitudinal analysis of Kansas' change in revenues relative to counterfactuals mirrors this pattern (see Appendix Figure G1). Overall, the fact that our results align with prior case-studies—studies that relied on different methodologies and counterfactuals—lends credibility to our analytic strategy.

Per Pupil State and Local Revenues. Two key questions in school finance literature are: (1) how do SFRs increase revenue, for example, by increasing revenues directly via state contributions or by regulating the ability of districts to increase local effort (Hoxby, 2001), and (2) are the effects of SFRs partially offset by districts reducing local revenues (e.g., by reducing local tax rates, which are unobserved in our data sets)? Figure 3 addresses these questions by estimating state-specific ATTs for state and local revenues, respectively. On average, state revenues for tercile 1 districts increased by 11 percent ($p = 0.02$), and 90 percent of the estimated effect sizes are expected to fall within a range of -22 to 44 percent, assuming normality. On average, local revenues decreased by 7 percent ($p = 0.15$), and 90 percent of the estimated effect sizes are expected to fall within a range of -35 to 21 percent. Regarding substitution, a 10 percent increase in state revenues is associated with a 3.5 percent increase in total revenues ($p < 0.01$) and a 3.9 percent reduction in local revenues ($p = 0.17$).[15]

Thus, addressing question one, the observed increase in total revenues is driven by increases from state revenues and not local revenues. Responding to question two, our results mirror those of Baicker and Gordon (2006) and Card and Payne (2002), who find some evidence of fiscal substitution resulting from SFRs, though the amount of fiscal

[15] We estimate state-specific ATTs for federal revenues and display them in Appendix: H: Additional Heterogeneity Results for Terciles 1 and 3; see Figure H2. For tercile 1 districts, we estimate a meta-analytic average of the ATTs of 0 percent with an associated p-value of 0.96, and for tercile 3 district, we estimate a meta-analytic average of -2 percent with an associated p-value of 0.21.

substitution we observe is small in magnitude and imprecisely estimated.

We note that Baicker and Gordon (2006) and Card and Payne (2002) evaluate fiscal substitution differently from each other and from our approach. The benefit of our approach is that we can directly estimate the state-specific relationships between total revenues, local revenues, and state revenues, and we can recover similar estimates from both Baicker and Gordon (2006) and Card and Payne (2002). For example, Baicker and Gordon (2006) estimate separate regressions using state revenues and local (county) revenues as outcomes. The authors observe a modest, statistically insignificant change in local revenues, suggesting little to no fiscal substitution. We find comparable results, in that local revenues decreased modestly relative to state revenues and are estimated less precisely.

Card and Payne (2002) take a structural approach to estimate the degree of fiscal substitution between state aid and local revenues. The authors show that the elasticity between total spending and state revenues, which they define as $e + \lambda$, can be estimated according to the following relationship: $\Delta\beta_2 \approx (e + \lambda)\Delta\beta_1$, where $\Delta\beta_2$ is the change in the income gradient of per pupil expenditures and $\Delta\beta_1$ is the change in the income gradient of state revenues.[16] In contrast to Card and Payne (2002), the elasticities we report are the coefficients α_1 and α_2, which come from the following regressions estimated by OLS: $\widehat{\beta_{3s}} = \alpha_1\widehat{\beta_{1s}} + \eta$ and $\widehat{\beta_{2s}} = \alpha_2\widehat{\beta_{1s}} + \nu$, where $\widehat{\beta_{3s}}$ is the state-specific ATT for total revenues, $\widehat{\beta_{2s}}$ is the state-specific ATT for local revenues, and $\widehat{\beta_{1s}}$ is the state-specific ATT for state revenues. We can approximate the approach of Card and Payne (2002), however, by taking the ratio of the meta-analytic averages of $\widehat{\beta_{1s}}$ and $\widehat{\beta_{3s}}$. By doing so, we obtain elasticities of 0.71 and 0.45 for total expenditures and total revenues, respectively, which are comparable to the elasticities of 0.53 to 0.66 the authors obtained.[17]

[16] Card and Payne (2002) estimate the elasticity using instrumental variables—indicators for court-ordered SFRs and formula changes are instruments for $\Delta\beta_1$.

[17] Despite some fiscal substitution, we find no strong evidence of Tiebout sorting. See Appendix B: Additional Details about Dependent Variables for data description and Appendix I: Tiebout Sorting following SFRs for empirical results.

[Figure 3 about here.]

Per Pupil Total Expenditures. Consideration of expenditures allows us to describe how states spent revenues. To begin, we show total expenditures for terciles 1 and 3 in Figure 4. The meta-analytic average of the ATTs for terciles 1 and 3 districts are 7 percent ($p < 0.01$) and 0 percent ($p = 0.93$), respectively, meaning that, on average, SFRs increased spending in bottom-income districts but not top-income districts (results are again identical whether SFRs occurring during the Great Recession are included or not). As before, we observe heterogeneity in the state-specific ATTs. For tercile 1 districts, the estimated standard deviation of the true effect sizes τ is 8 percent, indicating that 90 percent of the estimated effect sizes are expected fall within a range of -6 to 20 percent, assuming normality. Estimated effect sizes range between an 8 percent reduction in total expenditures (Arizona) and a 20 percent increase in total expenditures (New York).[18] For tercile 3 districts, 90 of the estimated effect sizes are expected to fall within a range between -7 to 7 percent, with estimated effect sizes ranging from -16 percent (Arizona) to 9 percent (New York). In total, using 10 percent significance, SFRs increased total expenditures among low-income districts in 15 of 26 states.

[Figure 4 about here.]

Per Pupil Instructional, Total Salary, and Capital. We now turn to expenditure sub-categories, focusing on instruction, total district salaries, and capital outlays for tercile 1 districts, displayed in Figure 5.[19]

For instructional expenditures, the meta-analytic average of the ATTs is 2 percent ($p = 0.22$), and the estimated τ implies that 90 percent of the estimated effect sizes are expected to fall within a range of -8 to 12 percent. For salaries, the meta-analytic average

[18] Mirroring out results for total revenues, Idaho substantially decreased total expenditures and Wyoming substantially increased total expenditures. The Spearman rank correlation coefficient between tercile 1 total revenues and expenditures is 0.87.

[19] Results for capital and salary expenditures in tercile 3 districts are shown in Appendix H; see Figure H3.

is 5 percent ($p < 0.01$), and the estimated τ implies that 90 percent of the estimated effect sizes are expected to fall within a range of -3 to 13 percent. For capital outlays, the meta-analytic average is 15 percent ($p = 0.09$), and the estimated τ implies that 90 percent of the estimated effect sizes are expected to fall within a range of -55 to 86 percent.

A 10 percent increase in total expenditures is associated with a 6.3 percent increase in instructional expenditures ($p = 0.01$), a 4.3 percent increase in total salary expenditures ($p = 0.03$), and a 28 percent increase in capital expenditures ($p = 0.01$). Thus, marginally, low-income districts in states undergoing SFRs tend to increase capital spending more than salary spending.

[Figure 5 about here.]

Role of Capital Expenditures. Among bottom-income districts, Jackson et al. (2016) observe increases in teacher salaries, teachers per student, and capital spending. As the authors explain, "the results suggest that the positive effects are driven, at least in part, by some combination of reductions in class size, having more adults per student in schools, increases in instructional time, and increases in teacher salary that may have helped attract and retain a more highly qualified teaching workforce," (Jackson et al., 2016, p 211). The results from the ASCM suggest that the mechanism driving improvements to academic outcomes is capital spending and not increases to total salaries or reductions in class size.[20]

Capital as a mechanism for increasing academic achievement is not impossible, as recent studies (e.g., Lafortune & Schönholzer, 2017) have shown new school construction can increase student achievement by increasing attendance. New school construction can also plausibly be linked to increased cognitive performance by reducing exposure to environmental contaminants (e.g., Persico & Venator, 2019) or simply by adding air

[20] We note here that in Candelaria and Shores (2019), we also find that high-poverty districts increased capital outlays and not total salary spending as a result of court-ordered SFRs, though we did not report these results in the published paper. A similar result is found in Lafortune et al. (2018). Taken together, these results suggest that the differences between our results here and Jackson (2018) have less to do with methodology and more to do with the SFR era being studied, as Jackson et al. (2016) mostly leverage pre-1990 SFRs and these other studies leverage post-1990 SFRs.

conditioning (Park, Goodman, Hurwitz, & Smith, 2020). Finally, recent work has cast doubt on the universality of positive effects resulting from reduced class sizes, which would be made possible by increasing instructional or salary expenditures (Leuven & Løkken, 2020), meaning that we need not expect academic returns to spending to come exclusively from hiring more instructors. Nevertheless, as Jackson (2018) details, the evidence regarding capital's effects on learning are mixed—indeed, Baron (2021) shows that bonds earmarked for capital do not produce academic gains, whereas bonds earmarked for instruction do—suggesting the focus on capital spending among SFR states in lieu of instructors may be attenuating the potential academic benefits of these expenditures.

Predictors of SFR Efficacy

We provide a detailed overview of our descriptive analysis in Appendix J: Predictors of SFR Efficacy and Appendix Figure J1. Overall, we find that states that included equalization plans, or included categorical aid or spending limits after the SFR tended to have more progressive outcomes, meaning that increases to tercile 1 revenues exceeded tercile 3 revenues. Other factors that positively predict progressivity include states having multiple SFRs and labor union strength. The first predictor suggests that litigation serves as a tool for ensuring the state maintains fidelity with its constitutional obligations (Weishart, 2019). The second predictor suggests that teacher unions have an important role in shaping both the level and distribution of K-12 spending (Brunner et al., 2020). The timing of an SFR (when it occurred from 1990 to 2014 or whether it occurred before or after NCLB or during the Great Recession) does not predict effect size heterogeneity.

Discussion

How often do SFRs "Work?"

In total, 15 states with an SFR either reduced revenues to low-income districts or failed to increase revenues relative to controls and noise in the data. Specifically, our results indicate that SFRs decreased revenues among low-income districts in two states ($p < 0.10$), generated effect sizes below zero but statistically insignificant ($p \geq 0.10$) in six additional states, and generated impacts greater than zero but also not distinguishable from zero in seven additional states. In this section, we provide some explanation for why SFRs failed to increase revenues—in both the substantive and statistical senses—in so many states. Our analysis proceeds by providing a description of outcomes and processes in states without SFRs.

We begin with a set of permutation tests that answer the following question: how did revenues change in low-income districts among states without documented SFRs once they are matched to their own counterfactuals? To answer this question, we assign each state among the set of control states a "placebo" treatment event, as if it had implemented an SFR at the same time as the treated SFR state. For example, Alaska's SFR took place in 2000. Thus, in separate estimations, we assign the placebo treatment event to each of the control states to take place in 2000, allowing each of the remaining control states to serve as counterfactuals (the treated state is excluded from serving as a comparison state). We then implement the ASCM procedure for each control state separately and calculate the proportion of control states that have ATTs greater than the ATT of the state with the true SFR. The box plot and percentages of non-SFR states with ATTs greater than the actual SFR state are shown in Figure 6.

In 10 states—Alaska, Maryland, Massachusetts, Montana, New Jersey, New Mexico, New York, North Dakota, Ohio, and Wyoming—the treatment effect for the true SFR is larger than the treatment effects of more than 90 percent of states without SFRs. The treatment effects for California, Missouri, and Vermont are larger than 86 percent of treatment effects for states without SFRs. In total, these 13 states comprise the set of states with SFRs where there is little ambiguity about the effect the SFR had on changing educational revenues for low-income school districts.

However, for 13 of the 26 states with a documented SFR, revenues in low-income districts did not increase more than revenues in states without an SFR. Specifically, among seven SFR states—Arkansas, Colorado, Indiana, Kansas, New Hampshire, Tennessee, and Washington—between one-quarter to one-half of states without SFRs have ATTs greater than the SFR state. And in six states—Arizona, Idaho, North Carolina, Pennsylvania, Texas, and West Virginia—fifty percent or more non-SFR states have an ATT greater than the estimated ATT for the SFR state.

[Figure 6 about here.]

What are the policy similarities and differences between states with and without SFRs?

The results in Figure 6 indicate that, in many cases, states without documented SFRs increased revenues to low-income districts in magnitudes similar to states with SFRs. For these results to be credible, it should be the case that non-SFR states mirrored SFR states in two areas most relevant for progressive school funding policy: the funding formulae and state contributions to K-12 education in low-income districts. Figure 7 documents trends in these features of school funding policy for states with and without documented SFRs.

In the left panel of Figure 7, we show the proportion of states that exclusively had foundation plans as funding formulae and the proportion of states that had foundation plans plus other components (e.g., a state has a foundation plan, an equalization plan, and provides categorical aid—what might be called a hybrid plan). We plot these proportions over time for states with and without documented SFRs. As shown, there is nearly perfect congruence between SFR and non-SFR states in terms of the broadly defined type of funding formulae they adopted. For instance, beginning in 1996–97, nearly all states switched from having exclusive foundation plans to having so-called hybrid plans, irrespective of an officially documented SFR.[21]

[21] Card and Payne (2002) also noted that states with and without court-ordered SFRs were changing their funding formulae over time but did not provide an explanation. Here, we update the time series and

Though Figure 7 shows that states without documented SFRs are changing their funding formulae over time in ways similar to states with SFRs, these funding formulae indicators do not tell us how much actual aid states provided to low-income districts. For instance, states without SFRs may have foundation plans but guarantee less money to low-income districts relative to states with SFRs. Results shown in the right panel of Figure 7 suggest that this is not the case and that states with and without documented SFRs provide similar levels of state aid to low-income districts. We plot average log total state revenues per pupil and average log total local revenues per pupil over time for states with and without documented SFRs, in bottom tercile districts, relative to a constant in 1989–90.[22] This graph shows that average state contributions per pupil in states with documented SFRs followed nearly identical trajectories (in terms of both levels and trends) as states without documented SFRs, once normalized to a constant base period. Though state contributions for SFR states begin to pull away from non-SFR states in 2003–04, it is hard to attribute this to SFRs directly, as only 5 of 26 states had SFRs after 2003–04.

[Figure 7 about here.]

A question that naturally emerges from these placebo tests and descriptive results is how states can change their funding formulae and increase revenues to low-income districts without having a school finance reform? We can describe numerous potential mechanisms at the state level that could explain increases in revenues among low-income districts in the absence of an SFR. Perhaps the most well known mechanism first occurred in Michigan, in

incorporate legislative SFRs and plot the funding formulae that states adopted over time, thus demonstrating that there is indeed overlap in the general types of funding formulae states adopt, regardless of whether the state was ordered to change its funding system by the courts, passed legislation to overturn its funding system, or did something else not observed.

[22] Specifically, we calculate total state revenues, total local revenues, and total Fall enrollment in tercile 1 for state s and year t; from that, we can compute the total state revenues contributions per pupil and total local revenues contributions per pupil in state s and year t and take the natural logarithm. This procedure generates two variables for 49 states (excluding Hawaii). We normalize these values in a regression by estimating $Y_{st} = \alpha_0 + SFR + SFR \times \Delta_t + \Delta_t + \varepsilon_{st}$, where SFR corresponds to whether the state had a documented SFR or not and Δ_t is a vector of year effects from 1989–90 to 2008–09, where 1989–90 is the reference year, corresponding to average log revenues for states without documented SFRs.

1994, whereby changes to the state constitution were implemented via referendum, thereby circumventing the legislature and courts (Chaudhary, 2009; Cullen & Loeb, 2004; Hyman, 2017; Papke, 2008; Roy, 2011). Similarly, in Florida, in 2002, the citizens voted to amend the state constitution, setting limits on class sizes, which of course increased spending (Chingos, 2012). In other cases, SFRs are implemented early on, sometimes pre-dating our data, but the effect of the SFR is not felt until later. For example, in Georgia, the State Supreme Court found their education finance system unconstitutional in 1981; however, Georgia schools did not see any increase in revenues until legislation was passed to earmark dollars from the state lottery in 1993 (Dee, 2004). In fact, many states earmark K-12 educational dollars with taxes from multiple sources, such as cigarettes and gaming, and separate legislative acts may be passed for each taxed item (see, for example, Sielke, Dayton, Holmes, Jefferson, & Fowler, 2001a, for a discussion of Illinois). These individual acts will not consistently be recorded as school finance reforms but can nevertheless result in monetary increases. Finally, holding local and state tax policy constant, resource shocks (e.g,. the discovery of shale gas) can be capitalized into K-12 revenues and distributed in ways favoring low-income districts (J. Marchand & Weber, 2020; J. T. Marchand, Weber, et al., 2015). These resource shocks would also fail to be recorded as an SFR despite having substantial impacts on the level and distribution of educational revenues.

We must acknowledge that these non-SFR pathways are likely incomplete, somewhat idiosyncratic, and may not explain the broad patterns in funding formulae adoption and state contributions documented in Figure 7. Aside from the handful of examples we identified above, we could find no systematic explanation for how progressivity could increase without a court ruling or legislative act.

Thus, one of the pressing questions that emerges from these findings is "what exactly is an SFR?" This question comes from the fact that, for more than half of states with documented SFRs, states without documented SFRs increased revenues to low-income districts at least as much. This result suggests that non-SFR states are able to increase revenues in ways not previously documented. This hypothesis is further supported by two pieces of evidence. First, states with and without SFRs experienced a strikingly similar pattern of funding formulae adoption. Second, we show that states' contributions to educational revenues followed very similar paths regardless of whether the state had a documented SFR or not. Taken together, these results raise a serious question as to what specific mechanisms states use to increase state aid and change funding formulae to increase revenues to low-income districts outside of the context of formally recognized SFR events.

Conclusion

Consistent with recent studies in the public finance of education literature, this paper finds that school finance reforms (SFRs), on average, increased revenues and spending per pupil more in low-income districts relative to high-income districts (Candelaria & Shores, 2019; Jackson et al., 2016; Lafortune et al., 2018). More importantly, this paper provides novel, compelling evidence about the substantial state-level heterogeneity in terms of how districts respond to SFRs. We estimate effect sizes at the state-income tercile level for each state that had an SFR. This enables us to quantify how revenues and expenditure allocations varied by income tercile across the states. Among low-income districts, we find that nine states increased revenues, two decreased revenues, and 15 saw no change in revenues. We see some evidence of fiscal substitution: total revenues increased by 3.5 percent for every 10 percent increase in state revenues while the decrease in local revenues in not statistically significant. Districts in SFR states further varied in their spending preferences and programmatic implementation. Districts, on average, increased spending more to capital than to salaries. And, in particular, low-income districts did not increase salary expenditures. One important takeaway from this analysis is that average effects mask heterogeneity; therefore, leveraging methods that provide state-specific estimates, such as ridge ASCM, is useful to better understand the distribution that underlies the average.

Because SFRs are costly and consequential for both educational and non-educational expenditures (Baicker & Gordon, 2006; Liscow, 2018), it is useful to know which reforms worked and to be able to describe the contexts in which SFRs were most productive. With more evidence suggesting that money matters for educational outcomes, researchers will need to better understand the conditions and contexts in which money is most productive. By unmasking the heterogeneity underlying an average treatment effect, research will be able to better inform policy discussions.

References

Abadie, A. (in press). Using synthetic controls: Feasibility, data requirements, and methodological aspects [Journal Article]. *Journal of Economic Literature*.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505.

Atchison, D. (2017). *The impact of school finance reform on equity in the state of New York.* Conference paper, Association for Education Finance and Policy. Retrieved from https://aefpweb.org/

Baicker, K., & Gordon, N. (2006). The effect of state education finance reform on total local resources. *Journal of Public Economics*, *90*, 1519–1535.

Baron, E. J. (2021). School spending and student outcomes: Evidence from revenue limit elections in wisconsin. *American Economic Journal: Economic Policy*.

Ben-Michael, E., Feller, A., & Rothstein, J. (2020). The augmented synthetic control method. *arXiv*. (1811.04170v3)

Brunner, E., Hyman, J., & Ju, A. (2020). School finance reforms, teachers' unions, and the allocation of school resources. *Review of Economics and Statistics*, *102*(3), 473–489.

Burtless, G. T. (1997). Does money matter? *Policy Studies Journal*, *25*(3), 489–492.

Candelaria, C. A., & Shores, K. A. (2019). Court-ordered finance reforms in the Adequacy era: Heterogeneous causal effects and sensitivity. *Education Finance and Policy*, *14*(1), 31-60. (DOI: 10.1162/EDFP_a_00236)

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.

Cellini, S. R., Ferreira, F., & Rothstein, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, *125*(1), 215–261.

Chaudhary, L. (2009). Education inputs, student performance and school finance reform in Michigan. *Economics of Education Review*, *28*(1), 90–98.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *The Quarterly Journal of Economics*, *126*(4), 1593–1660.

Chingos, M. M. (2012). The impact of a universal class-size reduction policy: Evidence from florida's statewide mandate. *Economics of Education Review*, *31*(5), 543–562.

Chung, I. H. (2015). Education finance reform, education spending, and student performance: Evidence from maryland's bridge to excellence in public schools act. *Education and Urban Society*, *47*(4), 412–432.

Clark, M. A. (2003). Education reform, redistribution, and student achievement: Evidence from the Kentucky Education Reform Act. *PhD Dissertation: Princeton University*. (Source: http://www.mathematica-mpr.com/~/media/publications/pdfs/education/edreform_wp.pdf)

Corcoran, S., & Evans, W. (2015). [Book Chapter]. In H. Ladd & M. Goertz (Eds.), *Handbook of research in education finance and policy, 2nd edition.* New York, NY: Routledge.

Cullen, J. B., & Loeb, S. (2004). School finance reform in Michigan: Evaluating proposal A. In J. Yinger (Ed.), *Helping children left behind: State aid and the pursuit of educational equity* (pp. 215–250). Cambridge, MA: The MIT Press.

Dee, T. S. (2004). Lotteries, litigation, and education finance. *Southern Economic Journal*, 584–599.

Dee, T. S., & Levine, J. (2004). The fate of new funding: Evidence from Massachusetts' education finance reforms. *Educational Evaluation and Policy Analysis*, *26*(3), 199–215.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, *7*(3), 177–188.

Domina, T., Pharris-Ciurej, N., Penner, A. M., Penner, E. K., Brummet, Q., Porter, S. R., & Sanabria, T. (2018). Is free and reduced-price lunch a valid measure of educational disadvantage? *Educational Researcher*, *47*(9), 539–555.

Doudchenko, N., & Imbens, G. W. (2017). *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis* (arXiv Working Paper No. 1610.07748v2). arXiv.org. (arXiv: 1610.07748v2)

Downes, T. (2004). School finance reform and school quality: Lessons from vermont. *Helping children left behind: State aid and the pursuit of educational equity*, 284–313.

Duncombe, W., & Johnston, J. M. (2004). The impacts of school finance reform in Kansas: Equity is in the eye of the beholder. In J. Yinger (Ed.), *Helping children left behind: State aid and the pursuit of educational equity* (pp. 147–192). Cambridge, MA: The MIT Press.

Evans, W. N., Schwab, R. M., & Wagner, K. L. (2019). The great recession and public education. *Education Finance and Policy*, *14*(2), 298–326.

Fredriksson, P., Öckert, B., & Oosterbeek, H. (2012). Long-term effects of class size. *The Quarterly Journal of Economics*, *128*(1), 249–285.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*.

Goncalves, F. (2015). The effects of school construction on student and district outcomes: Evidence from a state-funded program in ohio.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). Interpreting research on school resources and student achievement: A rejoinder to hanushek. *Review of Educational Research*, *66*(3), 411–416.

Guryan, J. (2001). *Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts* (NBER Working Paper No. 8269). National Bureau of Economic Research. (DOI: 10.3386/w8269)

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational evaluation and policy analysis*, *19*(2), 141–164.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods*, *1*(1), 39–65.

Hightower, A. M., Mitani, H., & Swanson, C. B. (2010a). *State policies that pay: A survey of school finance policies and outcomes.* Editorial Projects in Education.

Hightower, A. M., Mitani, H., & Swanson, C. B. (2010b). *State policies that pay: A survey of school finance policies and outcomes.* Editorial Projects in Education.

Hoxby, C. M. (2001). All school finance equalizations are not created equal. *The Quarterly Journal of Economics*, *116*(4), 1189–1231. (DOI: 10.1162/003355301753265552)

Hyman, J. (2017). Does money matter in the long run? Effects of school spending on educational attainment. *American Economic Journal: Economic Policy*, *9*(4), 256–280.

Jackson, C. K. (2018). *Does school spending matter? the new literature on an old question* (Tech. Rep.). Northwestern Mimeo.

Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, *131*(1), 157–218. (DOI: 10.1093/qje/qjv036)

Johnston, J. M., & Duncombe, W. (1998). Balancing conflicting policy objectives: The case of school finance reform. *Public Administration Review*, 145–158.

Joondeph, B. W. (1994). The good, the bad, and the ugly: An empirical analysis of litigation: Prompted school finance reform. *Santa Clara L. Rev.*, *35*, 763.

Kaul, A., Klößner, S., Pfeifer, G., & Schieler, M. (2015). Synthetic control methods: Never use all pre-intervention outcomes together with covariates.

Koski, W. S., & Reich, R. (2006). When adequate isn't: The retreat from equity in educational law and policy and why it matters. *EmORy lJ*, *56*, 545.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The quarterly journal of economics*, *114*(2), 497–532.

Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, *10*(2), 1-26.

Lafortune, J., & Schönholzer, D. (2017). Do school facilities matter? measuring the effects of capitalexpenditures on student and neighborhood outcomes [Working Paper].

Langer, L., & Brace, P. (2005). The preemptive power of state supreme courts: Adoption of abortion and death penalty legislation. *Policy Studies Journal*, *33*(3), 317–340.

Leuven, E., & Løkken, S. A. (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources*, *55*(1), 309–348.

Liscow, Z. (2018). Are court orders sticky? evidence on distributional impacts from school finance litigation. *Journal of Empirical Legal Studies*, *15*(1), 4–40.

Marchand, J., & Weber, J. G. (2020). How local economic conditions affect school finances, teacher quality, and student achievement: Evidence from the texas shale boom. *Journal of Policy Analysis and Management*, *39*(1), 36–63.

Marchand, J. T., Weber, J., et al. (2015). *The labor market and school finance effects of the texas shale boom on teacher quality and student achievement* (Tech. Rep.). University of Alberta, Department of Economics.

Martorell, P., Stange, K., & McFarlin Jr, I. (2016). Investing in schools: capital spending, facility conditions, and student achievement. *Journal of Public Economics*, *140*, 13–29.

McEachin, A., Domina, T., & Penner, A. (2020). Heterogeneous effects of early algebra across california middle schools. *Journal of Policy Analysis and Management*.

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

Neilson, C. A., & Zimmerman, S. D. (2014). The effect of school construction on test scores, school enrollment, and home prices. *Journal of Public Economics*, *120*, 18–31.

Orr, L. L., Olsen, R. B., Bell, S. H., Schmid, I., Shivji, A., & Stuart, E. A. (2019). Using the results from rigorous multisite evaluations to inform local policy decisions. *Journal of Policy Analysis and Management*, *38*(4), 978–1003.

Papke, L. E. (2008). The effects of changes in michigan's school finance system. *Public Finance Review*, *36*(4), 456–474.

Park, R. J., Goodman, J., Hurwitz, M., & Smith, J. (2020). Heat and learning. *American Economic Journal: Economic Policy*, *12*(2), 306–39.

Persico, C. L., & Venator, J. (2019). The effects of local industrial pollution on students and schools. *Journal of Human Resources*, 0518–9511R2.

Petropoulou, M., & Mavridis, D. (2017). A comparison of 20 heterogeneity variance estimators in statistical synthesis of results from studies: a simulation study. *Statistics in medicine*, *36*(27), 4266–4280.

Resch, A. M. (2008). The effects of the Abbott school finance reform on education expenditures in New Jersey. *PhD Dissertation: The University of Michigan*. (Source: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/61592/aresch_1.pdf)

Roch, C. H., & Howard, R. M. (2008a). State policy innovation in perspective: Courts, legislatures, and education finance reform. *Political Research Quarterly*, *61*(2), 333–344.

Roch, C. H., & Howard, R. M. (2008b). State policy innovation in perspective: Courts, legislatures, and education finance reform. *Political Research Quarterly*, *61*(2), 333–344.

Roy, J. (2011). Impact of school finance reform on resource equalization and academic performance: Evidence from Michigan. *Education Finance and Policy*, *6*(2), 137–167.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

Shores, K., & Candelaria, C. (2019). Get real! inflation adjustments of educational finance data. *Educational Researcher*, 0013189X19890338.

Shores, K., & Steinberg, M. P. (2019). The great recession, fiscal federalism and the consequences for cross-district spending inequality. *Journal of Education Finance*, *45*(2), 123–148.

Sielke, C. C., Dayton, J., Holmes, C. T., Jefferson, A., & Fowler, W. (2001a). Public school finance programs of the united states and canada: 1998-99. *National Center for Education Statistics, US Department of Education*.

Sielke, C. C., Dayton, J., Holmes, C. T., Jefferson, A. L., & Fowler, W. J. (2001b). *Public school finance programs of the United States and Canada: 1998–99* (Report No. NCES 2001-309). U.S. Department of Education, National Center for Education Statistics.

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. (DOI: 10.1162/EDFP_a_00044)

Verstegen, D. A. (2017). State finance policies for english language learners: New findings from a 50-state survey. *Journal of Education Finance*, *42*(3), 338–355.

Weishart, J. E. (2019). Rethinking constitutionality in education rights cases. *Ark. L. Rev.*, *72*, 491.

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

Winkler, A. M., Scull, J., & Zeehandelaar, D. (2012). How strong are us teacher unions? a state-by-state comparison. *Thomas B. Fordham Institute*.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., … others (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369.

Tables

Table 1
Mean Outcome Measures by Tercile and Treatment Status in 1989–90

	Tercile 1			Tercile 3		
	Control	Treatment	Diff.	Control	Treated	Diff.
Revenues						
log(Per Pupil Total Revenues)	9.06	9.13		9.03	9.09	
	(0.22)	(0.36)		(0.24)	(0.29)	
log(Per Pupil State Revenues)	8.34	8.47		8.12	8.12	
	(0.36)	(0.45)		(0.39)	(0.58)	
log(Per Pupil Local Revenues)	8.01	7.95		8.33	8.38	
	(0.52)	(0.56)		(0.54)	(0.60)	
log(Per Pupil Federal Revenues)	6.57	6.60		5.74	5.71	
	(0.41)	(0.73)		(0.45)	(0.51)	
Expenditures						
log(Per Pupil Total Expenditures)	9.07	9.12		9.05	9.10	
	(0.22)	(0.36)		(0.24)	(0.28)	
log(Per Pupil Instructional Expenditures)	8.44	8.51		8.42	8.46	
	(0.24)	(0.33)		(0.26)	(0.26)	
log(Per Pupil Total Salary Expenditures)	8.53	8.61		8.51	8.58	
	(0.21)	(0.33)		(0.24)	(0.25)	
log(Per Pupil Capital Expenditures)	6.12	6.18		6.27	6.34	
	(0.65)	(0.68)		(0.68)	(0.55)	
Enrollment						
Percent Black	29.3	16.9	+	10.5	5.57	+
	(26.9)	(18.0)		(11.9)	(5.87)	
Percent Hispanic	4.55	13.9	*	2.17	5.71	
	(5.73)	(18.4)		(2.19)	(10.3)	
Percent Free Lunch Eligible	40.8	40.2		16.3	13.6	
	(15.8)	(12.3)		(9.39)	(7.46)	
Percent White	61.5	56.8		84.7	85.2	
	(23.7)	(26.1)		(11.3)	(10.8)	

Notes: Comparison states are those that never had an SFR between the sample period of 1989–90 to 2013–14 and treated states are those that had a least one reform during the sample period. Parentheses contain sample standard deviation in 1989–90. Stars indicate the test of whether Treated State summary statistics are equal to Control State summary statistics for each dependent variable within Tercile.
$+ < 0.1$; $* < 0.05$; $** < 0.01$

Figures

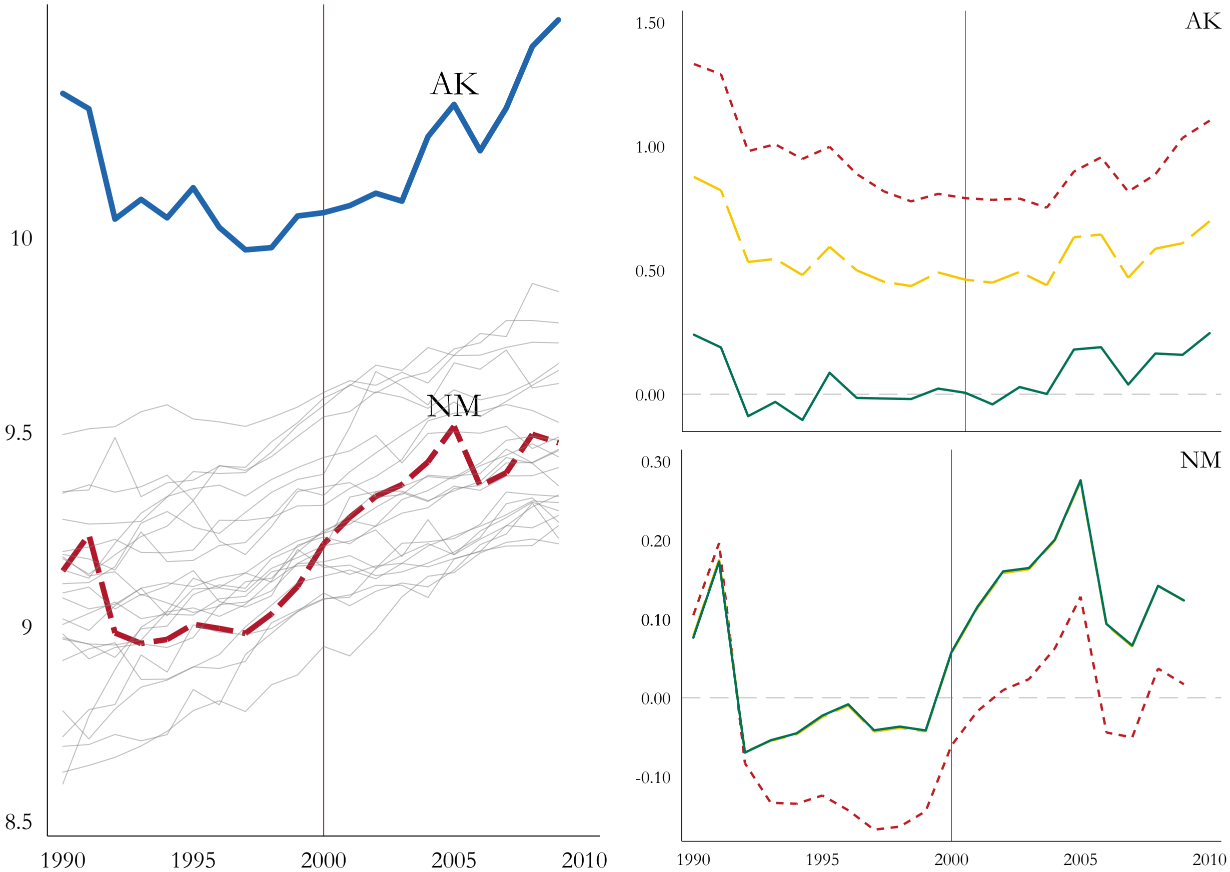
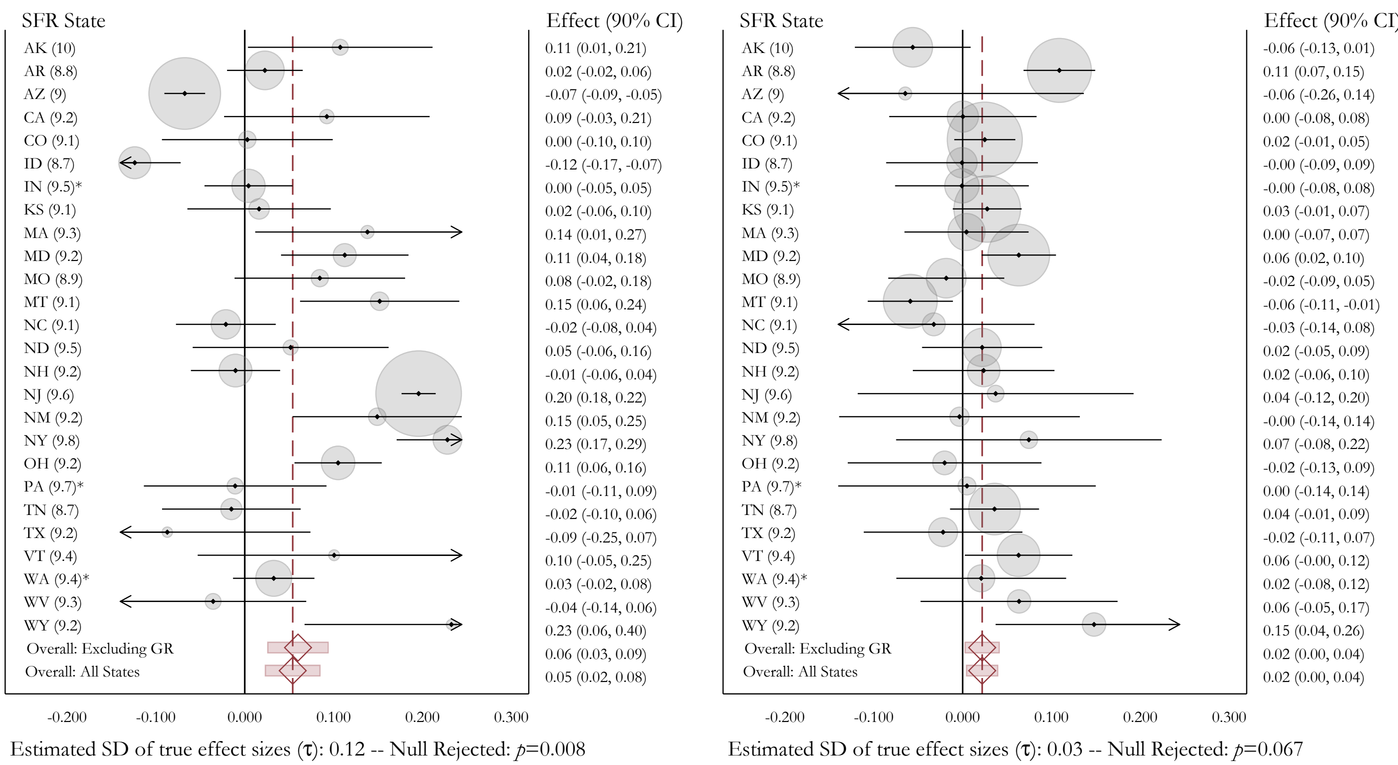


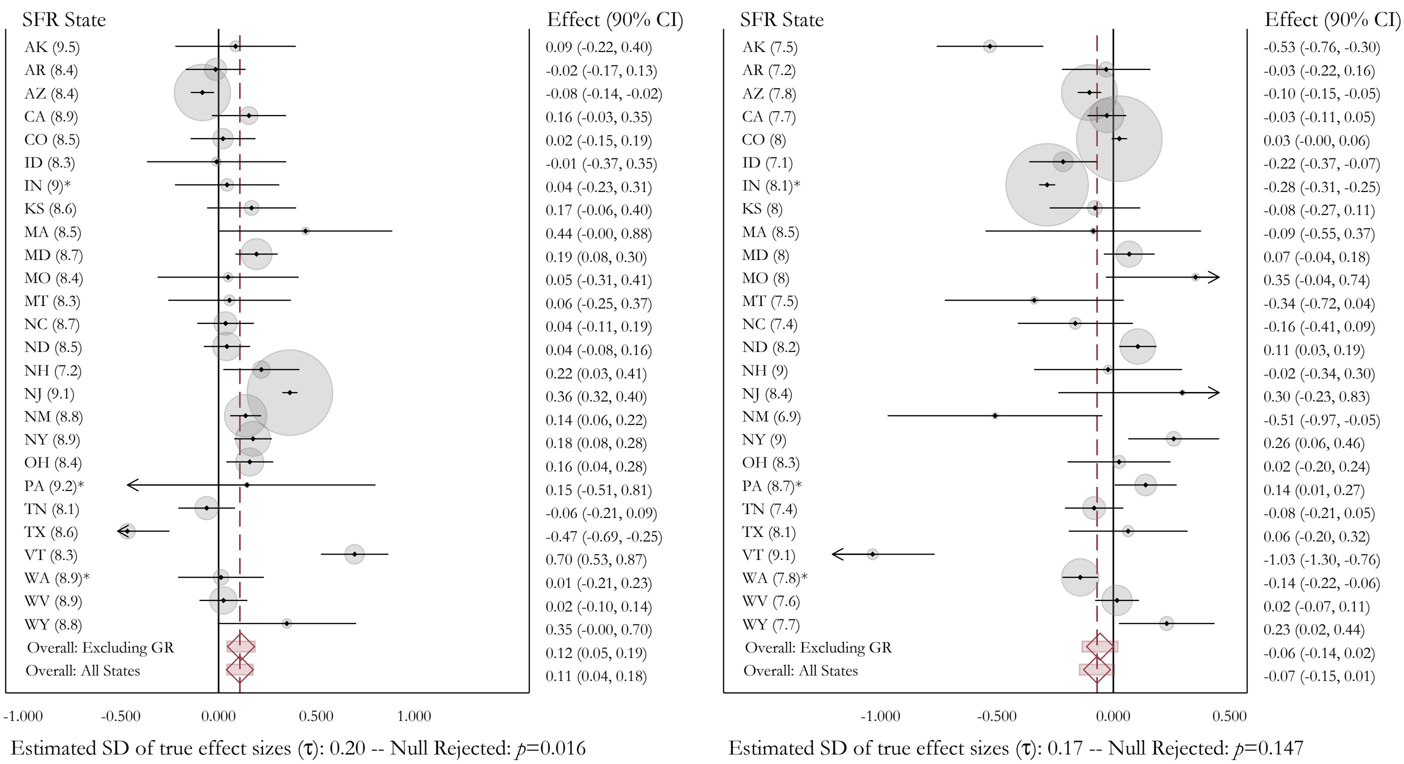
Figure 1. Approaches for Generating Counterfactuals for Individual Units: Alaska and New Mexico's School Finance Reform
Note: Left panel shows Alaska (blue solid) and New Mexico (red dash) log per pupil total revenues in tercile 1 districts against the same outcome for all states without an SFR (gray). The vertical line represents the timing of Alaska's and New Mexico's court cases in the year 1999-00. The right panel shows three different effect sizes for log per pupil total revenues in tercile 1 districts for Alaska (top-right) and New Mexico (bottom-right). In short-dashed red, the effect size is calculated by subtracting revenues in the treated state from mean revenues of all non-treated states, weighting each non-treated state equally. In long-dashed yellow, the effect size is calculated by subtracting the treated state's revenues from the weighted mean revenues of non-treated states, where weights are constructed from traditional SCM. In solid green, the effect size is calculated using weights from ridge ASCM.



(a) Tercile 1 (b) Tercile 3

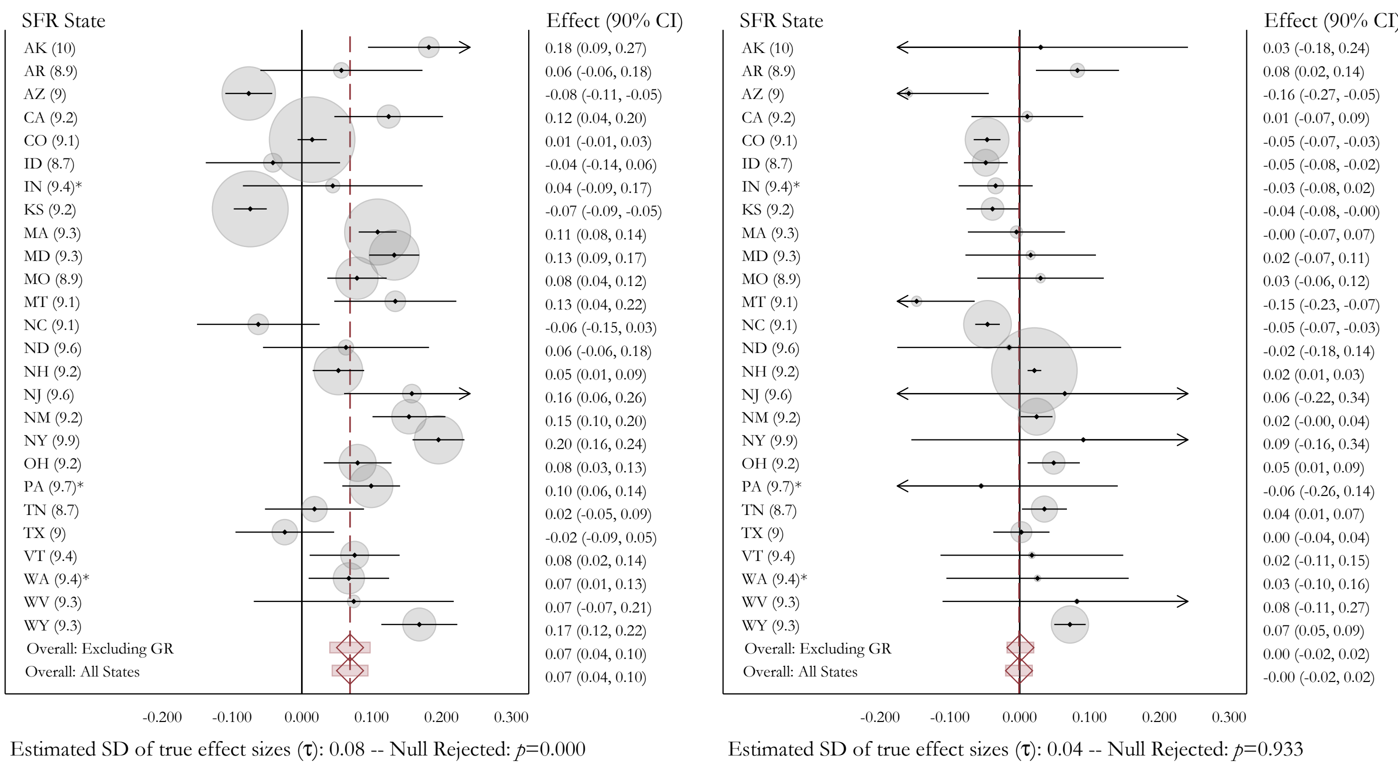
Figure 2. Per Pupil Total Revenues

Note: Dependent variable is log of per pupil total revenues. Results for terciles 1 and 3 are shown in the first and second panels, respectively. The vertical dashed line and bottom-most diamond show the meta-analytic average of the state-specific ATTs. For IN, PA, and WA—states with an SFR during the Great Recession, marked with an *—we include data from 1989–90 to 2013–14; for all other states, we use data from 1989–90 to 2008–09. The meta-analytic average that excludes IN, PA, and WA is shown for reference as the penultimate diamond. The size of the shaded circle represents each state's contribution to the meta-analytic average. Error bars reflect 90% CIs; ⟶ or ⟵ indicate the CI exceeds the axis range. The right-hand column displays the effect size and 90 percent CI. Each number in parentheses after state names is the pre-SFR log value of the dependent variable.



(a) Tercile 1: State Revenues (b) Tercile 1: Local Revenues

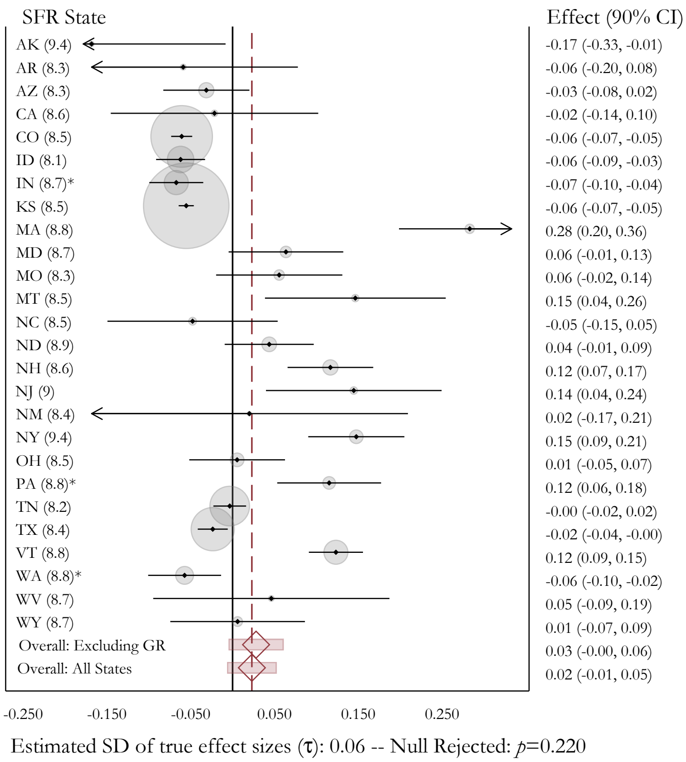
Figure 3. Per Pupil State and Local Revenues
Note: Dependent variables are log of per pupil states revenues (a) and log of per pupil local revenues (b) for tercile 1. The vertical dashed line and bottom-most diamond show the meta-analytic average of the state-specific ATTs. For IN, PA, and WA—states with an SFR during the Great Recession, marked with an *—we include data from 1989–90 to 2013–14; for all other states, we use data from 1989–90 to 2008–09. The meta-analytic average that excludes IN, PA, and WA is shown for reference as the penultimate diamond. The size of the shaded circle represents each state's contribution to the meta-analytic average. Error bars reflect 90% CIs; ⟶ or ⟵ indicate the CI exceeds the axis range. The right-hand column displays the effect size and 90 percent CI. Each number in parentheses after state names is the pre-SFR log value of the dependent variable.



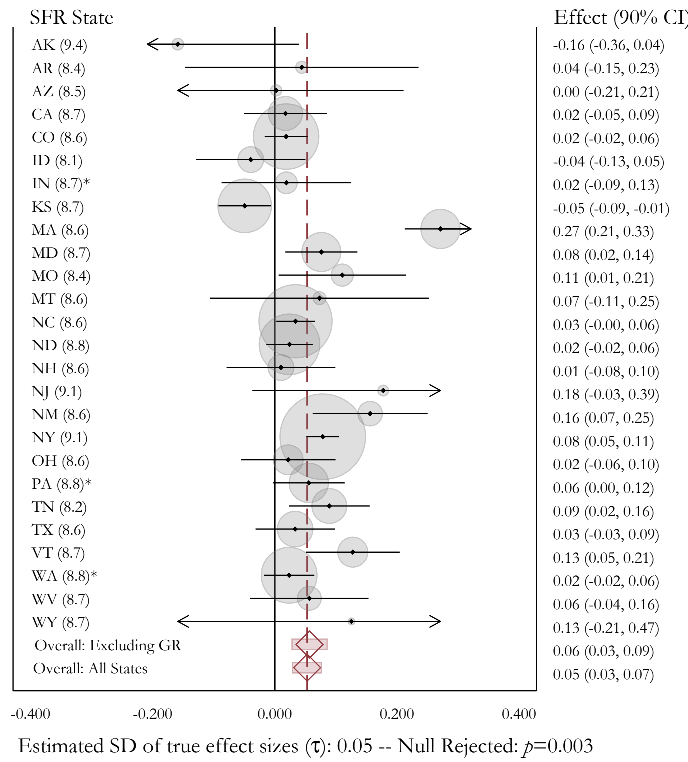
(a) Tercile 1 (b) Tercile 3

Figure 4. Per Pupil Total Expenditures

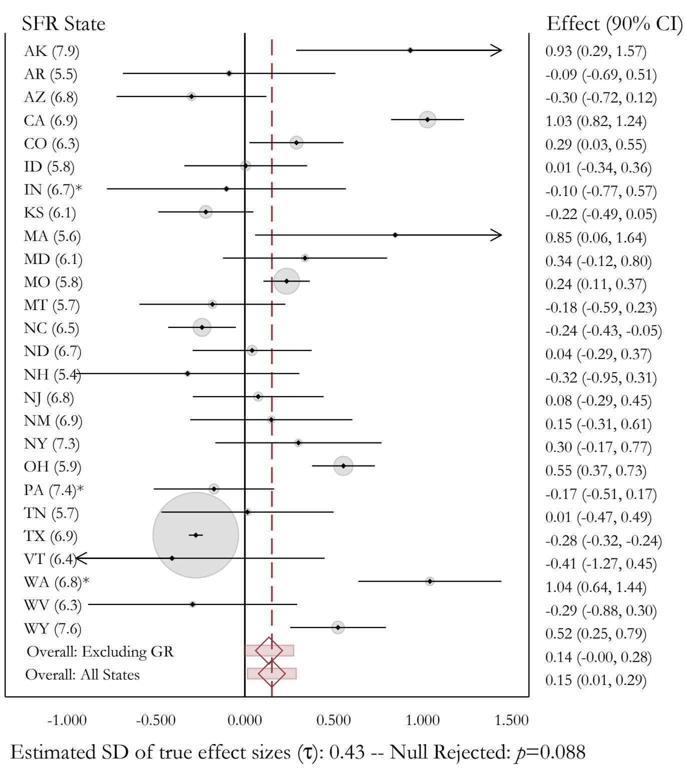
Note: Dependent variable is log of per pupil total expenditures. Results for terciles 1 and 3 are shown in panels (a) and (b), respectively. The vertical dashed line and bottom-most diamond show the meta-analytic average of the state-specific ATTs. For IN, PA, and WA—states with an SFR during the Great Recession, marked with an *—we include data from 1989–90 to 2013–14; for all other states, we use data from 1989–90 to 2008–09. The meta-analytic average that excludes IN, PA, and WA is shown for reference as the penultimate diamond. The size of the shaded circle represents each state's contribution to the meta-analytic average. Error bars reflect 90% CIs; ⟶ or ⟵ indicate the CI exceeds the axis range. The right-hand column displays the effect size and 90 percent CI. Each number in parentheses after state names is the pre-SFR log value of the dependent variable.



(a) Tercile 1: Instructional Expenditures



(b) Tercile 1: Salary Expenditures



(c) Tercile 1: Capital Outlays

Note: Dependent variables are log of per pupil instructional expenditures (a), log of per pupil local revenues (b), and log of per pupil capital (c) for tercile 1. The vertical dashed line and bottom-most diamond show the meta-analytic average of the state-specific ATTs. For IN, PA, and WA—states with an SFR during the Great Recession, marked with an *—we include data from 1989–90 to 2013–14; for all other states, we use data from 1989–90 to 2008–09. The meta-analytic average that excludes IN, PA, and WA is shown for reference as the penultimate diamond. The size of the shaded circle represents each state's contribution to the meta-analytic average. Error bars reflect 90% CIs; ⟶ or ⟵ indicate the CI exceeds the axis range. The right-hand column displays the effect size and 90 percent CI. Each number in parentheses after state names is the pre-SFR log value of the dependent variable.

Figure 5. Per Pupil Instructional, Salary, and Capital Expenditures

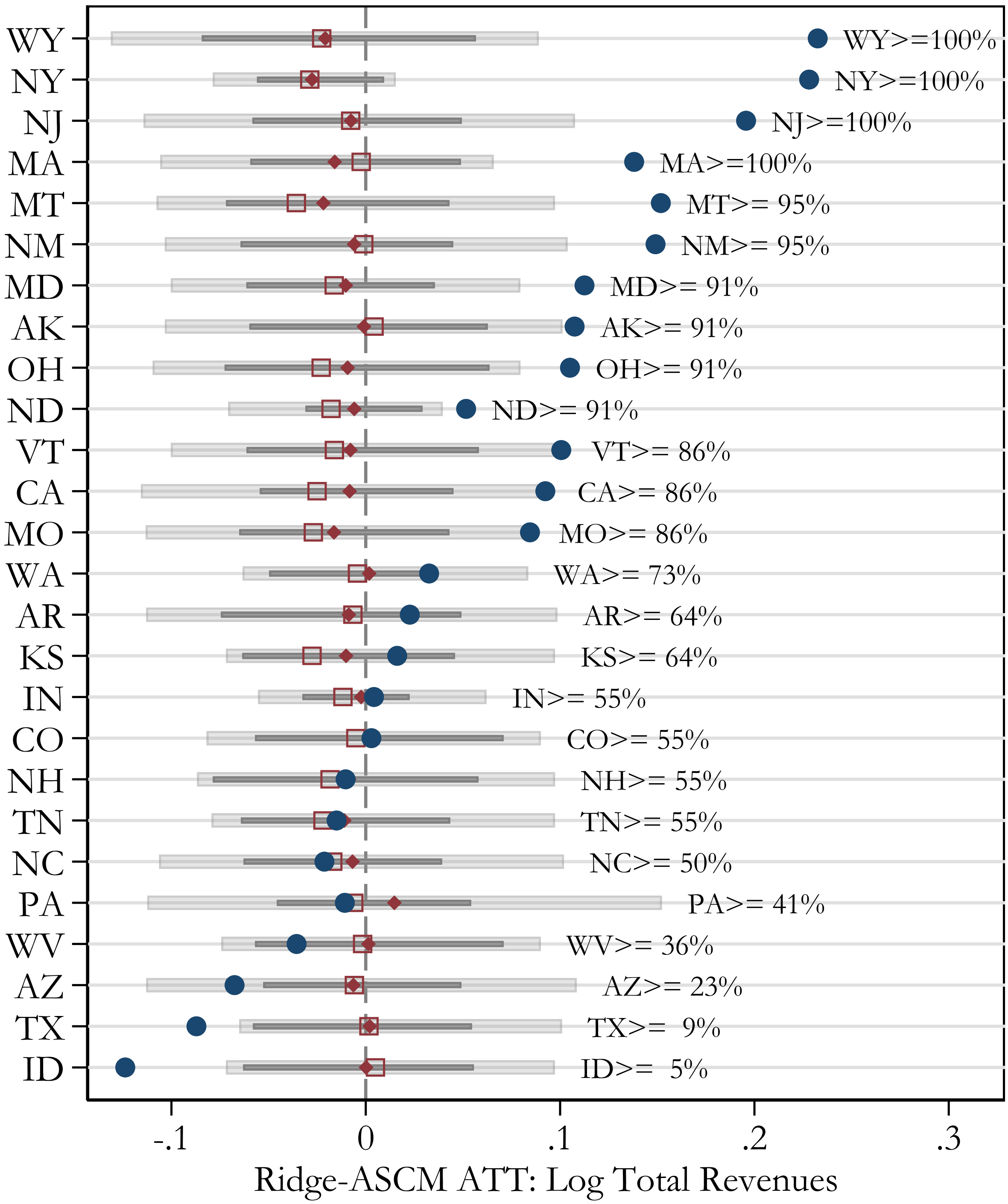


Figure 6. Placebo Effect Sizes for States with SFRs: Tercile 1 Total Revenues
Note: For each state with an SFR, all non-SFR states are assigned a placebo treatment event concurrent with the treated state. We then calculate effect sizes for these placebo states using ASCM; the outcome is log per pupil total revenues in tercile 1 districts. Each state's ATT is a solid circle. The distribution of these placebo ATTs is shown as follows: the transparent gray box represents the 10th and 90th percentiles; the solid gray box represents the IQR; the maroon box represents the median; the solid diamond represents the mean. The in-graph text indicates the percentage of non-SFR states with an ATT that is less than or equal to the ATT of the state with an SFR.

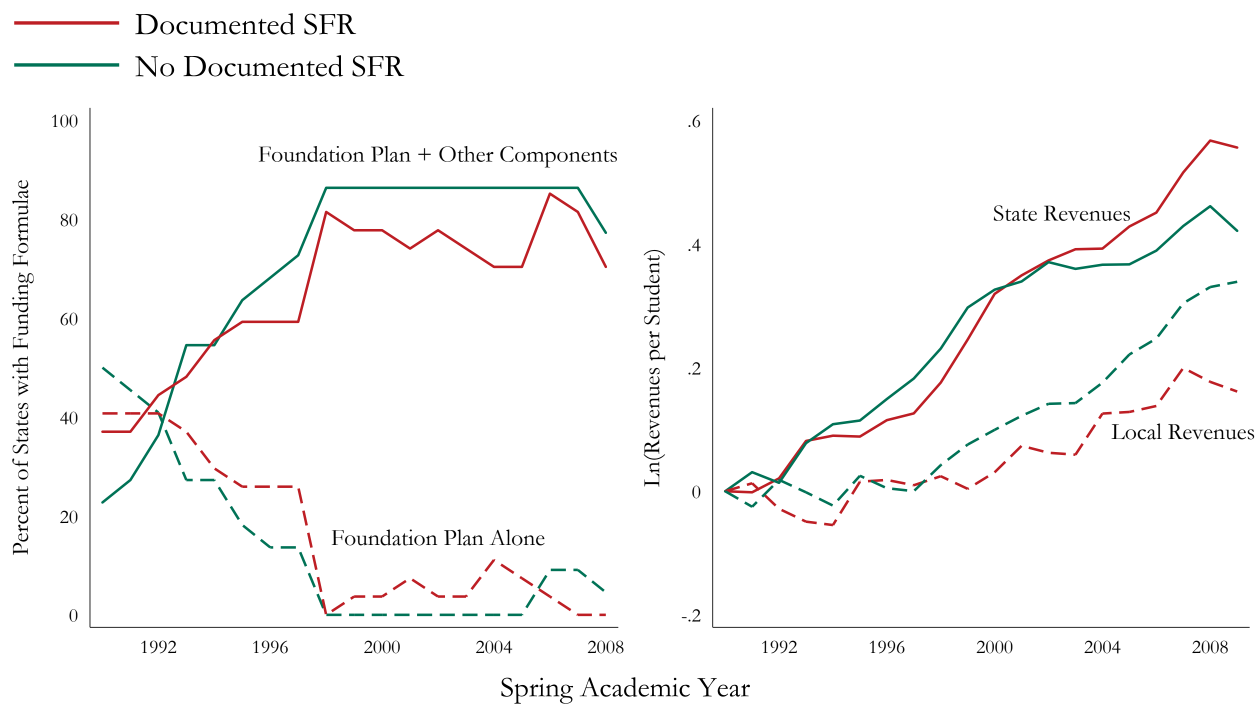


Figure 7. Similarities between SFR and non-SFR State: Funding Formula Type and State Revenues

Note: The left panel shows the proportion of SFR and non-SFR states with Foundation Plans only and Foundation Plans plus other components, over time. Funding formulae components are described in Appendix: Generating Panel Dataset of Funding Formula. The right panel shows the average log total state revenues per pupil and log total local revenues per pupil in SFR and non-SFR states, over time. Total state revenues, total local revenues, and total Fall enrollment in tercile 1 are calculated for state s and year t; then, the natural logarithm is taken. This procedure generates two variables for 49 states (excluding Hawaii). We normalize these values in a regression by estimating $Y_{st} = \alpha_0 + SFR + SFR \times \Delta_t + \Delta_t + \varepsilon_{st}$, where SFR corresponds to whether the state had a documented SFR or not and Δ_t is a vector of year effects from 1989–90 to 2008–09, where 1989–90 is the reference year.

Appendix A

List of Reforms

Table A1
Court-Ordered and Legislative School Finance Reforms

State	Court Case or Legislative Bill	Historical Decision Date	Converted Academic Year
Alaska	Kasayulie v. State of Alaska	1-Sep-99	2000
Arizona	Roosevelt v. Bishop	21-Jul-94	1995
Arizona	Hull v. Albrecht	23-Dec-97	1998
Arizona	Hull v. Albrecht	18-Feb-98	1998
Arkansas	Lake View v. Arkansas	1-Dec-94	1995
Arkansas	Approved Equitable School Finance Plan (Acts 917, 916, and 1194)	1-Feb-95	1996
Arkansas	Lake View v. Huckabee	21-Nov-02	2003
Arkansas	Lake View v. Huckabee	5-May-05	2005
Arkansas	Various acts resulting from Master's Report findings	5/31/07	2007
California	Leroy F. Greene School Facilities Act of 1998	27-Aug-98	1999
California	Senate Bill 6, Senate Bill 550, Assembly Bill 1550, Assembly Bill 2727, and Assembly Bill 3001	1-Aug-04	2005
Colorado	Bill 181; ; Various Other Acts	1-Jul-00	2001
Idaho	Idaho Schools for Equal Educational	18-Mar-93	1993
Idaho	Opportunity v. Evans (ISEEO) Senate Bill 1560	1-Mar-94	1994
Idaho	Idaho Schools for Equal Educational Opportunity v. Evans (ISEEO V)	21-Dec-05	2005
Indiana	HB 1001 (Pl229)	1-Jul-11	2012
Kansas	The School District Finance and Quality Performance Act	1-Jul-92	1993
Kansas	Montoy v. State; Montoy v. State funding increases	3-Jan-05	2005
Kentucky	**Rose v. Council for Better Education, Inc.**	**28-Sep-89**	**1990**
Kentucky	**Kentucky Education Reform Act (HB 940)**	**24-Mar-90**	**1990**
Maryland	Bradford v. Maryland State Board of Education	18-Oct-96	1997

Continued on next page.

Table A1 – continued from previous page.

State	Court Case or Legislative Bill	Historical Decision Date	Converted AY
Maryland	EducationBridge to Excellence in Public Schools	6-May-02	2002
Massachusets	McDuffy v. Secretary of the Executive Office of Education; Massachusetts Education Reform Act	15-June-93 (court); 18-June-93 (bill)	1994
Missouri	Committee for Educational Equality v. State of Missouri	1-Jan-93	1993
Missouri	Outstanding Schools Act (S.B. 380)	1-Aug-93	1994
Missouri	Senate Bill 287	29-Jun-05	2006
Montana	House Bill 667	1-Apr-93	1993
Montana	Columbia Falls Elementary School v. State	22-Mar-05	2005
Montana	M.C.A. § 20-9-309	1-Oct-07	2008
Montana	Montana Quality Education Coalition v. Montana	15-Dec-08	2009
New Hampshire	Claremont New Hampshire v. Gregg	30-Dec-93	1994
New Hampshire	Claremont School District v. Governor	17-Dec-97	1998
New Hampshire	Claremont v. Governor (Claremont III); RSA chapter 193-E	15-Oct-99	2000
New Hampshire	Opinion of the Justices–School Financing (Claremont VI)	7-Dec-00	2001
New Hampshire	Claremont School District v. Governor	11-Apr-02	2002
New Hampshire	Londonderry School District v. New Hampshire	8-Sep-06	2007
New Hampshire	SB 539	21-Apr-08	2008
New Jersey	**The Quality Education Act; Abbot v. Burke**	**05-Jun-90 (court); July-90 (law)**	**1991**
New Jersey	Abbott v. Burke	12-Jul-94	1995
New Jersey	Comprehensive Educational Improvement and Financing Act of 1996	20-Dec-96	1997
New Jersey	Special Master's Report; Abbott v. Burke	14-May-97	1997
New Jersey	Abbott v. Burke	21-May-98	1998
New Jersey	Abbott v. Burke	7-Mar-00	2000
New Jersey	The School Funding Reform Act of 2008	1-Jan-08	2008
New Mexico	Zuni School District v. State	14-Oct-99	2000
New Mexico	Deficiencies Corrections Program; Public School Capital Outlay	5-Apr-01	2001

Continued on next page.

Table A1 – continued from previous page.

State	Court Case or Legislative Bill	Historical Decision Date	Converted AY
New York	Act Campaign for Fiscal Equity, Inc. v. State	26-Jun-03	2004
New York	Campaign for Fiscal Equity, Inc. v. State	20-Nov-06	2007
New York	Education Budget and Reform Act	1-Apr-07	2007
North Carolina	Leandro v. State	24-Jul-97	1998
North Carolina	Hoke County Board of Education v. State	30-Jul-04	2005
North Dakota	SB 2200	3-May-07	2007
Pennsylvania	Act 61	9-July-08	2009
Ohio	DeRolph v. Ohio	25-Apr-97	1997
Ohio	DeRolph v. Ohio	11-May-00	2000
Ohio	Increased school funding (see 93 Ohio St.3d 309)	14-Sep-00	2001
Ohio	DeRolph v. Ohio	11-Dec-02	2003
Tennessee	**The Education Improvement Act**	**11-Mar-92**	**1992**
Tennessee	Tennessee Small School Systems v. McWherter	22-Mar-93	1993
Tennessee	Tennessee Small School Systems v. McWherter	16-Feb-95	1995
Tennessee	Tennessee Small School Systems v. McWherter	8-Oct-02	2003
Texas	**Edgewood Independent School District v. Kirby**	**22-Jan-91**	**1991**
Texas	**Carrolton-Farmers Branch ISD v. Edgewood Independent School District**	**30-Jan-92**	**1992**
Texas	Senate Bill 7	31-May-93	1993
Vermont	Brigham v. State	5-Feb-97	1997
Vermont	Revisions to Act 68; H.480	18-Jun-03	2004
Washington	McCleary v. State	1-Feb-10	2010
West Virginia	Tomblin v. Gainer	1-Aug-00	2001
Wyoming	Campbell County School District v. State	8-Nov-95	1996
Wyoming	The Education Resource Block Grant Model	April, 1997	1997
Wyoming	Wyoming Comprehensive Assessment System	June, 1997	1998
Wyoming	Campbell II; Recalibration of the MAP model	23-Feb-01	2001

Notes: Cases and bills in bold typeface are excluded from the analysis, as they are early in the sample. We require the first case to occur in academic year 1992–1993 (i.e., academic year 1993) in order to establish a sufficient baseline trend for synthetic control matching.

Appendix B

Additional Details about Dependent Variables

B.1 Addressing Volatility

To address volatility in our financial outcome measures, induced by large fluctuations in district enrollment from one year to the next, we follow Lafortune et al. (2018) and apply sample restrictions directly to district enrollment *before* scaling our fiscal variables by enrollment: first, we remove district-year observations in which district total enrollment is less than 100; second, we remove district-year observations in which enrollment exceeds mean district enrollment by scale factor 2; third, we remove the entire district from the panel if a district-year observation is removed according to the preceding two steps.

We then take each of the outcome measures that are to be scaled by district enrollment and divide by this new restricted enrollment variable, effectively removing districts from the data that are small or have an observation that is twice as large as the district mean. Because NCES sampled large districts in 1991–92, 1992–93, and 1994–95, we linearly interpolate the missing values for districts not included in the sample only for these years. All fiscal variables are then log transformed. The two sets of outcome variables are then subjected to an outlier procedure that trims each variable based on its state average in a given year. Specifically, if a given district observation is less than 20 percent or more than 500 percent of the state average, it is dropped (Lafortune et al., 2018; Murray, Evans, & Schwab, 1998).

B.2 Demographic Variables

In addition to these revenues and expenditures variables, we also include free lunch eligibility, and measures of race/ethnicity, including Black, Hispanic, and White to assess whether there is evidence of Tiebout sorting after an SFR. We obtain these demographic variables from the school-level, non-fiscal universe surveys from the NCES Common Core of Data for years 1989–90 to 2013–14. We sum these school-level demographic variables to form a district by year panel data set. For each district, we then calculate the percentage of the district population that is free lunch eligible, Black, Hispanic, or White, where the total enrollment has been adjusted for volatility, as was done for the financial variables.

For several years, New York City public schools reported enrollment counts as 33 geographic districts instead of as a single district; we aggregated the 33 districts to form a single New York City public schools district. Due to extensive missingness in the earlier years of the panel, we replace missing values using the NCES School Universe Survey Longitudinal 13-year data file. This longitudinal file contains imputed values for non-fiscal variables that were missing between years 1986 and 1998. And because missingness still persists in the later years of the panel, we linearly interpolate missing values at the district level only if the district is not missing three consecutive years of data; otherwise, the district is dropped.

For each district, we then calculate the percentage of the district population that is free lunch eligible, Black, Hispanic, or White, where the total enrollment has been adjusted for volatility, as was done for the financial variables. We then take the weighted mean of these variables for each tercile. The mean is preferred here instead of the median because the statistics are bounded between 0 and 100 and are less prone to outliers. Results that leverage these measures are nearly identical whether the mean or median is favored. Even with interpolation, some states are missing data, as more than three consecutive years of data are missing. Percent Free Lunch Eligibility is missing for AK, AZ, and TN among the treated states and for CT, NV, and WI among the comparison states. Percent enrollments for Black, Hispanic, and White are missing from TN.

Appendix C

Generating Panel Dataset of Funding Formula

Because funding formula terminology varies by study and has changed over time, we develop a funding formula dictionary comprised of five common definitions of funding formula components: foundation plan, flat grant, equalization, power equalization, and centralization. We identify two additional "add-on" components of the state funding formulae that are always used in conjunction with one or more of the five core formula: spending limits and categorical aid. States generally adopt "hybrid" funding formula, combining elements from each. For instance, at the time of a state's first SFR, 14 unique funding formula combinations were in place. Despite this heterogeneity, 22 of 26 states included a foundation plan as at least one component of their funding formula. Funding formulae in states without SFRs are similarly hybridized and reliant on foundation plans: in 2014, 16 unique funding formula combinations are present in the 23 states without an SFR, and 19 of these states include at least a foundation plan as part of their formula.

Linking the adopted funding formula to the time at which an SFR takes place requires a panel dataset of funding formulae, one that varies by state and year. We construct such a state-by-year dataset of funding formulae compiled from multiple studies and reports (Card & Payne, 2002; Hightower, Mitani, & Swanson, 2010a; Jackson et al., 2016; Lafortune et al., 2018; Sielke, Dayton, Holmes, Jefferson, & Fowler, 2001b; Verstegen, 2017).

To build the panel dataset of state funding formula, we catalog state funding formulae identified by the sources (Card & Payne, 2002; Hightower et al., 2010a; Jackson et al., 2016; Lafortune et al., 2018; Sielke et al., 2001b; Verstegen, 2017). These studies and reports include funding formula information across time from 1989-2008. Obtaining contemporary descriptions of state-level funding formula is challenging because public records of these formula vary in their degree of specificity and availability online. Obtaining these descriptions for the entire sample period is more challenging due to the limited availability of these formula in the public record. To obtain a historical record of state-level funding formula, we access prior studies and surveys that have indexed these formula. We do not access information directly from state archives; rather, we evaluate reports of funding formula that have been published over time, and additionally take advantage of studies that have collated these publications and historical archives.

Jackson et al. (2016) provide the only compendium of funding formula for multiple states and years. We use this data source for all states and years included in our study. We complement this database with additional years and states by taking state funding formula descriptions from Card and Payne (2002); Hightower, Mitani, and Swanson (2010b); Sielke et al. (2001b); Verstegen (2017). Specifically, we obtain descriptions of funding formula for all available states in 1990 from Card and Payne (2002) and for 1998 from Sielke et al. (2001b) and for 2006 from Verstegen (2017) and for 2008 from Hightower et al. (2010b).

Survey methods for collecting data vary across studies and lead to inconsistencies in information and terminology. In some cases this is minor, such as differences in describing foundation plans as“minimum foundation plan” (Card & Payne, 2002), “foundation program” (Verstegen, 2017), or “foundation aid” (Hoxby, 2001). However, particularly when discussing equalization plans, there is a wider variety of terms used to describe what we call “equalization” and “power equalization.” For example, power equalization is referred to as “local effort equalization” (Hightower et al., 2010b), “district power equalization” (Verstegen, 2017), and “variable guarantee” (Card & Payne, 2002). We observed some degree of inconsistent classifications in all of the funding formula we classified.

Although we find differences in how researchers label formula components, we are able to identify common definitions across studies. We reviewed the literature and collated funding formula terms and definitions across sources. Through an analysis of definitions and descriptions we were able to combine like terms into four main categories: flat grant, equalization, foundation, and other. Through this exercise we cross-checked definitions to ensure that, for example, the usage of “flat grant” or “foundation aid” was consistent across studies, regardless of specific nomenclature. Further analysis indicated that equalization needed to be split into two categories: equalization and power equalization, where power equalization indicates a direct relationship between level of local effort and level of state aid. The "other" category included full state funding, and what we came to call "add on" components, including spending limits and categorical aid. Next we provide the funding formula terms and definitions that emerged from our synthesis.

Foundation plans provide a guaranteed amount of funding per pupil in each district. Under this plan, the state utilizes block grants to supplement a district's expected contribution so that the guaranteed minimum is met. Flat grants are used to provide a statewide uniform dollar amount per pupil, with any additional spending provided by local revenue. Flat grants do not rely on local spending effort, and are often combined with other funding formula adjustments. Equalization is a block grant that varies based on district tax base or tax revenue. These are differentiated from foundation grants in that they are not meant to standardize to a minimum level of spending per pupil, but rather intended to add to district spending based on some observable characteristic, such as local income level. These are distinct from power equalization grants. Power Equalization is a matching grant, wherein local effort and tax levels are directly tied to additional state funding. Power equalization grants ensure that districts with the same tax rate have the same amount of money to spend per pupil, regardless of taxable wealth in their district. Finally, centralization plans are those in which the state assesses, levies, and distributes all tax funding related to school financing.

We identify two additional "add-on" components of the state funding formulae. These are always used in conjunction with one or more of the five core formula types. Spending limits create a ceiling on property taxes or the amount of funding that can be added to state foundation plans. Categorical aid components may be distributed on a per-pupil basis for students who fit in a defined category. Special education spending by the state is often in the form of categorical aid, where a set dollar amount is provided for each pupil served by the program.

Appendix D

Funding Formulae

Table D1
Funding Formula Distribution

State	1st SFR	Funding Formula Components	Num. of FF Changes post-SFR
AK	2000	FP + EQ	2
AR	1995	FP + EQ + SL	1
AZ	1995	FP + EQ + PE + SL	3
CA	1999	FP + FG	2
CO	2001	FP + EQ + CA	2
ID	1993	FP	3
IN	2012	FP + PE + CA + SL	0
KS	1993	FP + EQ + PE + SL	3
MA	1994	FP	3
MD	1997	FP + PE + FG	2
MO	1993	FP + EQ + PE + FG	3
MT	1993	FP + EQ + SL	2
NC	1998	FP + EQ + CA	3
ND	2007	EQ + PE	1
NH	1994	FP	4
NJ	1991	FP + EQ	2
NM	2000	FP + EQ + CA	2
NY	2004	EQ + FG	2
OH	1997	FP	4
PA	2009	FP + EQ + PE + CA	0
TN	1992	FP	3
TX	1991	FP + EQ + PE + SL	3
VT	1997	FP + EQ + SL	5
WA	2010	PE + CA + SL + CN	0
WV	2001	FP + PE + CA + SL	1
WY	1996	FP + EQ + SL	2

Notes: Funding formulae listed correspond to the funding formula in place at the start of the first reform. The number of FF changes post-SFR represents the number of changes to unique funding formulae components after the SFR. For example, Alaska had a foundation plan and equalization as part of its funding formulae components in 1999-2000, when it had its first SFR. The number 2 represents the fact that Alaska adopted a foundation plan and categorical aid (in 2005-2006) and a foundation plan, equalization, power equalization, categorical aid, and spending limits (in 2007-2008). If the state reverted back to the plan it had at the start of the SFR, it would not be recorded as a change. [Codes: FP=foundation plan; EQ=equalization; PE=power equalization; FG=flat grant; CN=centralization; CA=categorical aid; SL=spending limits]

Appendix E

Additional Predictors

In addition to state finance policy variables, we also gather state-level indicators of teacher union strength from the Thomas B. Fordham Institute and from Appendix Table 2 in (Brunner et al., 2020). Brunner et al. (2020) show that states with SFRs and greater teacher union strength increased expenditures more relative to states with SFRs and weaker teacher unions, which translated into larger effect sizes on student achievement. For the Fordham Institute, their index of teacher union strength is generated through a combination of factors including union resources and membership, involvement in politics, the scope of collective bargaining strength, state policies, and perceived union influence. Higher values on the index indicate stronger union status; these values come Winkler, Scull, and Zeehandelaar (2012) and are only available for the year 2011-12. Following Brunner et al. (2020), we also generate three additional variables using 1987 values, since those predate our SFRs. The first is a continuous measure for whether a state mandated, allowed, or prohibited collective bargaining agreements (CBA); the second is an indicator for whether the state was a "right-to-work" (RTW) state, coded as 0 if it was and 1 if not; the third is a continuous measure that combines the values from CBA and RTW. A state mandating CBA without RTW laws is highest and coded as 4 in this measure, and a state that prohibits CBA and has RTW laws is lowest and coded as 0. Further information is available in Appendix Table 2 in Brunner et al. (2020).

Appendix F

Methodological Details

F.1 SCM Notation and Setup

Using notation from Abadie, Diamond, and Hainmueller (2010) and Ben-Michael et al. (2020), we define the SCM in the SFR context. We observe data for $S+1$ states, where $s \in \{1, 2, \ldots, S+1\}$, and among these states, we designate the first state (i.e., $s = 1$) to be the treatment state that had an SFR.[23] The remaining S states serve as a pool of control group states, as they did not have SFRs. Given that a total of 27 states had SFRs during our analysis period, we examine each treated state as its own case study to examine the heterogeneity of effect sizes across states.[24] Each of these 27 case studies then draws from a pool of $S = 50 - 27 - 1 = 22$ never-treated control states, as we exclude Hawaii from the control sample since it is a one-district state.

With respect to time, we have a total of 25 academic years of fiscal data, spanning 1989–90 to 2013–14. The number of years before an SFR (i.e., pre-treatment years) varies by treated state. Following Abadie et al. (2010), we denote the number of these pre-treatment years as T_0, where $1 \leq T_0 < 25$. We operationalize T_0 to include the academic year of an SFR decision date; thus, we assume that treatment does not begin until the following academic year after an SFR court decision or legislative bill.

In matrix form, our data set is structured as follows:

$$\begin{pmatrix} Y_{1,1} & Y_{1,2} & \ldots & Y_{1,T_0} & Y_{1,T_0+1} & \ldots & Y_{1,25} \\ Y_{2,1} & Y_{2,2} & \ldots & Y_{2,T_0} & Y_{2,T_0+1} & \ldots & Y_{2,25} \\ \vdots & & & & & & \vdots \\ Y_{23,1} & Y_{23,2} & \ldots & Y_{23,T_0} & Y_{23,T_0+1} & \ldots & Y_{23,25} \end{pmatrix} \equiv \left(\begin{array}{cccc|ccc} X_{1,1} & X_{1,2} & \ldots & X_{1,T_0} & Y_{1,T_0+1} & \ldots & Y_{1,T_{25}} \\ \hline X_{2,1} & X_{2,2} & \ldots & X_{2,T_0} & Y_{2,T_0+1} & \ldots & Y_{2,T_{25}} \\ \vdots & & & & & & \vdots \\ X_{23,1} & X_{23,2} & \ldots & X_{23,T_0} & Y_{23,T_0+1} & \ldots & Y_{23,T_{25}} \end{array}\right).$$

The first matrix shows a dataset of dimension 23×25 for outcome Y_{st} (e.g., log per pupil

[23] In reality, we observe data for $S+1$ state-by-district income terciles, but because treatment occurs at the state level, and for simplicity, we describe the data as if it is state-by-year.

[24] In presenting results, we exclude Kentucky from the treated group because we do not have sufficient pre-treatment data for the state. The first year after the *Rose* decision coincides with our first year of data.

total expenditures in tercile 1 districts). Each row corresponds to a state—the first row is the treated state—and columns correspond to time. The second matrix partitions the first and denotes each pre-treatment outcome using X_{st}; the pre-treatment outcomes are covariates in our SCM. Collapsing the rows and columns of the partitioned matrix, we arrive at the following simplified data structure:

$$\left(\begin{array}{c|c} X_{1\cdot} & Y_{1\cdot} \\ \hline X_{0\cdot} & Y_{0\cdot} \end{array}\right),$$

where $Y_{1\cdot}$ is a row vector of outcomes for the state that had an SFR; $Y_{0\cdot}$ is the of matrix of outcomes among control states; and $X_{1\cdot}$ and $X_{0\cdot}$ are vectors of pre-treatment dependent variables (i.e., pre-treatment outcomes) among the treated state and control states, respectively. Though we observe 25 academic years of data for each state, the number of pre- and post-treatment years of data will vary depending on when the SFR took place; as previously mentioned, T_0 will depend on when the SFR takes place.

We discuss our outcome of interest, Y_{st}, by leveraging the potential outcomes framework (Rubin, 1974). Specifically, the potential outcome for state s in year t under the assumption of no SFR is $Y_{st}(0)$, and the potential outcome under the assumption of having an SFR is $Y_{st}(1)$. Because $s = 1$ denotes the treated state, we can formally define our treatment effects of interest (ATTs) as $\gamma_{1t} = Y_{1t}(1) - Y_{1t}(0)$; however, the fundamental problem of causal inference is that we observe $Y_{1t}(1)$ but not $Y_{1t}(0)$. That is, we only observe outcomes $Y_{st}(1)$ if the state experienced an SFR; we do not observe its outcome if it did not experience an SFR. Such data environments can be cast as missing data problems, where the missing data are the outcomes of treated units in the absence of treatment.

Taking the unweighted average on non-treated states is one approach to filling in the missing potential outcomes for $Y_{1t}(0)$; however, as previously discussed, this approach does not yield a valued counterfactual. In contrast, SCM leverages $X_{0\cdot}$ to find and weight control units that most resemble $X_{1\cdot}$.

F.2 Traditional SCM

To estimate $Y_{1t}(0)$ from Y_0, SCM implements a minimization procedure that estimates w_s^{SCM}, a time-invariant weight for each state s in the control group. The minimization procedure attempts to satisfy the following conditions:

$$\sum_{s=2}^{23} w_s^{\text{SCM}} X_{s1} = X_{11}$$
$$\sum_{s=2}^{23} w_s^{\text{SCM}} X_{s2} = X_{12}$$
$$\vdots$$
$$\sum_{s=2}^{23} w_s^{\text{SCM}} X_{sT_0} = X_{1T_0},$$

where the system of equations above shows that weights are determined by matching exclusively on the pre-treatment outcomes (Doudchenko & Imbens, 2017; **?**)—for each $t \in \{1, \ldots, T_0\}$—with the purpose of setting differences between treatment and control equal to zero.[25] Then, we use the estimated weights, $\hat{w}_s^{\text{SCM}}$, and apply them to the outcomes of the $S = 22$ members of the control group, which gives us

$$\hat{Y}_{1t}^{\text{SCM}}(0) = \sum_{s=2}^{23} \hat{w}_s^{\text{SCM}} Y_{st}. \tag{1}$$

The estimate of $Y_{1t}(0)$ at time t is a weighted average of control outcomes in the treatment period, and it characterizes the counterfactual outcome of the treated state in year t if it had not undergone reform. This "synthetic" counterfactual group enables us to estimate dynamic treatment effects for $s = 1$ as

$$\hat{\gamma}_{1t}^{\text{SCM}} = Y_{1t} - \hat{Y}_{1t}^{\text{SCM}}(0) \quad \text{for } t > T_0.$$

[25] Multiple papers have pointed out that including all lagged dependent variables effectively cancels out any additional lagged covariates (e.g., Kaul, Klößner, Pfeifer, and Schieler (2015)).

F.3 Ridge Augmented Synthetic Control Method (Ridge ASCM)

Bias in the SCM estimator can be introduced when the SCM weights (w_s^{SCM}) do not achieve good balance in the pre-treatment period. As suggested by Ben-Michael et al. (2020), we address this problem by specifying the ridge ASCM as follows:

$$\hat{Y}_{1t}^{\text{aug}}(0) = \underbrace{\sum_{s=2}^{23} \hat{w}_i^{\text{SCM}} Y_{it}}_{\text{(1) SCM estimate}} + \underbrace{\underbrace{\left(X_{1\cdot} - \sum_{s=2}^{23} \hat{w}_s^{\text{SCM}} X_{s\cdot}\right)}_{\text{(a) SCM match quality}} \cdot \underbrace{\hat{\eta}_t^{\text{r}}}_{\substack{\text{(b) Ridge} \\ \text{coefficient} \\ \text{vector}}}}_{\text{(2) bias correction}}, \tag{2}$$

where there are two additive terms: (1) The SCM estimate for $Y_{1t}(0)$ and (2) a bias correction to address poor match quality in the SCM estimate. Given we have previously discussed the SCM estimator, we now focus our attention on the bias correction term.

To describe the bias correction term, we first explain the notation for (a) the SCM match quality component and (b) the ridge coefficient vector component. For the SCM match quality component, $X_{1\cdot}$ is a 1-by-T_0 row vector of pre-treatment outcomes for treated state $s = 1$; $X_{s\cdot}$ is also a 1-by-T_0 row vector of pre-treatment outcomes but for control state s, where $s \in \{2, 3, \ldots, 23\}$. When SCM match quality is good, pre-treatement differences between the treated unit and the SCM pre-treatment counterfactual are small for each pre-treatment year t; when these differences are large, the match quality is bad.[26]

The ridge coefficient vector describes the estimated relationship between pre-treatment outcomes and post-treatment outcomes for the control group. Formally, $\hat{\eta}_t^{\text{r}}$ is a T_0-by-1 vector of coefficients for post-treatment year t that estimated using a multivariate ridge regression of centered, control post-treatment outcomes, $\tilde{Y}_{st}$, on centered, control pre-treatment outcomes, $\tilde{X}_{s\cdot}$, with penalty parameter λ^{r}. The full ridge coefficient

[26] Given that pre-treatment match is based on a comparison of the distance between two vectors, statistics such as an the L^2-norm are appropriate for describing match quality.

matrix, $\hat{\eta}^{\mathrm{r}}$, comes from the following minimization problem:

$$\min_{\eta} \frac{1}{2} \sum_{s=2}^{23} \sum_{t=T_0+1}^{T} \left(\tilde{Y}_{st} - \tilde{X}'_{s\cdot}\eta \right)^2 + \lambda^{\mathrm{r}} \|\eta\|_2^2. \tag{3}$$

Ridge augmentation can reduce bias by increasing pre-treatment fit, but it can also increase bias by over-fitting to noisy pre-treatment outcome data. The penalty parameter λ^{r} regulates this trade-off between improved pre-treatment fit and approximation error. It does so by affecting the magnitude of the $\hat{\eta}_t^{\mathrm{r}}$ coefficients, which appear in the bias correction term in Equation (2). As $\lambda^{\mathrm{r}} \to \infty$, ridge ASCM converges to traditional SCM because the $\hat{\eta}_t^{\mathrm{r}}$ coefficients shrink toward zero. In cases where $\lambda^{\mathrm{r}} \to 0$ and traditional SCM match quality is bad, the bias correction term in Equation (2) becomes large. When this occurs, pre-treatment fit will be nearly perfect with the consequence that extrapolation error is more likely.

Because different values of λ^{r} can influence estimates of $\hat{Y}_{1t}^{\mathrm{aug}}(0)$, following Ben-Michael et al. (2020), we use cross-validation to provide guidance in selecting its value. Specifically, for treatment state $s = 1$, we estimate pre-treatment time period t using the following ridge ASCM model:

$$\hat{X}_{1t}^{\mathrm{aug}} = \underbrace{\sum_{s=2}^{23} \hat{w}_i^{\mathrm{SCM}} X_{i(-t)}}_{\text{SCM estimate}} + \underbrace{\left(X_{1(-t)} - \sum_{s=2}^{23} \hat{w}_s^{\mathrm{SCM}} X_{s(-t)} \right) \cdot \hat{\eta}_{(-t)}^{\mathrm{r}}}_{\text{bias correction}}, \tag{4}$$

where the SCM estimate and and the bias correction terms are estimated from data that *excludes* time period t.

Then, for a given value of λ^{r}, we perform leave-one-out cross validation across pre-treatment time periods (i.e. $t \leq T_0$) and compute the mean-squared error as follows:

$$CV(\lambda^{\mathrm{r}}) = \sum_{t=1}^{T_0} \left(X_{1t} - \hat{X}_{1t}^{\mathrm{aug}} \right)^2. \tag{5}$$

In our analyses, we select λ^{r} as the maximum λ^{r} within 1 standard error of the λ^{r} that minimizes Equation (5).

Because our full panel data set is uniquely defined by state, year, and income tercile, we estimate state-by-district income tercile effects for all states undergoing an SFR. For each outcome of interest, we compute the dynamic treatment effects (ATTs) as

$$\hat{\gamma}_{1t}^{\text{aug}} = Y_{1t} - \hat{Y}_{1t}^{\text{aug}}(0) \quad \text{for } t > T_0.$$

In our heterogeneity analyses, we compute the average of the dynamic ATTs, $\bar{\hat{\gamma}}_1^{\text{aug}}$ (i.e., ATT), to compare effect sizes across states. Our reported standard errors are based on a row-based jackknife to allow for autocorrelation within states (Doudchenko & Imbens, 2017).

A final question is what to do with states that underwent an SFR during the Great Recession. Three states underwent an SFR concurrently with the Great Recession—Indiana (2011–12), Pennsylvania (2008–09), and Washington (2009–10). As is known, the Great Recession reduced educational revenues and spending heterogeneously both between and within states Evans, Schwab, and Wagner (2019); Shores and Steinberg (2019). In principle, the ASCM should identify states with similar trends in revenues and spending as these states, even if the states were differentially affected by the Great Recession. However, because of the uniqueness of the Great Recession in terms of the shock and patterns of subsequent recovery, ASCM may perform poorly in this environment. We therefore estimate synthetic controls for all other SFR states restricting the dataset to include observations from 1989–1990 to 2008–09, before the fiscal shock of the recession was felt (Shores & Candelaria, 2019). We then estimate a separate set of synthetic controls for Indiana, Pennsylvania, and Washington using all available years of data from 1989–90 to 2013–14. We generate two meta-analytic averages, one which includes only the pre-Great Recession SFRs and one that includes all SFRs.

Appendix G

Assessing Ridge ASCM Estimates

We summarize the quality and stability of our synthetic control efforts in four ways. First, we show the ATTs for each SFR state-tercile over time. These provide a visual image of how well the pre-treatment match appears for each state-tercile, as well as the dynamics of the estimated effects following the SFR. These plots are shown in Figure G1. As shown, the "common trends" assumption is met for all states, as effect sizes are oriented around zero. Effect size estimates are somewhat volatile, especially for a few states, subsequent to the SFR, indicating both noise in the data as well as real fluctuations in revenues from the NCES F-33 data.

Second, we show model fit statistics for the ridge ASCM and SCM methods for log per pupil total revenues in terciles 1 and 3. These model fit statistics indicate cumulative pre-treatment effect size deviation from zero (leveraging the L^2 norm) using the weights from ASCM and SCM respectively. These model fit statistics are scaled to be proportional to the model fit one would obtain if uniform weights were applied to control units. Thus, as values approach one, ASCM or SCM obtain equivalent pre-treatment match quality as applying uniform weights to all control units; as values approach zero, ASCM or SCM have perfect match quality relative to applying unifom weights. In Figure G2, we show the scaled model fit statistics for log per pupil total expenditures in terciles 1 and 3.

As is evident (and expected), ASCM weights always perform as well than SCM weights. In many cases, however, SCM weights achieve identical pre-treatment match as ASCM weights, meaning that pre-treatment match quality cannot be improved with the de-biasing from ridge regression. In a few cases cases (e.g., Alaska, North Dakota, West Virginia, Wyoming), ASCM weights greatly improve upon SCM weights, meaning that pre-treatment match quality was improved with ridge augmentation.

Third, we plot the distribution of synthetic control weights for each SFR state-tercile. A key difference between SCM and ASCM is that ASCM leverages negative weights when control states are outside the convex hull of SFR states; thus, we will observe more negative weights when ASCM is leveraged relative to ASCM. The distribution of weights is shown in Figure G3. Comparing Figures G3 and G2, we can see that the states where ASCM dramatically improves model fit (e.g., Alaska), ASCM also leverages negative weights. In states where ASCM and SCM are identical, negative weights are never used, which is by construction for SCM but is not necessary for ASCM.

Fourth, because the choice of λ^{r} can be consequential for estimating the ATT, we illustrate stability of our estimates by re-estimating the ASCM model with imposed values of λ^{r} between 1×10^{-7} and 9×10^{5}. In total, we estimate 108 alternative specifications of the ASCM. To facilitate comparison to the preferred ATT, we estimate $\widehat{ATT}^{\mathrm{ratio}}$ as $\widehat{ATT}^{L(\lambda)}/\widehat{ATT}^{CV(\lambda)}$, where $\widehat{ATT}^{L(\lambda)}$ is the estimated ATT for specification L given penalty parameter λ^{r} and $\widehat{ATT}^{CV(\lambda)}$ is the estimated ATT obtained from the cross-validation procedure described above. The statistic $\widehat{ATT}^{\mathrm{ratio}}$ indicates the ratio of the ATT for a specified λ^{r} relative to our preferred cross-validation estimate.

In Figures G4 and G5, we show the median $\widehat{ATT}^{\mathrm{ratio}}$ for log total revenues and expenditures (plus the revenues and expenditures sub-categories) terciles 1 and 3. For most states, the choice of λ^{r} is largely irrelevant, as the median of $\widehat{ATT}^{\mathrm{ratio}}$ are equal to one. For total revenues, in all state-terciles, the median $\widehat{ATT}^{ratio}$ is equal to or very close to one, meaning that in most cases $\widehat{ATT}^{L(\lambda)} \approx \widehat{ATT}^{CV(\lambda)}$. In some cases and for some sub-categories (e.g., state revenues in Alaska among tercile 1 districts) choice of λ^{r} is very consequential and the median $\widehat{ATT}^{\mathrm{ratio}}$ is much greater than 1. In such cases, the cross-validation objective function is useful and mitigates potential bias from researcher degrees of freedom (Gelman & Loken, 2013).

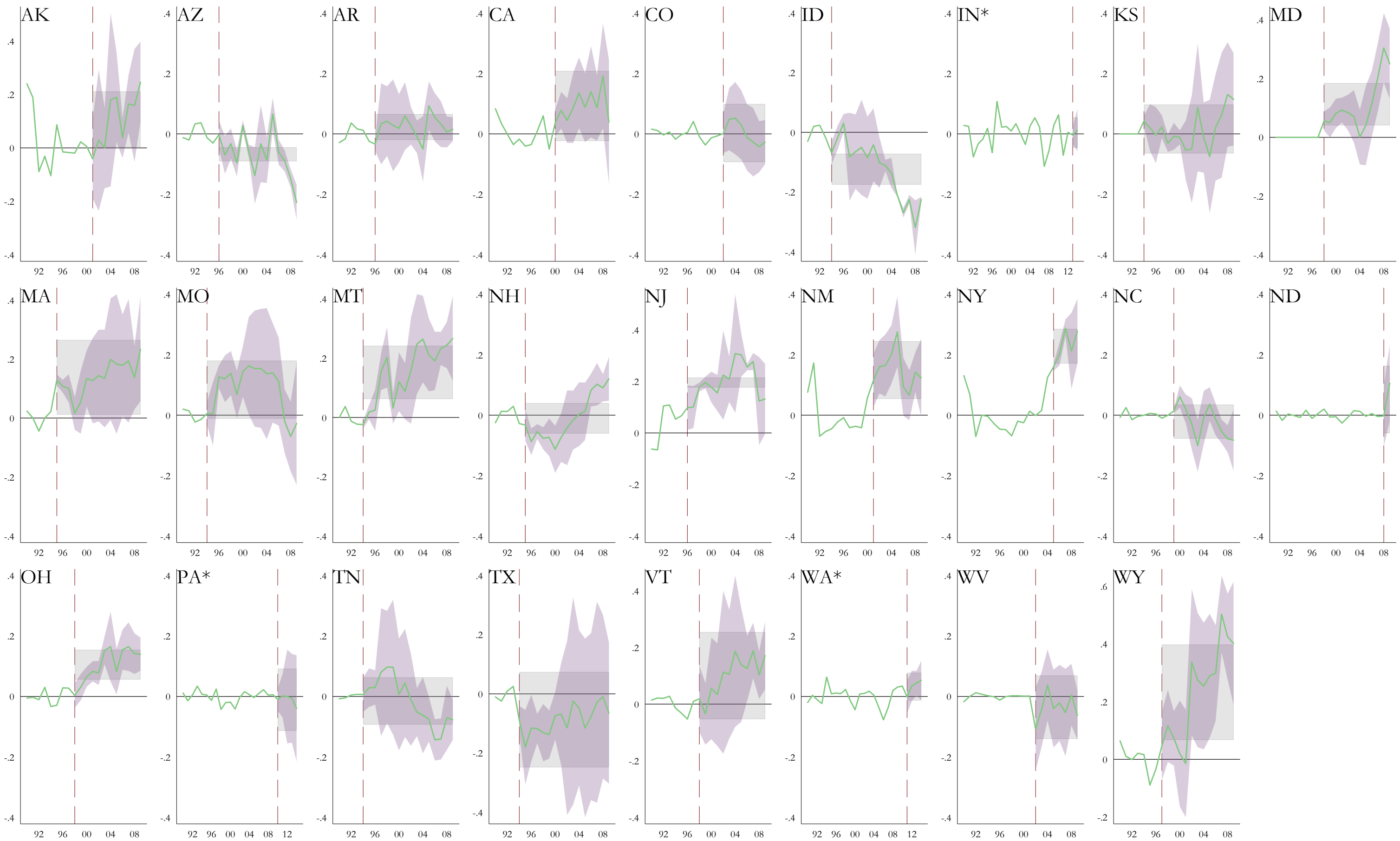


Figure G1. ASCM Effect Sizes by Year: Tercile 1 Log Per Pupil Total Revenues

Note: This figure plots the ATT of log per pupil total revenues for tercile 1 for each state-year relative to its ASCM counterfactual. The year-specific row-based jackknife 90% CIs are represented by the fluctuating darker range area; the average row-based jackknife 90% CIs are represented by the uniform lighter range area. All states except Wyoming have a common y-axis. For states with SFR after concurrent with the Great Recession—Indiana, Pennsylvania, and Washington—years 1990 to 2014 are shown; for all other states, years 1990 to 2008 are shown.

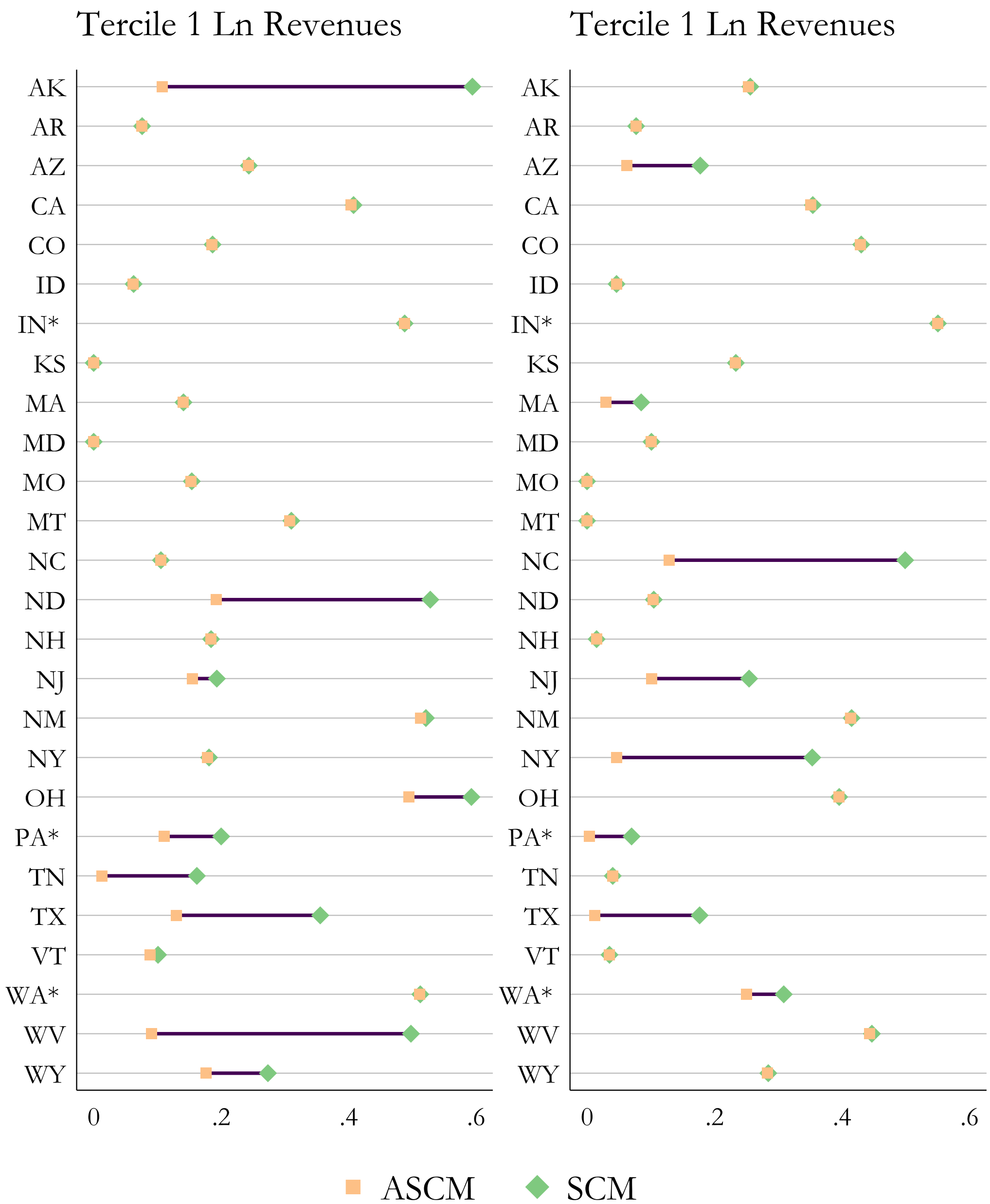


Figure G2. Model Fit Statistics for ASCM and SCM: Log(Revenues per Pupil)
Note: Model fit statistics are shown for traditional synthetic controls methods (SCM) and ridge augmented synthetic controls methods (ASCM). Match quality is defined as the cumulative pre-treatment effect size deviation from zero (the L^2-norm) using the weights from the respective approaches and scaled relative to an approach that applies uniform weights to all non-treated states. A value of 0 indicates that there is no cumulative deviation from zero in the pre-treatment match quality; a value of 1 indicates that the weights from SCM or ASCM are no better than applying uniform weights.

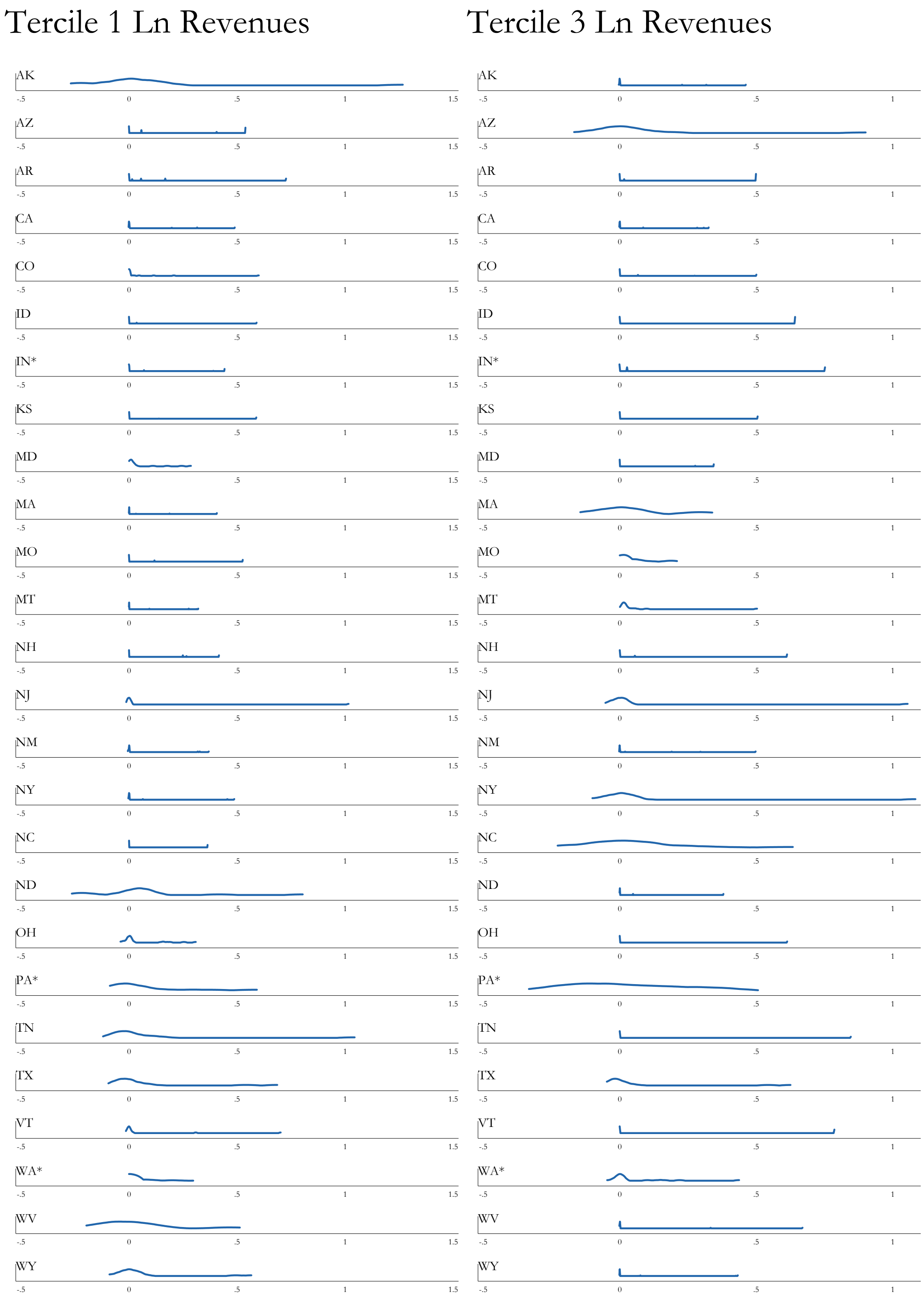


Figure G3. Distribution of Weights Assigned to Counterfacutals
Note: Kernel density estimates for the weights assigned to counterfactual states from the set of states that never had an SFR. Weights can be less than zero. Bunching at zero indicates that ASCM did not leverage negative weights; results in such cases are nearly identical to traditional SCM.

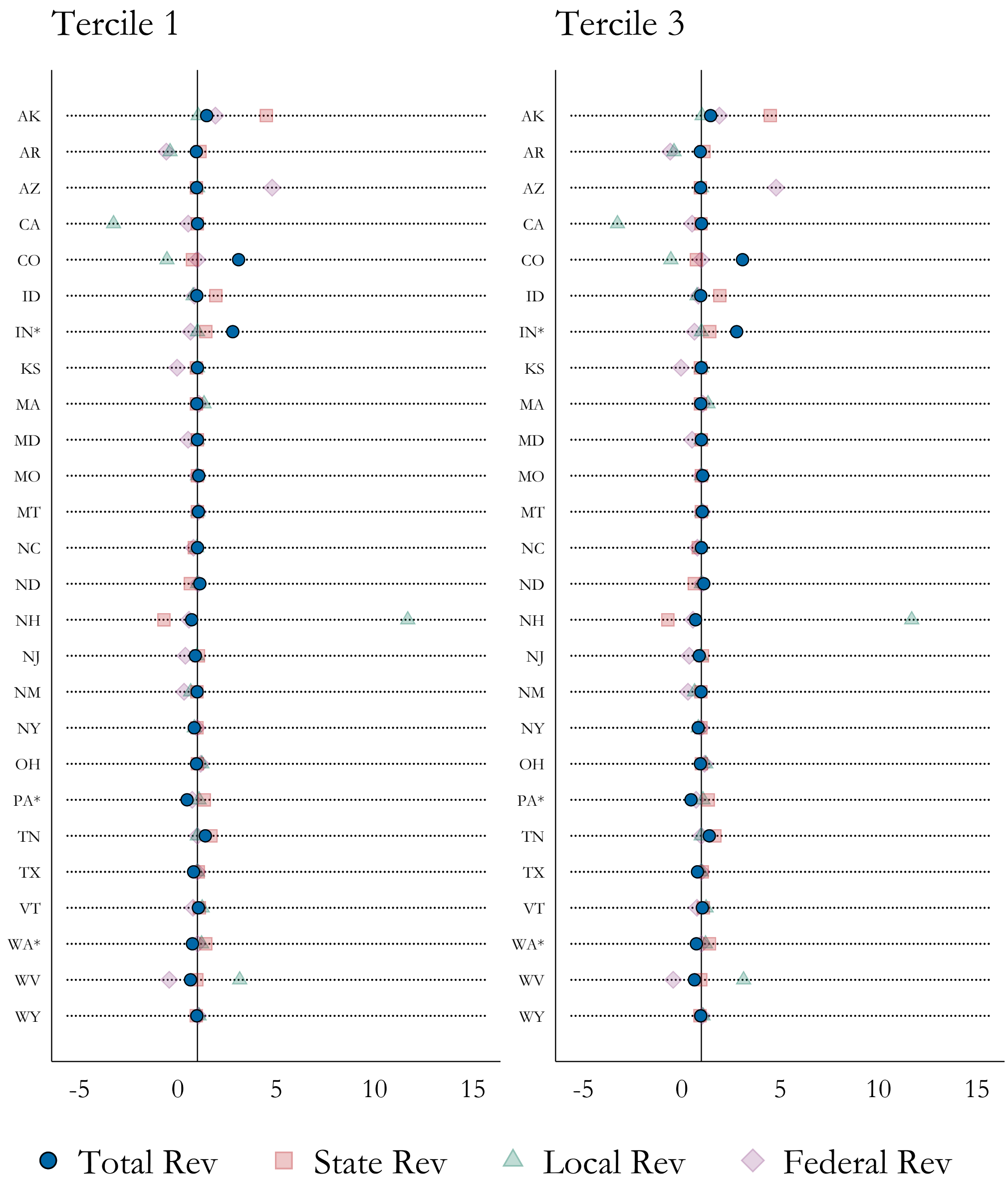


Figure G4. Stability of ATT as Function of λ^{r}: Log(Revenues per Pupil) and Revenues Sub-Categories

Note: Each dot corresponds to one of four values of $\widehat{ATT}^{\text{ratio}} = \frac{\widehat{ATT}^{L(\lambda)}}{\widehat{ATT}^{CV(\lambda)}}$, where L indexes one of 108 values of λ^{r} between 1×10^{-7} and 9×10^{5}, and CV indexes the value of λ^{r} obtained via cross-validation. The four values presented here represent the median of this ratio for log total revenues, log state revenues, log local revenues, and log federal revenues in tercile 1 districts. The vertical solid line indicates a value of 1, which implies that the median of these ATTs is equal to the ATT generated from λ chosen by cross-validation.

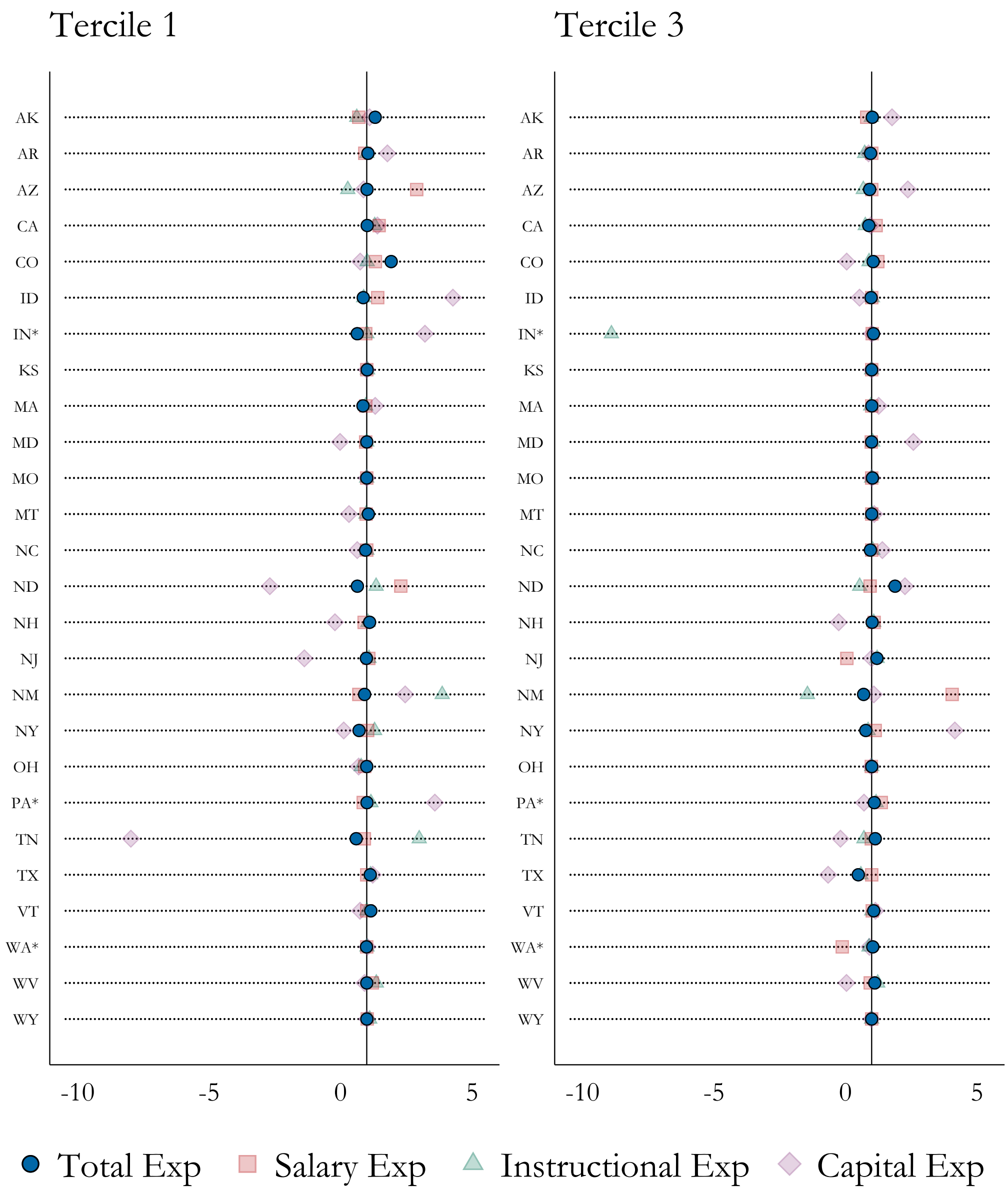
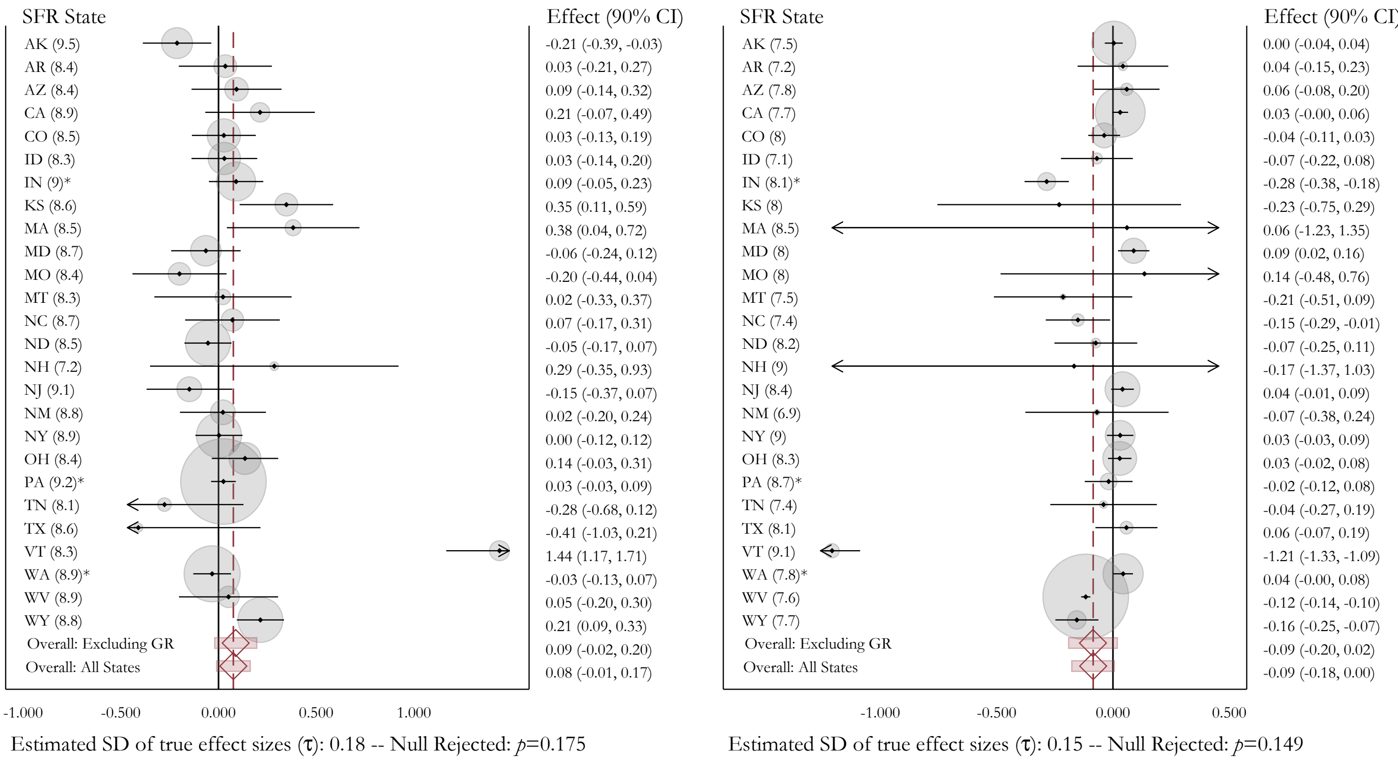


Figure G5. Stability of ATT as Function of λ^{r}: Log(Expenditures per Puipil) and Expenditures Sub-Categories

Note: Each dot corresponds to one of four values of $\widehat{ATT}^{\text{ratio}} = \frac{\widehat{ATT}^{L(\lambda)}}{\widehat{ATT}^{CV(\lambda)}}$, where L indexes one of 108 values of λ^{r} between 1×10^{-7} and 9×10^{5}, and CV indexes the value of λ^{r} obtained via cross-validation. The four values presented here represent the median of this ratio for log total revenues, log state revenues, log local revenues, and log federal revenues in tercile 1 districts. The vertical solid line indicates a value of 1, which implies that the median of these ATTs is equal to the ATT generated from λ chosen by cross-validation.

Appendix H

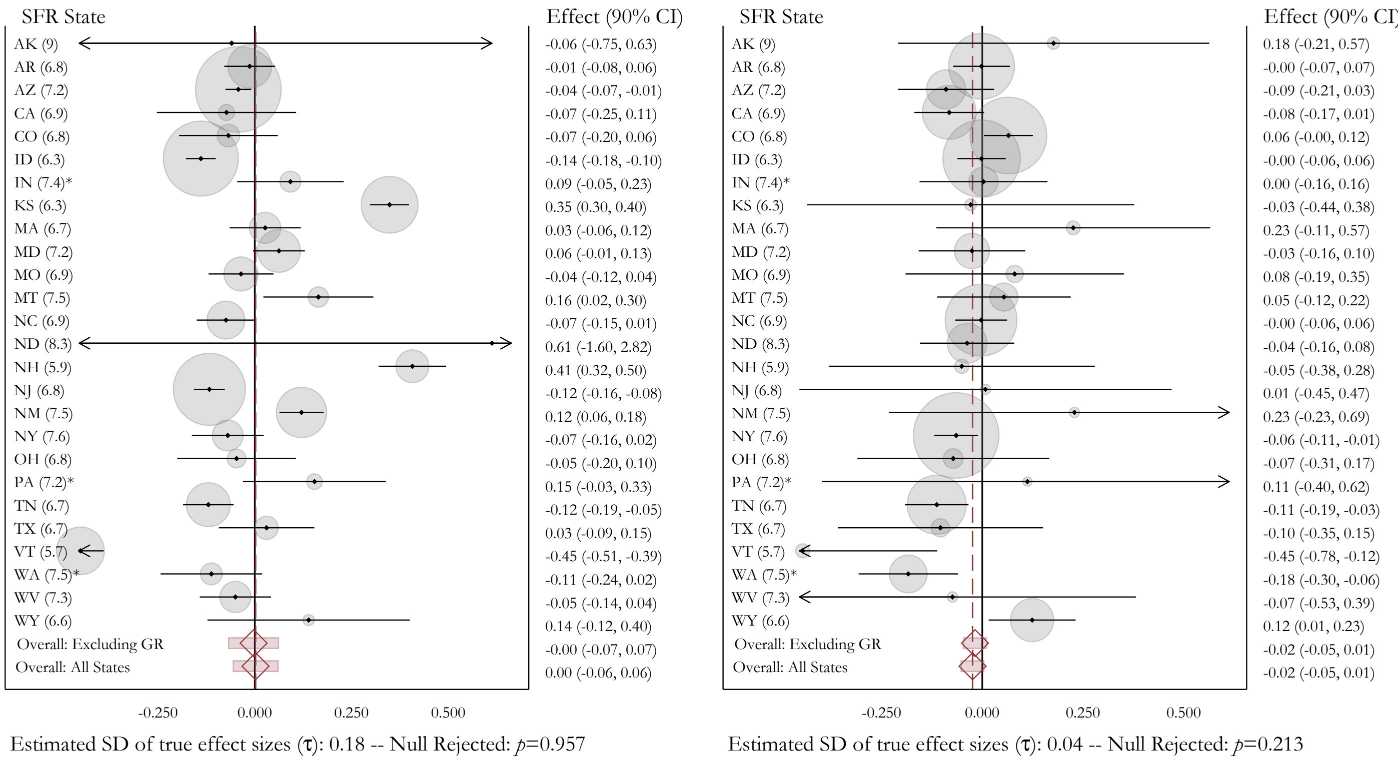
Additional Heterogeneity Results for Terciles 1 and 3



(a) Tercile 3: State Revenues

(b) Tercile 3: Local Revenues

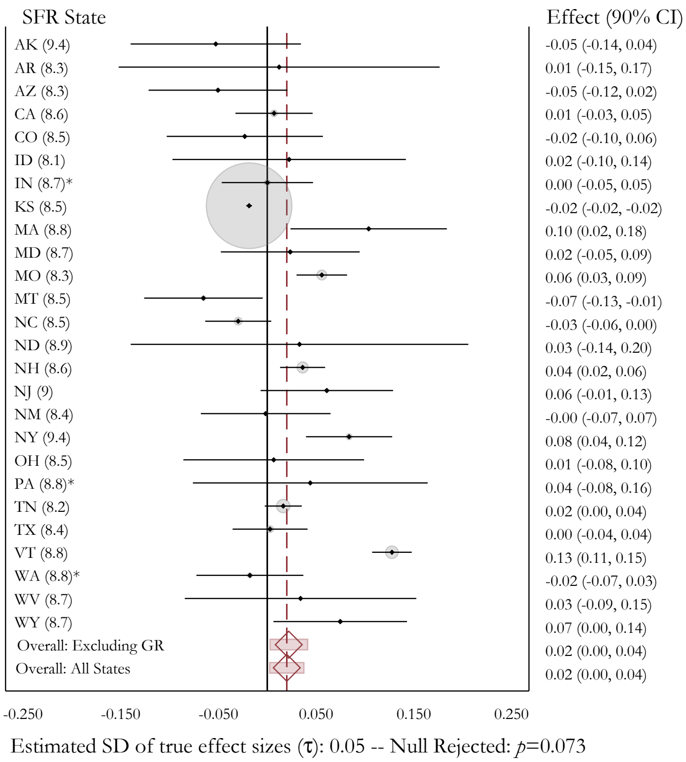
Figure H1. Tercile 3 Per Pupil Total State and Local Revenues
Note: Dependent variables are log of per pupil state and local revenues. Results for tercile 3 state and local revenues are shown in the first and second panels, respectively. The vertical dashed line and bottom-most diamond show the meta-analytic average of the state-specific ATTs. For IN, PA, and WA—states with an SFR during the Great Recession, marked with an *—we include data from 1989–90 to 2013–14; for all other states, we use data from 1989–90 to 2008–09. The meta-analytic average that excludes IN, PA, and WA is shown for reference as the penultimate diamond. The size of the shaded circle represents each state's contribution to the meta-analytic average. Error bars reflect 90% CIs; ⟶ or ⟵ indicate the CI exceeds the axis range. The right-hand column displays the effect size and 90 percent CI. Each number in parentheses after state names is the pre-SFR log value of the dependent variable.



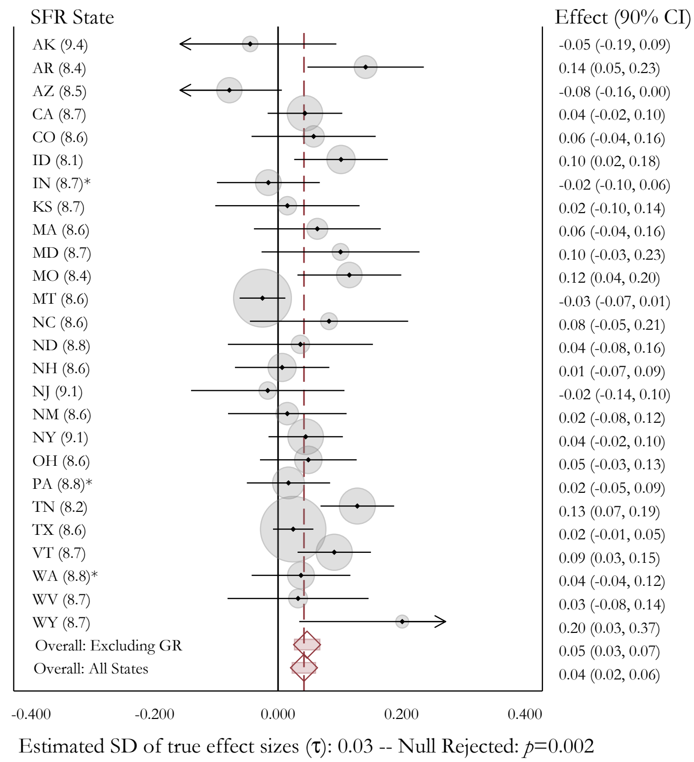
(a) Tercile 1

(b) Tercile 3

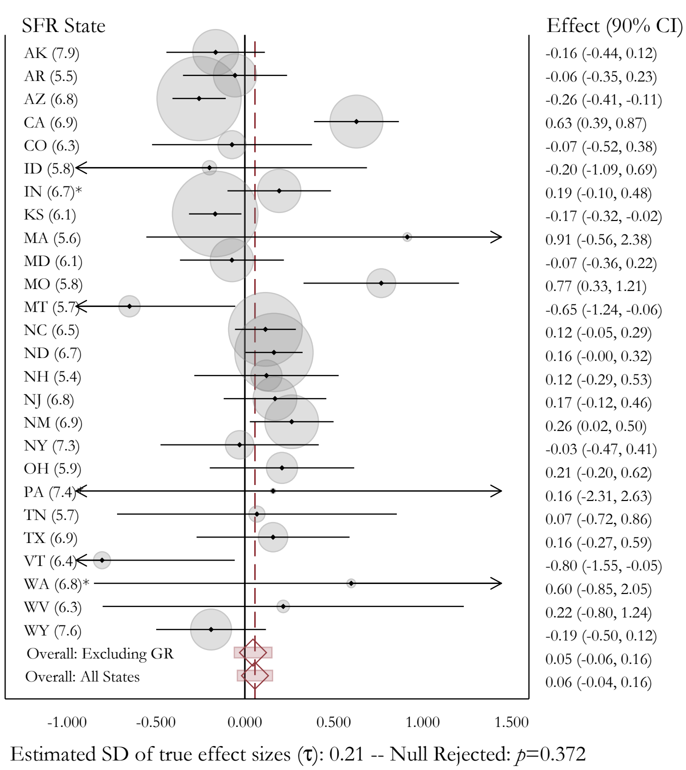
Figure H2. Per Pupil Federal Revenues
Note: Dependent variable is log of per pupil federal revenues. Results for terciles 1 and 3 are shown in the first and second panels, respectively. The vertical dashed line and bottom-most diamond show the meta-analytic average of the state-specific ATTs. For IN, PA, and WA—states with an SFR during the Great Recession, marked with an *—we include data from 1989–90 to 2013–14; for all other states, we use data from 1989–90 to 2008–09. The meta-analytic average that excludes IN, PA, and WA is shown for reference as the penultimate diamond. The size of the shaded circle represents each state's contribution to the meta-analytic average. Error bars reflect 90% CIs; ⟶ or ⟵ indicate the CI exceeds the axis range. The right-hand column displays the effect size and 90 percent CI. Each number in parentheses after state names is the pre-SFR log value of the dependent variable.



(a) Tercile 3: Instructional Expenditures



(b) Tercile 3: Salary Expenditures



(c) Tercile 3: Capital Outlays

Note: Dependent variables are log of per pupil instructional expenditures (a), log of per pupil local revenues (b), and log of per pupil capital (c) for tercile 3. The vertical dashed line and bottom-most diamond show the meta-analytic average of the state-specific ATTs. For IN, PA, and WA—states with an SFR during the Great Recession, marked with an *—we include data from 1989–90 to 2013–14; for all other states, we use data from 1989–90 to 2008–09. The meta-analytic average that excludes IN, PA, and WA is shown for reference as the penultimate diamond. The size of the shaded circle represents each state's contribution to the meta-analytic average. Error bars reflect 90% CIs; ⟶ or ⟵ indicate the CI exceeds the axis range. The right-hand column displays the effect size and 90 percent CI. Each number in parentheses after state names is the pre-SFR log value of the dependent variable.

Figure H3. Per Pupil Instructional, Salary, and Capital Expenditures

Appendix I

Tiebout Sorting following SFRs

After an SFR, we might expect parents to selectively move to areas where school spending has increased. To assess whether SFRs induce population composition changes, on average, within each tercile, we estimate the ASCM for terciles 1 and 3 using percent Black, percent Hispanic, percent White, and percent Free Lunch Eligible as outcomes. Relevant estimates are shown in Table I1. There is no strong evidence of Tiebout sorting. The percentage of Black students decreased by 0.30 percentage points in states with SFRs in tercile 1 districts ($p = 0.07$), representing about 1.67 percent of a standard deviation of the percent Black among treatment states in 1989–90. Likewise, the 0.38 percentage point decline in Hispanic enrollment in tercile 3 districts is statistically significant ($p = 0.08$) but practically trivial, representing 3.6 percent of a standard deviation of the percent Hispanic among treatment states in 1989–90.

Table I1
Assessing Tiebout Sorting

	Tercile 1				Tercile 3			
	ATT	Std. Error	τ	p-value	ATT	Std. Error	τ	p-value
Percent Black	-0.30	0.16	0.51	0.070	-0.13	0.11	0.39	0.25
Percent Hispanic	-0.0100	0.18	0.53	0.96	-0.38	0.21	0.60	0.080
Percent White	0.10	0.38	1.43	0.79	-0.41	0.27	0.78	0.14
Percent Free Lunch Eligible	-0.74	1.07	6.24	0.50	-0.68	0.58	2.87	0.26

Notes: Comparison states are those that never had an SFR between the sample period of 1989–90 to 2013–14 and treated states are those that had a least one reform during the sample period. The ATT is reported in percentage points (scale is 0 to 100).

Appendix J

Predictors of SFR Efficacy

Leveraging the 26 point estimates of SFR impacts on per pupil total revenues for districts in terciles 1 and 3, we now perform a descriptive analysis to assess the extent to which SFR-related policies predict variation in effect sizes among states. The descriptive analysis is conducted as a sequence of regressions weighted by the inverse variance of the effect size, where variance is the square of the row-based jackknife standard errors for the point estimates.

For funding formula variables, we include the modal funding formula used by state following the SFR. Each funding formula component (foundation plan, flat grant, equalization, power equalization, centralization, categorical aid, and spending limits) is entered as a binary indicator variable, with foundation plans as the reference category—all but three or six states, depending on whether the funding formula is contemporaneous or subsequent to the SFR, had a foundation plan. In separate regressions, we include indicator variables for whether the SFR was a single SFR induced by (a) the courts or (b) the legislature, or (c) whether the state experienced multiple SFRs. Then, in a separate regression limited to states with multiple SFRs, we test whether a statute following a court order differs from having multiple statutes or multiple court orders. To test whether SFRs vary in effectiveness over time, in separate regressions, we include a variable indicating what year the SFR took place, whether the SFR took place after the No Child Left Behind Act (NCLB), and whether the SFR took place during the Great Recession. Finally, following Brunner et al. (2020), we assess whether SFR effect sizes are larger in states with collective bargaining agreements or in states with teacher bargaining power is greater.

We report the results of our descriptive analysis in Figure J1, where we display point estimates and 90 percent confidence intervals. The top panel shows that the modal funding formulae the state adopted subsequent to its SFR are predictive of heterogeneity. Specifically, states that included equalization plans, or included categorical aid or spending limits tended to have more progressive outcomes, meaning that increases to tercile 1 revenues exceeded tercile 3 revenues. States that adopted categorical aid (and to a lesser extent spending limits) increased revenues in both tercile 1 and tercile 3 districts, but gains were larger in tercile 1 districts. In contrast, states that adopted equalization plans decreased revenues to tercile 3 districts, thus resulting in the observed progressivity. States adopting power equalization plans experienced regressive revenues allocations, as tercile 1 revenues decreased on average and tercile revenues were unchanged.

The second panels shows that the context in which the SFR was initiated has, in general, little bearing on outcomes. Progressivity estimates (the third column) are all indistinguishable from zero. Looking at tercile 1 and tercile 3 estimates separately, states with multiple SFRs were able to increase revenues to low-income districts on average, whereas revenues increased in high-income districts in states where a legislative statute followed a court order. These results provide some empirical support for Weishart (2019), who argues that litigation serves as a tool for ensuring the state maintains fidelity with its constitutional obligations. In other words, it is the act of litigating and not the resolution to litigation that ensures the state provides sufficient funding for low-income students. Similarly, our results suggest that statutes are less effective instruments for increasing spending to low-income students relative to litigation itself and court action.

The fourth through sixth panels (Timing of SFR, SFR after NCLB, and SFR coincides GR) show that the timing of SFRs is not associated with effect size heterogeneity. Variables indicating linear time and indicators for whether the SFR took place after NCLB or concurrently with the Great Recession are all close to zero and not statistically significant.

Finally, panels seven through ten (Fordham Institute Union Strength, CBA, RTW, and Union Index (Brunner et al., 2020)) test whether state-level union strength predicts effect size heterogeneity. Like Brunner et al. (2020), we find consistent evidence that collective bargaining strength predicts SFR progressivity, increasing revenues in bottom-income districts, decreasing or not affecting revenues in top-income districts, and increasing progressivity overall. These results are consistent for each of the measures of union strength we include.

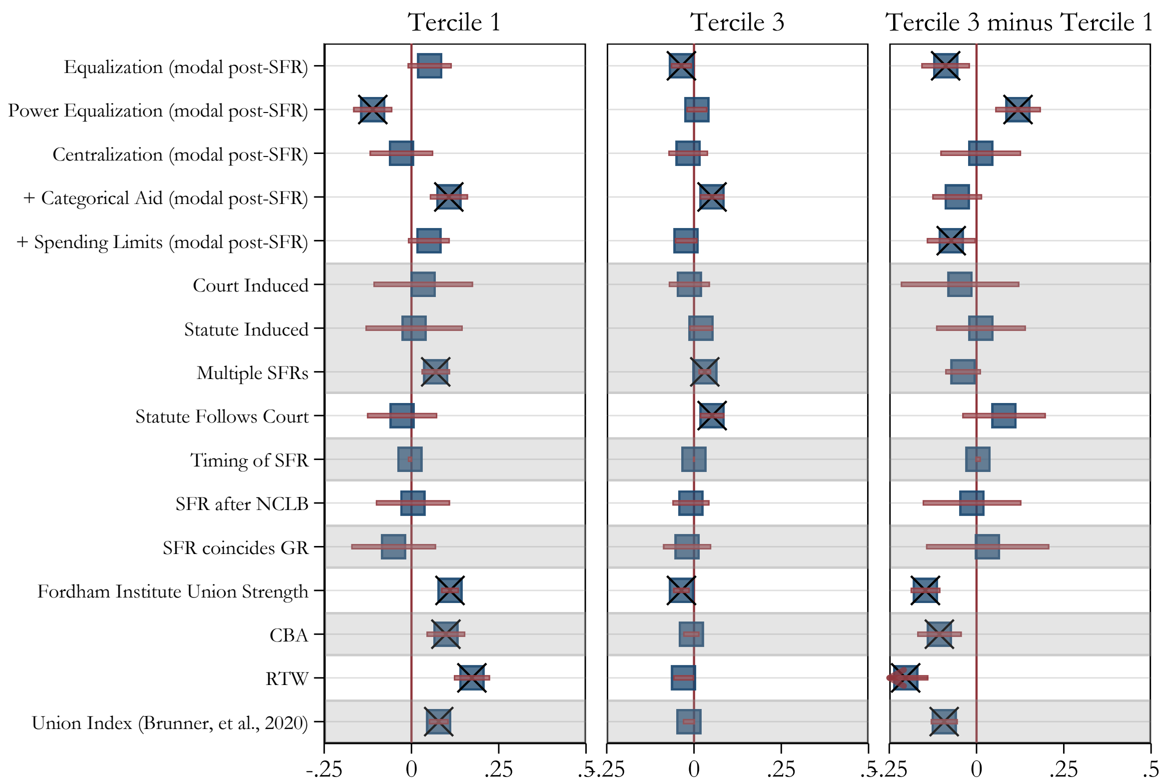


Figure J1. SFR policy variables and estimated effect sizes for log total expenditures
Note: List of predictors and data sources are taken from Appendix: Additional Predictors. Individual regressions are demarcated by alternating gray/white panels. From the top, the first **white** panel includes indicators for funding formulae components: Equalization, Power Equalization, Flat Grant, Centralization, +Categorical Aid, and +Spending Limits, where Foundation Plans are the reference category. The second **gray** panel includes indicators for whether the SFR had a single court order, a single statute, or had multiple court orders and/or statutes; the reference category is omitted. The third **white** panel includes an indicator for whether a statute follows a court order, restricted to states with multiple SFRs; the reference category includes states with only court rulings. Panels 4 through 6 (Timing of SFR, SFR after NCLB, SFR coincides GR) test whether the timing of the SFR predict variation in ATT effect sizes. For Timing of SFR (**gray**), the event year is entered continuously, for SFR after NCLB (**white**), an indicator is included if the SFR takes place after 2001, and for SFR coincides GR (**gray**), an indicator is included if the SFR takes place after 2008. Panels 7 through 10 (Fordham Institute, CBA, RTW, Union Index) test whether a state's union strength predict variation in ATT effect sizes. For Fordham Institute (**white**), a continuous descriptor from Fordham Institute is included; for CBA (**gray**), a continuous indicator of the state's collective bargaining climate in 1987 is included (Brunner et al., 2020); for RTW (**white**), an indicator of whether the state had right-to-work laws in 1987 is included; and for Union Index (**gray**), we combine the CBA and RTW variables to create a continuous indicator of union strength (Brunner et al., 2020). Range caps are 90% confidence intervals, and an X indicates $p \leq 0.10$. Results for tercile 1 districts, tercile 3 districts, and differences between tercile 1 and 3 districts are shown in the left, middle, and right panels, respectively.