3-11-2021

Did Spending Cuts During the Great Recession Really Cause Student Outcomes to Decline?

Jessica Goldstein
University of Arkansas, Fayetteville

Josh B. McGee
University of Arkansas, Fayetteville, jbm04@uark.edu

Did Spending Cuts During the Great Recession Really Cause Student Outcomes to Decline?

3/11/21

By JESSICA GOLDSTEIN AND JOSH B. MCGEE *

Jackson, Wigger, and Xiong (2020a, JWX) provide evidence that education spending reductions following the Great Recession had widespread negative impacts on student achievement and attainment. This paper describes our process of duplicating JWX and highlights a variety of tests we employ to investigate the nature and robustness of the relationship between school spending reductions and student outcomes. Though per-pupil expenditures undoubtedly shifted downward due to the Great Recession, contrary to JWX, our findings indicate there is not a clear and compelling story about the impact of those reductions on student achievement. Moreover, we find that the relationship between K-12 spending and college-going rates is likely confounded with contemporaneous higher education funding trends. While we believe that K-12 spending reductions may have negative impacts on student outcomes, our results suggest that estimating generalizable causal effects remains a significant challenge.

* Goldstein: University of Arkansas, 201 Graduate Education Building, University of Arkansas, Fayetteville, AR 72701 (e-mail: jsgoldst@uark.edu). McGee: University of Arkansas, 201 Graduate Education Building, University of Arkansas, Fayetteville, AR 72701 (e-mail: JoshMcGee@uark.edu).

I. Introduction

Since Coleman et al. (1966), researchers have sought to understand the relationship between school expenditures and educational outcomes. Early studies cast doubt on the idea that school spending is strongly linked with student performance (Clark, 2003; Hanushek, 2003; Papke, 2008; Roy, 2011). More recently, however, researchers have used causal identification strategies to provide evidence that overall spending plays a more influential role than previously thought (Jackson et al., 2016; Candelaria & Shores, 2017; Hyman, 2017; Miller, 2017; Gigliotti & Sorensen, 2018; Lee & Polachek, 2018; LaFortune et al., 2018; Jackson et al., 2020). The extent to which school expenditures impact student outcomes is of particular interest today as the COVID-19 pandemic creates significant state budget shortfalls while simultaneously increasing fiscal demands on schools.

Evidence from previous recessions may provide pertinent information that can be used to ease the impact of the COVID-19 pandemic on schools and students. Prior studies have claimed to uncover significant causal relationships between recessionary changes and important educational factors such as teacher quality and retention (Nagler, Piopiunik, & West, 2017; Fuchsman & Zamarro, 2019). The focus of this paper is one such study, a forthcoming *American Economic Journal: Economic Policy* article titled "Do School Spending Cuts Matter? Evidence from The Great Recession," that investigates the impact of spending cuts related to the Great Recession on test scores and college-going rates (Jackson, Wigger, and Xiong, 2020a; JWX hereafter). The paper exploits state-level variation in education spending using an instrumental variables (IV) approach to identify the causal effect of recessionary spending cuts on student outcomes.

JWX uses pre-recession (i.e., 2008) K-12 state revenue share to instrument for the effect of expenditures on educational outcomes. This approach relies on the assumption that states which depend more heavily on state-generated education funding are more vulnerable to recessionary changes. The authors argue state-share captures recession vulnerability because it is subject to crowd-out from other more-pressing budgetary demands, such as Medicare and unemployment benefits, and because state income and sales tax revenues are more susceptible to the business cycle than local revenue, namely property taxes.

We replicate JWX as a first step in modeling the potential effects of the pandemic on K-12 spending and outcomes. This paper describes our process of duplicating JWX and highlights a variety of tests we employ to investigate the nature and robustness of the relationship between school spending reductions and student outcomes. Specifically, we explore how the findings differ when we implement different modeling choices around data source and state groupings. We also investigate the extent to which K-12 spending may be confounded with higher education spending and the implications this may have on capturing changes in college enrollment. While we are able to reproduce the JWX findings, we do not find consistent evidence that spending cuts resulting from the Great Recession had a causal effect on student outcomes. Instead, we find:

- the JWX group IV model results are sensitive to reasonable changes in data source for state revenue share and per-pupil expenditures;

- the JWX group IV model categorizes very few states as having low-reliance (***three states***) or high-reliance (***four states***) on state revenue to fund K-12 education, and the results are sensitive to small changes in group composition; and
- post-recession K-12 and higher education funding trends are related, and as a result, the JWX finding that K-12 spending reductions impacted state college-going rates is likely confounded with contemporaneous higher education funding trends.

The rest of the paper proceeds as follows. First, we briefly describe the JWX study and findings. Section II describes the data used to reproduce JWX. Section III motivates our robustness checks. Our analytical approach is described in Section IV. Results are outlined in Section V and Section VI concludes the paper, providing implications of our findings.

Description of "Do School Spending Cuts Matter?

Evidence from The Great Recession"

JWX uses an instrumental variables (IV) analysis to causally link spending reductions resulting from the Great Recession to student outcomes. Since unobserved factors influence both per-pupil expenditures (PPE) and student outcomes, PPE is considered endogenous. Therefore, researchers cannot directly investigate the causal impact of PPE on student outcomes. Isolating the effect of changes in PPE requires finding an instrument that is causally linked to PPE, has no impact on the outcome measure except through PPE, and is not caused by any factors that also impact student outcomes.

JWX exploits variation in state-appropriations to public education as an instrument for per-pupil spending. Specifically, the authors utilize the share of public-school revenues generated from state sources just prior to the Great Recession in place of per-pupil spending (i.e., 2008 state share of K-12 revenue). This identification strategy relies on the fact that state-level taxes are more susceptible to the business cycle than either local property taxes or federal funding. Therefore, states which are more reliant on state funding are likely to experience larger recession-induced cuts to school spending.

JWX presents evidence that spending reductions cause declines in both student achievement and college matriculation rates. The authors find that for every $1,000 reduction in per-pupil spending attributed to the recession, scores on the National Assessment of Educational Progress (NAEP) exam decline by 3.85 percent of a standard deviation. Results are similar for college-going rates, where a $1,000 reduction in per-pupil expenditures results in a decrease in college enrollment of 1.24 percentage-points.[1] In addition to the main analysis, JWX finds that spending reductions also increased achievement gaps by both race and socioeconomic status, and that states that cut K-12 spending hired fewer personnel and reduced capital expenditures rather than core K-12 expense categories.

The JWX findings suggest that both student achievement and attainment may be broadly responsive to fluctuations in aggregate education spending. Moreover, their results imply that high-poverty districts – where state aid makes up a larger proportion of revenues – may be the most at-risk to changes in school funding. Naturally, these findings have important implications

[1] Very few states saw per-pupil spending reductions of $1,000 or more in the years after the Great Recession (i.e., only 3-5 states, depending on data source, between 2007 and 2017. As a result, readers should use caution when using these estimates to predict the impact of expenditure changes of $1,000 or more.

for policymakers interested in understanding the relationship between education finance and student outcomes.

II. Data

Our dataset, which mirrors the sources and strategies employed in JWX, contains data on all 50 states and the District of Columbia spanning the years between 2001 and 2019. We closely match the mean, standard deviation, and number of observations for all data elements used in JWX's main analyses.[2] Table 1 provides summary statistics for our data and compares them to the JWX data.[3]

Unadjusted test scores for the National Assessment of Educational Progress (NAEP) were obtained from publicly available data aggregated and hosted by the Urban Institute.[4] To match the JWX data we use public school state-year average scores for 4th and 8th grade math and reading assessments between 2002 and 2017. All scores are standardized to a base year of 2003 using the national public school NAEP test means and standard deviations.[5]

College-going data come from the Integrated Postsecondary Education Data System (IPEDS). The data include information reported by institutions on the number of first-time college freshmen who graduated from high school in the past 12 months, aggregated by state of origin.[6] This portion of the survey is only administered during even years. To compute a college-going rate, we obtain population estimates from the U.S. Census Bureau for the number of 17-year-olds and 18-year-olds in each state the year prior to enrollment.[7] Our college-going rate divides the number of enrollees from each state by the average of the number of 17-year-olds and 18-year-olds in the state the previous year.[8] This rate is then associated with the year which matches the population estimates and NAEP administration years. For example, the enrollment from 2004 is divided by population in 2003 and then the resulting rate is associated with 2003 in our dataset.

We obtain school finance data from the Census Local Education Agency (School District) Finance Survey (F-33) and the National Center for Education Statistics (NCES) National Public Education Financial Survey (NPEFS) both of which are available at the Common Core of Data (CCD). The data contain information on school revenue and expenditures (both total and current), as well as the number of students enrolled in each district and state.[9] We CPI-adjust all dollar variables to constant 2015 dollars.

[2] While the JWX data and code should be available through the journal once it is published (see www.openicpsr.org/openicpsr/search/aea), it is not yet publicly available. However, because all of the study data comes from public sources, it is relatively straightforward to recreate the study dataset.

[3] See JWX Table 1 for summary statistics.

[4] Five states do not have 2002 NAEP scores: AK, CO, NH, NJ, and SD, and Iowa did not report an 8th grade reading score. To generate a balanced panel of data, we impute the mean reading z-score for those same states in 2003. Results are consistent both including and excluding these observations.

[5] Data obtained from the NAEP Data Tool (https://www.nationsreportcard.gov/ndecore/landing). 2003 National NAEP Exam Statistics: 4th grade reading – Mean = 216, SD = 37; 8th grade reading – Mean = 261, SD = 35; 4th grade mathematics – Mean = 234, SD = 28; 8th grade mathematics – Mean = 276, SD = 36

[6] We use files labeled "Residence and migration of first-time freshmen: Fall 2018" and use the associated downloadable STATA .do file provided by IPEDS to organize and summarize the data.

[7] Data from 2000-2010 come from the State Intercensel Estimates (https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html) and data from 2010-2019 come from the Vintage 2019 Estimates (https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html).

[8] JWX reports using the average number of 17- and 18-year-olds in their Data Appendix.

[9] The F33 does not collect information on charter schools held by non-governmental organizations or private companies. Data for states with large charter sectors will be affected by this choice.

Census provides F33 data both at the district and state levels. For the state-level data, Census harmonizes the information with both past years and the NPEFS. Instead of using the state level data provided by Census, JWX aggregates the district-level data, leading to differences between JWX's state-level data and Census state-level data.[10] We refer to the data aggregated to the state level from the F-33 district level file as "Census District," and the state data released by Census as "Census State."

We also retrieve data on economic conditions in each state and year from the Bureau of Labor Statistics (BLS). Following JWX, we use these data to construct Bartik economic control variables for unemployment and average wage. This process is described in A.2 of the online appendix for JWX. We obtain state employment shares for 2007 – the first year of the recession – using BLS "CSVs By Industry" and "Annual Averages."[11][12] Average wage by industry is obtained from BLS Occupational Employment Statistics data. Unemployment data by industry and year are obtained from the BLS Economic News Release.[13][14]

Information on higher education finance is obtained from two sources. State share of higher education funding comes from the IPEDS finance surveys. Data on state and local appropriations per FTE and tuition and fees per FTE are obtained from the State Higher Education Executive Officers Association (SHEEO, 2019).

III. Motivation for Robustness Checks

Econometric models can be constructed in a variety of ways, and many modeling choices may be somewhat arbitrary or theoretically unimportant. Simmons et al. (2011) called the set of these choices "researcher degrees of freedom." If the model's estimates represent the true causal impact, they should be consistent across many different reasonable ways of constructing the data and model. On the other hand, if ambiguous data or modeling decisions make a large impact in the results, then we should question the validity of the findings.

JWX presents several robustness checks for their analysis; however, they do not present tests in a few key areas where there are significant researcher degrees of freedom. As part of our replication, we test the sensitivity of the analysis to various data and modeling choices to ascertain the impact of these choices on the results. The remainder of this section describes the theoretical motivation that underlies our approach and describes the specific tests we use to investigate the nature and robustness of the relationship between school spending reductions due to the Great Recession and student outcomes.

[10] JWX winsorizes extreme values for per-pupil expenditures by capping values of districts with expenditures great that 200 percent the 99th percentile of per-pupil revenues or less than 50 percent of the first percentile. Our analysis shows that state-level means and standard deviations of spending data are not significantly changed when winsorizing; therefore, we utilize the non-winsorized data in our study. Our non-winsorized spending data is not statistically different from JWX winsorized data.

[11] https://www.bls.gov/cew/downloadable-data-files.htm

[12] Census industry code crosswalk used for the BLS unemployment rate can be found here: https://www.census.gov/topics/employment/industry-occupation/guidance/code-lists.html

[13] For example, data for 2019 and 2020 can be found at: https://www.bls.gov/news.release/empsit.t14.htm.

[14] While JWX does not provide any summary statistics for these Bartik control variables, the 2019 NBER working paper version (Jackson et al., 2019) does provide Figure A4 depicting yearly averages for the unemployment rate Bartik. The figure is not well labeled but our values generally match this figure. Our average Bartik control for the unemployment rate is: 2001-4.66%, 2002-5.74%, 2003-5.94%, 2005-4.85%, 2007-4.35%, 2009-9.12%, 2011-8.32%, 2013-6.80%, 2015-4.99%, 2017-3.93%.

Instrument Data

Both the Census and NCES provide data on K-12 revenue by source. JWX uses Census District data to define their instrument.[15] These two institutions each make their own unique determinations regarding what is state versus local revenue, leading to significant differences in revenues attributed to each category for some states. For example, Census State lists Arkansas' state revenue share as 76 percent, counting the state mandated minimum local property tax effort as state revenue (see note on page B-1 of the Census 2008 survey documentation). On the other hand, NCES data for 2008 show Arkansas' state revenue share at 56.7 percent, nearly 20 percentage points lower than Census State. Given the purpose of the instrument is to capture recession vulnerability, the NCES data for state share is, at least for Arkansas, a better fit for the analysis because it counts property tax revenue as local.

Table 2 summarizes the 2008 state revenue share data from JWX (Column 1), Census State (Column 2), and NCES (Column 3).[16] In total, there are 13 states where JWX and NCES data for 2008 state revenue share differ by more than two percentage points. Six states differ by more than five percentage points and three states vary by more than 15 percentage points.[17] These differences are potentially important given how few observations are included in the analysis (i.e., the 50 states and Washington D.C.).

For some states, similar to the case of Arkansas discussed above, the official data sources do a poor job capturing the unique nature of their school funding systems. In Vermont, for example, public education is largely funded through a dedicated statewide property tax.[18] In total, between 60 and 70 percent of Vermont's K-12 education funding comes from property taxes.[19] Because the property tax is statewide and distribution is managed through the state, both Census State and NCES data list Vermont's state share as between 85 percent and 90 percent. JWX lists Vermont as having 68 percent state revenue share. The state share instrument is intended to measure reliance on types of taxes and an allocation process that would make education funding more susceptible to cuts during a recession. However, that is far from the case for Vermont, which relies in large part on dedicated, stable property taxes. Both Census State and NCES shows that Vermont increased its spending per pupil after the Great Recession. NCES data shows that Vermont's constant dollar per-pupil spending increased by $1,345 and $3,038 over the 5- and 10-year periods beginning in 2007. The 10-year increase represents the fifth largest increase among the states (see Table 3 column 7).

California, like Arkansas and Vermont, has a state mandated local property tax effort. JWX argues that California's property tax dollars should be categorized as local revenue because they are not subject to crowd-out and are more stable than other state-level taxes. However, the same is true in Vermont and Arkansas, whose property taxes are counted as state revenue for the purposes of the JWX instrument. JWX does not explain why state-mandated/collected property tax revenue is handled differently for California than it is for Arkansas and Vermont, or explore

[15] The JWX state revenue share data can be found in Table A2 of the online appendix for the forthcoming version (Jackson et al., 2020a). The 2008 Census state share data can be downloaded at https://www.census.gov/data/tables/2008/econ/school-finances/secondary-education-finance.html. There are differences between the JWX state revenue share and Census State revenue share that are the result of JWX aggregating from Census District data.

[16] NCES also collects information property tax revenue, which is a plausible substitute for the JWX instrument. The 2005-07 state-level property tax share calculated using NCES data (Table 2, column 5) has a 0.65 correlation with the state revenue share documented in table A2 of JWX (Table 2, column 1).

[17] In addition to how state mandated local tax effort is categorized, Census and NCES also handle pension payments somewhat differently.

[18] A description of Vermont's school funding system is available at EdBuild. http://funded.edbuild.org/state/VT

[19] https://ljfo.vermont.gov/assets/Uploads/d76744a36e/The-Education-Fund-and-Education-Finance-v2.pdf

the model's sensitivity to handling these states differently. The latter two states are particularly important for identifying JWX's results because of the way states are grouped in the econometric model, which we will discuss in more detail below.

Washington D.C. also presents a challenge when categorizing revenue by source. Like other states, education in D.C. is primarily financed through a combination of local property taxes, general fund appropriations, and federal funds. However, the D.C. government combines state and local functions under one umbrella.[20] Since it is not technically a state, neither NCES nor Census categorize D.C.'s general fund appropriations as state revenue. Both Census and NCES simply lists Washington D.C.'s state revenue as "not applicable." JWX assigns Washington D.C.'s state revenue share as zero. Given the theoretical justification for the instrument and D.C.'s reliance on general fund appropriations for a large share of school funding, a value greater than zero is more appropriate. In fact, since the entire school budget is appropriated by the council, there is a reasonable argument that D.C.'s state share should be 100 percent. Because of this uncertainty around how to handle Washington D.C.'s state K-12 revenue share, we drop the jurisdiction from many of our analyses.[21]

For their main analysis, JWX uses 2008 state K-12 revenue share to group states into high-, medium-, and low-reliance categories. Because of the data differences discussed above, applying the same categorization rules to different data sources yields different groupings. Table 2 includes color coding to indicate high- and low-reliance groups across data sources. Cells colored green are states which are below the 0.33 threshold JWX uses to define low-reliance, and cells colored red are above the 0.67 threshold JWX uses to define high-reliance. We will discuss these thresholds in more detail below. Simply changing data sources from JWX to NCES results in four states changing categories.

K-12 Spending and Outcomes Trends

To motivate their analysis, JWX compares trends over time for PPE and student outcomes, demonstrating that they seemingly move together. We follow suit, graphing trends in PPE, NAEP scores, and the college-going rate. Figures 1 and 2 depict these trends. Our figures show that PPE declined during the Great Recession but increased sharply after 2013. We also include NCES's projections for 2018 and 2019 (dashed blue lines), which show that PPE has likely continued to climb in the years after the JWX study period.

As indicated in our graphs, both NAEP scores and the college-going rate increased with PPE during the years leading up to the recession. However, it is less apparent whether PPE and student outcomes have been moving together after the outset of the recession. Figure 1 shows that while PPE rebounded a few years after the recession, NAEP scores have continued to decline. The national college-going rate, depicted in Figure 2, leveled off somewhat following the recession but generally the trend has been positive over the entire period despite significant movement in average PPE.[22] The recent divergence between PPE and student outcomes raises

[20] A description of Washington D.C.'s school funding system is available at NCES and EdBuild, both linked below.
https://nces.ed.gov/edfin/pdf/StFinance/District.pdf
http://funded.edbuild.org/state/DC

[21] JWX provides estimates dropping Washington D.C., Hawaii, and/or California, and finds that their results are robust to these changes. We estimate models dropping all four problematic states, AR, D.C., CA and VT, finding attenuated and insignificant point estimates.

[22] JWX does not provide a figure comparing changes in PPE to changes in the college-going rate, which is their outcome variable. Instead, in Figure 1, they compare PPE to changes in the number of first time college enrollees, which, unlike the college-going rate, appears to dip after the recession onset.

the question of whether the seeming correlation between the variables in early 2000's was simply a coincidence.

The JWX identification strategy relies on the relationship between pre-recession state revenue share and differences between states in post-recession spending. To illustrate how state-level K-12 expenditures changed following the Great Recession, Table 3 provides constant dollar PPE differences over the 10-year period from 2007 to 2017 using our three different data sources: Census District, Census State, and NCES.

While average PPE decreased following the great recession, not all states experienced declines, and many saw sizeable increases. Both Census State and NCES data show that more than half of states, 34 and 32 respectively, experienced per-pupil spending increases over the 10-year period beginning in 2007. Even in the immediate aftermath of the recession (i.e., between 2007 and 2012), only half of states, 25 for both Census State and NCES, experienced spending reductions.

Figures 3-6 show the relationship between these spending changes and pre-recession state revenue share from Table 2. Each of these figures uses the same data source for both variables (e.g., Census State for both state revenue share and PPE change). Vertical red lines depict JWX's 0.33 and 0.67 thresholds for low- and high-reliance on state revenue. We have also included a linear fit line along with the slope and p-value. The three panels represent the period before the Great Recession (i.e., 2003-2007), the period after (i.e., 2007-2017), and the trend change.

Our Figure 3 is equivalent to JWX Figure 3, depicting a strong negative relationship between JWX state revenue share and Census District PPE change after the recession. However, it's apparent that D.C. is contributing mightily to this linear fit. Figure 4 uses the same data but drops D.C. This change nearly halves the slope and results in an insignificant relationship both before and after the recession.

Figure 5 uses Census State data and drops D.C. While the relationship is still negative it is much weaker and not statistically significant. Lastly, Figure 6 depicts NCES data and drops D.C., and the relationship is even further attenuated and highly insignificant.

In this section, we have shown that in the post-recession period, aggregate NAEP scores and college going rates do not appear to move closely with PPE; that there are meaningful differences in measures of spending across data sources; and after dropping D.C., changes in spending appear to only be weakly correlated with pre-recession state K-12 revenue share. The combination of these descriptive results raises concerns that JWX's proposed instrument is weak and highlights how data source is likely influential in the analysis. Sections IV and V more rigorously investigate these issues.

State-Group Definitions

The primary analysis in JWX relies on grouping states into high-, medium-, and low-reliance categories. In addition to the data issues outlined in the previous sections, the thresholds JWX use to categorize states result in very few states being placed in the high- and low-reliance categories. JWX does not provide a theoretical or empirical justification for the specific thresholds used to categorize states and does not test different groupings.

As noted in the previous sections, JWX defines high-reliance states as those that depend on state appropriations for more than 67 percent of total revenue and low-reliance as less than 33 percent. Of the 51 jurisdictions included in the analysis (i.e., 50 states and Washington D.C.), the

JWX categorization rules result in only ***three*** states being placed in the low-reliance group and ***four*** in the high reliance group. All other states are relegated to the middle-reliance category.

Of the seven states included in the high or low groupings, four of them are based on questionable data, including three of the four states in the high-reliance category. Outlier jurisdictions Washington D.C. (low-reliance with zero percent state share) and Hawaii (high-reliance with 85 percent state share) are included in the low- and high-reliance categories, respectively.[23] In addition, Arkansas (high-reliance with 76 percent state share) and Vermont (high-reliance with 68 percent state share), who's property tax revenues are arguably misclassified as state revenue, are included in the high-reliance category.

Looking back at Figures 3-6, it is also clear that there are not a significant number of outliers on the left and right sides of the distribution that might make the proposed thresholds obvious. Instead both the high and low threshold are close to the mass of states in the middle of the graph. This is especially true on the left side where the line cuts right through several states, including South Dakota which would have been classified as low-reliance if the JWX threshold value were carried out one more decimal place (i.e., 0.333 instead of 0.33).[24]

It is also not the case that states categorized as high/low-reliance experienced the largest changes in PPE undermining the argument that these extreme thresholds increase treatment contrast. Table 3 columns 5-7 provide each state's rank for PPE change between 2007 and 2017. The rankings are ascending from largest gain to largest loss. JWX low-reliance states D.C. and Illinois experience some of the largest gains, but most of the other low/high reliance states are in the middle of the pack. Vermont, on the other hand, despite being categorized as high-reliance saw some of the largest PPE gains following the recession.

The small number of states in the high- and low-reliance groups, the presence of states with problematic data in both groups, and the arbitrary nature of the thresholds raises the question of how sensitive the JWX results are to different state-group definitions.

JWX provides results from robustness tests dropping any combination of one, two, or three states from their analysis (see figure A3 in the online appendix). However, this does little to test the robustness of their results to different state groupings given that the vast majority of states (i.e., 44) are in the middle-reliance category. The results from the test JWX performs would primarily be generated by dropping states in the large middle reliance category, which we would expect to have a limited effect on the estimate.

We test the impact of state-group definitions by estimating results using four symmetric percentile sets to categorize states, as depicted in the columns of Table 4 for JWX state revenue share. Column 1 of Table 4 depicts the grouping used in JWX, and columns 3-6 represent additional groupings based on the 10th and 90th percentile, 15th and 85th percentile, 20th and 80th percentile, and 25th and 75th percentile, respectively.[25] We also estimate results using [1/3 and 2/3] as the high/low thresholds (Column 2). This latter threshold set represent the stated thresholds in JWX, but in implementation they truncated the threshold values at two decimal places, moving South Dakota out of the low-reliance group.

In Table 4, states highlighted in red text are those which are different from the JWX grouping (Column 1). States not listed in either the low- or high-reliance groups are in the medium-reliance category. The number of states in each category is listed in parenthesis at the

[23] The school districts in both Washington D.C. and Hawaii combine state and local government levels, making attempts to separate state vs. local revenue challenging.

[24] Dropping DC and classifying SD as low-reliance results in attenuated and insignificant point estimates for both NAEP scores and college-going (see Tables 6A and 6B).

[25] The percentile values are determined based on data source for state revenue share and are slightly different between sources.

top of each column. Increasing the threshold for low-reliance from the JWX value (33 percent) to the 10th , 15th, 20th and 25th percentiles adds 2, 4, 7, and 10 states to that group, respectively. Decreasing the threshold for low-reliance from the JWX value (67 percent) to the 10th , 15th, 20th and 25th percentiles adds 1, 3, 6, and 8 states to that group, respectively. Overall, using the 10th and 90th percentiles to define state groups is a very small change, adding only 3 states total to the low/high groups

Given that there are no strong theoretical reasons to use the specific thresholds used in JWX and that this is an area with significant researcher degrees of freedom, our empirical approach investigates how state group definitions influence the JWX results. We believe using percentiles to define state groups is empirically justifiable and testing slightly different and more inclusive groupings will help us better understand the influence of the groupings on the results. Sections IV and V outline our analytical approach and present findings for different specifications using each grouping.

Potential Confounding of the JWX College-Going Result

For K-12 state revenue share to be a valid instrument for PPE, it must not affect the college-going rate through any other pathway except through PPE, and it must not be influenced by any other factors which are also causally linked to college-going. We believe that potential confounding between K-12 and higher education funding trends may violate the exclusion restriction necessary to causally estimate the effect of PPE on college-going rates. We also believe there may be economic, political, and institutional factors that influence both pre-recession K-12 state revenue share and post-recession college-going rate. Figure 7 provides a directed acyclic graph (DAG) illustrating our concerns.

To investigate confounding, we first examine time trends to see if K-12 and higher education funding measures move in tandem. Figure 8 depicts the relationship among PPE, state and local higher education appropriations per full-time equivalent (FTE) student, and tuition revenue per FTE. Per-pupil expenditures and higher education appropriations move closely together after the recession, both falling sharply between 2008 and 2012, while tuition revenue steadily increases over time. Figure 9 shows these same three variables indexed to their 2002 values to display percentage change. The magnitude of the changes in higher education appropriations and tuition dwarf those for PPE, raising the prospect that higher education finance might have greater potential to influence college-going rates.

These graphs lead us to believe that K-12 and higher education spending may be confounded, and even if a relationship between K-12 PPE and college-going rates exists, it may be overshadowed by the impact of changing higher education appropriations and tuition revenue. It is reasonable to theorize that higher education funding may influence college-going more than K-12 appropriations. Recent studies have shown that college outcomes, including enrollment, are responsive to higher education funding changes (Deming and Walters, 2017 and Bond et al., 2019). While we do not attempt to identify the true effect of higher education spending on college enrollment, we investigate the possibility that higher education appropriations and tuition confound the JWX college-going findings. Section IV provides additional details regarding our approach.

IV. Empirical Approach

Our empirical methods follow JWX closely, with variable definitions matching those described in the Identification Strategy section of JWX. To our knowledge, we deviate only in areas where we test the robustness of the JWX results. The sections below present our empirical approach for three different analyses: 1) event study graphs, 2) group IV, 3) randomization inference and 4) tests for higher education confounding.[26] Each section describes our methods and approach for robustness checks.[27]

Event Study Graphs

JWX presents event study graphs as suggestive evidence that per pupil expenditures, NAEP scores, and the college-going rates declined more in states with a high reliance on state revenues relative to states with a low reliance on state revenue. Equation 1 describes the event study estimating equation.

$$(1) \qquad Y_{st} = \sum_{t=2001}^{2017} \beta_t \left(I_{\Omega_s > q(50),s} \times I_{T=t}\right) + \rho\left(I_{post}\right) + \alpha_s + \theta_t + \upsilon_{st}$$

Subscripts *s* and *t* indicate state and year, respectively. The outcome variable, Y_{st}, is either PPE, average NAEP scores, college-going rates, or high school graduation rates. $I_{\Omega_s > q(50),s}$ is an indicator for states where the 2008 K-12 state share is above the national median of 0.48. We use the JWX state share data for this analysis. $I_{T=t}$ are year-specific indicators. I_{post} is an indicator for years after 2008. The α_s are state fixed effects, and the θ_t are year fixed effects. β_t is the coefficient of interest and represents the year-specific difference between states above versus below the median. We omit the year 2007, so the calculated differences are relative to that year.

Group Instrumental Variables Model

Following JWX, we estimate a group IV in which states are categorized as being high-, medium-, or low-reliance based on 2008 state share of K-12 revenue. These category variables are then used to define the instrument. For the group IV analysis, we estimate the equations below using two stage least squares (2SLS).

[26] We also investigate the linear IV presented in the JWX online appendix and previous versions of JWX. This analysis is included in the appendix.

[27] Errors are clustered at the state level in all models.

$$(2)\quad PPE_{st} = \sum_{g=2}^{3}\left[\pi_{1g}\left(I_{gs} \times I_{post} \times (T - 2008)\right)\right] + \sum_{g=2}^{3}\left[\emptyset_{1g}\left(I_{gs} \times I_{post}\right)\right] +$$

$$\rho_{12}\left(I_{post}\right) + \delta\boldsymbol{C}_{st} + \alpha_{1s} + \theta_t + (\tau_{1s} \times T) + \varepsilon$$

$$(3)\quad Y_{st} = \beta\left(\widehat{PPE}_{st}\right) + \sum_{g=2}^{3}\left[\emptyset_{1g}\left(I_{gs} \times I_{post}\right)\right] + \rho_{22}\left(I_{post}\right) + \delta\boldsymbol{C}_{st} + \alpha_{2s} + \theta_t +$$

$$(\tau_{2s} \times T) + \varepsilon$$

The subscript g refers to the reliance categories, with 1 representing low-reliance, 2 representing medium-reliance, and 3 representing high-reliance. I_{gs} is a state specific indicator for group membership, I_{post} is an indicator for years after 2008, and $(T - 2008)$ indicates current year, T, relative to 2008. $\boldsymbol{C}_{st}$ is a vector of state and year-specific Bartik economic controls for the unemployment rate and average annual wage, which are defined in JWX. State (α_{*s}) and year (θ_{*t}) fixed effects are included, along with the pre-recession, state-specific time trends (τ_{*s}). The second stage outcome variable, Y_{st}, is either a) average standardized NAEP scores or college-going rate. β, the coefficient on the predicted values from the first stage, $\widehat{PPE}_{st}$, is the coefficient of interest.

We test the sensitivity of this model to 1) the use of different data sources for the instrument and PPE and 2) different state group definitions. We estimate the model using the three data sources discussed previously: Census District, Census State, and NCES. In addition to total expenditures from each source, we also estimate the model using current spending from Census State and NCES. To test the impact of state group definitions, we vary state-reliance group membership as described in the motivation section above and outlined in Table 4.

Randomization Inference

When working with finite samples it is often the case that a small number of data points or clusters are highly leveraged. This leverage can result in volatile standard errors and over-rejection of the null (Young, 2019a). Young (2019b) shows that in the presence of heteroscedasticity and leverage, conventional F-statistics are likely to indicate a strong first stage even when instruments are weak at much higher probability than previously understood. The following excerpt from Young (2019b) captures the issue well:

> In a world in which economists experiment with plausible instruments in the privacy of their offices, publicly reported results could easily be filled with instruments which, while legitimately exogenous in the population, are nevertheless irrelevant or very nearly so, with the strong reported F being the result of an unfortunate finite sample correlation with the endogenous disturbances, producing undesirably biased estimates.

In the face of these challenges, randomization inference can be used to help us better understand the potential distribution of test statistics and their likelihood of occurring by chance (Imbens and Rosenbaum 2005). Randomization inference starts by defining the sharp null under which outcome values are the same regardless of treatment or control status. Treatment can then be repeatedly randomly assigned to units and the model estimated to generate a distribution for the test statistic. At its simplest, the estimated p-value using randomization inference is the percent of test statistics greater than the test statistic generated using the true data.

For this paper, we repeatedly randomize states to two groups, one with three states and one with four states, mimicking JWX low/high reliance groups, and estimate the group IV model described above. We use the JWX 2008 state revenue share and Census District spending in this exercise. We randomize 100,000 times for each outcome, and for each iteration, we capture the coefficient on per-pupil spending, its standard error, and the first stage F-statistic. We then use the coefficient estimates to plot the coefficient distribution and calculate the percent of estimates greater than the estimate from the true data (i.e., the estimated p-value).

Higher Education Confounding

We test for higher education confounding in two ways. First, we add state and local higher education appropriations per FTE and tuition and fees per FTE into the group IV as control variables. Next, we estimate the group IV replacing the dependent variable with state and local higher education appropriations per FTE and tuition and fees per FTE. If controlling for higher education finance variables affects the estimated impact of PPE on college going or instrumented PPE predicts these higher education funding variables, then the exclusion restriction may be violated in the JWX model.

JWX provides results from a similar robustness test (see Table A18 in the online appendix). However, while they use levels of tuition, fees, and aid we use per FTE measures analogous to per-pupil spending for K-12. Using a per-FTE measure adjusts for state size, similar to PPE K-12 spending. These per-FTE measures are not mechanically linked to the college-going rate because the enrollment used to calculate the rate occurs in the year following our per-FTE measures.

V. Results

Event Study Graphs

Figures 10-12 provide the event study graphs from our analysis. These figures include a dot indicating the point estimate for the year-specific difference between states that are above versus below the median for 2008 K-12 state revenue share, as well as 95 percent confidence intervals around the point estimates. Point estimates are relative to 2007, the omitted year in the model.

We find some evidence that, following the recession, PPE declined more in states where pre-recession K-12 state revenue share was above the median (Figure 10). However, for NAEP scores (Figure 11) and college-going rates (Figure 12) we find no statistically significant differences between states that are above versus below the median K-12 state revenue share.

Our event study results are different from JWX, which finds significant divergence after 2007 in NAEP scores and college-going rates for states that are above versus below the state revenue share median.

Data Source

Tables 5A and 5B present the results using different data sources for the instrument and spending. All results presented in these tables use the low/high-reliance thresholds of 0.33 and 0.67. We have three different sources for 2008 state revenue share and six different sources for per-pupil spending, resulting in 18 possible combinations for each outcome.

The letters A and B in the table names correspond to our two outcome variables: A) NAEP scores and B) college-going rates. We estimate results both including (Column 2) and excluding (Columns 4-8) Washington D.C. Columns 1 and 3 provide the original results from JWX for comparison. Cells highlighted in yellow are not statistically insignificant at the 5 percent level. Standard errors and F-statistics are provided below point estimates.

For both NAEP (Table 6A) and college-going (Table 6B), our replication estimates using the original JWX categorization and data source (Row 1, Columns 1 & 3) are similar in magnitude and significance to the JWX results both including and excluding D.C.

Out of the 18 total combinations for NAEP, only 5 are statistically significant. When D.C. is excluded, only 2 out of 15 are statistically significant. Those two require the use of the JWX state share for the instrument.

Results are similar for college-going. Out of the 18 total combinations for the college-going rate, only 5 are statistically significant. When D.C. is excluded, only 3 out of 15 are statistically significant. Those three require the use of the JWX state share for the instrument.

For both outcome variables, all iterations that exclude D.C. and use Census State or NCES data to define the instrument are insignificant. These tables show that finding a significant result is contingent on making very specific data choices, and that small deviations from these choices leads to insignificant results.

State Group Definitions

Tables 6A and 6B present results using different state group definitions. The general table setup is the same as Table 5. All results in these tables use the JWX 2008 state revenue share to define the instrument.

We model each of the six state categorization frameworks presented in Table 4: the JWX thresholds (Row 1 in Tables 6A & B) as well as categorizations based on 1/3 and 2/3 (Row 2), the 10^{th} and 90^{th} percentiles (Row 3), 15th and 85^{th} percentiles (Row 4), 20^{th} and 80^{th} percentiles (Row 5), and 25th and 75^{th} percentiles (Row 6). As in the previous section, we also provide results for six different sources for per-pupil spending. This results in a total of 36 total combinations.

For both outcome variables, when we apply categorizations based on percentiles of state K-12 revenue share, the point estimates shrink toward zero and lose significance. Somewhat surprisingly, even the most modest grouping change, using thresholds of 1/3 and 2/3, results in insignificant point estimates in all but one instance. For NAEP, only 4 of 36 combinations yield statistically significant results, and when D.C. is excluded, only 2 out of 30 are significant. For college-going, only 4 out of 36 results are significant, and when D.C. is excluded, only 3 out of 30 are significant.

Our estimates indicate that the JWX group IV results are highly sensitive to changes in state group composition. Using thresholds of 1/3 and 2/3 assigns South Dakota to the low-reliance group but results in no other changes (see column 2 of Table 4). However, even this minor change results in insignificant results for both outcome variables and all six PPE data sources except for one (i.e., NAEP using Census District and including D.C.). Similarly, using the top and bottom decile of pre-recession state K-12 revenue share only adds two states to the low-reliance group and one state to the high-reliance group relative to the JWX categorization (see column 2 of Table 4), but again, even this small change yields statistically insignificant results for both NAEP and college-going.

Randomization Inference

Figures 13 and 14 present the results of our randomization inference exercise for NAEP scores and college-going, respectively. Each of the figures provides two panels. The first panel shows the distribution for all 100,000 estimates, and the second shows the distribution for the subsample with first stage F-statistics greater than 10. The vertical red lines on the graph represents the estimated coefficient using the true data. The estimated p-value is provided at the top of each red line.

For both outcomes, randomization inference implies p-values that are much larger than those estimated via conventional means and that are far from statistical significance. This is true for the both the wider distribution of all results and the narrower distribution generated from the subsample with F-statistics greater than 10. For NAEP, the estimated p-value is 0.36 for the full sample and 0.23 for the subsample. For college-going, the estimated p-values are closer together – 0.41 for the full sample and 0.38 for the subsample. These results indicate that conventional estimation may be overstating the statistical significance of the estimates generated by the JWX model.

Higher Education Confounding

As outlined in IV, we test for confounding in two ways. First we include state and local higher education appropriations per FTE and tuition and fees per FTE as control variables in the group IV model for college-going. Second, we replace the dependent variable in the group IV model with our two higher education revenue variables. Table 7 provides the results for these two tests using both the JWX thresholds for high/low-reliance and [1/3, 2/3] thresholds. We use the JWX 2008 state revenue share and Census District spending in this exercise.

We find that, for both threshold sets, controlling for our two higher education revenue variables yields insignificant point estimates for college-going and that instrumented PPE is statistically significant and positively related to state and local higher education appropriations per FTE. We do not find a statistically significant relationship between instrumented PPE and tuition and fees per FTE.

Based on the results of our tests, we believe that the college-going finding is confounded with contemporaneous higher education trends. While it may be feasible that both K-12 spending and higher education spending impact college-going, we do not think it is possible to disentangle those impacts using a JWX-style IV approach.

VI. Conclusions

As states face looming budget shortfalls and schools grapple with the effects of COVID-19, it has become increasingly important to understand the potential impact of education funding reductions on student outcomes. Using data from the years around the Great Recession, JWX suggests that K-12 spending reductions have widespread negative impacts on both student achievement and attainment.

While we believe that spending reductions may have negative impacts on student outcomes, our results suggest that there is not a clear and compelling story about the impact of school spending reductions related to the Great Recession on student achievement. Moreover, we find that the relationship between K-12 spending and college-going rates is likely confounded with contemporaneous higher education funding trends. Specifically, we find:

- the JWX group IV model results are sensitive to reasonable changes in data source for state revenue share and per-pupil expenditures;
- the JWX group IV model categorizes very few states as having low-reliance (***three states***) or high-reliance (***four states***) on state revenue to fund K-12 education, and the results are sensitive to small changes in group composition; and
- post-recession K-12 and higher education funding trends are related, and as a result, the JWX finding that K-12 spending reductions impacted state college-going rates is likely confounded with contemporaneous higher education funding trends.

Our results highlight the challenges of estimating causal effects for loosely defined interventions (e.g., funding changes) at the level of states and the importance in exercising caution when making policy recommendations based on such studies. It is not surprising that it is difficult to empirically estimate the impact of school funding on student outcomes. K-12 education is a complex system and how schools deal with funding changes varies widely both within and across states. Unless those changes are substantial and have a direct, sizeable impact on classrooms, student outcomes are unlikely to respond quickly enough to be captured in short timeframes. Therefore, it is questionable whether rigorous research can consistently show that school spending matters for educational outcomes on such a broad scale.

REFERENCES

Bound, J., Braga, B., Khanna, G., & Turner, S. (2019). Public universities: The supply side of building a skilled workforce. RSF: The Russell Sage Foundation Journal of the Social Sciences, 5(5), 43-66.

Candelaria, C.A. and Kenneth A. Shores, K. (2017). Court-Ordered Finance Reforms in The Adequacy Era: Heterogeneous Causal Effects and Sensitivity. Education Finance and Policy, pages 1–91, 6

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

Coleman, J. S. Equality of Educational Opportunity. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 1966.

Deming DJ, Walters CR. (2017) The Impacts of Price and Spending Subsidies on U.S. Postsecondary Attainment. Working Paper.

Fuchsman, D., & Zamarro, G. (2019). Local Labor Market Conditions, Principals' Leadership, Conscientiousness, and Beginning Teacher Turnover: A Study During the Great Recession. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3322023

Gigliotti, P. and Lucy Sorensen, L. (2018). Education Expenditures and Student Performance: Evidence from the Save Harmless Provision in New York State. Economics of Education Review, pages 167–182, 10

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The economic journal*, *113*(485), F64-F98.

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, 11.

Imbens, G., & Rosenbaum, P. (2005). Randomization inference with an instrumental variable. Journal of the Royal Statistical Society, Series A, 168(1).

Jackson, C. K. "Rebuttal_AEJ_arkansas_post.pdf" (2020). Available at: http://works.bepress.com/c_kirabo_jackson/41/

Jackson, C. K., Wigger, C., & Xiong, H. (2020a). Do school spending cuts matter? Evidence from the great recession. *American Economic Journal: Economic Policy*. forthcoming.
Northwestern University, February 27,2020b
National Bureau of Economic Research, August 2019

Jackson, C. K. (2020). *Does school spending matter? The new literature on an old question*. American Psychological Association.
https://psycnet.apa.org/record/2020-30963-008

Jackson, C. K., Johnson, R. C., & Persico, C. (2015). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms (No. w20847). *Cambridge, MA: National Bureau of Economic Research.* https://www.nber.org/papers/w20847

Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, *10*(2), 1-26. https://www.aeaweb.org/articles?id=10.1257/app.20160567

Lee, D. L., McCrary, J., Moreira, M. J., & Porter, J. (2020). Valid t-ratio Inference for IV. *arXiv preprint arXiv:2010.05058*.

Lee, K. G., & Polachek, S. W. (2018). Do school budgets matter? The effect of budget referenda on student dropout rates. *Education Economics*, *26*(2), 129-144.

Miller, C. (2017). The Effect of Education Spending on Student Achievement: Evidence from Property Wealth and School Finance Rules.

Nagler, M., Piopiunik, M., & West, M. R. (2015). *Weak markets, strong teachers: Recession at career start and teacher effectiveness* (No. w21393). National Bureau of Economic Research.

Papke, L.E. "The Effects of Spending on Test Pass Rates: Evidence from Michigan." Journal of Public Economics 89, no. 5–6 (2005): 821–39. doi:10.1016/j.jpubeco.2004.05.008.

Roy, Joydeep. Impact of School Finance Reform on Resource Equalization and Academic Performance: Evidence from Michigan. Education Finance and Policy 6, no. 2 (2011): 137–67.

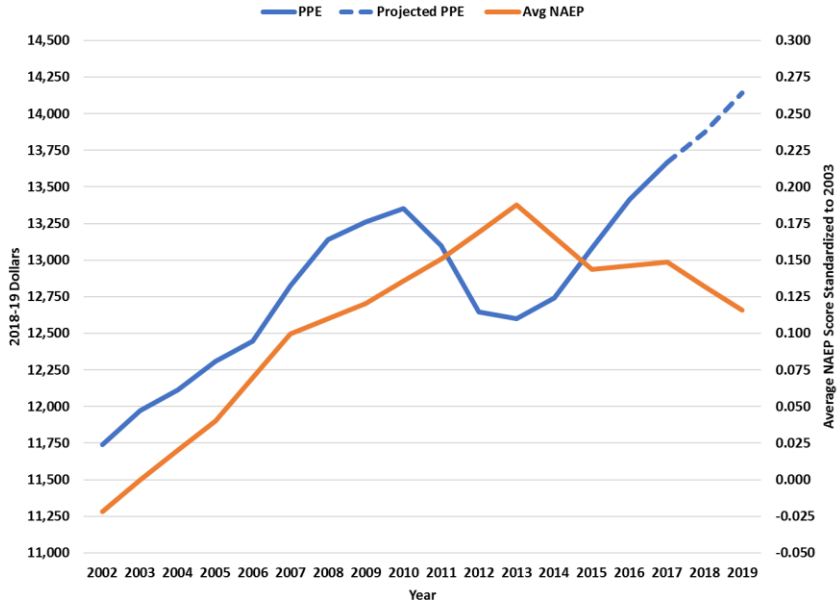
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological science, 22(11), 1359-1366.

Young, A. (2019a). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. The Quarterly Journal of Economics, 134(2), 557-598.

Young, A. (2019b). Consistency without inference: Instrumental variables in practical application.

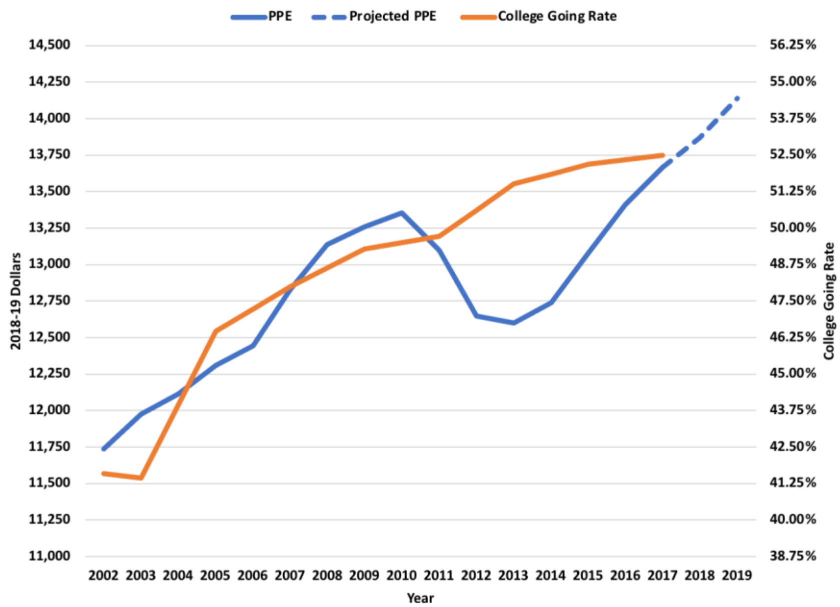
Tables and Figures

Figure 1: Per-Pupil Current Expenditures (PPE) and Average NAEP Scores



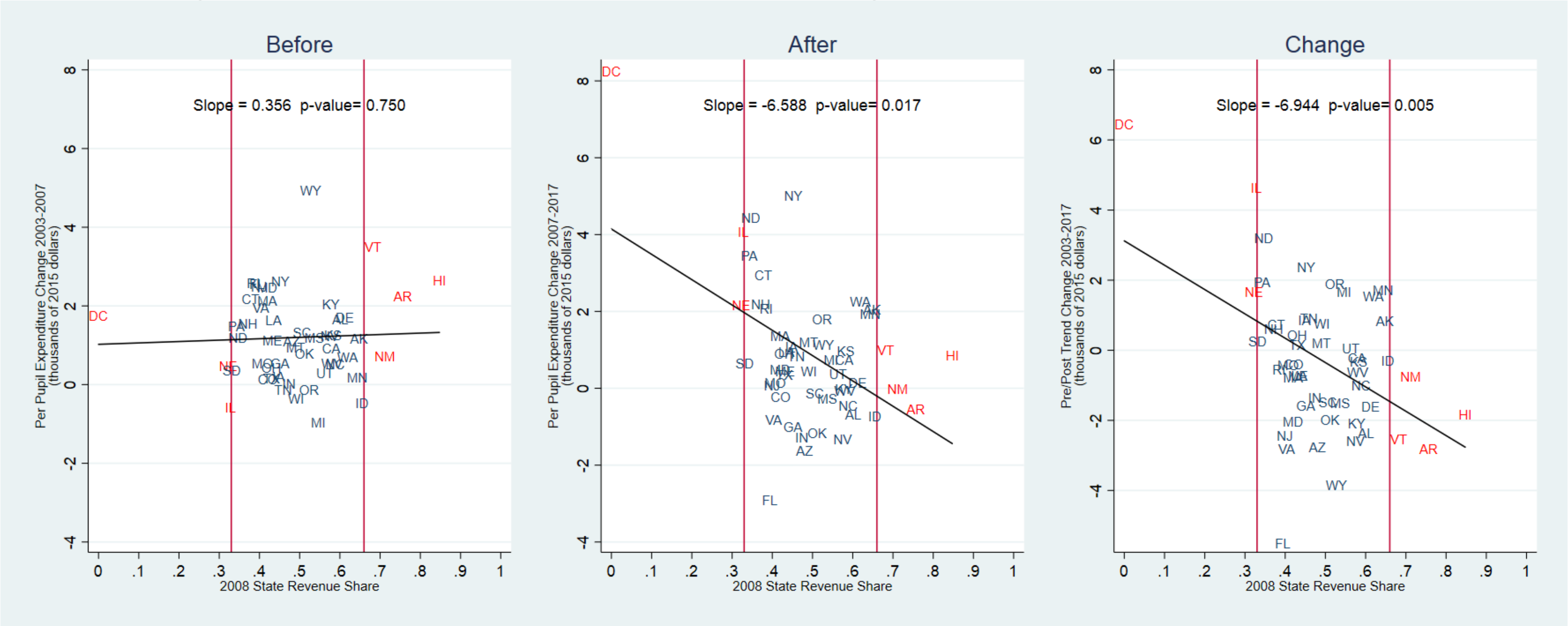
Notes: PPE was collected from the 2019 NCES Digest of Education Statistics table 236.55. Average NAEP is the national public average score on reading and mathematics tests standardized to 2003. NAEP data were collected from the NAEP Data Explorer. Only the reading test was administered in 2002. Mathematics test scores for that year were imputed using linear interpolation

Figure 2: Per-Pupil Current Expenditures (PPE) and College-Going Rate



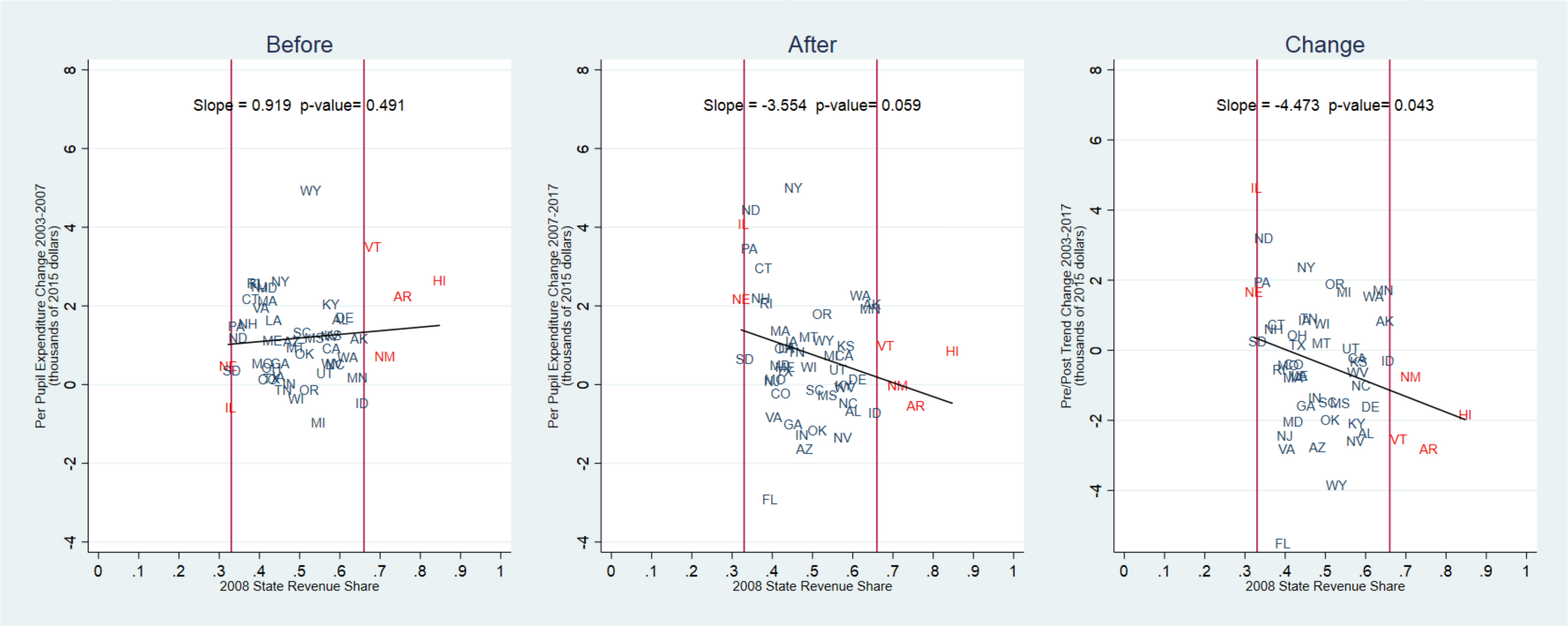
Notes: PPE was collected from the 2019 NCES Digest of Education Statistics table 236.55. The college-going rate is the number of first-time enrollees in the fall semester who graduated high school in the past 12 months divided by the Census estimate for the number of 17-year-olds in the previous year. The national rate was calculated by summing the numerator and denominator across states in each year. See the Data section for more details on the data and calculation.

Figure 3: Relationship between Per-Pupil Expenditure Change and State Revenue Share – Census District



Notes: The red lines represent the percentile labeled at the top of the figure. For illustrative purposes, Washington D.C. is included in the figure with a value of zero despite being listed as N/A in the NCES data.

Figure 4: Relationship between Per-Pupil Expenditure Change and State Revenue Share – Census District dropping D.C.



Notes: The red lines represent the percentile labeled at the top of the figure. For illustrative purposes, Washington D.C. is included in the figure with a value of zero despite being listed as N/A in the NCES data.

Figure 5: Relationship between Per-Pupil Expenditure Change and State Revenue Share – Census State

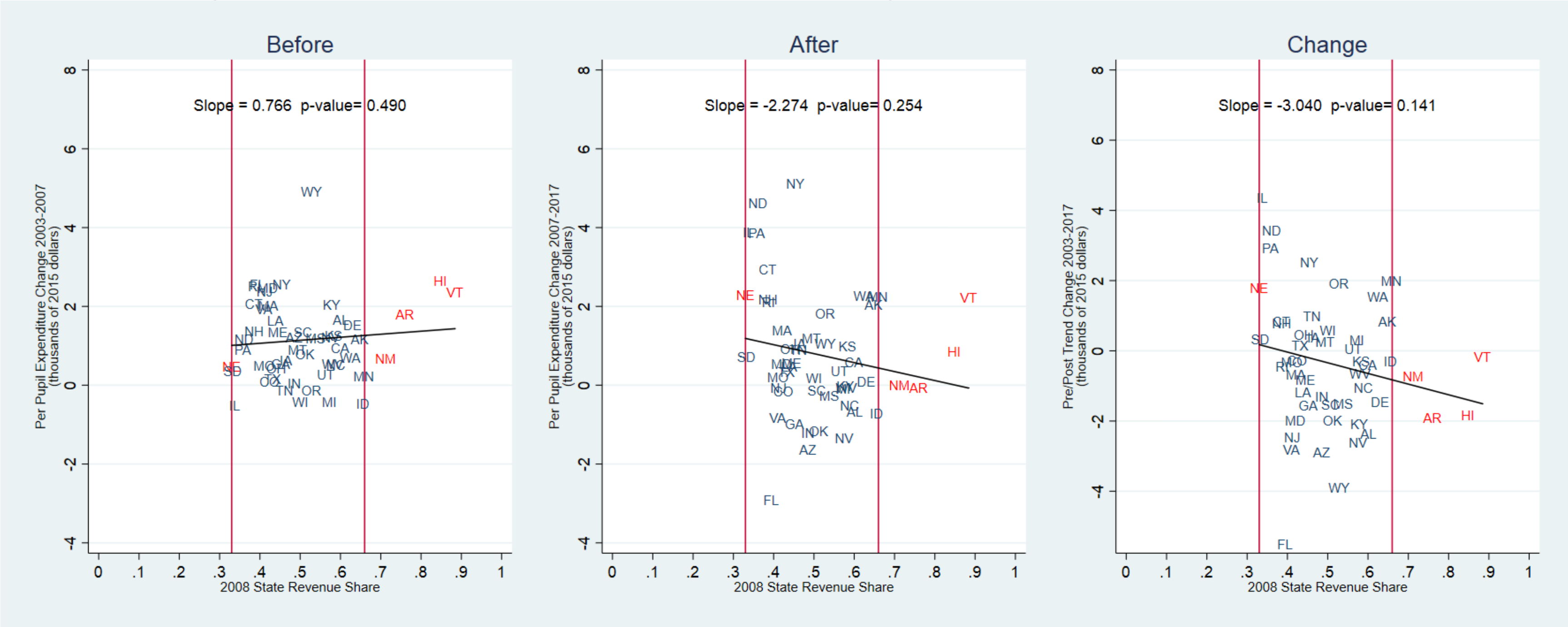


Figure 6: Relationship between Per-Pupil Expenditure Change and State Revenue Share – NCES

Figure 7: College-Going IV Directed Acyclic Graphs (DAGs)

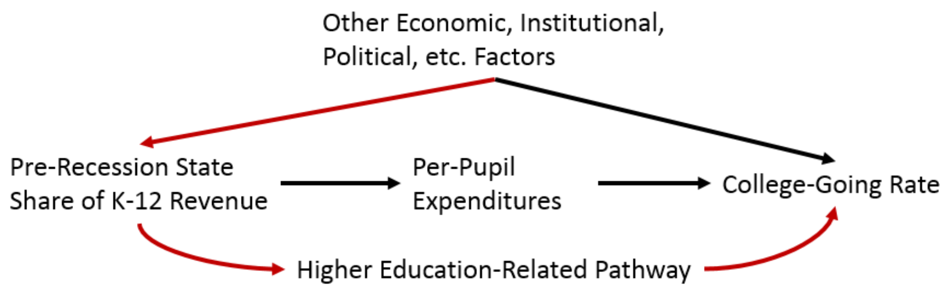
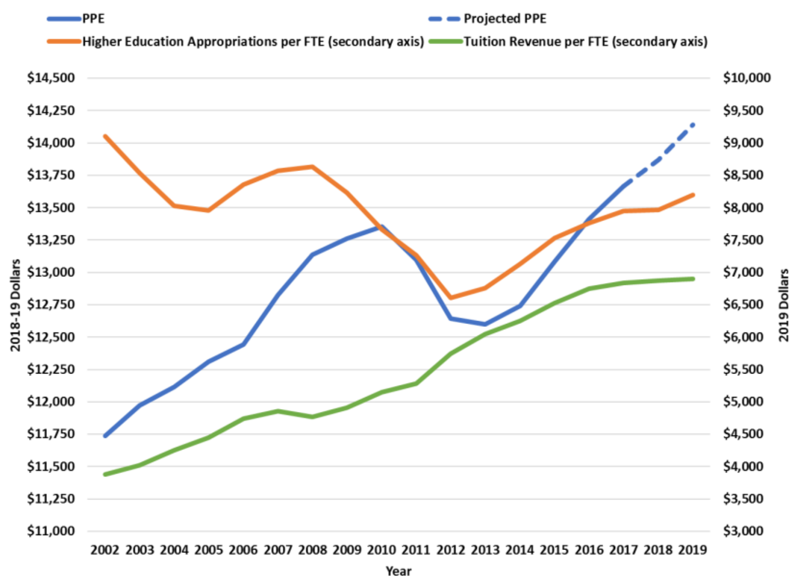
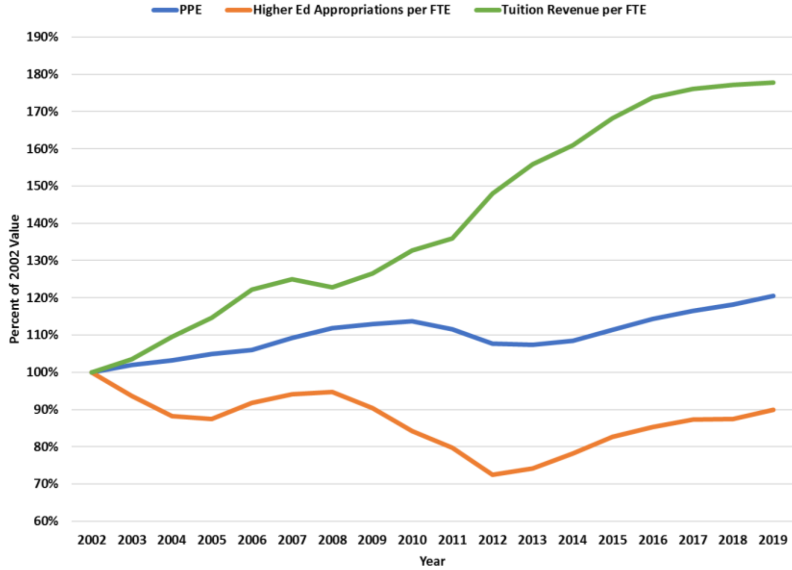


Figure 8: Per-Pupil Current Expenditures (PPE) and Higher Education Funding Trends



Notes: PPE was collected from the 2019 NCES Digest of Education Statistics table 236.55. State and local appropriations per FTE and tuition and fees per FTE are obtained from the State Higher Education Executive Officers Association (SHEEO, 2019).

Figure 9: 2002 Indexed Per-Pupil Current Expenditures (PPE) and Higher Education Funding Trends



Notes: PPE was collected from the 2019 NCES Digest of Education Statistics table 236.55. State and local appropriations per FTE and tuition and fees per FTE are obtained from the State Higher Education Executive Officers Association (SHEEO, 2019).

Figure 10: Per-Pupil Expenditures Event Study Graph Using the Median as Percent Threshold for High-Reliance

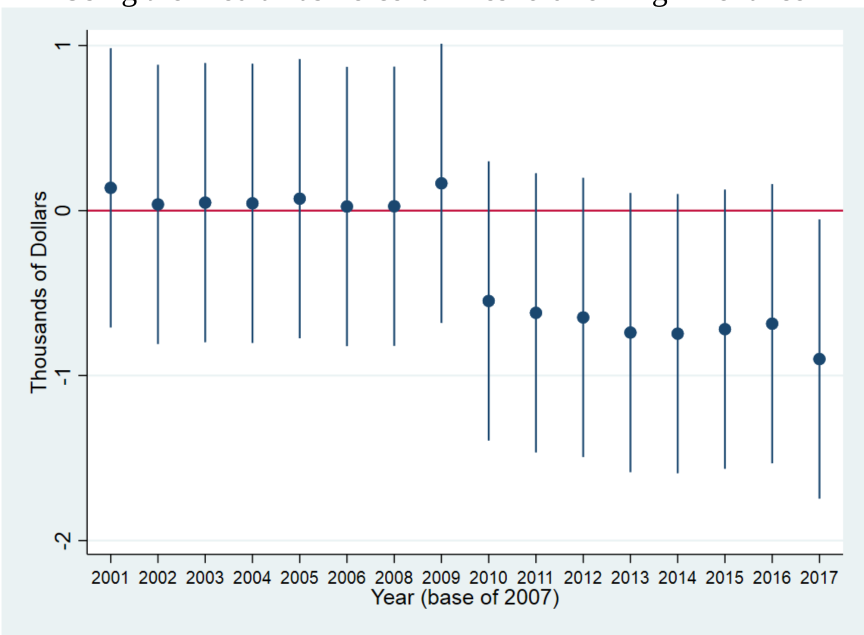


Figure 11: NAEP Scores Event Study Graph
Using the Median as Percent Threshold for High-Reliance

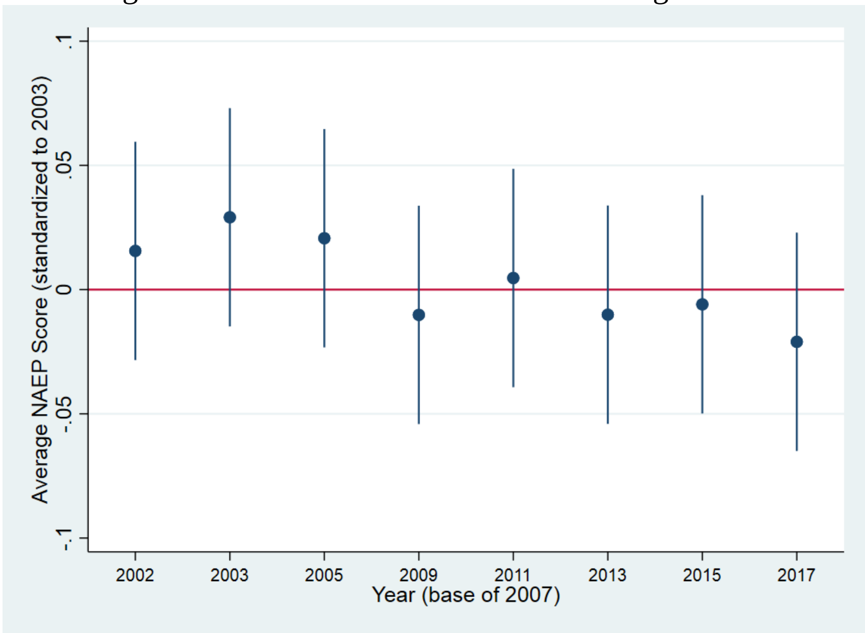


Figure 12: College-Going Rate Event Study Graph
Using the Median as Threshold for High-Reliance

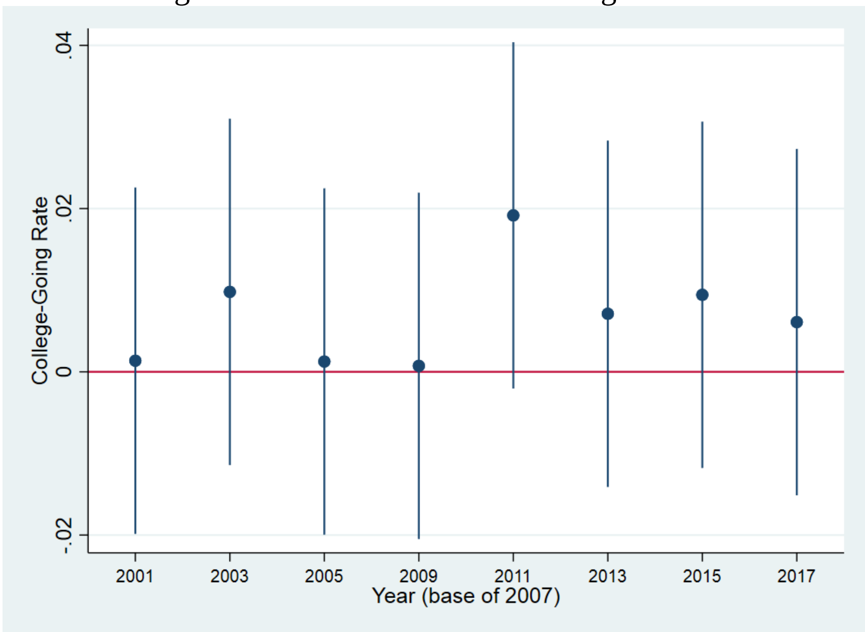


Figure 13: Point Estimate Distribution and P-Value from Randomization Inference Exercise – NAEP

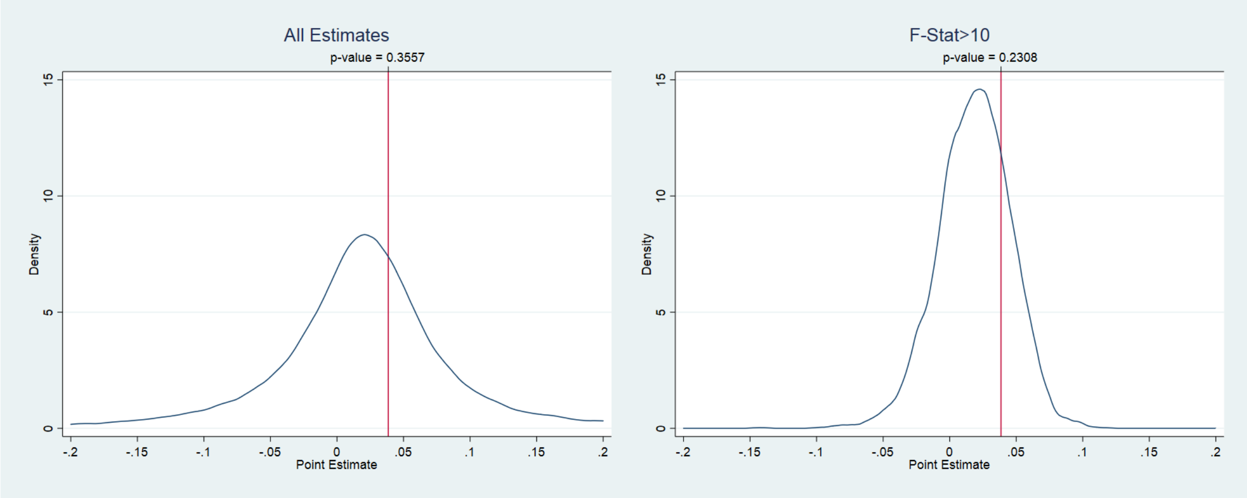


Figure 14: Point Estimate Distribution and P-Value from Randomization Inference Exercise – College Going Rate

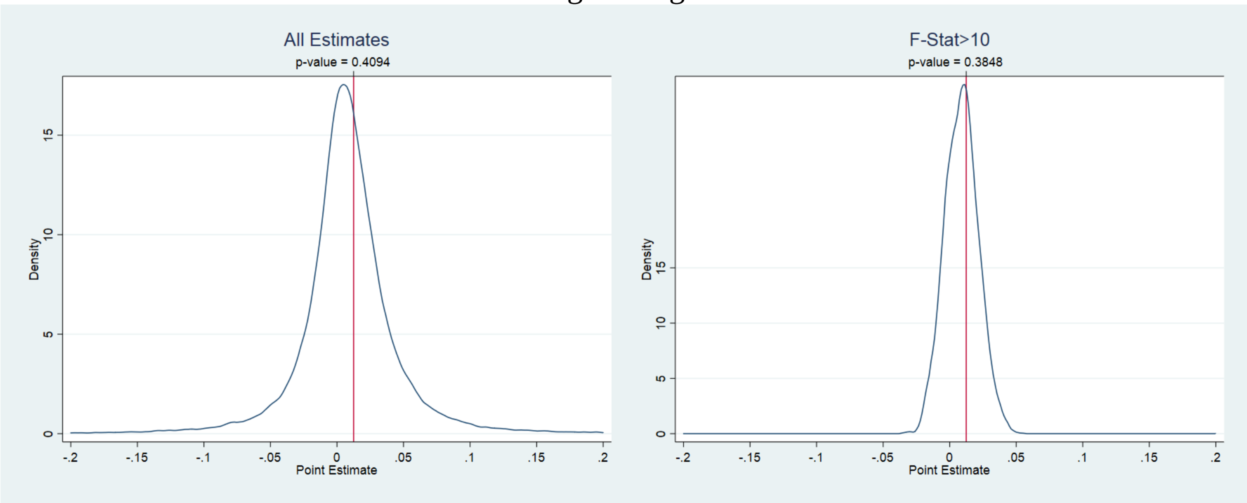


Table 1: Study Data Summary Statistics

	JWX Values			Replication Values			P-Value for t-test of Difference in Means
	N	Mean	SD	N	Mean	SD	
Total Per-pupil spending: Census District (2015 dollars)	510	13,208	3,807	510	13,214	3,804	0.97
Total Per-pupil spending: Census State (2015 dollars)				510	12,922	3,593	
Total Per-pupil spending: NCES State (2015 dollars)				510	12,583	3,440	
Current Per-pupil spending: Census State (2015 dollars)				510	11,067	2,927	
Current Per-pupil spending: NCES State (2015 dollars)				510	11,192	3,043	
Share of revenue from state sources in 2008 (JWX)	51	0.493	0.137	51	0.493	0.137	
Share of revenue from state sources in 2008 (Census State)				51	0.502	0.144	
Share of revenue from state sources in 2008 (NCES)				51	0.496	0.144	
Average NAEP Z-Score (standardized to 2003)	459	0.125	0.211	459	0.124	0.211	0.94
College Enrollment Rate (17 & 18-year-olds)	459	0.475	0.078	459	0.475	0.078	0.90
Bartik Unemployment Rate				510	5.870	1.655	
Bartik Average Annual Wage (2015 dollars)				510	43,140	5,465	
Tuition per FTE (2015 dollars)				857	5,664	2,705	
Appropriations per FTE (2015 dollars)				857	7,335	2,456	

Notes: Variables were collected for each state and the District of Columbia for the following years: 2001, 2002, 2003, 2005, 2007, 2009, 2011, 2013, 2015, and 2017. College-going variables are available for odd numbered years starting in 2001. We use NAEP from 2002 matching JWX. NAEP test scores are standardized to 2003, and then averaged for each state and year. Only the reading test was administered in 2002, and five states do not have 2002 NAEP scores: AK, CO, NH, NJ, and SD. We used 2003 average reading z-score as the 2002 value for these five states. We model the college-going rate as the number of first-time college enrollees divided by the average of the number of 17 & 18-year-olds in a state.

Table 2: 2008 State K-12 Revenue Share

		(1)	(2)	(3)
		JWX	Census State	NCES
State	Abbreviation	2008 State Revenue Share		
Alabama	AL	60.2%	60.2%	60.6%
Alaska	AK	64.9%	64.9%	66.3%
Arizona	AZ	48.1%	48.5%	51.7%
Arkansas	AR	75.7%	76.0%	56.7%
California	CA	57.9%	59.9%	61.3%
Colorado	CO	42.1%	42.4%	42.2%
Connecticut	CT	37.8%	38.5%	39.6%
Delaware	DE	61.2%	63.0%	62.0%
District of Columbia	DC	00.0%	N/A	N/A
Florida	FL	39.4%	39.4%	38.8%
Georgia	GA	45.1%	45.2%	45.4%
Hawaii	HI	84.8%	84.8%	84.8%
Idaho	ID	65.5%	65.5%	67.1%
Illinois	IL	32.9%	33.8%	31.2%
Indiana	IN	47.3%	48.5%	53.5%
Iowa	IA	44.8%	46.5%	46.5%
Kansas	KS	58.4%	58.4%	57.5%
Kentucky	KY	57.9%	57.9%	57.3%
Louisiana	LA	43.6%	43.9%	44.8%
Maine	ME	43.2%	44.5%	44.9%
Maryland	MD	42.0%	42.0%	42.1%
Massachusetts	MA	41.8%	42.1%	41.9%
Michigan	MI	54.6%	57.3%	57.5%
Minnesota	MN	64.4%	65.8%	65.9%
Mississippi	MS	53.7%	53.8%	54.5%
Missouri	MO	40.8%	41.1%	33.3%
Montana	MT	49.0%	49.4%	49.7%
Nebraska	NE	32.3%	33.0%	33.1%
Nevada	NV	57.5%	57.5%	30.8%
New Hampshire	NH	37.1%	38.6%	38.6%
New Jersey	NJ	40.0%	41.3%	42.1%
New Mexico	NM	71.2%	71.2%	70.8%
New York	NY	45.2%	45.4%	44.8%
North Carolina	NC	58.8%	58.8%	65.7%
North Dakota	ND	34.7%	36.1%	36.3%
Ohio	OH	43.0%	44.1%	45.6%
Oklahoma	OK	51.2%	51.2%	54.2%
Oregon	OR	52.4%	52.8%	52.3%
Pennsylvania	PA	34.3%	35.8%	36.5%
Rhode Island	RI	38.5%	38.7%	39.9%
South Carolina	SC	50.6%	50.7%	50.8%
South Dakota	SD	33.1%	33.2%	33.9%
Tennessee	TN	45.9%	46.1%	45.6%
Texas	TX	43.1%	43.2%	44.8%
Utah	UT	56.3%	56.3%	56.7%
Vermont	VT	68.3%	88.5%	85.9%
Virginia	VA	40.3%	41.0%	41.0%
Washington	WA	61.9%	62.4%	62.5%
West Virginia	WV	58.1%	58.1%	59.1%
Wisconsin	WI	49.2%	50.1%	50.0%
Wyoming	WY	52.8%	52.9%	52.8%

Note: Census values are provided because JWX differs from the Census state-level data. Green is used to highlight states that have 2008 state revenue share less than 0.33 and red is used to highlight states that have 2008 state revenue share greater than 0.67.

Table 3: Total Per Pupil Expenditure Change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
State	JWX Categorization	2007-17 (10yr) Real Total Per-Pupil Expenditure Change			Rank from largest positive change to largest negative change		
		Census District	Census State	NCES	Census District	Census State	NCES
Alabama	M	-643	-646	-672	43	43	43
Alaska	M	2,081	2,081	2,156	11	13	11
Arizona	M	-1,599	-1,604	-1,038	50	50	49
Arkansas	H	-505	-44	-117	42	37	37
California	M	762	621	822	25	24	22
Colorado	M	-199	-127	-246	39	40	40
Connecticut	M	2,983	2,980	2,518	6	6	6
Delaware	M	175	125	-159	32	32	38
District of Columbia	L	8,287	8,346	6,197	1	1	1
Florida	M	-2,873	-2,873	-2,525	51	51	51
Georgia	M	-972	-964	-998	46	46	48
Hawaii	H	887	887	881	22	22	21
Idaho	M	-690	-690	-832	44	44	47
Illinois	L	4,108	3,911	3,870	4	4	4
Indiana	M	-1,253	-1,184	-728	48	48	45
Iowa	M	1,125	1,079	1,030	17	18	17
Kansas	M	1,013	1,029	309	19	19	29
Kentucky	M	10	9	411	36	34	27
Louisiana	M	969	513	596	20	27	25
Maine	M	474	586	956	28	25	20
Maryland	M	520	568	472	27	26	26
Massachusetts	M	1,402	1,413	2,314	14	15	9
Michigan	M	779	-56	-739	24	38	46
Minnesota	M	1,955	2,278	2,247	12	9	10
Mississippi	M	-240	-240	-16	40	41	33
Missouri	M	165	241	-19	33	30	34
Montana	M	1,233	1,228	1,037	15	16	16
Nebraska	L	2,203	2,320	1,445	9	7	13
Nevada	M	-1,304	-1,318	-1,107	49	49	50
New Hampshire	M	2,222	2,218	2,418	8	11	7
New Jersey	M	118	-27	-212	34	35	39
New Mexico	H	15	27	-66	35	33	35
New York	M	5,043	5,141	4,142	2	2	3
North Carolina	M	-430	-473	-630	41	42	42
North Dakota	M	4,467	4,649	4,745	3	3	2
Ohio	M	943	957	258	21	20	30
Oklahoma	M	-1,132	-1,124	-675	47	47	44
Oregon	M	1,821	1,849	1,441	13	14	14
Pennsylvania	M	3,493	3,898	2,365	5	5	8
Rhode Island	M	2,101	2,140	1,270	10	12	15
South Carolina	M	-92	-112	-110	38	39	36
South Dakota	M	673	755	673	26	23	24
Tennessee	M	863	927	967	23	21	19
Texas	M	379	371	335	31	29	28
Utah	M	403	397	216	30	28	32
Vermont	H	1,027	2,245	3,038	18	10	5
Virginia	M	-790	-805	-622	45	45	41
Washington	M	2,298	2,289	2,049	7	8	12
West Virginia	M	-31	-35	774	37	36	23
Wisconsin	M	470	215	240	29	31	31
Wyoming	M	1,153	1,092	1,012	16	17	18

Note: In column 1, green is used to highlight states are categorized as low-reliance in JWX and red is used to highlight states that are categorized as high-reliance in JWX.

Table 4: Categorizations for Group IV using JWX State Share

	(1) JWX Groups	(2) 1/3 and 2/3	(3) 10th and 90th Percentiles	(4) 15th and 85th Percentiles	(5) 20th and 80th Percentiles	(6) 25th and 75th Percentiles
Low-Reliance	**(3 States)** ***D.C.*** IL NE	**(4 States)** ***D.C.*** IL NE **SD**	**(5 States)** ***D.C.*** IL NE **PA** **SD**	**(7 States)** ***D.C.*** IL **ND** NE **NH** **PA** **SD**	**(10 States)** **CT** ***D.C.*** **FL** IL **ND** NE **NH** **PA** **RI** **SD**	**(13 States)** **CT** ***D.C.*** **FL** IL **MO** **ND** NE **NH** **NJ** **PA** **RI** **SD** **VA**
High-Reliance	**(4 States)** AR ***HI*** NM VT	**(4 States)** AR ***HI*** NM VT	**(5 States)** AR ***HI*** **ID** NM VT	**(7 States)** **AK** AR ***HI*** **ID** **MN** NM VT	**(10 States)** **AK** **AL** AR **DE** ***HI*** **ID** **MN** NM VT **WA**	**(12 States)** **AK** **AL** AR **DE** ***HI*** **ID** **KS** **MN** **NC** NM VT **WA**

Notes: Percentile groups are determined using NCES average 2005-07 state revenue share. States that are highlighted in red are different from the JWX categorization (column 1). D.C. and Hawaii are in bold and italics because their education funding systems are fundamentally different from other states. We drop D.C. from some analyses. Dropping D.C. results in MO moving into the bottom decile and PA into the bottom 15th percentile, otherwise the groups remain the same. Based on the text of JWX the group cutoffs are state share less than 1/3 for low reliance and state share greater than 2/3 for high reliance. When implementing these rules, JWX appears to have simplified to 0.33 and 0.66. If the 1/3 rule had been applied strictly, South Dakota should also have been included in the low-reliance group.

Table 5A: Group IV Using Different Data Sources – NAEP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Including Washington D.C.		**Excluding Washington D.C.**					
	Original JWX Results	**Census District Total Exp PPE**	**Original JWX Results**	**Census District Total Exp PPE**	**Census State Total Exp PPE**	**Census State Current Exp PPE**	**NCES Total Exp PPE**	**NCES Current Exp PPE**
JWX State Share	0.0385** [0.0110] 21.71	0.0393** [0.0115] 21.57	0.0283* [0.0120] 24.08	0.0317* [0.0143] 22.75	0.0333* [0.0153] 24.13	0.0455 [0.0235] 13.76	0.0475 [0.0304] 5.071	0.0587 [0.0445] 3.076
Census State Share		0.0478** [0.0103] 14.94		0.0379 [0.0191] 17.62	0.0411 [0.0207] 17.89	0.0532 [0.0334] 13.88	0.0575 [0.0457] 3.790	0.0397 [0.0456] 2.036
NCES State Share		0.0358** [0.0117] 2.124		0.0308 [0.0252] 0.773	0.0338 [0.0288] 0.778	0.0491 [0.0463] 0.758	0.0279 [0.0216] 0.904	0.0403 [0.0354] 0.532

Notes: Robust standard errors in brackets. F-Stat provided below standard errors. Cells with insignificant point estimates are highlighted in yellow. Negative point estimates are in red text. ** p<0.01, * p<0.05

Table 5B: Group IV Using Different Data Sources – College-Going Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Including Washington D.C.		**Excluding Washington D.C.**					
	Original JWX Results	**Census District Total Exp PPE**	**Original JWX Results**	**Census District Total Exp PPE**	**Census State Total Exp PPE**	**Census State Current Exp PPE**	**NCES Total Exp PPE**	**NCES Current Exp PPE**
JWX State Share	0.0124** [0.00387] 36.97	0.0125** [0.0039] 36.56	0.0100 [0.00501] 24.54	0.0100* [0.0050] 24.14	0.0102* [0.0050] 27.64	0.0135* [0.0060] 11.92	0.0142 [0.0073] 5.494	0.0180 [0.0100] 3.745
Census State Share		0.0138* [0.0055] 26.86		0.0112 [0.0084] 18.73	0.0114 [0.0085] 20.32	0.0151 [0.0095] 10.67	0.0199 [0.0156] 4.113	0.0213 [0.0162] 2.891
NCES State Share		0.0263 [0.0155] 3.511		0.0413 [0.0391] 1.467	0.0409 [0.0372] 1.527	0.0523 [0.0486] 1.062	0.0461 [0.0435] 1.464	0.0665 [0.0780] 0.622

Notes: Robust standard errors in brackets. F-Stat provided below standard errors. Cells with insignificant point estimates are highlighted in yellow. Negative point estimates are in red text. ** p<0.01, * p<0.05

Table 6A: Group IV Using Different State Groupings – NAEP

	(1)	(2)	(3)	(4)	(5)	(6)
	Including Washington D.C.	**Excluding Washington D.C.**				
High and Low Group Cutoffs	**Census District Total Exp PPE**	**Census District Total Exp PPE**	**Census State Total Exp PPE**	**Census State Current Exp PPE**	**NCES Total Exp PPE**	**NCES Current Exp PPE**
[0.33, 0.66]	0.0393** [0.0115] 21.57	0.0317* [0.0143] 22.75	0.0333* [0.0153] 24.13	0.0455 [0.0235] 13.76	0.0475 [0.0304] 5.071	0.0587 [0.0445] 3.076
[1/3, 2/3]	0.0329* [0.0127] 21.57	0.0235 [0.0146] 13.25	0.0244 [0.0156] 14.28	0.0345 [0.0266] 5.032	0.0356 [0.0231] 7.531	0.0455 [0.0386] 2.988
[10pctl, 90pctl]	0.0221 [0.0199] 6.111	0.0089 [0.0250] 5.459	0.0069 [0.0245] 10.44	0.0138 [0.0368] 4.423	0.0130 [0.0364] 3.271	0.0210 [0.0529] 2.040
[15pctl, 85pctl]	0.0093 [0.0242] 5.352	-0.0081 [0.0291] 4.519	-0.0096 [0.0251] 7.545	-0.0072 [0.0438] 4.222	-0.0175 [0.0361] 3.204	-0.0154 [0.0560] 1.991
[20pctl, 80pctl]	0.0264 [0.0319] 1.461	0.0204 [0.0480] 0.824	0.0085 [0.0411] 1.067	0.0043 [0.0657] 0.957	0.0329 [0.0961] 0.320	0.0166 [0.1363] 0.218
[25pctl, 75pctl]	0.0484 [0.0477] 0.861	0.0714 [0.0977] 0.430	0.0352 [0.0699] 0.493	0.0352 [0.1265] 0.247	0.2806 [0.7168] 0.0829	0.4713 [1.4658] 0.0527

Notes: Robust standard errors in brackets. F-Stat provided below standard errors. Cells with insignificant point estimates are highlighted in yellow. Negative point estimates are in red text. ** p<0.01, * p<0.05

Table 6B: Group IV Using Different State Groupings – College-Going Rate

	(1)	(2)	(3)	(4)	(5)	(6)
	Including Washington D.C.	**Excluding Washington D.C.**				
High and Low Group Cutoffs	**Census District Total Exp PPE**	**Census District Total Exp PPE**	**Census State Total Exp PPE**	**Census State Current Exp PPE**	**NCES Total Exp PPE**	**NCES Current Exp PPE**
[0.33, 0.66]	0.0125** [0.0039] 36.56	0.0100* [0.0050] 24.14	0.0102* [0.0050] 27.64	0.0135* [0.0060] 11.92	0.0142 [0.0073] 5.494	0.0180 [0.0100] 3.745
[1/3, 2/3]	0.0088 [0.0059] 29.81	0.0053 [0.0070] 20.82	0.0051 [0.0072] 23.60	0.0073 [0.0096] 5.991	0.0085 [0.0088] 7.778	0.0124 [0.0113] 3.837
[10pctl, 90pctl]	0.0110 [0.0063] 7.442	0.0081 [0.0082] 5.313	0.0067 [0.0074] 9.603	0.0120 [0.0101] 4.222	0.0111 [0.0105] 4.025	0.0154 [0.0126] 2.451
[15pctl, 85pctl]	0.0103 [0.0100] 5.649	0.0071 [0.0130] 3.796	0.0058 [0.0118] 5.409	0.0110 [0.0193] 3.566	0.0075 [0.0158] 2.924	0.0131 [0.0230] 1.763
[20pctl, 80pctl]	0.0063 [0.0155] 1.765	0.0024 [0.0210] 1.084	0.0011 [0.0197] 1.271	-0.0010 [0.0302] 1.364	0.0028 [0.0307] 0.579	-0.0034 [0.0506] 0.402
[25pctl, 75pctl]	0.0025 [0.0198] 0.970	-0.0022 [0.0292] 0.533	-0.0063 [0.0282] 0.672	-0.0185 [0.0530] 0.451	-0.0058 [0.0670] 0.127	-0.0118 [0.1601] 0.0340

Notes: Robust standard errors in brackets. F-Stat provided below standard errors. Cells with insignificant point estimates are highlighted in yellow. Negative point estimates are in red text. ** p<0.01, * p<0.05

Table 7: Predict Tuition and State Appropriations Per FTE for Higher Education

	(1)	(2)	(3)	(4)	(5)	(6)
	Low/High Thresholds [0.33, 0.67]			Low/High Thresholds [1/3, 2/3]		
	College-Going Rate	Tuition per FTE	State Appropriations per FTE	College-Going Rate	Tuition per FTE	State Appropriations per FTE
Per-Pupil Spending (thousands)	0.0141	190.4263	752.6623**	0.0065	280.6700	688.4743**
	[0.0080]	[127.5699]	[260.5614]	[0.0098]	[158.0761]	[248.3484]
States	50	50	50	50	50	50
State-Year Observations	450	850	850	450	850	850
F-Stat	42.16	28.61	28.61	25.07	20.94	20.94

Notes: ** p<0.01, * p<0.05