EdWorkingPaper No. 19-52

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

Kenneth A. Shores
University of Delaware

Christopher A. Candelaria
Vanderbilt University

Sarah E. Kabourek
University of Chicago

Sixty-seven school finance reforms (SFRs) in 27 states have taken place since 1990; however, there is little empirical evidence on the heterogeneity of SFR effects. We provide a comprehensive description of how individual reforms affected resource allocation to lowand high-income districts within states. We then examine whether characteristics of the SFR, such as the funding formula that was adopted, predict effect size heterogeneity. Taken together, this research aims to provide a rich description of variation in states' responses to SFRs, as well as explanation of this heterogeneity as it relates to contextual factors.

VERSION: May 2020

Suggested citation: Shores, Kenneth A., Christopher A. Candelaria, and Sarah E. Kabourek. (2020). Spending More on the Poor? A Comprehensive Summary of State-Specific Responses to School Finance Reforms from 1990–2014. (EdWorkingPaper: 20-52). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/5s1v-yr69

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: May 11, 2020

Author Note

For correspondence, please contact co-first authors Shores and Candelaria at kshores@udel.edu and chris.candelaria@vanderbilt.edu.

Abstract

Sixty-seven school finance reforms (SFRs) in 27 states have taken place since 1990; however, there is little empirical evidence on the heterogeneity of SFR effects. We provide a comprehensive description of how individual reforms affected resource allocation to low- and high-income districts within states. We then examine whether characteristics of the SFR, such as the funding formula that was adopted, predict effect size heterogeneity. Taken together, this research aims to provide a rich description of variation in states' responses to SFRs, as well as explanation of this heterogeneity as it relates to contextual factors.

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

There has been a spate of school finance reforms (SFRs) since 1990: sixty-seven reforms in 27 states. One reason for this activity is that SFRs are an effective policy for increasing spending to lower-income districts. As recent national studies show, SFRs increase spending in poorer districts (Candelaria & Shores, 2019; Jackson, Johnson, & Persico, 2016; Lafortune, Rothstein, & Schanzenbach, 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 national audits of SFRs, however, are likely to overlook important variation among states.

We expect heterogeneity across states because the history of SFRs, beginning in the 1990s, is one of diversity. Some states changed their finance system through court order; others changed it because of legislative activity; still others changed it 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 kept the funding formula structure but changed its components or weights; and, in general, funding formulas varied substantially across states. Finally, some states were sued because facilities were deemed inadequate; others were sued because aggregate spending was inadequate.[1]

This paper examines state-level variation in the effect sizes of SFRs on school 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 (e.g., its adopted funding formula, whether it was sued multiple times, or the number of times the funding formula was changed) are predictive of SFR impacts. Of 27 states with an SFR during this period, only only five—Kansas, Kentucky, Maryland, Massachusetts,

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

and Vermont—have been evaluated. This small sample of state-level case studies provides an incomplete picture the SFR heterogeneity landscape.

Our study is motivated by the idea that understanding the variability of SFR effects across different contexts can be useful for policymakers. As we demonstrate, SFRs exhibit substantial variability in terms of their effects on spending and resource allocation; therefore, pursuing a reform can be a risky option for policymakers, 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.

Low-income districts are differentially impacted by SFRs relative to high-income districts, on average (Candelaria & Shores, 2019; Lafortune et al., 2018; Sims, 2011); therefore, we describe 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. Specifically, we compute the average level of resources across districts in the bottom tercile of the income distribution and across districts in the top tercile within each state. This approach accounts for heterogeneity of SFR impacts within states while also providing a way to assess the progressivity of reforms between terciles.

Our analytic strategy involves two steps: first, obtain causal estimates of effect size heterogeneity at the state-by-district income tercile level; second, link these estimated effect sizes to variables describing the SFR policy environment for purposes of descriptive analysis. To estimate state-by-district income tercile impacts, we use a de-biased synthetic controls approach—the "augmented" synthetic control method, which was recently developed by Ben-Michael, Feller, and Rothstein (2019). This method enables us to construct a unique comparison group—that is, a synthetic control group—for each state-by-district tercile that was exposed to an SFR. Using the augmented synthetic control method, we obtain estimates that describe how a state's SFR changed resource patterns, in terms of expenditure levels and expenditure category, at the state-by-district income tercile level for each of the 26 states that had SFRs between 1990 and 2014. [2] With these estimated effects, we then conduct descriptive analyses by linking these effects to variables describing the SFR policy environment. Based on a review of the literature, we examine the state's funding formula, whether the state changed funding formula multiple times, whether the state experienced multiple SFRs, and whether legislation was passed following a ruling from the state supreme court.

The short version of our findings is as follows: in the aggregate, summarizing all estimated responses to SFRs, expenditures increased by 6 percent in low-income districts and 1 percent 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. Across low-income districts, a 10 percent increase in total spending corresponds to a 19 percent increase in capital spending and a 6 percent increase in salary spending.

At the same time, the heterogeneity of responses may temper enthusiasm for SFRs. First, 2 of 26 SFRs in this period decreased spending in low-income districts, and in 14 states there was no change in low-income spending. Further, our results show that low-income districts increased capital spending versus increasing salaries or reducing class sizes. 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.

Despite the prevalence of SFR activity, expectations that SFRs will have

[2] 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. See details in the Research Methods section.

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 sophisticated methodological approaches can be used to evaluate treatment effect variation in settings where randomization is impossible. Thus, our study complements 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 (a) an aggregate effect combining SFRs across states or (b) an SFR-specific effect, based on a single reform in a given state. In the aggregate, recent studies leveraging the timing of an SFR as an exogenous shock to school spending have found consistently 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 aggregate 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 specifically target capital expenditures, resulting in capital expenditure increases (Jackson et al., 2016), which have no direct impact on class sizes. However, capital expenditures 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.[3]

SFR-specific impact evaluations would be useful for understanding the heterogeneity of impacts on total spending, as well as for understanding variability in possible mechanisms through which different resource allocations could affect student outcomes. Unfortunately, only 5 (of 27) states have been evaluated during this period, and among these studies, there has been little attention to mechanisms. 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).[4] Among these, results range from moderate spending increases with little improvement to student outcomes (Kansas, Kentucky, and Maryland) to increases in both spending and academic outcomes (Massachusetts and Vermont).

[3] 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 nothing.

[4] 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).

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 expenditures, teacher salaries, capital expenditures, and class sizes.

In addition to understanding the heterogeneity of SFR-impacts among states, we also conduct a descriptive analysis to assess whether SFR-related policy variables influence SFR progressivity. We do this by identifying factors that characterize the school finance context in the reform states. Variability in funding formulas will determine how much aid is allocated to low-income districts, as some formulas, 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), and whether the state changed its funding formula multiple times.

Data

To understand how SFRs impacted resources and to understand how these impacts varied by state-by-district 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 provided supplemental research on case histories and because they have a more recent list. 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 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 expenditure subcategories: current expenditures on elementary and secondary education, instructional staff support services, capital outlays, and teacher salaries. The panel data set of fiscal outcomes we assemble spans academic years 1989–90 to 2013–14.[5] In our analyses, we scale 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).

We also draw data from the Public Elementary and Secondary School Universe Survey, which is distributed by NCES. From this survey, extract full-time equivalent counts of teachers in each school across the United States between 1989–90 to 2013–14. Using local education agency codes, we aggregate these data to the district-by-year level and merge it with our fiscal data. We then generate a ratio of teachers per 100 students for analysis.

To address volatility in our 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. Specifically, we:

1. Remove small districts in which the total enrollment is less than 100.
2. Remove district-year observations in which enrollment exceeds mean district enrollment by scale factor 2.
3. Remove district-year observations in which enrollment is greater than 15% of the district average.
4. Remove district-year observations in which enrollment is greater than 10% above or below the district's constant growth rate trend.
5. Remove an entire district from the analytic sample if the restrictions (1) to (4) above cause the district to have more than 33% of its observations removed.

[5] 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.

We take each of the outcome measures that are to be scaled by district enrollment and divide by this new restricted enrollment variable. All fiscal variables are then log transformed and non-fiscal variables remain in levels. 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).

To assess heterogeneity across the income distribution, we place districts into income terciles based on their state's 1989 median income levels, which 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-by-district income tercile and year, we then compute the weighted median of our outcome variables of interest, where the weights are based on the annual district restricted enrollment described above. Finally, because identifying synthetic counterfactuals can be biased if there is measurement error or volatility in the dependent variable (Abadie, Diamond, & Hainmueller, 2015; Powell, 2018), we smooth the data by taking three-year moving averages as a final data transformation.

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.

Summary statistics for all outcome variables are shown in Table 1. Sample means and standard deviations are computed among states that had a court-ordered or legislative reform between 1989–90 and 2013-14. We provide statistics for terciles 1 and 3, corresponding to lower-income and higher-income districts, respectively. Outcome means between the terciles are statistically different from each other at the 5 percent level, except for log per pupil capital expenditures; the mean difference for capital expenditures between terciles 1 and 3 is statistically significant at the 10 percent level.

[Table 1 about here.]

Predictors of SFRs

To explain heterogeneity based on the nature of the SFR context, 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.

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, centralization, spending limits, and categorical aid. We identify two additional "add-on" components of the state funding formulas 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, as at least one component of their funding formula, a foundation plan. Funding formula 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. Additional details about the construction of the funding formula panel can be found in Appendix B; tabulations of states with SFRs and the funding formulas present in the state following an SFR are shown in Appendix C.

Research Methods

To estimate effect sizes for each state that underwent an SFR, we use the ridge augmented synthetic control method (ridge ASCM) developed by Ben-Michael et al. (2019). In what follows, we discuss how to construct counterfactuals in a case-study environment such as ours using synthetic control methods (SCM), introduce notation for these approaches, and then introduce and contrast ridge ASCM.

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

Our goal is to provide individual effect sizes (ATTs) for each state-by-district income tercile that experienced an SFR.[6] 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 Colorado's SFR in 2001 as a case study.

In the left panel of Figure 1, in blue, tercile 1 log per pupil total expenditures are shown for Colorado and, in gray, the same data are shown for all other states not experiencing an SFR. Colorado's 2001 legislative act "Bill 181" is displayed as a vertical line. If we only compared Colorado's spending before and after 2001 (i.e., option 1), ignoring the non-SFR comparison states, it would appear as if Colorado experienced a short increase in spending before returning to baseline trends. However, we worry that the short-term increase in spending may have been due to a common shock experienced by other states; thus, it would be useful to have a comparison group.

[Figure 1 about here.]

[6] The reader can feel free to skip ahead to a more formal treatment of the problem in the next sub-section if desired.

Options (2) through (4) represent different approaches to constructing comparison groups and are shown in the right panel of Figure 1. The short-dashed red line allows each of the non-treated states to serve equally as counterfactual, or control, states to Colorado (option 2). Here, we simply take the difference between Colorado's tercile 1 log per pupil total expenditures from the mean tercile 1 log per pupil expenditures 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 Colorado's pre-SFR spending levels or trajectories, as they tend to have lower spending overall and their spending was increasing at greater rates than Colorado prior to its SFR; this result suggests that non-SFRs do not uniformly make for good counterfactuals to Colorado.

Options (3) and (4) use matching methods to identify states that most resemble Colorado'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. Traditional SCM does much better than naïvely assigning each control state the same weight, as was done with option (2) and shown in short-dashed red. Yet, we see some evidence of poor match quality, especially in years 1990–91, 1998–99, and 1999–00, evidenced by the long-dashed yellow deviations from zero in those years. Ridge ASCM, option (4), provides further improvement and shows no sign of deviations from zero in the pre-period.

The consequences for these different approaches can be seen in the post-period beginning in 2001. Option (2) shows that Colorado's SFR resulted in spending losses relative to control states. Options (3) and (4) tell a different story: a sharp increase in spending immediately following the SFR and then a dramatic decline beginning in 2005.

We now turn to more formal notation to explain the SCM for estimating unit-specific effect sizes and how ridge ASCM improves on the SCM estimator.

SCM Notation and Setup

Using notation from Abadie, Diamond, and Hainmueller (2010) and Ben-Michael et al. (2019), 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.[7] The remaining S states serve as a pool of control group states, as they did not have SFRs. Given that 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.[8] 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 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

[7] 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.

[8] 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.

and denotes each pre-treatment outcome using X_{st}; the pre-treatment outcomes are covariates in our SCM. Collapsing the rows and columns 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 $Y_{1t}(0)$. In contrast, SCM leverages $X_{0\cdot}$ to find and weight control units that most resemble $X_{1\cdot}$.

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 satify 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 (Ben-Michael et al., 2019; Doudchenko & Imbens, 2017)—for each $t \in \{1, \ldots, T_0\}$—with the purpose of setting differences between treatment and control equal to zero.[9] Then, we use the estimated weights, $\hat{w}_s^{\text{SCM}}$, and apply them to the outcomes of the $S+1=23$ 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

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

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 and when those pre-treatment outcomes

[9] 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)).

are reliable predictors of post-treatment outcomes. As suggested by Ben-Michael et al. (2019), 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.[10]

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

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

Ridge augmentation can reduce bias by increasing pre-treatment fit, but it can also

[10] 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.

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. (2019), 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, $\bar{\hat{\gamma}}_1^{\text{aug}}$, to compare effect sizes across states. Our reported standard errors are based on a row-based jackknife to allow for autocorrelation within states (Ben-Michael et al., 2019; Doudchenko & Imbens, 2017).

Assessing Ridge ASCM Estimates

We summarize the quality and stability of our synthetic control efforts in two ways. First, we show model fit statistics for the ridge ASCM and SCM methods for log per pupil total expenditures 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 2, we show the scaled model fit statistics for log per pupil total expenditures in terciles 1 and 3.

[Figure 2 about here.]

As is evident (and expected), ASCM weights always perform as well or better than SCM weights. In some instances (e.g., for Idaho, Kansas, Massachusetts), ASCM weights are as good as SCM weights, meaning that pre-treatment match quality cannot be improved with the de-biasing from ridge regression. In many cases (e.g., Alaska, California, North Dakota), ASCM weights greatly improve upon SCM weights, meaning that pre-treatment match quality was improved with ridge augmentation.

Second, 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 $\widehat{ATT}^{\text{ratio}}$ as $\widehat{ATT}^{L(\lambda)} / \widehat{ATT}^{CV(\lambda)}$, where $\widehat{ATT}^{L(\lambda)}$ is the estimated average ATT for specification L given penalty parameter λ^{r} and $\widehat{ATT}^{CV(\lambda)}$ is the estimated average ATT obtained from the cross-validation procedure described above. The statistic $\widehat{ATT}^{\text{ratio}}$ indicates the ratio of the average ATT for a specified λ^{r} relative to our preferred cross-validation estimate.

In Figure 3, we show the mean $\widehat{ATT}^{\text{ratio}}$, as well as its extremes at the 1st and 99th percentiles, for log expenditures terciles 1 and 3. For most states, the choice of λ^{r} is largely irrelevant, as the mean, and 1st and 99th percentiles of $\widehat{ATT}^{\text{ratio}}$ are equal to one. In all but three cases, for tercile 1 log per pupil total expenditures, the mean $\widehat{ATT}^{ratio}$ is equal to or very close to one, meaning that in most cases $\widehat{ATT}^{L(\lambda)} \approx \widehat{ATT}^{CV(\lambda)}$. Yet, in some cases (e.g., Alaska, Indiana, and Washington among tercile 1 districts) choice of λ^{r} is very consequential and the average ATT fluctuates widely. In such cases, the cross-validation objective function is useful and mitigates potential bias from researcher degrees of freedom (Gelman & Loken, 2013).

[Figure 3 about here.]

Results

Our results proceed as follows: (1) we first show dynamic ATT effect sizes for each state-by-district income tercile undergoing an SFR. We then (2) compute average ATTs for terciles 1 and 3 in each state and use these averages to describe heterogeneity for multiple resource types. Finally, (3) to better understand the SFR environment in which SFRs are productive, we leverage the state-by-district income average ATTs as outcome variables in prediction models with covariates derived from the SFR political landscape.

Effect Sizes over Time

Results from ASCM for log per pupil total expenditures terciles 1 (low-income districts) and 3 (high-income districts) are shown in Figure 4. The solid line corresponds to tercile 1; the dashed line to tercile 3. This figure illustrates both the quality of the pre-treatment match—as the effect size for nearly all states is perfectly balanced around zero and, in most cases, does not deviate far from zero—and the dynamic change in tercile 1 and 3 spending following a state's SFR.

As is evident, there is heterogeneity both within and across state-district income terciles. Within state treatment effect heterogeneity is notable in states like Missouri, in which an initial positive treatment effect reverses five years after the SFR. Such dynamism is important to model, and is a known source of bias in traditional difference-in-differences models de Chaisemartin and d'Haultfoeuille (2019); Goodman-Bacon (2018). Our case study approach to estimating treatment effects avoids this problem. Of greater interest to us is the effect size heterogeneity between states, both with respect to the tercile 1 estimates, and the differences between the tercile 1 and 3 estimates. Below, we take a more formal approach to describing effect size heterogeneity.

[Figure 4 about here.]

Effect Size Heterogeneity

To distinguish between within study variance (i.e., sampling variability) and between study variance, 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). With these statistics, heterogeneity can then be quantified with the I^2 statistic ($I^2 = 100 \times (Q - K - 1)/Q$), which describes the total proportion of variance due to heterogeneity (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003).

State-district income tercile effect sizes for per pupil total expenditures, capital expenditures, salary expenditures, and teachers per 100 students are shown in Figures 5, 6, 7, and 8 below. Results for tercile 1 are shown in the first panel, and results for tercile 3 are shown in the second panel; for each outcome variable, states are sorted alphabetically. The vertical dashed line and diamond shows the precision-weighted average of the ATTs from each SFR state, and the magnitude of the state-specific weights is shown on the right column. The magnitude of the pooled ATT is shown in the last row of the right column, and the I^2 statistic is shown in the last row of the left column. The displayed error bars indicate 95% 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. Finally, unweighted correlations between tercile-1 and -3 effect sizes are shown parenthetically in the sub-titles.

Per Pupil Total Expenditures. In Figure 5, the population average ATTs for terciles 1 and 3 districts are 6% and 1%, respectively, meaning that, on average, SFRs increased spending in bottom-income districts but not top-income districts. However, this average belies considerable heterogeneity. For instance, the I^2 statistic for terciles 1 and 3 are 82.3% and 74.7%, respectively, indicating considerable heterogeneity among states (Julian & Higgins, 2019). For tercile 1 districts, estimated effect sizes range between a 17% reduction in total spending (Arizona) and a 29% increase in total spending (Wyoming), and the precision-weighted 95% confidence interval ranges between 2% and 11%. For tercile 3 districts, the implied 95% confidence interval is between -2% and 4%. In total, SFRs increased spending in 10 of 26 state-district income terciles ($p < .1$); therefore, spending among low-income districts in more than a one-half of states undergoing SFRs did not increase or decrease relative to synthetic counterfactuals. Effect sizes in terciles 1 and 3 are moderately correlated ($\rho = 0.69$), with an implied elasticity of 0.46 (i.e., a 1% increase in tercile 1 spending is associated with a 0.46% increase in tercile 3 spending).

[Figure 5 about here.]

Per Pupil Capital Expenditures. For capital expenditures, in Figure 6, the I^2 ranges between 70.7% to 78.7%, again demonstrating real variation among states. The implied 95% confidence intervals are between 2% and 41% and -8% and 21% for districts in terciles 1 and 3, respectively. On average, capital spending increased by 22% and by 6% in district terciles 1 and 3, respectively. Thus, SFRs have a progressive effect on capital expenditures. In total, SFRs increased capital spending in tercile 1 districts in 6 states ($p < .1$), meaning that 20 states either did not increase or decreased capital spending. The implied elasticity between total and capital spending is 1.9%, meaning that a 1% increase in tercile 1 total spending is associated with a 1.9% increase in tercile 1 capital spending.

[Figure 6 about here.]

Per Pupil Salary Expenditures. In Figure 7, for salary expenditures, the I^2 ranges between 55.8% to 79.9%. The implied 95% confidence intervals are between -2% and 6% and 2% and 11% for districts in terciles 1 and 3, respectively. On average, tercile 1 districts increased salary expenditures by 2% and, in tercile 3 districts, by 6%. Thus, SFRs have a *regressive* effect on salary expenditures. In total, SFRs increased salary spending in tercile 1 districts in 6 states ($p < .1$), meaning that 20 states either did not increase or decreased salary spending. The implied elasticity between total and salary spending is 0.6%, meaning that a 1% increase in tercile 1 total spending is associated with a 0.6% increase in tercile 1 salary spending.

[Figure 7 about here.]

Teachers per 100 Students (Class Sizes). Our estimates for class size effects are very imprecise, as shown in Figure 8; the I^2 is 0% for both tercile 1 and 3 districts, meaning that all of the variation in the effect sizes comes from sampling variation. The precision-weighted average of the ATTs is 0.13 and 0.30 for terciles 1 and 3, respectively. Thus, for both salary expenditures and class sizes, we observe greater improvements, on average, in tercile 3 districts relative to tercile 1 districts.

[Figure 8 about here.]

Comparison to Earlier Case Studies. The results for total expenditures described above are mostly aligned with the limited number of prior case studies conducted on state-specific SFRs. For Massachusetts, Maryland and Vermont, the positive impact of their states' respective SFRs match work by Chung (2015, Maryland), Dee and Levine (2004) and Guryan (2001) (Massachusetts) and Downes (2004, Vermont). Studies of Kansas (Duncombe & Johnston, 2004; Johnston & Duncombe, 1998) identified limited impacts of their respective SFRs. Our results differ from these earlier studies in that the ATT for Kansas is not different from zero ($p < 0.1$). Note that our time-series extends beyond these earlier studies, and from Figure 4, Kansas did experience a brief uptick in tercile 1 spending coinciding with a brief decline in tercile 3 spending, which mostly corroborates these earlier studies. However, shortly after Kansas' SFR, tercile 1 total spending declined dramatically and tercile 3 total spending increased dramatically, thus eliminating the initial progressive effects of Kansas' SFR. In general, the fact that our results mostly align with prior case-studies—studies that relied on different methodologies and counterfactuals—lend credibility to our analytic strategy.

Two aspects of these results are worth considering in more detail, as they deviate from our current understanding of SFR impacts. The first relates to the allocation of expenditures following an SFR, and the second relates to the non-significant effect sizes for many state-district income terciles.

Discussion of Heterogeneity Results: SFRs Increase 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 salaries or class sizes.[11]

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, for example increasing attendance. Moreover, depending on the level of environmental contaminants present in existing older schools, new school construction can plausibly be linked to increased cognitive performance by reducing exposure to those contaminants (e.g. Persico & Venator, 2019). Finally, recent work has cast doubt on the universality of positive effects resulting from reduced class sizes Leuven and Løkken (2020), meaning that we need not expect the 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, a recent working paper by Baron (2019) shows that bonds earmarked for capital do not produce academic gains, but 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.

[11] We note here that in Candelaria and Shores (2019), we also find that high-poverty districts increased capital and not salary spending as a result of court-ordered SFRs, though we did not report these results in the published paper. Taken together, these results suggest the divergence may have less to do with methodology and more to do with the SFR era being studied, since Jackson et al. (2016) mostly leverage pre-1990 SFRs.

Discussion of Heterogeneity Results: Non-Significant Impacts of SFRs. Many studies have leveraged SFRs (e.g., Brunner, Hyman, Ju, et al., 2018; Candelaria & Shores, 2019; Jackson et al., 2016; Lafortune et al., 2018) to recover exogenous variation in spending that can, in turn, be linked to student outcomes. One possible conclusion from this literature might be that SFRs are an especially useful way to increase spending to low-income districts. Our results suggest that they are effective in the aggregate, but low-income districts in SFR states do not consistently increase spending after a reform.

Among tercile 1 districts, SFRs did not increase spending in more than half of the states that had a reform relative to synthetic counterfactuals. Because 20 of 26 states have an estimated ATT greater than zero, the consideration is about precision, as many of these ATTs have 90% confidence intervals that overlap zero. How can we understand the relatively large confidence intervals for so many states undergoing SFRs?

To build intuition, consider placebo tests, which, for a given state undergoing an SFR, assign that treatment date to each of the states in the donor pool. The placebo test takes each of these states and calculates an ATT for them using the same methods of ASCM. The placebo p-value corresponds to the percentage of these control states that have ATTs as large as the actually treated state (in absolute terms), thereby answering the question of whether a randomly selected state is as likely to increase spending to low-income districts as one that actually had an SFR. As shown in the left panel of Figure 9, these placebo p-values behave very similarly to the p-values generated from the row-based jackknife; indeed, the p-values are correlated at 0.95 and 0.89 among districts in terciles 1 and 3, respectively. Therefore, we interpret the (lack of) precision obtained for these estimates as an indication that many states without SFRs increased spending to low-income districts as much as or greater than states with SFRs.

[Figure 9 about here.]

A comprehensive explanation for how states without SFRs were able to increase spending at levels on par with states undergoing SFRs is beyond the scope of this paper.

Nevertheless, we can describe numerous potential mechanisms at the state level that could explain increases in spending among low-income districts in the absence of an SFR. Perhaps the most well known mechanism first occurred in Michigan, in 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 dollars were earmarked 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 (Sielke, Dayton, Holmes, Jefferson, & Fowler, 2001a, e.g., for Illinois). These individual acts will not consistently be recorded as school finance reforms but can nevertheless result in monetary increases.

Note, finally, that regardless of SFR status, there is overlap in the generic type of funding formulae states have in a given year. Using the taxonomy of funding formulae we compiled (described above in Section Data), Figure 10 shows the percentage of states with a funding formula of type F in year t, separated into SFR and non-SFR states. Prior to 1998–99, unmodified foundation plans were used by a plurality of states for both SFR and non-SFR states. Then, beginning in 1999–00, states began to adopt increasingly complex funding formulae, by adding components, such as categorical aid, to the base foundation plan. By 2004–05, the majority of states, again regardless of having an SFR, had a funding formula that included a combination of a foundation plan, categorical aid, and additional components. Though the trajectory from simple to complex funding formulae is evident in both SFR and non-SFR states, Figure 10 does suggest that SFR states are more experimental with their funding formulae, as SFR states have used 44 distinct funding formulae in our data compared to 34 among non-SFR states.

[Figure 10 about here.]

Predictors of SFR

Leveraging the 26 point estimates (and their standard errors) for log per pupil total expenditures for districts in terciles 1 and 3, we now perform a descriptive analysis to assess the extent to which SFR-related policies predict the heterogeneity in effect sizes across states. The descriptive analysis is conducted as a sequence of regressions weighted by the inverse variance of the effect size. For funding formula variables, we include the funding formula in use at the time the SFR was implemented and 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 dummy variable, with foundation plans as the reference category, since 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 dummy 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. Finally, to test whether SFRs vary in effectiveness over time, we include a variable indicating what year the SFR took place, and, in a separate regression, whether the SFR took place after the No Child Left Behind Act (NCLB). Results are displayed in Figure 11.

States that include categorical aid have consistently progressive outcomes, meaning that increases to tercile 1 expenditures exceed tercile 3 expenditures. States with power equalization plans and spending limits are consistently regressive, meaning that increases to tercile 3 expenditures exceed tercile 1 expenditures. In general, the range of estimated effect sizes is smaller among the funding formula categories the states used at the time of the SFR relative to the funding formulae states adopted subsequently to the SFR. This result provides further evidence that states undergoing an SFR became more experimental with their funding formulae, resulting in greater effect size variability.

Among states in which SFRs were initiated by a court order and among states that have experienced multiple suits, tercile 1 expenditures have increased and tercile 3 expenditures have remained constant. However, when a statute follows a court order increases to tercile 3 expenditures exceed increases to tercile 1 expenditures. 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.

Finally, we see no evidence that SFRs occurring earlier or later in our sample are more/less effective, as the point estimate on the timing of the SFR is a precisely estimated zero. Further, states with SFRs occurring after NCLB are no more or less effective at increasing tercile 1 spending than states with SFRs prior to NCLB.

Conclusion

Consistent with recent studies in the public finance of education literature, this paper finds that school finance reforms (SFRs), on average, increased 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-by-income tercile level for each state that had an SFR. This enables us to quantify how expenditure allocations varied by income tercile across the states. Among low-income districts, we find that 10 states increased spending, two decreased spending, and 14 saw no change in spending. Comparing low- and high-income districts, we find that a 1% increase in spending among low-income districts is associated with a 0.46% increase in spending among high-income districts. 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 or reduce class sizes. 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.

Finally, states that had or adopted categorical aid consistently increased spending more to low-income districts than to high-income districts, increasing the progressivity of spending allocation in that state. Further, the courts alone are more likely to increase low-income spending than the legislature alone, and states with multiple SFRs increase spending to low-income districts more than states that have experienced only one SFR.

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, researchers should be able to better guide policy.

References

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.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495–510.

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. (2019). School spending and student outcomes: Evidence from revenue limit elections in wisconsin. *Available at SSRN 3430766*.

Ben-Michael, E., Feller, A., & Rothstein, J. (2019). The augmented synthetic control method. *arXiv*. (1811.04170v2)

Brunner, E., Hyman, J., Ju, A., et al. (2018). *School finance reforms, teachers' unions, and the allocation of school resources* (Tech. Rep.).

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.

de Chaisemartin, C., & d'Haultfoeuille, X. (2019). *Two-way fixed effects estimators with heterogeneous treatment effects* (Tech. Rep.). National Bureau of Economic Research.

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.

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.

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.

Goodman-Bacon, A. (2018). *Difference-in-differences with variation in treatment timing* (Tech. Rep.). National Bureau of Economic Research.

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.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, *21*(11), 1539–1558.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557–560.

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.

Julian, P., & Higgins, T. (2019). *Cochrane handbook for systematic reviews of interventions.* Wiley-Blackwell.

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

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*.

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.

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.

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.

Powell, D. (2018). *Imperfect synthetic controls did the massachusetts health care reform save lives?* (Tech. Rep. No. WR-1246). RAND. doi: 10.7249/WR1246

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.

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.

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

Dependent Variable Summary Statistics Among States with SFRs

	Tercile 1		Tercile 3	
	Mean	S.D.	Mean	S.D.
log(Expenditures per Pupil)	9.34	0.34	9.28	0.27
log(Instructional Expenditures per Pupil)	8.71	0.33	8.64	0.27
log(Tch. Salary Expenditures per Pupil)	8.73	0.27	8.69	0.24
log(Capital Expenditures per Pupil)	6.43	0.77	6.50	0.64
Teachers per 100 Students	6.53	1.15	6.02	0.85

Notes: Summary statistics are computed for all states that had a court-ordered or legislative school finance reform. All variables span academic years 1989–90 to 2013–14.

Figures

Figure 1. Colorado “Bill 181” (2001): Log Total Spending Tercile 1
Note: Left panel shows Colorado log per pupil total spending in tercile 1 districts against the same outcome for all states without an SFR. The vertical line represents the timing of Colorado’s legislative “Bill 181.” The right panel shows three different effect sizes for log per pupil total spending in tercile 1 districts. In short-dashed red, the effect size is calculated by subtracting Colorado’s spending from the mean spending of all non-treated states, weighting each non-treated state equally. In long-dashed yellow, the effect size is calculated by subtracting Colorado’s spending from the weighted mean spending of non-treated states, where weights are constructed from traditional SCM. In solid green, the effect size is calculated using weights from ridge ASCM.

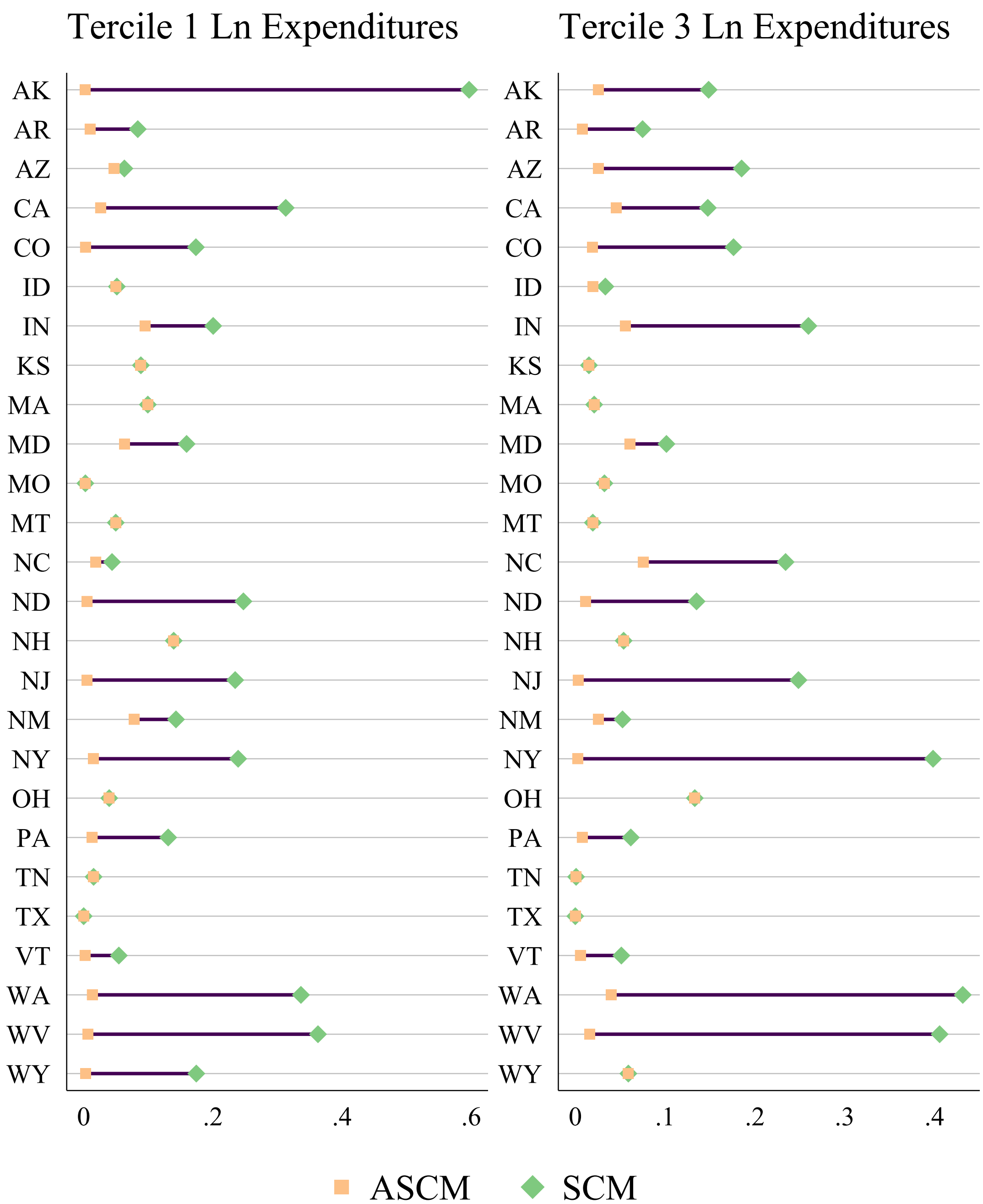


Figure 2. Model Fit Statistics for ASCM and SCM: Log(Expenditures per Pupil)
Note: model fit statistics are shown for traditional synthetic controls methods (SCM) and ridge augmented synthetic controls methods (ASCM). Match quality is defined as the cumulative pre-treatment effect size deviation from 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.

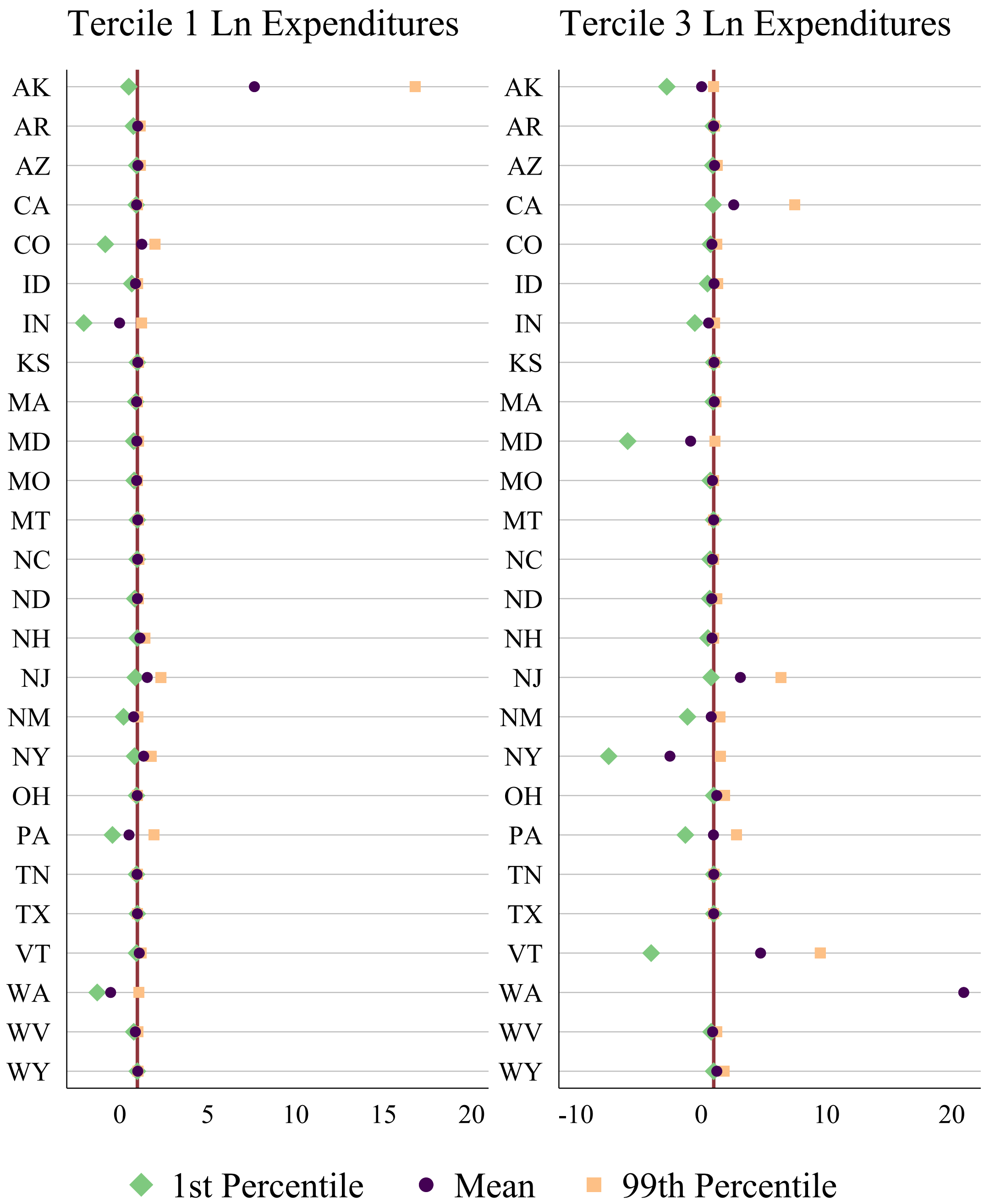


Figure 3. Stability of ATT as Function of λ^{r}: Log(Expenditures per Puipil)

Note: Each dot corresponds to one of three values of $\widehat{ATT}^{\mathrm{ratio}} = \frac{\widehat{ATT}^{L(\lambda)}}{\widehat{ATT}^{CV(\lambda)}}$, 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 three values presented here are the mean, the 1st percentile, and the 99th percentile.

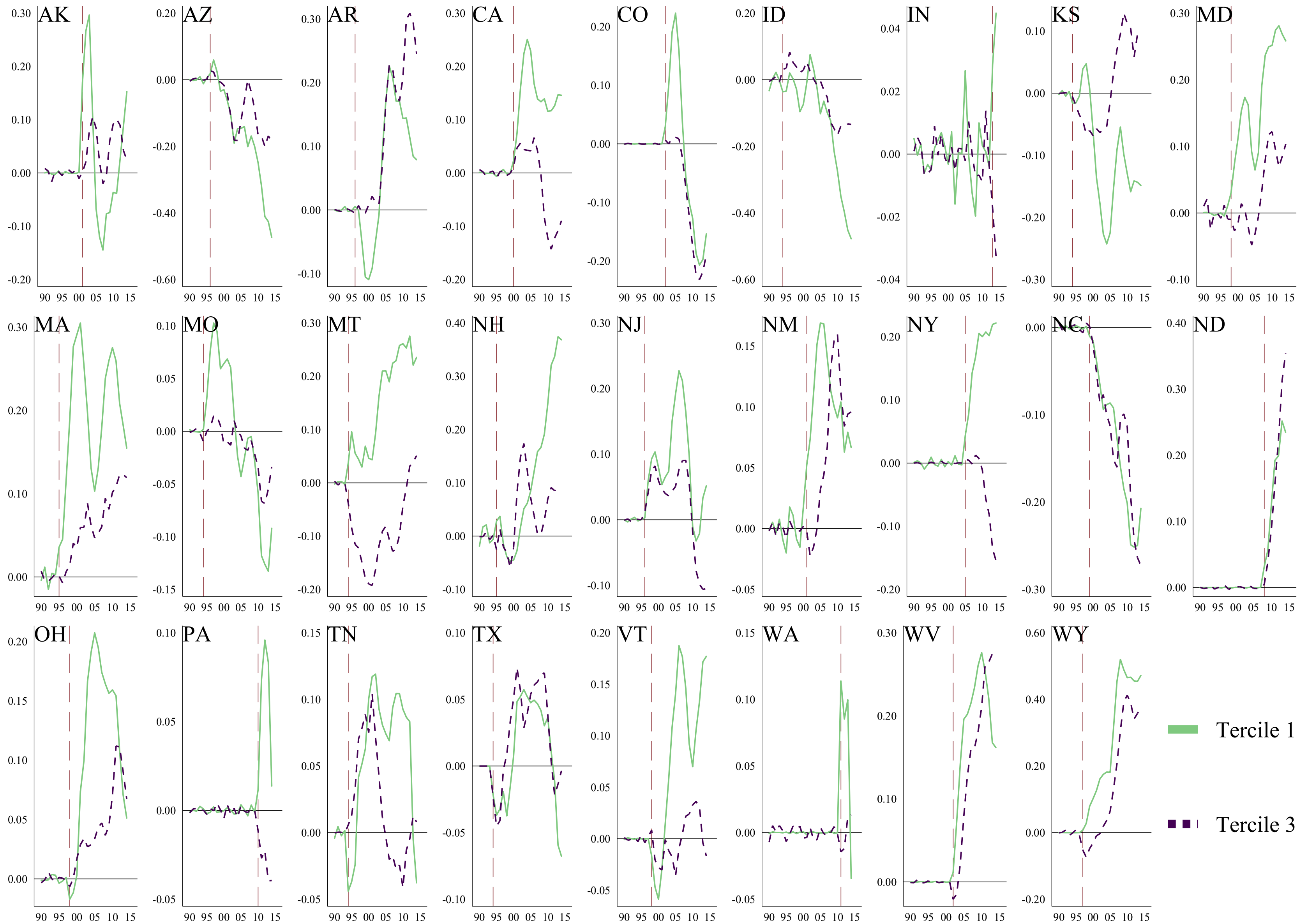
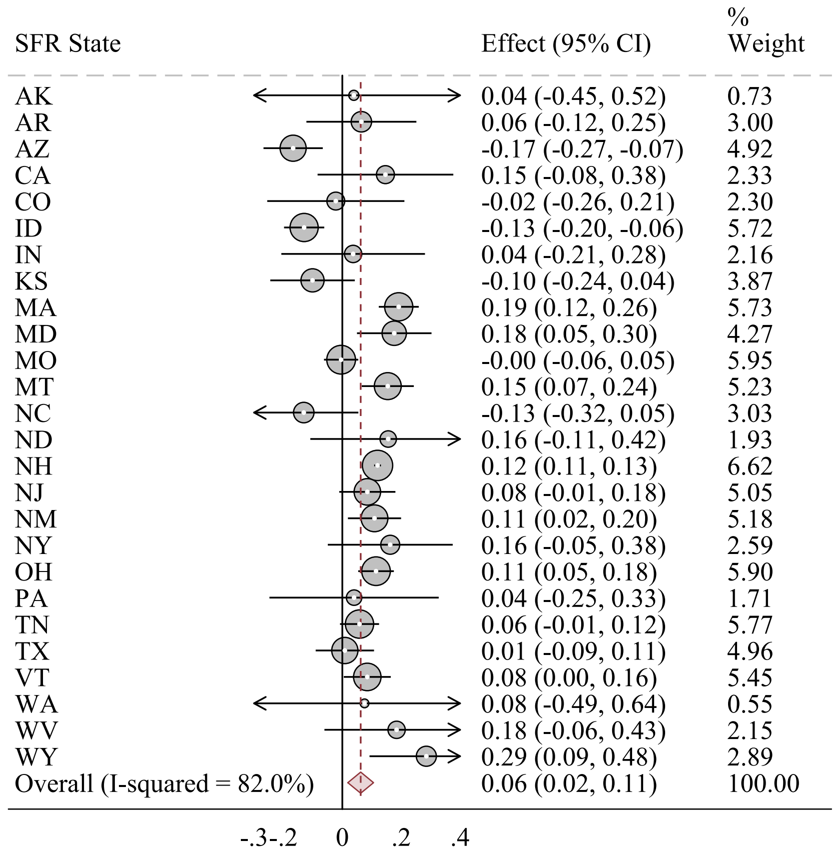
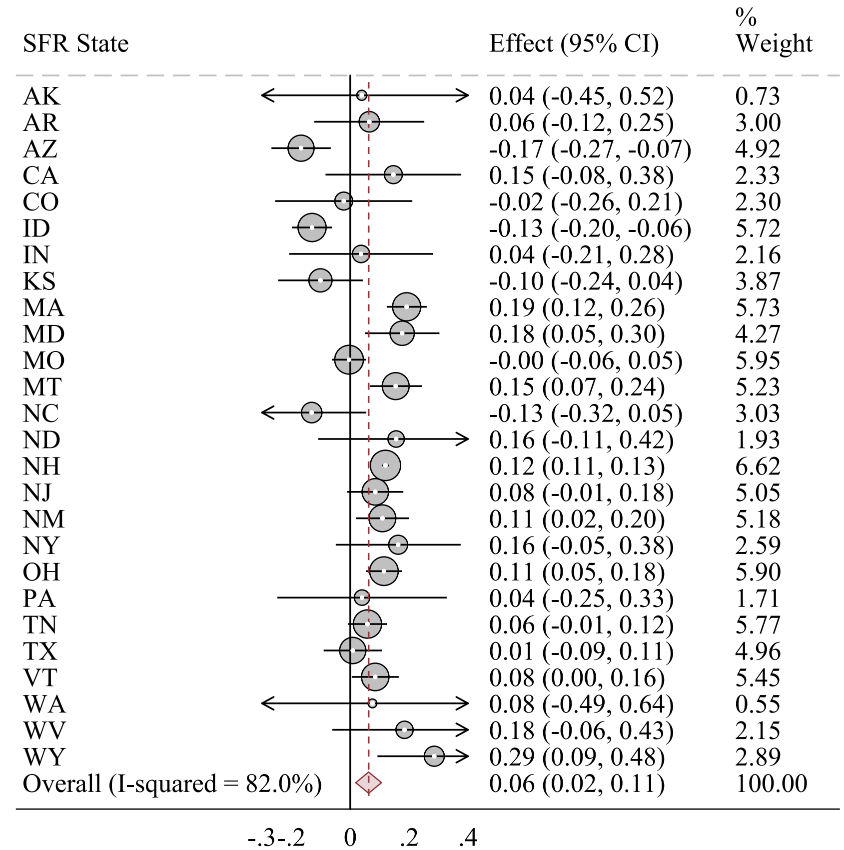


Figure 4. ASCM Effect Sizes: Terciles 1 and 3 Log Per Pupil Total Expenditures
Note: This figure plots the ATT of log per pupil total expenditures for each state-year-income tercile relative to its ASCM counterfactual.



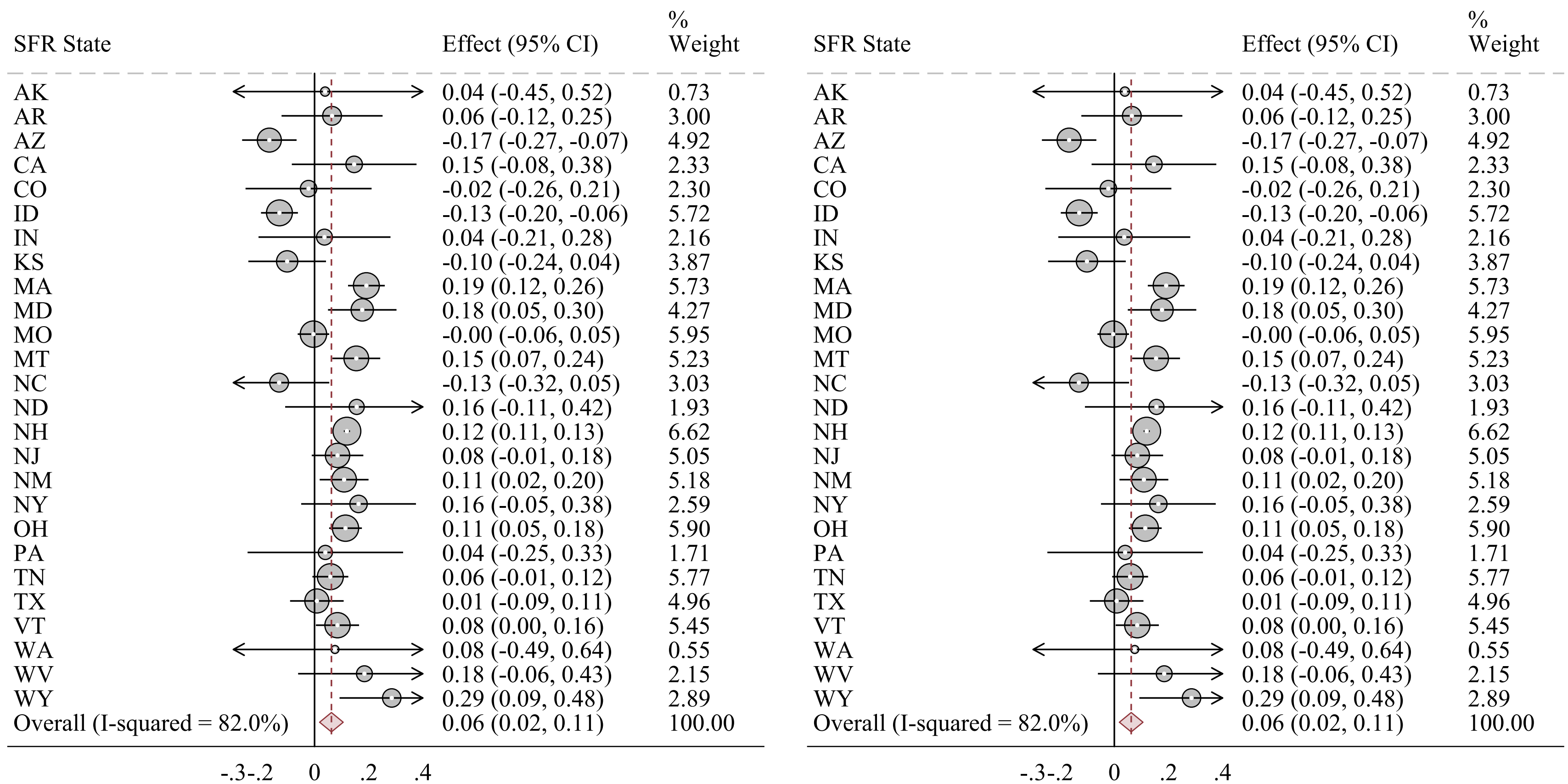
(a) Tercile 1



(b) Tercile 3

Figure 5. Per Pupil Total Expenditures ($\rho_{T1,T3} = 0.69$)

Note: Results for terciles 1 and 3 log total expenditures are shown in the first and second panels, respectively. The vertical dashed line and diamond shows the precision-weighted average of the ATTs from each SFR state, and the magnitude of the state-specific weights is shown on the right column. The magnitude of the pooled ATT is shown in the last row of the right column, and the I^2 statistic is shown in the last row of the left column. The displayed error bars indicate 95% confidence intervals; $\longrightarrow$ or $\longleftarrow$ indicate the confidence interval exceeds the axis range. Unweighted correlations between tercile 1 and 3 effect sizes are shown parenthetically in the sub-titles.

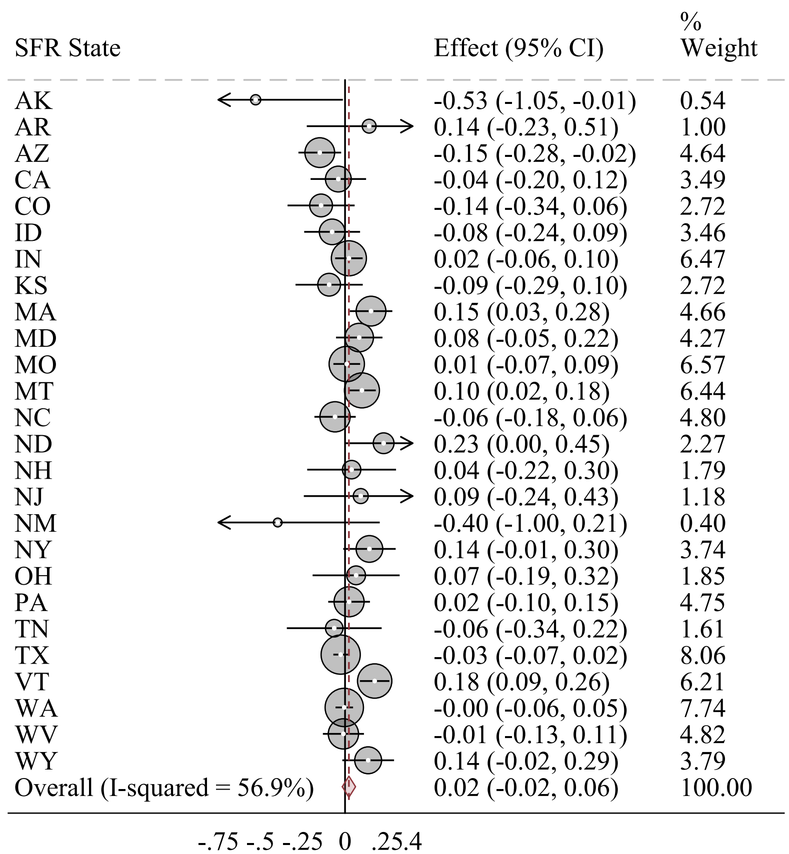


(a) Tercile 1

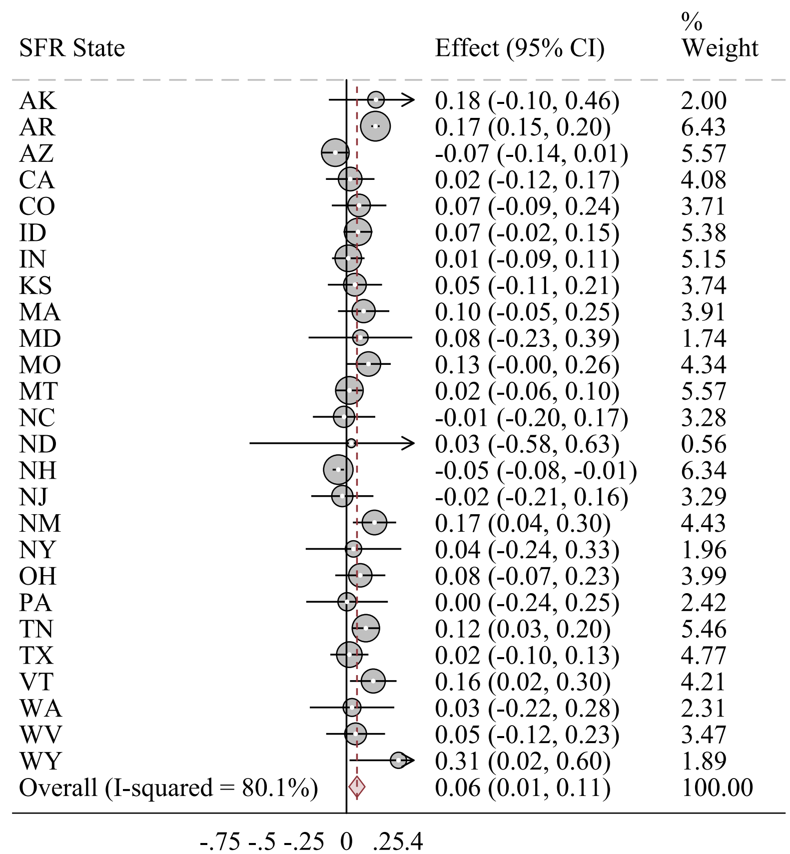
(b) Tercile 3

Figure 6. Per Pupil Capital Expenditures ($\rho_{T1,T3} = 0.39$)

Note: Results for terciles 1 and 3 log capital expenditures are shown in the first and second panels, respectively. The vertical dashed line and diamond shows the precision-weighted average of the ATTs from each SFR state, and the magnitude of the state-specific weights is shown on the right column. The magnitude of the pooled ATT is shown in the last row of the right column, and the I^2 statistic is shown in the last row of the left column. The displayed error bars indicate 95% confidence intervals; $\longrightarrow$ or $\longleftarrow$ indicate the confidence interval exceeds the axis range. Unweighted correlations between tercile 1 and 3 effect sizes are shown parenthetically in the sub-titles.

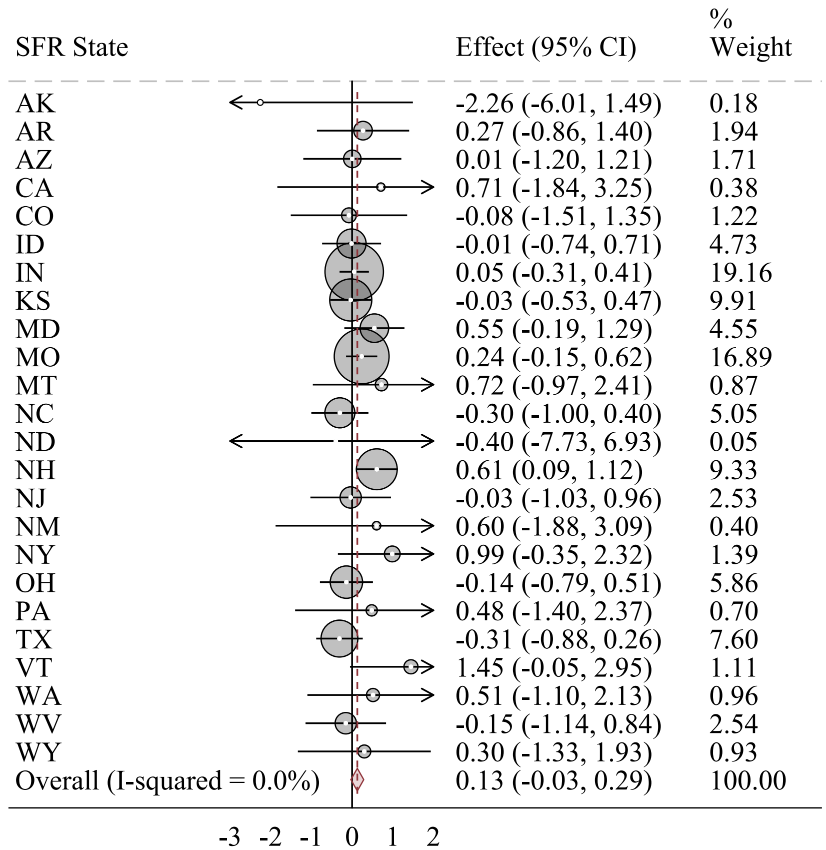


(a) Tercile 1

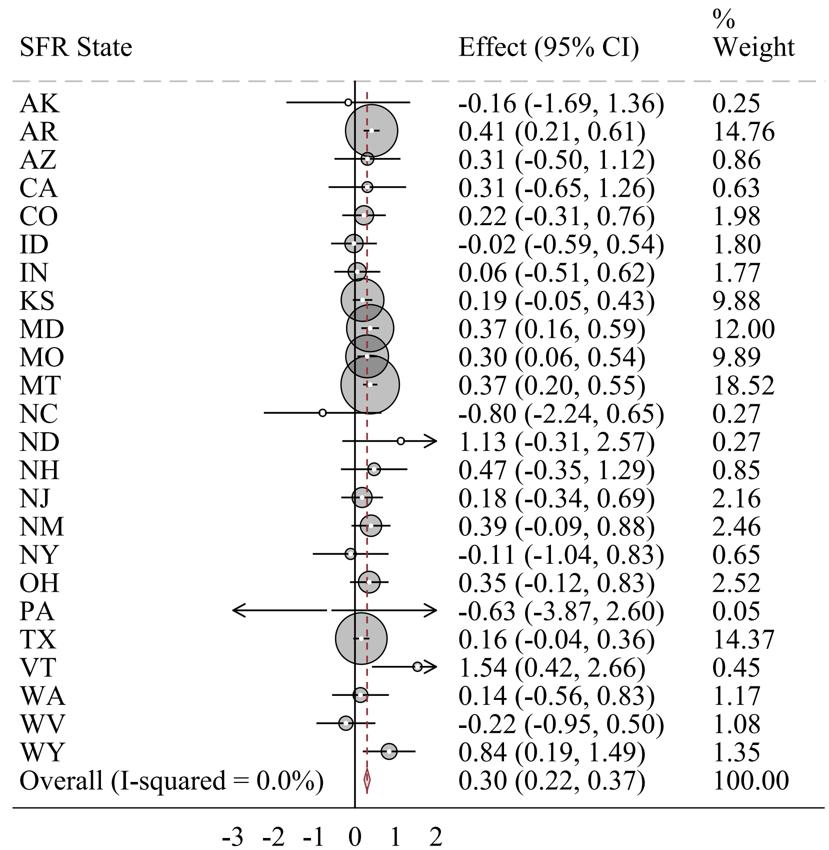


(b) Tercile 3

Figure 7. Log Per Pupil Salary Expenditures ($\rho_{T1,T3} = 0.48$)
Note: Results for terciles 1 and 3 log salary expenditures are shown in the first and second panels, respectively. The vertical dashed line and diamond shows the precision-weighted average of the ATTs from each SFR state, and the magnitude of the state-specific weights is shown on the right column. The magnitude of the pooled ATT is shown in the last row of the right column, and the I^2 statistic is shown in the last row of the left column. The displayed error bars indicate 95% confidence intervals; $\longrightarrow$ or $\longleftarrow$ indicate the confidence interval exceeds the axis range. Unweighted correlations between tercile 1 and 3 effect sizes are shown parenthetically in the sub-titles.



(a) Tercile 1



(b) Tercile 3

Figure 8. Teachers per 100 Students ($\rho_{T1,T3} = 0.57$)

Note: Results for terciles 1 and 3 teacher to student ratios are shown in the first and second panels, respectively. The vertical dashed line and diamond shows the precision-weighted average of the ATTs from each SFR state, and the magnitude of the state-specific weights is shown on the right column. The magnitude of the pooled ATT is shown in the last row of the right column, and the I^2 statistic is shown in the last row of the left column. The displayed error bars indicate 95% confidence intervals; $\longrightarrow$ or $\longleftarrow$ indicate the confidence interval exceeds the axis range. Unweighted correlations between tercile 1 and 3 effect sizes are shown parenthetically in the sub-titles.

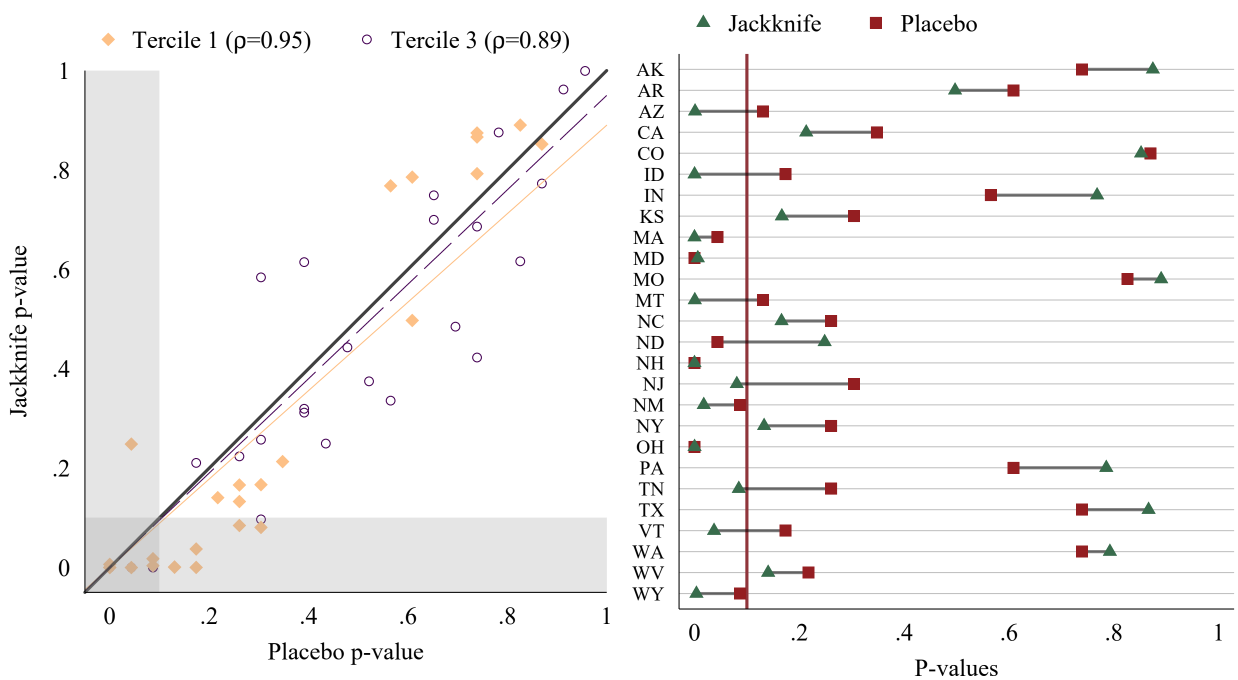


Figure 9. Jackknife and Placebo p-values: Log Per Pupil Total Expenditures
Note: Left panel shows p-values from the row-based jackknife and placebo tests for log per pupil expenditures in terciles 1 and 3. Correlations between row-based jackknife and placebo tests are shown in parentheses. Right panel shows tercile 1 log per pupil expenditures for p-values from the row-based jackknife and placebo tests for each state with an SFR.

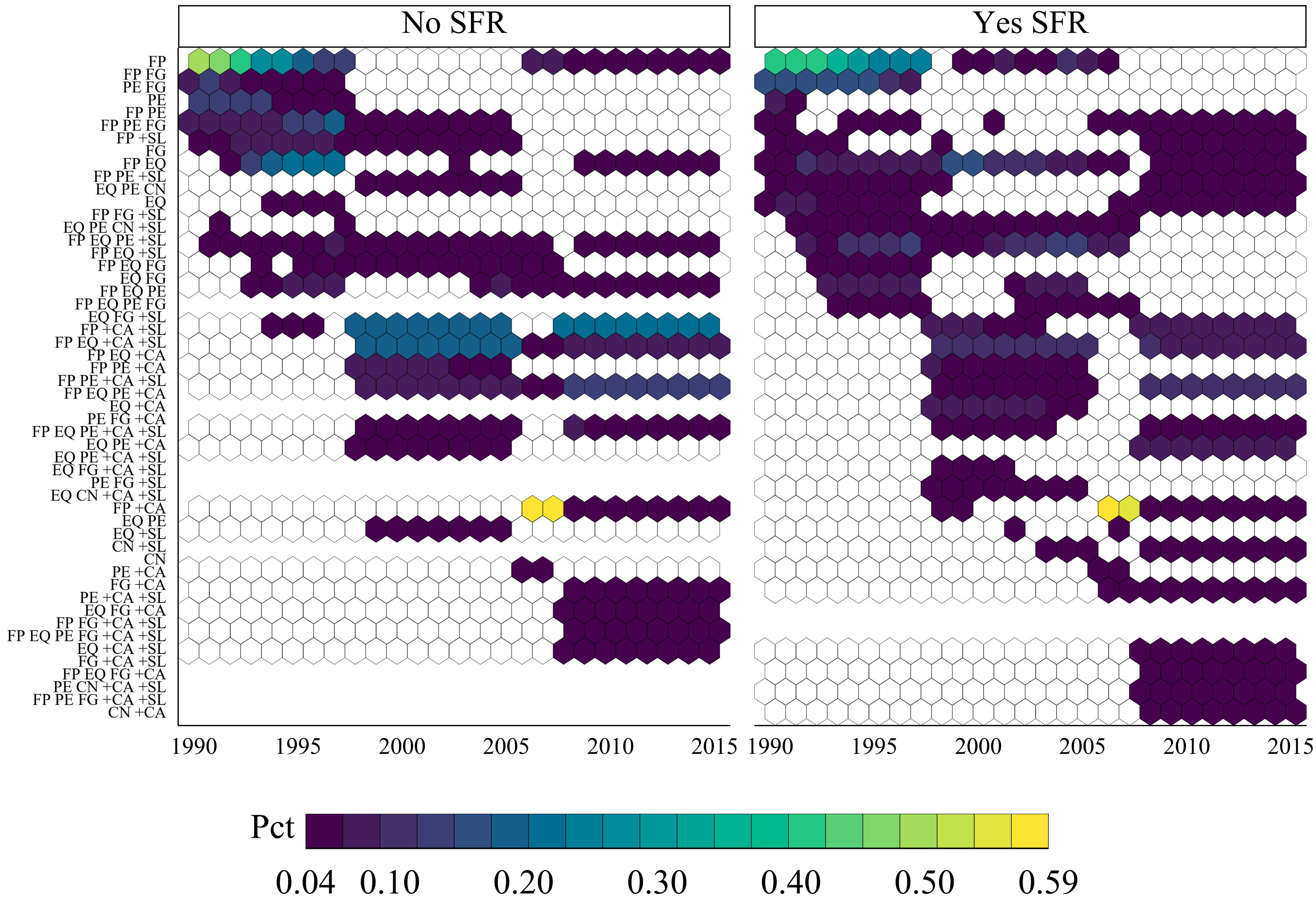


Figure 10. Distribution of Funding Formula Type, by SFR and non-SFR States

Note: This figure shows the percentage of states with a funding formula of type F (displayed categorically on the y-axis) in year y, separated by states with and without an SFR. Funding formula categories are described in Section Data. An empty hexagon indicates that 0 percent of states had that funding formula in a given year; if the cell is completely empty, it means that no states, among either non-SFR or SFR states, ever had a funding formula of that type.

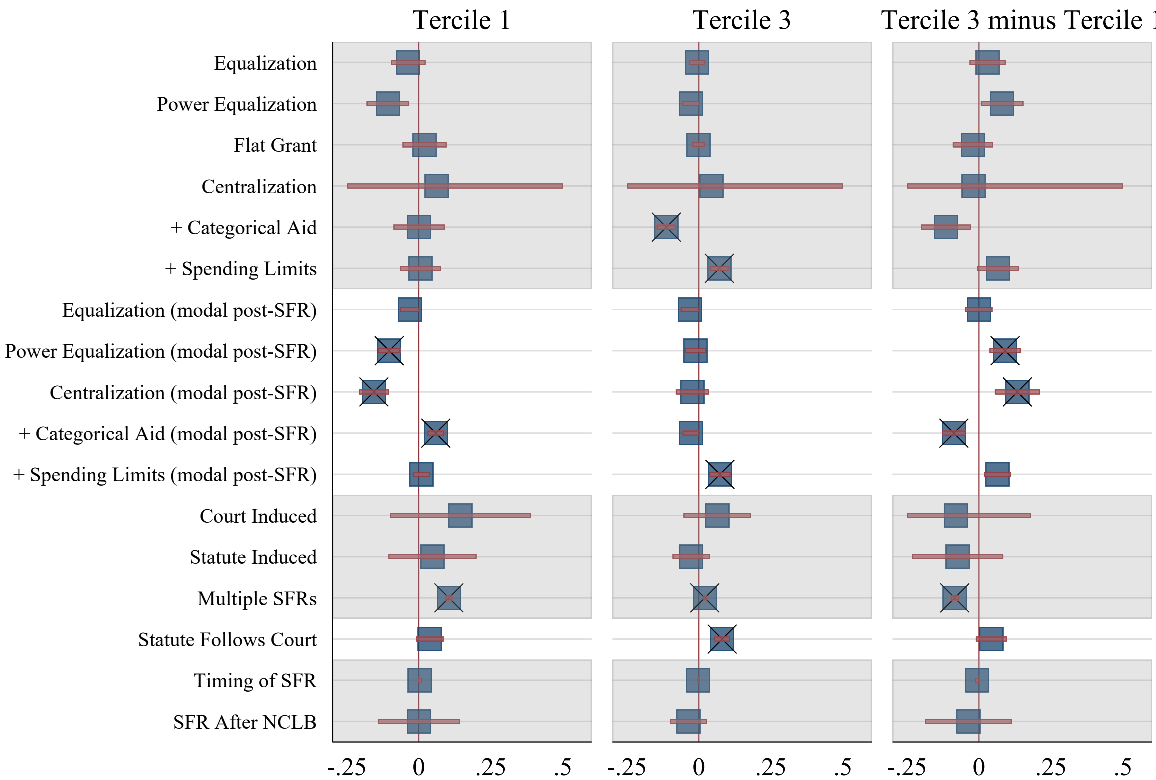


Figure 11. SFR policy variables and estimated effect sizes for log total expenditures
Note: List of predictors and data sources are taken from Table **??**. Each gray and white panel indicates a separate regression (e.g., results in the top gray panel, which includes the variables Equalization, Power Equalization, Flat Grant, Centralization, +Categorical Aid, and +Spending Limits are from a single regression). Range caps are 68.2% confidence intervals (i.e., +/1 standard error), 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.

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

Generating Panel Dataset of Funding Formula

Linking the adopted funding formula to the time at which an SFR takes place requires a panel dataset of funding formulas, one that varies by state and year. We construct such a state-by-year dataset of funding formulas 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 formulas 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 formulas. 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 C

Funding Formulas

Table C1

Funding Formula Distribution

State	Funding Formula Components	Num. of FF Changes post-SFR
AK	FP+EQ	0
AR	FP+EQ+SL	0
AZ	FP+EQ+PE+SL	0
CA	FP+FG	0
CO	FP+EQ+CA	0
ID	FP	0
IN	FP+PE+CA+SL	0
KS	FP+EQ+PE+SL	0
MA	FP	0
MD	FP+PE+FG	0
MO	FP+EQ+PE+FG	0
MT	FP+EQ+SL	0
NC	FP+EQ+CA	0
ND	EQ+PE	0
NH	FP	0
NJ	FP+EQ	3
NM	FP+EQ+CA	0
NY	EQ+FG	0
OH	FP	0
PA	FP+EQ+PE+CA	0
TN	FP	4
TX	FP+EQ+PE+SL	3
VT	FP+EQ+SL	0
WA	PE+CA+SL+CN	0
WV	FP+PE+CA+SL	0
WY	FP+EQ+SL	0
WY	FP+EQ+SL	0

Notes: Funding formulas listed correspond to the funding formula in place at the start of the first reform.