THE MACROECONOMIC RETURNS TO SCHOOL FINANCE REFORMS: MICRO-CALIBRATED PREDICTIONS MEET CAUSAL EVIDENCE*

C. Kirabo Jackson†

August 21, 2026

A large literature posits a central role for human capital in determining macroeconomic output, while another documents substantial private returns to educational investment. However, whether private gains translate into aggregate gains remains uncertain, as signaling, displacement, or general-equilibrium adjustment may drive a wedge between private and aggregate returns. This paper bridges these literatures with causal evidence that state-level school finance reforms (SFRs) increase macroeconomic output. First, I combine design-based micro estimates of SFR effects on private earnings with a cohort replacement framework to simulate aggregate state earnings. The simulation predicts little change for 10 to 15 years, then gradual gains as treated cohorts enter the labor force, reaching 7.8 percent after 60 years. Next, I test this prediction using the staggered implementation of SFRs across U.S. states. In an event-study framework, aggregate wages per worker track the predicted path, remaining flat through about 15 years before rising roughly 4.7 percent above counterfactual levels after 30 years — consistent with private earnings gains aggregating to the broader economy. Finally, I show that these effects extend to broader economic output, with GDP per capita declining modestly initially before rising more than 5 percent after 30 years.

JEL Codes: [E24, H52, I22, I26, J24, O47]

Keywords: school finance reform, human capital, macroeconomic growth, event study, cohort aggregation.

**No funding was received for this project.* All errors are my own.
†*Northwestern University, kirabo-jackson@northwestern.edu*

I. Introduction

Do human-capital investments increase macroeconomic outcomes, and can microeconomic evidence predict these increases? A large literature has established that educational investments improve the educational attainment and earnings of directly affected students (Card and Krueger, 1992; Goldin and Katz, 2019; Duflo, 2001; Jackson et al., 2016; Rothstein and Schanzenbach, 2022). However, policymakers invest in education not simply to raise the earnings of treated individuals, but to improve aggregate economic performance. Indeed, in canonical models, human capital plays a central role in determining the level of output and potentially in sustaining growth (Lucas, 1988; Barro, 1991; Mankiw et al., 1992). Despite evidence linking cognitive skills to economic growth (e.g., Hanushek and Woessmann (2012, 2015); Angrist et al. (2021)), and calibrations linking micro evidence to macroeconomic outcomes (e.g., Bils and Klenow, 2000), credible causal evidence that educational investments increase aggregate economic performance remains scarce.

This paper seeks to bridge the gap between these literatures by providing causal estimates of the macroeconomic returns to a human-capital intervention, specifically state-level school finance reforms in the U.S. The analysis proceeds in two parts. First, I use design-based micro estimates to simulate the evolution of state-level macroeconomic outcomes implied by these reforms (aggregate wages per worker). Second, I exploit quasi-experimental variation in reform implementation to identify their effects on these outcomes and causally test the predictions of the simulation.

The difficulty in identifying the macroeconomic effects of human-capital investments is that the most credible evidence comes from microeconomic estimates of individual outcomes in partial equilibrium (Heckman et al., 1998), but these private returns need not aggregate mechanically. That is, private returns may overstate aggregate effects due to signaling (Spence, 1973) or general-equilibrium effects from large-scale educational investments (e.g., Duflo (2004); Khanna (2023)). Conversely, positive wage spillovers may cause aggregate earnings gains to exceed the sum of the private gains (e.g. (Moretti, 2004)). As such, whether credible microeconomic estimates translate into measurable gains in aggregate labor earnings remains an open empirical question.

Even if private earnings gains aggregate, broader macroeconomic effects remain uncertain. Fiscal costs and transitional adjustment may initially depress income or output, while higher productivity may gradually induce complementary investment. Whether these forces ultimately generate measurable gains in GDP is a separate and equally important open empirical question.

Accordingly, the analysis addresses two questions. First, do the private earnings gains documented in the microeconomics literature aggregate into measurable increases in aggregate labor earnings? Second, do these gains translate into broader improvements in aggregate output? To answer these questions, I exploit staggered school finance reforms (SFRs) across U.S. states between 1973 and 2005 and estimate their long-run effects on state-level wages and GDP.

To discipline the empirical analysis and make an explicit link between the micro estimates and aggregate outcomes, the first part of the paper asks what aggregate labor earnings would look like if the micro evidence aggregated entirely through a human-capital channel. To this aim, I develop a framework that combines quasi-experimental estimates of the private returns to school spending with life-cycle earnings profiles and demographic turnover. The framework simulates or predicts how individual earnings gains accumulate as reform-exposed cohorts replace older workers, generating an *ex ante* path for aggregate labor earnings (under the assumption of no spillovers or crowd-out). I calibrate the framework using estimates of the effects of school finance reforms on private earnings from Jackson et al. (2016) and Rothstein and Schanzenbach (2022). Because the prediction is constructed without reference to the aggregate data, it provides a falsifiable benchmark against which the macroeconomic evidence can be evaluated.

The key finding from the first section is that the simulation predicts a gradual transition. The micro estimates suggest that aggregate earnings per worker (or average wages) remain essentially unchanged for the first 10 to 15 years, then rise steadily as reform-exposed cohorts replace older workers, reaching approximately 3 percent by year 30 and converging to a steady-state increase of roughly 7.8 percent after about 60 years. Because the simulation aggregates only the *private* earnings gains of exposed cohorts, it provides a benchmark for the aggregate response in the absence of spillovers or displacement. Earnings below the predicted path might suggest signaling, displacement, or negative spillovers, whereas a steeper path might imply positive wage spillovers.

The second part of the paper takes the simulation as given and tests it empirically, asking whether the causal macroeconomic evidence matches the path predicted by the micro evidence. Identification comes from exploiting the staggered implementation of reforms across states and comparing the evolution of aggregate outcomes in reform states to that in non-reform states over the same periods. Specifically, I estimate the effects of SFRs on state-level average wages from the Quarterly Census of Employment and Wages (QCEW) and obtain dynamic treatment effects using the Callaway and Sant'Anna (2021) estimator. With both the simulated path and the causally estimated path (and measures of their uncertainty), I can formally test the predictions of the simulation from the first part of the paper. Although the full steady state takes roughly 60 years to materialize, many reforms allow effects to be estimated through 30 years post-reform, capturing much of the predicted transition and providing a meaningful test of the framework.

The first main finding from the second part of the paper is that the *observed* evolution of average wages matches the *predicted* path. Post-reform effects remain near zero through approximately year 13, the horizon at which the first fully treated cohorts begin reaching working age, before rising gradually over subsequent decades. By year 30 average wages have increased by approximately 4.7 percent. The timing and magnitude of the estimated response track the micro-calibrated prediction. Formal path tests reject the hypothesis that the two paths are unrelated ($p < 0.01$ in all specifications) and fail to reject that they are equal. These findings establish that investments in elementary and secondary education generate economically meaningful long-run increases in aggregate labor earnings. Moreover, they show that a simple aggregation of previously estimated microeconomic returns accurately predicts both the timing and magnitude of the macroeconomic response. While formal path tests are consistent with a pure human capital aggregation explanation, consistent with Acemoglu and Angrist (2001), the data cannot rule out modest positive spillovers of the magnitude implied by Moretti (2004) or Baum-Snow et al. (2026).

Having established that the micro estimates predict aggregate labor earnings (that is, that the micro estimates aggregate roughly in line with what the average cohort-level exposure to SFRs in the micro studies would predict), I next examine whether these gains extend to broader measures of economic performance. Specifically, I consider Personal Income (PI), which captures a broader set of income sources beyond wages, and GDP per capita, which measures the value of production within the state rather than labor earnings alone.[1] Because official state GDP data begin only in 1997, I backcast GDP using its stable post-1997 relationship with PI and, *importantly*, show that the results are robust to using only observed GDP data for a narrower set of reforms.

To help frame this analysis of GDP, I use as a benchmark a simple competitive economy in which labor receives a constant share of output. Under this benchmark, aggregate earnings and GDP should rise proportionately once the economy fully adjusts in the long run. In the short run, however, GDP may rise by less than aggregate earnings, or even decline, as fiscal and adjustment costs are incurred immediately while productivity gains emerge gradually over time.

[1] I find no detectable effect on the employment-to-population ratio, making the per-worker wage response comparable to the per-capita output response.

The last key finding is that both PI per-capita and GDP per-capita follow the transition path implied by the micro-based benchmark. Each outcome exhibits a modest initial decline after reform, though the early estimates are imprecise, before rising steadily beginning approximately fifteen years after adoption, consistent with fiscal and other transition costs being incurred before the productivity gains fully materialize. By approximately year 30, personal income per-capita has increased by roughly 4.5 percent and real GDP per-capita by slightly more than 5 percent – somewhat above the simulated path, but not statistically significantly so. As with average wages, formal tests reject the hypothesis that the evolution of PI and GDP per capita is unrelated to the simulated earnings path and fail to reject that they move one-for-one with it.

I show that the simulation results are robust to alternative assumptions regarding life-cycle effects and to using estimates from either underlying study separately. Empirically, I demonstrate robustness across alternative estimation strategies and to dropping groups of states with unusually high levels of GDP or school spending.

Because school finance reforms are not implemented at random, there is always a concern that alternative mechanisms may drive the results. However, the micro-founded estimates are derived from identifying variation entirely different from that underlying the aggregate estimates and are therefore subject to a different set of potential confounds. As such, any alternative explanation would have to generate bias in the within-time comparisons across cohorts underlying the micro estimates that tracks reform-induced increases in school spending. *At the same time*, a distinct source of bias must influence the aggregate estimates to produce no effects for 10 to 15 years, begin producing gains just as exposed children come to comprise a meaningful share of the labor market, accumulate at roughly the rate of cohort replacement. Moreover, the magnitude of these aggregate gains would have to converge toward that independently predicted by the micro estimates. As a result, the close correspondence between the simulated and observed paths (i.e., the formal path test) serves as a form of overidentification test, making a broad class of biases arising from contemporaneous fiscal effects, migration, and other place-based confounds highly unlikely.

The paper makes two main contributions. First, it develops and tests a simple framework for translating credible microeconomic estimates into testable quantitative predictions for long-run macroeconomic outcomes. By combining established micro estimates with demographic accounting, the cohort-aggregation approach generates transparent, *ex ante*, and falsifiable predictions for the evolution of aggregate earnings. The predicted path matches the observed evolution of aggregate wages, income, and output following school finance reforms, demonstrating that credible microeconomic causal estimates can successfully predict macroeconomic dynamics. More broadly, the framework may apply to policies whose aggregate effects emerge gradually as treated individuals enter the labor force, markets, or other economic aggregates.

Second, the paper provides causal evidence that investments in elementary and secondary education generate large and persistent increases in aggregate labor earnings, personal income, and real GDP per-capita. While prior work establishes that school finance reforms improve outcomes for directly affected students, this paper shows that those gains accumulate over time into economically meaningful improvements in aggregate state economic performance. More broadly, while a large literature documents a strong relationship between human capital and aggregate economic output, this paper provides direct causal evidence linking investments in human capital to subsequent increases in macroeconomic output.

The remainder of the paper is organized as follows. Section II is the first part of the paper and develops the cohort-aggregation framework and simulates the aggregate earnings path implied by the micro estimates. The second part of the paper tests the predictions from the first part. Sections III and IV describe the data and empirical strategy. Section V presents the main results on average wages. Section VI extends the analysis to personal income and GDP per-capita, introducing the competitive benchmark linking aggregate labor earnings to broader macroeconomic outcomes. Section VII reports robustness checks. Section VIII interprets the dynamic aggregate response and considers alternative explanations. Section IX concludes.

II. Simulating the Aggregate Income Path Implied by Micro Estimates of School Finance Reform Effects

A key part of the analysis is deriving what the existing micro-level evidence on school finance reform implies for aggregate earnings. The logic is simple. The micro studies estimate how much a child's future adult earnings increase following exposure to a reform when measured at specific ages. If aggregate earnings respond only through this childhood channel, then the aggregate earnings effect in any year is a weighted average of these individual effects across workers in the labor force. This section lays out the assumptions and framework used to construct that weighted average from the micro evidence and traces how it evolves over time, yielding the predicted time path of the aggregate per-worker earnings effect of a reform. This predicted path serves as the central benchmark against which the aggregate empirical estimates are evaluated. Full derivations of each step in the simulation are provided in Appendix IX.

II.A. Age profile and population structure

The micro studies report the effect of school finance reform exposure on adult earnings averaged over relatively broad age ranges. Simulating how each exposed cohort contributes to aggregate earnings as it enters the labor force, however, requires estimates of the earnings effect at each individual age. Because the underlying studies do not provide age-specific estimates, I make a set of assumptions to approximate the age profile of the treatment effect.

Age Profile: Assuming the effects of school finance reform accumulate gradually over the life cycle (i.e., the effect on wages at age 25 is smaller than at age 35), I model the earnings effect of childhood exposure as a smooth, non-decreasing function of age that is zero at age 20, reaches its maximum level, P, at age 45, and remains constant thereafter. The age profile is constructed as a blend of a linear ramp and a concave quadratic ramp (see Appendix IX). Although motivated by standard life-cycle considerations, the predicted aggregate path over the first 30 post-reform years is largely insensitive to the assumed profile (see Appendix IX). Also note that the simulated path is largely unchanged when using worker age weights instead of population age weights. See Appendix Figure A10. Specifically, the age-specific earnings effect is $\varphi(a) = P \times g(a)$, where $g(a)$ is the normalized age profile satisfying $g(20) = 0$ and $g(a) = 1$ for all $a \geq 45$. Thus, $g(a)$ determines the shape of the earnings profile, while P determines its magnitude. As described next, the calibration pins down P.

Population Shares: Another key input into the simulation is the age distribution of the workforce. This distribution is used both to age-adjust the implied reform effects and to aggregate cohort-level treatment effects into an economy-wide earnings effect. Population age shares, denoted by (s(a)), are taken from the Census age distribution and normalized to sum to one over ages 18–65. I assume that these age shares remain constant over time. Holding the age distribution fixed isolates the dynamics of interest (the gradual replacement of untreated cohorts by treated cohorts) from changes in the age composition of the workforce.

II.B. Calibration

I calibrate the scale of the earnings profile using two benchmark estimates from the SFR literature. Jackson et al. (2016) (JJP) estimate that reform-induced spending increases log adult earnings by 0.078 (SE = 0.0196) for individuals exposed throughout all twelve school-age years and measured over ages 20–45, whereas Rothstein and Schanzenbach (2022) (RS) estimate an effect of 0.0036 (SE = 0.0015) per year of exposure to a post-1990 reform, with earnings measured over ages 26–39.

I place the two estimates on a common basis before combining them. First, assuming effects increase linearly with years of exposure, I rescale both to thirteen years of exposure (for K-12). I also harmonize the reporting age ranges. Because a 5 percent earnings effect observed at age 40 corresponds to a different underlying treatment effect than a 5 percent effect observed at age 25 when effects grow with age, estimates averaged over different age ranges are not directly comparable. I therefore apply a scaling factor that accounts for the concavity of the assumed age profile over the wider JJP window relative to the narrower RS window (see Appendix IX for the derivation). This minor adjustment amounts to multiplying the RS estimate by $K = 0.9555$. Together, these adjustments yield harmonized estimates of 0.0845 for JJP and 0.0447 for RS. I combine them using a simple average and treat the estimates as independent because the studies exploit largely non-overlapping sets of reforms (pre-1990 and post-1990) and therefore distinct treated cohorts. The resulting benchmark is $\bar{\beta} = 0.0646$ (SE = 0.0141).

As noted above, the plateau, P, is the key scale parameter in the simulation. I calibrate it to the benchmark empirical estimate, $\bar{\beta}$. Because $\bar{\beta}$ reflects an average effect over ages 20–45, I choose P so that the population-weighted average of the age-specific effects, $\varphi(a) = P \times g(a)$, over those ages equals $\bar{\beta}$. This calibration yields $P = 0.1105$, implying that the earnings effect reaches 11.05 percent at age 45 and remains at that level thereafter. Together, P and the normalized age profile $g(a)$ fully determine the age-specific earnings effects. The blue line in the top panel of Figure 1 shows the age profile for a cohort exposed to the reforms throughout all 13 years of schooling.

II.C. Cohort exposure and the aggregate path

The next step is to aggregate the age-specific earnings effects. A reform affects only cohorts whose school-age years (ages 5–17) overlap the post-reform period. I assign each cohort an exposure equal to the fraction of its thirteen school-age years occurring after the reform, with cohorts not yet in school at enactment receiving full exposure and those already finished receiving none. Thus, with effects linear in years of exposure, a cohort exposed for x years receives $x/13$ of the full effect. The red and green lines in the top panel of Figure 1 show the age profiles for cohorts exposed to the reforms for 8 and 4 years, respectively. The age profile for the cohort with 8 years of exposure lies below that for the cohort exposed for all 13 years of schooling, while the profile for the cohort with 4 years of exposure lies below that for the cohort with 8 years of exposure.

Let $e(a,t) \in [0,1]$ denote this exposure fraction for workers of age a in event year t. Assuming no spillovers or crowd-out, the aggregate log-earnings effect in event year t is the population-weighted sum of age-specific effects scaled by cohort exposure:

$$A(t) = \underbrace{P}_{\substack{\text{plateau}\\\text{effect}}} \sum_{a=18}^{65} \underbrace{s(a)}_{\substack{\text{population}\\\text{share}}} \underbrace{e(a,t)}_{\substack{\text{reform}\\\text{exposure}}} \underbrace{g(a)}_{\substack{\text{normalized}\\\text{age profile}}} . \tag{1}$$

Because $A(t)$ is linear in the benchmark estimate, $\bar{\beta}$, sampling uncertainty in $\bar{\beta}$ propagates proportionally to the predicted aggregate earnings path. I use this relationship to construct confidence intervals for $A(t)$ that reflect estimation uncertainty in the underlying micro estimates.

Simulation Results: The lower panel of Figure 1 presents the simulation results. The aggregate effect is essentially zero for the first decade, then grows as exposed cohorts replace unexposed ones, reaching about 1.7 percent after 24 years, 3.0 percent after 30 years, 4.4 percent after 36 years, 6.8 percent after 48 years, and a long-run population-weighted average of 7.8 percent after roughly 60 years. Even after propagating sampling uncertainty, the intervals are informative. At 24 years the effect ranges from 1.0 to 2.4 percent (95% CI), rising to between 1.7 and 4.3 percent at 30 years, between 2.5 and 6.2 percent at 36 years, and between 4.5 and 11.2 percent in the long run. The simulation therefore yields empirically meaningful predictions that can be taken to the data. Appendix Figure A8 confirms that this path is nearly identical under a simple linear age-earnings profile, and Appendix Figure A9 shows that calibrating to each source study separately yields similar timing and curvature, with the pooled calibration giving substantially tighter confidence intervals. Having established what the micro evidence implies for aggregate average earnings, the remainder of the paper assesses whether design-based estimates of the effects of SFRs on aggregate outcomes are consistent with this simulated path.

III. Data

Average wages per worker is the primary outcome for estimating the aggregate earnings effects, and is the aggregate counterpart to the micro earnings effects simulated in Section II. Data on average annual earnings per worker come from the Quarterly Census of Employment and Wages (QCEW), produced by the U.S. Bureau of Labor Statistics (2025) and available from 1975 through 2025. The QCEW is a near-universal establishment census covering essentially all jobs subject to state unemployment insurance laws—roughly 95 percent of U.S. employment—and reports total covered wages and average covered employment by state and year.[2] I measure average annual earnings per worker as total covered wages divided by average covered employment.

If the micro evidence captures the direct effects of school finance reform, they should first appear in average earnings. However, if they also generate broader equilibrium effects, the gains should be reflected in broader measures of economic activity. To examine this, I supplement the QCEW analysis with two broader outcomes from the U.S. Bureau of Economic Analysis (2026). I use per-capita personal income (BEA Regional Economic Accounts, SAINC, 1959–2025) and real GDP per-capita (BEA SAGDP, 1997–2025). Personal income broadens the measure beyond wage earnings to include returns to capital, proprietors' income, employer supplements, and government transfers, while GDP per-capita measures the total value added produced within the state. Because official state GDP data begin only in 1997, I backcast GDP for earlier years using each state's average post-1997 log GDP-to-PI ratio and the observed evolution of real personal income within that state.[3] I show that all results are robust to restricting the analysis to observed GDP data only. Examining these outcomes provides evidence on whether SFRs generated aggregate economic effects at all, while comparing the estimated effects across outcomes reveals whether those gains are confined to workers' earnings or extend to broader measures of economic activity.

The treatment for this analysis is the implementation of a school finance reform. The reform timing employed draws on Jackson et al. (2016) who examine early reforms before 1993 and Lafortune et al. (2018) who focus on reforms from the 1990s onward. Both studies identify a reform event as a court order or legislative act that compelled a material change to the state school

[2] State-level QCEW data are available from 1975 onward. Establishments were classified under SIC through 2000 and NAICS thereafter, with NAICS-based totals reconstructed to 1990. This change is immaterial for the statewide, all-industry wage measure: industry reclassification does not affect aggregate wages or employment, and any common discontinuity is absorbed by year fixed effects.

[3] See Appendix X for details.

finance formula. Applying the coding from these papers yields 35 reform states with first-reform dates ranging from 1973 to 2005. The full list of states, years, sources, and coding notes appears in Appendix Table A1. The remaining states that enacted no such reform over this period serve as the never-treated control group throughout the analysis.

Table 1 summarizes the underlying data samples. It reports outcomes both before and after the reforms for reform states and also provides a snapshot of outcomes that allows for comparisons between reform and non-reform states. For context, Panel A reports state-level per-pupil spending taken directly from the Jackson et al. (2016) school finance data (in 2012 dollars) and illustrates the increase in per-pupil spending following reform, from an average of $6,357 during the pre-reform period to $9,133 after reform. Panel B reports the QCEW wage panel, my measure of aggregate earnings, which spans 1975–2025 and contains 1,785 state-year observations for the 35 reform states and 814 observations for the never-treated control states. Panel C reports per-capita personal income from 1959–2025 (2,345 and 1,072 observations, respectively), while Panel D reports real GDP per-capita from 1997–2025 (1,015 and 464 observations, respectively).

Table 1 also provides descriptive evidence on the comparability of reform and never-treated states in the same year. In 1990, average per-pupil spending was similar across the two groups ($8,366 versus $8,090), as were average annual QCEW earnings ($22,205 versus $22,145) and per-capita personal income ($18,723 versus $19,071). While these comparisons do not establish the validity of the research design, the close correspondence across several key economic measures is reassuring. By contrast, reform states exhibit somewhat lower GDP per-capita in 1997—the first year of observed BEA GDP data—($42,798 versus $52,661). This difference is driven largely by a small number of high-income never-treated states, particularly the District of Columbia and Alaska, whose exceptionally high GDP per-capita inflates both the control-group mean and its dispersion.[4]

IV. Empirical Strategy

To assess whether the evolution of outcomes is consistent with the simulation patterns, I estimate dynamic treatment effects by comparing the evolution of outcomes in reform states following the implementation of school finance reform with the evolution of outcomes in states that either had not yet implemented a reform or never implemented one during the sample period. These dynamic

[4] Appendix Figure A7 shows that the main event study plots are virtually unchanged in models that exclude Alaska and the District of Columbia, and also additionally excluding Hawaii and Delaware which have atypical GDP.

treatment effects can be summarized by the event-study specification

$$\log Y_{st} = \alpha_s + \lambda_t + \sum_{k \neq -1} \beta_k \mathbf{1}\{t - T_s^* = k\} + \varepsilon_{st}, \tag{2}$$

where Y_{st} is the outcome (average wages, personal income per-capita or GDP per-capita), T_s^* is the year school finance reform first took effect in state s, α_s and λ_t are state and year fixed effects, and β_k denotes the treatment effect at event time k. Rather than estimating this equation using conventional two-way fixed effects, I estimate the event-study effects using the Callaway and Sant'Anna (2021) estimator, which compares the evolution of outcomes in reform states following school finance reform with outcomes in states that either had not yet implemented a reform or never implemented one during the sample period. This model identifies the event-time treatment effects under the standard parallel-trends assumption (i.e., that the treatment and comparison groups would have followed the same trajectory had treatment not occurred) while avoiding biases that can arise under staggered treatment timing. I report estimates for $k \in [-8, +30]$, with $k = -1$ omitted as the reference period. Standard errors are clustered at the state level.

Because the Callaway and Sant'Anna (2021) estimator requires the period immediately preceding treatment to be observed, the set of reforms contributing to each event-study estimate depends on the availability of pre-treatment data. The QCEW series is available annually from 1975 through 2025, so the analysis excludes the two reforms implemented before the start of the sample.[5] Consequently, the QCEW event study is based on 33 reforms. By contrast, the personal income panel (and the GDP panel, which uses imputed GDP prior to 1997 based on personal income growth) is available annually from 1959 through 2025 and so includes all 35 reforms. For both datasets, the 2025 endpoint yields a balanced event-study sample of all included reforms through event year 20. Beyond event year 20, the number of contributing reforms declines mechanically as later reforms have not yet accumulated thirty years of post-treatment observations. Consequently, estimates through event year 30 are identified using the 20 reforms implemented by 1995 in the QCEW analysis and the 22 reforms implemented by 1995 in the personal income and GDP analyses.

Identification in this setting rests on the standard parallel-trends assumption that, absent reform,

[5] These are the 1973 Michigan and New Jersey reforms. Both states are already treated when the QCEW series begins in 1975 and therefore cannot contribute either as treated cohorts or as not-yet-treated comparison units. California's initial court ruling (*Serrano I*, 1971) also predates the sample, but because the state's school finance system was not fundamentally restructured until *Serrano II* and the enactment of AB 65, I code California's reform as occurring in 1976, permitting its inclusion in the analysis.

treated and untreated states would have followed similar trajectories. I assess this directly using the pre-reform event-time coefficients. Beyond identification, the cohort-aggregation framework developed in Section II provides an additional, independently derived restriction on the dynamics of the treatment effects. The framework predicts that aggregate earnings remain close to zero for more than a decade and then rise gradually as increasingly exposed cohorts enter the workforce. Many alternative explanations—including fiscal stimulus, governance reforms, or migration—would instead produce immediate responses or very different dynamic patterns. While no single dynamic pattern can establish causality on its own, few alternative mechanisms would generate the same delayed, mechanically timed accumulation implied by cohort replacement. Consequently, a close correspondence between the estimated and predicted paths substantially strengthens the interpretation of the estimated effects as reflecting reform-induced human-capital accumulation rather than alternative explanations or state-level confounds. This observation motivates the formal path test developed below.

Testing the Predicted Path

The simulation in Section II delivers an *ex ante* prediction for the evolution of aggregate earnings following a school finance reform. A conventional joint test of the post-reform coefficients is valid but not well suited to this setting because the simulation predicts essentially no aggregate response during the first decade after reform – rendering a simple pre vs post test very underpowered even if the data tracked the simulated path exactly. The simulated path instead provides a prediction of both the timing and relative magnitude of the response, allowing inference to focus on the economically relevant patterns in the data. Taking this simulation seriously, rather than simply asking whether all post-reform coefficients differ from zero, I formally assess whether the *estimated* event-study path follows this *predicted* trajectory.

Let $\hat{\beta} = (\hat{\beta}_k)'$ denote the estimated post-reform event-study coefficients and let $A = (A(k))'$ denote the simulated path over the same horizons. I summarize the relationship between the two using the following model

$$\hat{\beta}_k = \theta A(k) + e_k. \tag{3}$$

The parameter θ summarizes how closely the estimated aggregate response follows the path predicted by the micro evidence. Three values are of particular interest. If $\theta = 0$, the aggregate response is unrelated to the predicted path. If $\theta = 1$, the estimated response matches the micro-calibrated prediction exactly, implying that the private earnings gains aggregate without net wage spillovers or crowd-out. Values of θ between 0 and 1 would be consistent with some crowding out or displacement (as in Duflo (2004); Khanna (2023)), while values of θ above 1 would be consistent with positive wage spillovers (as suggested by Moretti (2004); Baum-Snow et al. (2026)). To account for the correlations among the event-study coefficients, I estimate θ by generalized least squares using the full covariance matrix of the Callaway and Sant'Anna (2021) estimator rather than by ordinary least squares.[6] Moreover, to account for uncertainty in the predicted path, when testing for equality of predicted and actual effects, I augment the variance of $\hat{\theta}$ to account for uncertainty in the pooled micro estimate.[7] Because the simulated path remains close to zero during the first decade, the earliest event times contribute little identifying information while adding estimation noise. I therefore estimate θ over a variety of event windows and show that the main conclusions are robust to alternative estimation windows.

V. Effects on Actual Aggregate Outcomes

The top panel of Figure 2 presents the event-study estimates for average annual earnings per worker from the QCEW. The figure plots the estimated treatment effect for each event year relative to the year immediately preceding the reform, together with 95 percent confidence intervals. Several features are immediately apparent. First, there is no evidence of differential pre-trends in average wages between reform and comparison states (and the pretreatment estimates are each individually precisely estimated and close to zero), providing support for the identifying assumptions underlying the event-study design. Consistent with the absence of differential pre-trends, a formal test reported in Table 2 of a linear pre-treatment slope yields a point estimate of -0.064 and a p-value of 0.137.

[6] Algebraically, this is equivalent to a minimum-distance estimator:

$$\hat{\theta} = \frac{A'\Sigma^{-1}\hat{\beta}}{A'\Sigma^{-1}A}, \qquad \widehat{\mathrm{Var}}(\hat{\theta}) = (A'\Sigma^{-1}A)^{-1},$$

where Σ denotes the covariance matrix of the event-study estimates clustered at the state level.

[7] For tests of $\theta = 1$, I augment the variance of $\hat{\theta}$ to account for uncertainty in the pooled micro estimate,

$$\widehat{\mathrm{se}}(\hat{\theta})^2 = (A'\Sigma^{-1}A)^{-1} + \hat{\theta}^2 \left(\frac{\mathrm{se}(\bar{\beta})}{\bar{\beta}}\right)^2,$$

treating the micro and macro estimates as independent because they are obtained from distinct samples.

The second key pattern is that the estimated effects remain close to zero for roughly the first 15 years following the reform, after which average wages begin to rise steadily. Indeed, the table reports that the linear trend between years 0 and 15 is estimated to be -0.096 with a p-value of 0.279. From that point onward, the estimated effects increase approximately linearly, reaching about 4.7 percent (with a wide confidence interval) by 30 years after the reform. Consistent with the visual trend break, the linear trend between years 16 and 30 is estimated to be 0.286 (p<0.001).

This delayed response is informative for two reasons. First, plausible alternative explanations for increases in wages would not be expected to take more than a decade to materialize. Second, the timing of the increase matches the pattern predicted by the cohort-aggregation framework developed in Section II. Rather than producing an immediate increase following the reform, average wages remain essentially unchanged (both visually and statistically) until several fully exposed cohorts enter the labor force and partially exposed cohorts have aged sufficiently for a meaningful share of their wage gains to be realized. From that point onward, the estimated effects rise steadily.

To assuage concerns that this pattern is driven by the choice of estimator, I also present event-study estimates using the conventional (or naive) two-way fixed effects estimator and the stacked difference-in-differences estimator in Appendix Figure A2. Both alternative estimators yield the same basic pattern, indicating that the results are not sensitive to the estimation approach.

To assess the extent to which these patterns track the simulation, panel (b) of Figure 2 compares the estimated QCEW wage response directly with the no-spillover path implied by the micro evidence. The two paths are similar in their timing, shape, and level. Both remain close to zero for more than a decade after reform and begin rising only once increasingly exposed cohorts enter the labor force. The estimated wage response then follows the same gradual upward pattern as the simulation, although the point estimates generally lie somewhat above the no-spillover benchmark at the longest horizons. The visual evidence is therefore consistent with the no-spillover benchmark, while the point estimates leave open the possibility of positive equilibrium wage spillovers.

I now evaluate these comparisons formally by testing whether the estimated event-study path follows the predicted shape and whether its scale differs from the benchmark value of $\theta = 1$. Specifically, I estimate the scaling parameter relating the observed event-study coefficients to the simulated aggregate earnings path over several event-time windows, ranging from the full 0–30 year horizon to windows that begin 5, 10, and 15 years after the reform. Estimating the relationship over multiple horizons assesses whether the results are sensitive to the period over which the comparison is made, particularly given that the simulation predicts little aggregate response during the first decade following the reform.

The results, reported in Table 2, tell a consistent story across specifications. For every event-time window, the estimated scaling parameter is positive and statistically different from zero, while being statistically indistinguishable from 1 – indicating that the observed wage path tracks the simulated aggregate earnings path.[8] Indeed, one cannot reject the null hypothesis that $\theta = 1$ in any specification, with p-values ranging from 0.10 to 0.82. Notably, when the comparison begins five years after the reform, the estimated scaling parameter is $\hat{\theta} = 1.09$, close to the one-for-one relationship implied by the cohort-aggregation framework, and across all four windows the wage-based point estimates range from 1.092 to 2.01. Overall, while there is some variability in point estimates across time windows, the consistency of these findings across alternative event-time windows indicates that the close correspondence between the observed and simulated paths is not an artifact of the particular horizon chosen for the comparison.

V.A. Discussion of the QCEW Results

The relationship between the aggregate wage response and the path predicted by applying the micro estimates to a simple cohort-replacement model has several implications. First, the similarity between the two suggests that the private wage gains documented in JJP and RS do not primarily reflect a redistribution of earnings across workers. If increases in the supply of more-skilled workers substantially reduced the wages of incumbent workers, or if a substantial portion of the estimated private returns reflected signaling or other mechanisms that redistribute rather than increase aggregate earnings, the aggregate wage response would be expected to fall below that implied by the micro estimates. Instead, the aggregate response is at least as large as the no-spillover benchmark, although the difference is not statistically significant.

Second, the correspondence between the micro-predicted and aggregate wage paths provides independent corroboration of the interpretation of the micro estimates as reflecting genuine increases in human capital. In particular, the aggregate results largely rule out explanations based on compositional changes across school districts. If the micro estimates were driven by higher-ability students sorting into higher-spending districts, such sorting could generate apparent gains at the

[8]Note that the test of whether the slope equals zero does not incorporate estimation error from the simulation, because error in the simulation changes only its scale, not its shape. In contrast, the test of whether the two are equal accounts for estimation error in both the event-study estimates and the simulated path.

district level but would not mechanically increase average wages at the state level.

Finally, the fact that the observed wage path generally lies above the simulated no-spillover path is consistent with the possibility of positive equilibrium spillovers. However, the path tests are not sufficiently precise to distinguish modest positive spillovers from zero spillovers. I therefore interpret this part of the evidence as suggestive. Overall, the estimates provide little indication of negative displacement effects and leave open the possibility of positive wage spillovers, but the data are not precise enough to allow one to conclude that such spillovers are present.

VI. Personal Income and GDP per-capita

VI.A. Conceptual Framework

The preceding section shows that the simulated aggregate labor-earnings path implied by the micro estimates tracks observed aggregate labor earnings. A remaining question is whether these gains translate into broader improvements in aggregate output. The relationship between labor earnings and GDP is not mechanical. School finance reforms require substantial public investment today in exchange for productivity gains that materialize only gradually as exposed cohorts enter the labor market, so the transition path of GDP may differ substantially from the short-run effect on wages. Moreover, labor accounts for only about two-thirds of output, so the long-run effect on GDP depends on how capital responds and on the return it earns. I therefore organize the analysis around the macroeconomic response one might expect in a competitive economy in the long run, which provides a benchmark against which the estimated GDP path can be interpreted.

To formalize this benchmark, I adopt a competitive Cobb–Douglas production function as a reference point rather than a maintained assumption,

$$Y_{st} = A_{st} K_{st}^{\alpha} (H_{st} L_{st})^{1-\alpha}, \tag{4}$$

where K_{st} is physical capital, L_{st} labor, and H_{st} average human capital. Under competitive factor markets each factor is paid its marginal product, and each factor's share is independent of quantities, so that

$$W_{st} = (1 - \alpha) Y_{st}, \tag{5}$$

where W_{st} denotes aggregate labor earnings. The constant share benchmark has considerable empirical support (Gollin, 2002), though its constancy has been questioned in more recent work (Karabarbounis and Neiman, 2014).

Equation (5) has a direct implication for what wage data alone can reveal. Because wages equal labor's share $(1 - \alpha)Y$ of output, and $\alpha > 0$, capital necessarily earns a share of any increase in output. As a purely accounting matter, if all of the returns to school spending accrued to labor and were observed in higher wages, the increase in GDP would be the labor share times the percent increase in labor earnings, so GDP would rise by less (in percentage terms) than wages. This is a lower bound rather than a likely outcome. As labor becomes more productive (as implied by the increase in aggregate average wages) it raises the marginal product of capital, so capital earnings rise as well, and the reforms may also raise total factor productivity, which raises both labor and capital earnings. Under any of these channels output rises by more than the labor share times the wage gain, and a wage-only accounting understates the aggregate return. The benchmark thus implies that in the long-run labor earnings, capital earnings, and GDP increase proportionately. This prediction is testable.

This is a long-run prediction, not a description of the transition. Financing educational investment through taxation or borrowing may temporarily reduce activity, treated cohorts enter the labor market only gradually, and the capital stock adjusts to higher human capital with a lag. GDP may therefore fall in the short run even as average wages remain flat, before both rise over the longer horizon as cohorts enter, capital adjusts, and fiscal distortions fade.

Recovering the long-run estimand cleanly requires that the per-worker and per-capita responses be comparable, which in turn requires the two normalizations to move together. As shown in Figure 3, the employment-to-population ratio remains essentially unchanged throughout the transition. Although there is some evidence of a modest decline of approximately 0.4 percentage points on average between years 6 and 18, the estimates are imprecise, and by approximately year 20 the employment-to-population ratio is virtually indistinguishable from its pre-reform level. The increase in average wages therefore closely tracks the evolution of aggregate labor earnings per-capita, allowing the subsequent comparison between labor earnings and GDP per-capita to be interpreted largely as reflecting changes in productivity rather than changes in labor utilization.

To illustrate the framework conceptually, Figure 4 presents a stylized transition path for GDP under the benchmark model. The dashed line shows the simulated transition path of labor earnings per worker implied by the microeconomic estimates, while the solid line illustrates how GDP per capita evolves under the benchmark once the transitional costs of reform are taken into account. In the early years following a reform, no treated cohort has yet entered the labor force, so human capital and hence productivity are unchanged. Over this horizon the reforms generate financing and reallocation costs, with no offsetting productivity gains, so measured GDP may fall below its baseline. Only later, as treated cohorts enter the workforce and raise productivity, can output recover, eventually rising above baseline and converging to the labor-earnings benchmark as the transitional costs dissipate. While the magnitude and timing of any initial decline are uncertain and are not calibrated here, such a pattern may arise when transition costs precede the productivity gains (as is likely the case with SFRs).

Informed by this framework, the empirical analysis addresses two related questions. First, how do school finance reforms affect aggregate GDP over the transition to the new long-run equilibrium? In particular, does GDP initially decline as fiscal costs are incurred and resources are reallocated before rising as treated cohorts enter the labor market and complementary capital adjusts? Second, once these transition dynamics have largely dissipated, are the long-run gains in GDP similar (in percentage terms) to those observed for wages?

VI.B. Effect on PI and GDP per-capita

The event-study estimates for log personal income per-capita are presented in Figure 5a. Unlike the average wage measure examined in the previous section, personal income per-capita captures a broader measure of the income received by state residents, including labor earnings, proprietors' income, rental income, interest, and dividend income. Although it is not a direct measure of aggregate output, it provides a useful bridge between labor-market outcomes and GDP by capturing a substantially broader set of income sources than wages alone.

Figure 5a is consistent with the predictions from the framework. Unlike average wages—which remained close to zero for about fifteen years following the reform— PI per-capita exhibits a modest decline during the early post-reform period, although the estimates are not statistically distinguishable from zero (and one fails to reject that the trend between year 0 and year 15 is zero). Beginning about sixteen years after the reform, however, PI rises steadily and almost linearly, reaching an estimated increase of roughly 4.5 percent by year 30. As with average earnings, one rejects that the trend from year 16 onward is zero at conventional levels – indicative of a real difference in trend.

The figure also overlays the aggregate labor-earnings path implied by the micro estimates. While personal income per-capita initially lies somewhat below this benchmark, the two series converge approximately twenty years after the reform and thereafter evolve similarly. Consistent with this, across all four event-time windows, one rejects the null hypothesis that the personal income path is unrelated to the simulated benchmark ($\theta = 0$, $p < 0.05$ in each case), and in three of the four windows one fails to reject that the path follows the benchmark one-for-one ($\theta = 1$). This convergence suggests that the labor-market gains documented earlier ultimately extend beyond wages to broader measures of household income, consistent with the benchmark prediction that aggregate income should eventually grow proportionately with aggregate labor earnings once transitional forces have largely dissipated.

The dynamic pattern is also consistent with the conceptual framework developed in the previous section. In the short run, school finance reforms require substantial public investment while the productivity gains materialize only gradually as treated cohorts enter the labor market. Temporary fiscal costs, the opportunity cost of redirecting resources toward education, and other transitional adjustments may therefore dampen aggregate income during the early years following the reform. As these transitional forces dissipate and increasingly larger shares of the workforce benefit from the reform, the persistent earnings gains documented in the micro evidence translate into comparable proportional increases in personal income per-capita. Thus, while the early dynamics are consistent with transitional adjustment costs, the long-run convergence between personal income and the micro-based labor-earnings benchmark suggests that the aggregate income gains ultimately mirror the labor-market gains identified in the micro data.

While personal income provides evidence that the gains extend beyond labor earnings to broader measures of household income, it remains an income measure rather than a measure of aggregate production. I therefore next examine real GDP per-capita, which provides the most comprehensive measure of economic output. Figure 5b presents the event-study estimates for log real GDP per-capita together with the micro-based labor-earnings benchmark.

The estimated GDP path matches the predictions of the framework. Following the reform, GDP per-capita exhibits a modest initial decline before increasing steadily beginning approximately fifteen years after the reform, eventually reaching an estimated increase of over 5 percent by year 30. Between years 0 and 15, the slope for GDP is negative, but one fails to reject that the slope is zero (p-value=0.242). Although the early declines are imprecisely estimated, the overall pattern is consistent with the presence of short-run transition costs followed by gradually increasing productivity as treated cohorts enter the labor market and the economy adjusts to higher levels of human capital. Indeed, as with average earnings and PI per-capita, one rejects that the trend from year 16 onward is zero at conventional levels – indicative of real increases in GDP per-capita once more treated cohorts increasingly enter the labor market.

The GDP series uses imputations from personal income data prior to 1997. To rule out the possibility that this pattern is driven by the imputation, I present an analogous figure in Appendix Figure A6, where the series is restricted to years 0 through 26, **and no GDP data are imputed**. Reassuringly, this series shows the same basic pattern. As in the imputed series, GDP per-capita declines over the first several years, reaching a trough around year 7, and then rebounds, culminating in GDP that is about 5 percent higher by year 25. The rebound likely reflects some combination of two forces: the dissipation of the initial disruption associated with the reform and the productivity gains that arise as treated cohorts enter the labor market. The confidence intervals are wide, so the precise timing of the transition from decline to gains cannot be sharply identified, and the relative contribution of these two channels cannot be separated. The broad pattern of an initial decline followed by a sustained rebound is nonetheless clear.

The comparison with the labor-earnings benchmark is informative. While the GDP series tracks the benchmark closely throughout much of the post-reform period, the estimated GDP path lies modestly above the simulated earnings path beginning around eighteen years after the reform. Formal tests reject the null hypothesis that the GDP trajectory is unrelated to the simulated benchmark (the largest p-value across the estimation windows is $p = 0.002$), indicating that the micro-based earnings predictions strongly predict the observed evolution in aggregate output. At the same time, while the point estimates range from 1.86 to 2.17 across the four windows, the data fail to reject the hypothesis that the long-run GDP path has a one-for-one relationship with the simulated earnings path. Thus, the data support all three qualitative predictions of the benchmark framework: an initially muted response consistent with transition costs, a gradual increase as reform-treated cohorts enter the labor market, and long-run growth that is statistically consistent with proportional increases in aggregate labor earnings and GDP.

Taken together, the evidence from personal income and GDP paints a consistent picture. The early dynamics are consistent with modest short-run adjustment costs, followed by persistent improvements in aggregate income and production that mirror the labor-earnings gains implied by the micro evidence. While the point estimates suggest that GDP may ultimately increase somewhat more than labor earnings, this difference is not estimated precisely. The data therefore do not reject the benchmark prediction that the percentage increase in aggregate output is proportional to the increase in aggregate labor earnings, though the point estimates allow for somewhat larger effects on GDP.

VI.C. Implied Long-Run GDP Effect

Although the event studies span approximately 30 years following the reforms, the simulated aggregate earnings path does not converge to its long-run steady state until roughly 60 years after implementation. Consequently, the empirical analysis does not directly observe the full equilibrium implied by the simulation. Nevertheless, the correspondence between the simulated earnings path and the observed evolution of average wages, personal income per-capita, and real GDP per-capita over the first three decades provides considerable support for the predictive validity of the simulated transition path. Moreover, because all three event studies continue to rise at the end of the sample and track the simulated dynamics throughout the observed period, there is little empirical evidence that the economy has yet reached its new steady state.

The micro-based simulation predicts that school finance reforms ultimately increase aggregate labor earnings by approximately 7.8 percent in steady state. The preceding analysis further shows that both personal income per-capita and GDP per-capita evolve closely with the simulated earnings path throughout the observed transition. While the point estimates for GDP are modestly steeper than the simulated earnings benchmark, the difference is not statistically distinguishable from a one-for-one relationship. The evidence is therefore consistent with a proportional long-run relationship between aggregate labor earnings and GDP.

Under this benchmark interpretation, the simulation implies that school finance reforms ultimately increase real GDP per-capita by approximately 7.8 percent relative to the counterfactual without reform. Importantly, this is a long-run projection rather than a directly estimated treatment effect. Because real GDP per-capita has already increased by roughly 5 percent by year 30 (nearly two-thirds of the projected steady-state effect and the simulated transition has roughly another 30 years to run), a 7.8 percent steady-state increase by year 60 represents a relatively modest extrapolation from the observed trajectory.

VII. Robustness

As in all applied work, the results are a function of the modeling and estimation decisions made by the researcher. As such, it is important to assess the extent to which these decisions drive the results. In the simulation section, I show that the predicted aggregate earnings paths are very similar irrespective of how one assumes the gains from school finance reform evolve over the life cycle. In particular, the shape of the transition path for aggregate earnings is largely invariant to this decision and is primarily driven by the deterministic entry of progressively more exposed cohorts over time. Indeed, the projected earnings path is roughly equal to the average treatment effect multiplied by the predicted fraction of the working-age population that is exposed.

The robustness of the findings across measures and datasets is also worth noting. Average wages are measured using the QCEW, a Bureau of Labor Statistics establishment census constructed from state unemployment insurance records, while GDP is a Bureau of Economic Analysis product built on a separate national accounts framework. While they differ in important ways, they trace the same delayed, cohort-timed path, suggesting that the agreement reflects a common underlying economic response rather than any particular measurement approach.

Regarding the empirical results, I show that the evolution of average wages is essentially flat both before and for several years following the implementation of the reforms. This shows that the subsequent increase in average wages—which occurs precisely when the simulation predicts that treated cohorts begin entering the labor force—is likely not driven by pre-existing trends. To assess the robustness of this basic pattern, Appendix Figures A2 through A5 show that the results are similar for all three outcomes using either the naive DID estimator or the stacked DID estimator. While these alternative estimators exhibit some pre-trending in the full sample, the pre-trends are flat when excluding four resource-rich states that have unusually large economic swings during energy booms. Removing these states has little effect on the preferred and more robust Callaway and Sant'Anna (2021) estimates, but eliminates the pre-trends for the alternative estimators while preserving the main results. In the cleaner sample, all three estimators show the same basic pattern: flat pre-trends before reform, with improving outcomes beginning between years 10 and 15.

VIII. Discussion: Alternative Explanations

The analysis thus far interprets the aggregate responses as reflecting improvements in workforce human capital induced by school finance reforms. An alternative is that the increases in average wages, personal income, and GDP reflect migration, contemporaneous fiscal changes, or other place-based forces. Two features of the evidence help distinguish these explanations. The first is *timing*: aggregate increases are absent for roughly fifteen years following reform and then accumulate gradually, with the onset coinciding with the entry of substantially exposed cohorts into the labor market. The second is *micro-macro consistency in magnitude*: the aggregate response evolves with approximately the magnitude predicted *ex ante* by the individual-level estimates underlying the benchmark.

The second restriction is particularly informative because the benchmark is based on comparisons among individuals at the same time. That is, both Jackson et al. (2016) and Rothstein and Schanzenbach (2022) estimate the effects of SFR by comparing the outcomes of more- and less-exposed cohorts *at the same time*, essentially conditioning out place-based confounds such as other labor-market policies or general equilibrium adjustments. Thus, if the micro estimates are causal and the macro effects closely track what those estimates predict—that is, the appropriately weighted aggregation of the micro effects—this correspondence implies that such confounds play a limited role in explaining the aggregate results. Even if the micro estimates were themselves biased, however, the close correspondence would still be informative. For bias to explain the result, confounding in the micro estimates (which arises from a different source of variation and is largely insulated from contemporaneous place-level shocks) would have to generate treatment effects that, once aggregated across cohorts, mimic the effects of a different set of confounds operating in the aggregate data, in both their timing and magnitude. This coincidence seems unlikely. As such, the close correspondence between the observed macro effects and those implied by the micro estimates rules out a broad class of plausible alternative explanations.

Intuitively, the micro-founded benchmark provides an independent prediction for what the aggregate data should look like absent bias. When the observed aggregate data track that prediction despite being identified from different variation, this provides evidence against important biases in either design. Of course, one cannot rule out offsetting biases that happen to reproduce exactly the timing and magnitude predicted by the micro evidence. But the simpler interpretation of the close correspondence is that such biases are limited. While the path test suggests minimal bias in general, it is worth discussing why some specific alternative explanations are unlikely.

Contemporaneous fiscal effects. School finance reforms require substantial public expenditures financed through higher taxes, borrowing, or the reallocation of other resources. These costs could temporarily reduce economic activity and may explain the modest declines in PI and GDP during the first decade following reform. They are less able, however, to explain the persistent accumulation of gains beginning roughly fifteen years after reform. Standard fiscal effects tend to have output responses emerging within quarters rather than decades (Auerbach and Gorodnichenko, 2012), and there is no reason a fiscal-financing channel would generate gains matching the magnitude implied by childhood exposure to reform-induced spending. Fiscal effects may therefore contribute to the short-run dynamics but are unlikely to account for the long-run pattern.

Migration. Migration is another potential explanation if reforms induce households to relocate toward reform states. Three considerations weigh against this interpretation. First, existing evidence suggests that residential sorting responses to school finance reforms are modest and, where present, emerge within the first several years following reform (Chakrabarti and Roy, 2015; Aaronson, 1999)—well before the delayed labor-market effects documented here. Second, studies of migration responses to school finance reforms focus primarily on within-state moves, which are likely much more common than the cross-state moves that could bias the aggregate estimates. Finally, even if reforms induced higher-income households to move into reform states, such migration would presumably occur while their children are in school. There is no obvious reason why its effects would instead emerge on the schedule of treated cohorts entering the labor market, much less generate gains that closely match the individual-level returns to childhood exposure.

Place-based and other slowly accumulating confounds. A possibility is that reform states experienced some other slowly emerging advantage that coincided with the reforms. Such a confound could potentially reproduce the delayed timing of the aggregate response, but it would also need to generate gains of approximately the magnitude implied by the individual-level estimates (which condition on such effects). Some source of bias matching either feature alone is plausible, but matching both the cohort-entry timing and quantitative benchmark is highly unlikely.

IX. Conclusion

This paper provides quasi-experimental evidence that broad-based investments in K–12 education generate substantial long-run improvements in aggregate economic performance. Exploiting staggered school finance reforms across U.S. states, I show that the macroeconomic effects of educational investment emerge gradually through cohort replacement in the labor force. Reforms improve the human capital of exposed students, but aggregate effects materialize only as these cohorts enter and progressively replace older workers in the labor market.

A central contribution of the paper is to link credible microeconomic estimates of the returns to educational investment with long-run macroeconomic outcomes. Using cohort-specific earnings estimates derived from existing studies, I construct an *ex ante* simulation of the aggregate labor-earnings path implied by the reforms. The observed evolution of average wages follows this predicted path, providing evidence that the private returns identified in the micro data aggregate well to the state economy.

The paper then asks whether these labor-market gains translate into broader improvements in economic performance. Personal income per-capita and real GDP per-capita exhibit similar dynamics. Both display modest, though imprecisely estimated, declines during the first decade following reform (consistent with the short-run fiscal costs and transition dynamics associated with large educational investments), before rising steadily as increasingly larger shares of the workforce consist of treated cohorts. Throughout the observed transition, both series track the micro-based labor-earnings benchmark.

To interpret these patterns, I develop a simple benchmark based on a competitive Cobb–Douglas economy. The benchmark predicts that, after the economy has fully adjusted, aggregate labor earnings and GDP should increase proportionately. The empirical evidence is broadly consistent with this prediction, suggesting that the gains from school finance reforms are broadly shared across the economy. An implication of this result is that some of the social gains associated with school finance reforms are not observed in individual wages. Indeed, if the entire return accrued to labor, GDP would rise by only the labor share times the proportional wage gain—roughly two-thirds of the wage increase under competitive pricing. The confidence intervals for the path tests in all models reject the hypothesis that GDP rises only by the labor share times the micro-predicted earnings gain (i.e. all models reject a slope of 0.66 or lower at the 10 percent level, and the full 0-30 year window rejects at the 5 percent level). Taken together, these results indicate that the aggregate gains from school finance reforms extend beyond the increases observed in labor earnings.

Although the event studies extend roughly thirty years beyond the reforms, both the simulated and observed transition paths continue to rise at the end of the sample, suggesting that the economy has not yet reached its new steady state. Under the benchmark interpretation, the simulated long-run earnings effect therefore implies an eventual increase of approximately 7.8 percent in both aggregate labor earnings and real GDP per-capita.

Overall, the results highlight the importance of taking a long-run perspective when evaluating investments in education. The paper also demonstrates how credible microeconomic estimates can be combined with dynamic macroeconomic evidence to quantify the aggregate consequences of human-capital investments, providing a framework for connecting individual-level causal estimates to long-run economic outcomes. Finally, the results bridge a deep microeconomic literature on the private returns to educational investment and a long tradition of estimating the relationship between human capital and macroeconomic growth by providing compelling evidence that human-capital investments that raise individual earnings can translate into broader economic output.

References

Aaronson, D. (1999). The effect of school finance reform on population heterogeneity. *National Tax Journal 52*, 5–26.

Acemoglu, D. and J. Angrist (2001). How large are human capital externalities? evidence from compulsory schooling laws. *NBER Chapters*, 9–74.

Angrist, N., S. Djankov, P. K. Goldberg, and H. A. Patrinos (2021). Measuring human capital using global learning data. *Nature 592*, 403–408.

Auerbach, A. J. and Y. Gorodnichenko (2012, 5). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy 4*, 1–27.

Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics 106*, 407–443.

Baum-Snow, N., N. Gendron-Carrier, and R. Pavan (2026, February). Local human capital spillovers. Working Paper.

Bils, M. and P. J. Klenow (2000). Does schooling cause growth? *American Economic Review 90*, 1160–1183.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*, 200–230.

Card, D. and A. B. Krueger (1992, 2). Does school quality matter? returns to education and the characteristics of public schools in the united states. *Journal of Political Economy 100*, 1–40.

Chakrabarti, R. and J. Roy (2015). Housing markets and residential segregation: Impacts of the michigan school finance reform on inter- and intra-district sorting. *Journal of Public Economics 122*, 110–132.

Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review 91*, 795–813.

Duflo, E. (2004). The medium run effects of educational expansion: evidence from a large school construction program in indonesia. *Journal of Development Economics 74*, 163–197.

Goldin, C. and L. F. Katz (2019, 6). The race between education and technology. *The Race between Education and Technology*.

Gollin, D. (2002). Getting income shares right. *Journal of Political Economy 110*, 458–474.

Hanushek, E. A. and L. Woessmann (2012). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. *Journal of Economic Growth 17*, 267–321.

Hanushek, E. A. and L. Woessmann (2015). *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, MA: The MIT Press.

Heckman, J. J., L. Lochner, and C. Taber (1998). Explaining rising wage inequality: Explanations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics 1*, 1–58.

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

Karabarbounis, L. and B. Neiman (2014). The global decline of the labor share. *The Quarterly Journal of Economics 129*, 61–103.

Khanna, G. (2023). Large-scale education reform in general equilibrium: Regression discontinuity evidence from india. *Journal of Political Economy 131*, 549–591.

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

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*, 3–42.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics 107*(2), 407–437.

Moretti, E. (2004). Workers, education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review 94*, 656–690.

Rothstein, J. and D. W. Schanzenbach (2022). Does money still matter? attainment and earnings effects of post-1990 school finance reforms. *Journal of Labor Economics 40*, S141–S178.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics 87*, 355–374.

U.S. Bureau of Economic Analysis (2026). Regional economic accounts. State gross domestic product and personal income data.

U.S. Bureau of Labor Statistics (2025). Quarterly census of employment and wages. Annual average county and state employment and wages.

Tables and Figures

Table 1. Summary Statistics for School Spending and Aggregate Economic Outcomes

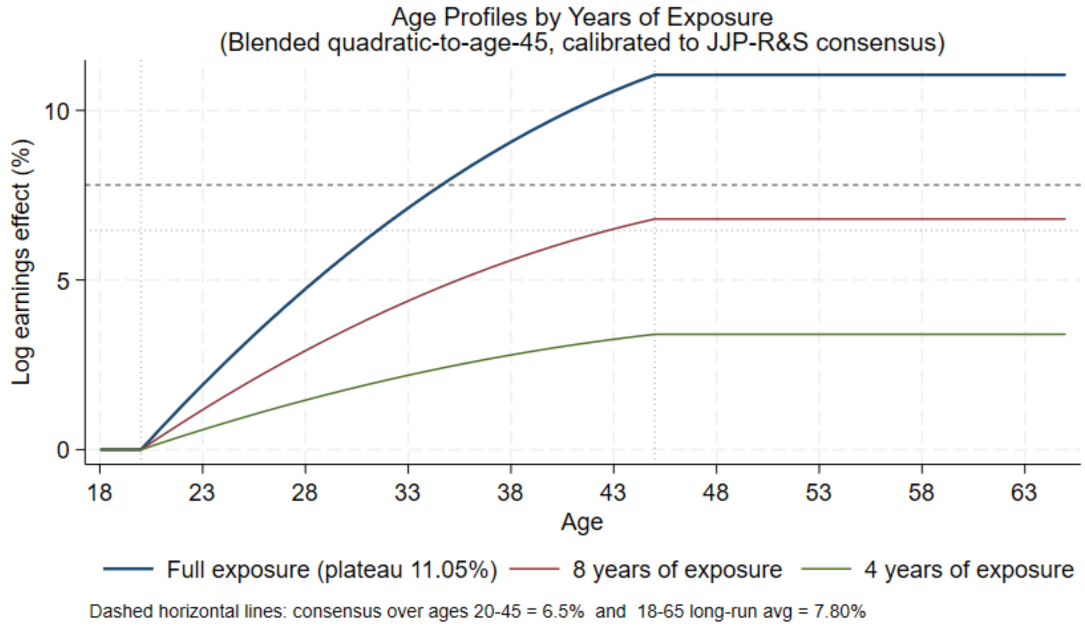
	States	Obs.	Mean	SD
Panel A: Per-Pupil Spending (1967–2010)				
All reform states	35	1,424	$7,737	$2,638
Pre-reform period	35	716	$6,357	$2,005
Post-reform period	35	708	$9,133	$2,461
Never-treated control states	16	584	$7,973	$2,895
Snapshot: 1990				
Reform states	33	33	$8,366	$2,136
Never-treated control states	12	12	$8,090	$1,884
Panel B: Average Annual Earnings per Worker (QCEW, 1975–2025)				
All reform states	35	1,785	$34,813	$18,457
Pre-reform period	33	557	$17,831	$6,613
Post-reform period	35	1,228	$42,515	$16,888
Never-treated control states	16	814	$34,206	$18,639
Snapshot: 1990				
Reform states	35	35	$22,205	$3,061
Never-treated control states	16	16	$22,145	$4,358
Panel C: Per-Capita Personal Income (1959–2025)				
All reform states	35	2,345	$24,862	$20,728
Pre-reform period	35	1,112	$8,632	$7,023
Post-reform period	35	1,233	$39,499	$17,912
Never-treated control states	16	1,072	$25,130	$20,773
Snapshot: 1990				
Reform states	35	35	$18,723	$2,957
Never-treated control states	16	16	$19,071	$3,332
Panel D: Real Per-Capita GDP (1997–2025)				
All reform states	35	1,015	$53,627	$11,561
Pre-reform period	7	31	$46,744	$9,829
Post-reform period	35	984	$53,844	$11,549
Never-treated control states	16	464	$63,650	$34,805
Snapshot: 1997				
Reform states	35	35	$42,798	$7,300
Never-treated control states	16	16	$52,661	$28,435

Notes: Panel A reports state-year per-pupil spending from Jackson, Johnson, and Persico (2016). Panel B reports average annual earnings per worker from the Quarterly Census of Employment and Wages (QCEW). Panels C and D report per-capita personal income and real per-capita GDP from the Bureau of Economic Analysis. Pre- and post-reform periods are defined relative to each state's first school finance reform. The 1990 snapshots compare reform and never-treated states before the major reform wave for Panels A–C; Panel D uses 1997 because BEA GDP data begin in that year. The sample includes 35 reform states and 16 never-treated control states. Reform years range from 1973 to 2005 (median 1993), with 6 reforms in the 1970s, 7 in the 1980s, 18 in the 1990s, and 4 in the 2000s.

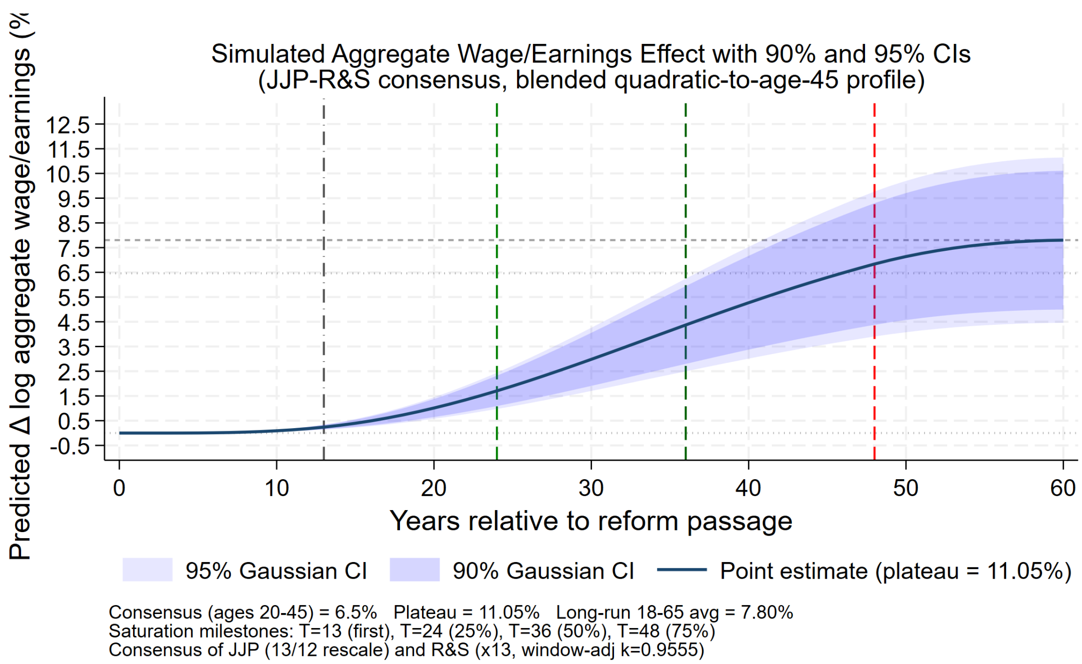
Table 2. Trend and Simulation-Path Tests

	Average wage	Personal income per-capita	GDP per-capita	GDP/Wage ratio
Panel A. Linear trend tests				
Pre-reform slope	-0.064	-0.035	-0.003	–
Standard error	(0.043)	(0.053)	(0.066)	
$p(0)$	0.137	0.501	0.961	
Slope, years 0–15	-0.096	-0.014	-0.116	–
Standard error	(0.088)	(0.090)	(0.099)	
$p(0)$	0.279	0.872	0.242	
Slope, years 16–30	0.286	0.271	0.502	–
Standard error	(0.078)	(0.099)	(0.123)	
$p(0)$	< 0.001	0.006	< 0.001	
Panel B. Simulation-path tests				
Years 0–30	1.522	2.512	2.149	1.412
Standard error	(0.273)	(0.362)	(0.516)	
$p(0)$	< 0.001	< 0.001	< 0.001	
$p(1)$	0.225	0.022	0.100	
Years 5–30	1.092	2.062	1.866	1.710
Standard error	(0.338)	(0.426)	(0.581)	
$p(0)$	0.001	< 0.001	0.001	
$p(1)$	0.824	0.087	0.222	
Years 10–30	1.468	1.413	1.952	1.329
Standard error	(0.372)	(0.542)	(0.636)	
$p(0)$	< 0.001	0.009	0.002	
$p(1)$	0.341	0.508	0.214	
Years 15–30	2.008	1.392	2.171	1.081
Standard error	(0.434)	(0.560)	(0.698)	
$p(0)$	< 0.001	0.013	0.002	
$p(1)$	0.103	0.539	0.165	

Notes: Panel A reports GLS linear trends fitted to the event-study coefficients over the indicated event-time windows. Estimates are annual changes in log points multiplied by 100 and can therefore be interpreted roughly as percentage-point changes per year. Standard errors in parentheses. The reported $p(0)$ values test the null hypothesis that the slope equals zero.
Panel B reports the GLS scaling parameter θ from $\widehat{\beta}_k = \theta A(k)$, where $A(k)$ denotes the simulated aggregate-earnings path. The parameter is estimated over the reported event-time window. Standard errors reported in parentheses treat the simulated path as fixed. The reported $p(0)$ values test the null hypothesis $H_0 : \theta = 0$ using these standard errors. The reported $p(1)$ values test the null hypothesis $H_0 : \theta = 1$ using a larger, conservative standard error that additionally incorporates sampling uncertainty in the simulated path.
The final column reports the ratio of the estimated GDP scaling parameter to the estimated average-wage scaling parameter. No inference is reported for this ratio because the separate outcome estimations do not recover the cross-outcome covariance required for a valid test. All within-outcome calculations use the full covariance matrix of the event-study coefficients.



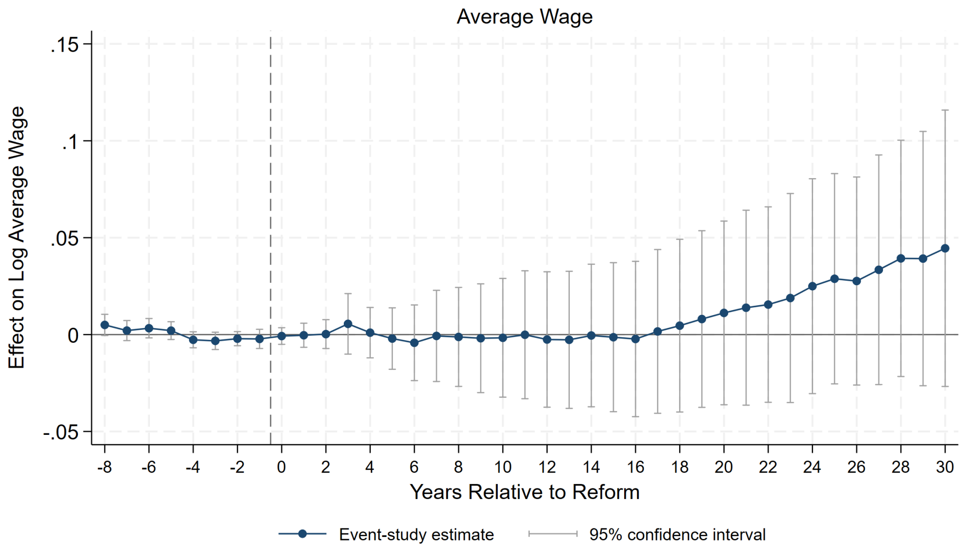
(a) Age profiles of earnings effects by years of exposure



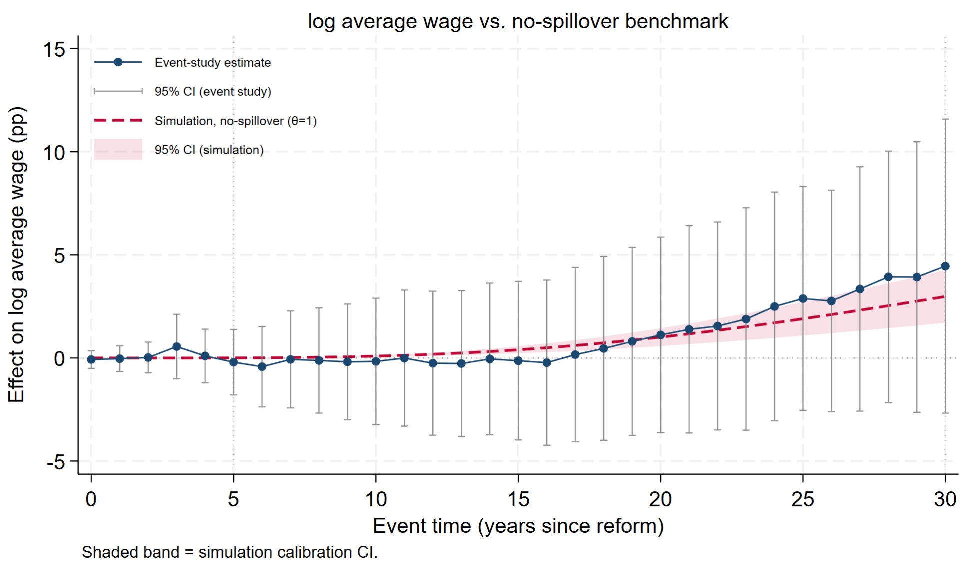
(b) Predicted aggregate earnings effect of school finance reform

Figure 1. Simulated Earnings Effects of School Finance Reform

Notes: Panel (a) plots the assumed age profile of the earnings effect of school finance reform for cohorts receiving thirteen, eight, and four years of exposure. The full-exposure profile is calibrated to the combined JJP–RS benchmark. Effects are linear in years of exposure, so the eight- and four-year profiles equal 8/13 and 4/13 of the full-exposure profile at each age. Panel (b) plots the corresponding simulated aggregate earnings effect by years since reform passage. The benchmark effect over ages 20–45 is 6.5 percent, implying an age-45 plateau of 11.05 percent and a long-run population-weighted average effect over ages 18–65 of 7.8 percent. Shaded regions denote 90 and 95 percent confidence intervals obtained by propagating sampling uncertainty in the combined benchmark. Vertical lines mark the entry of the first fully exposed cohort into the workforce and the horizons at which 25, 50, and 75 percent of the working-age population is fully exposed.



(a) Dynamic Effects on Average Wages



(b) Estimated and Simulated Wage Paths

Figure 2. Dynamic Effects of School Finance Reform on Average Wages

Notes: Panel (a) reports event-study estimates of the effect of school finance reform on log average annual earnings per worker measured in the Quarterly Census of Employment and Wages (QCEW). Event time is measured relative to the first year in which the reform took effect, with event year -1 as the reference period. The comparison group consists of states that had not yet implemented a reform or never implemented one during the sample period. Estimates are obtained using the Callaway and Sant'Anna (2021) estimator, with 95 percent confidence intervals based on standard errors clustered at the state level. Panel (b) compares the estimated wage response with the simulated earnings path implied by the micro estimates under the no-spillover benchmark, $\theta = 1$. The solid line reports the event-study estimates, with 95 percent confidence intervals shown by the vertical bars. The dashed line reports the simulated path, and the shaded region denotes the 95 percent confidence interval associated with uncertainty in the micro calibration.

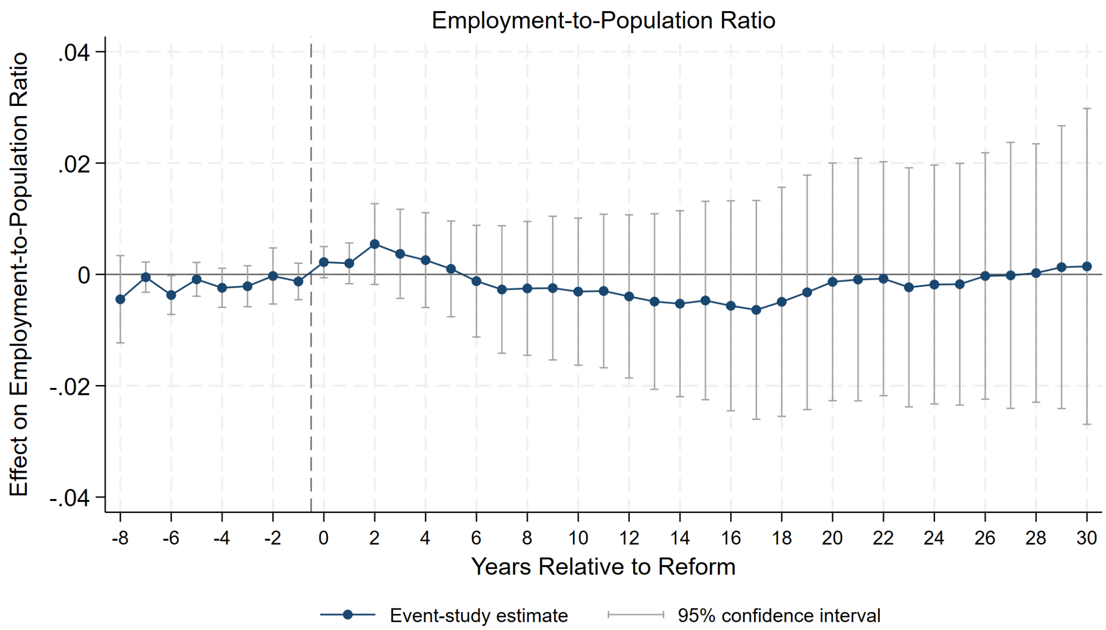


Figure 3. Dynamic Effects of School Finance Reform on the Employment-to-Population Ratio
Notes: The figure reports event-study estimates of the effect of school finance reform on the employment-to-population ratio. Event time is measured relative to the first year in which the reform took effect, with event year −1 as the reference period. The comparison group consists of states that had not yet implemented a reform or never implemented one during the sample period. Estimates are obtained using the Callaway and Sant'Anna estimator. Vertical bars denote 95 percent confidence intervals based on standard errors clustered at the state level. The estimates provide little evidence of a persistent effect on employment per-capita. This suggests that the initial decline in GDP per-capita is unlikely to be explained primarily by reduced employment, although the estimates do not distinguish among other potential transitional mechanisms.

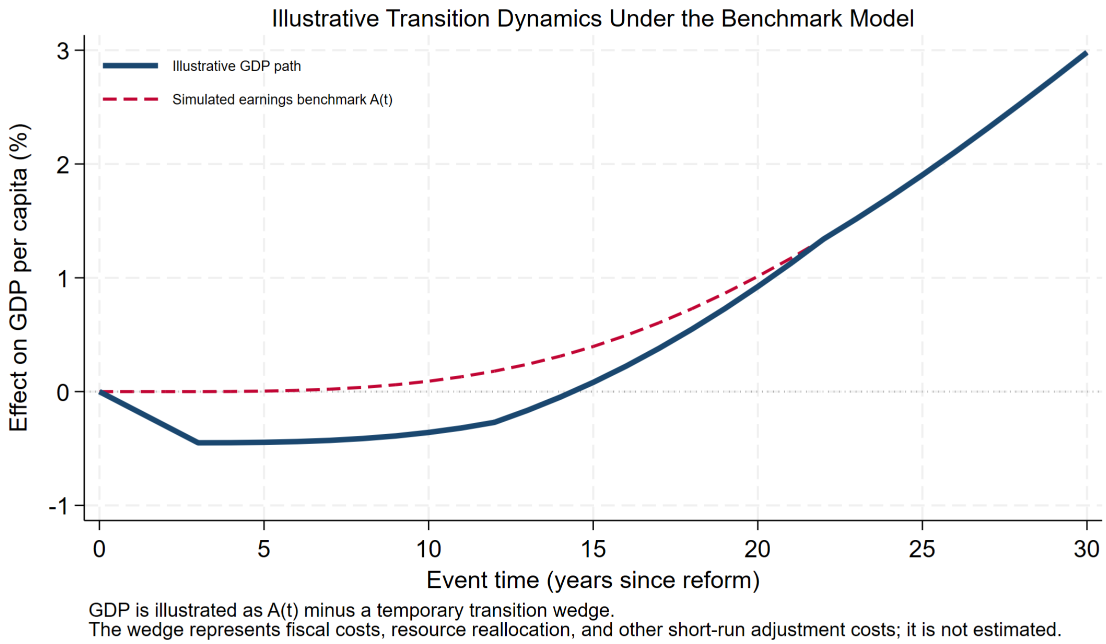
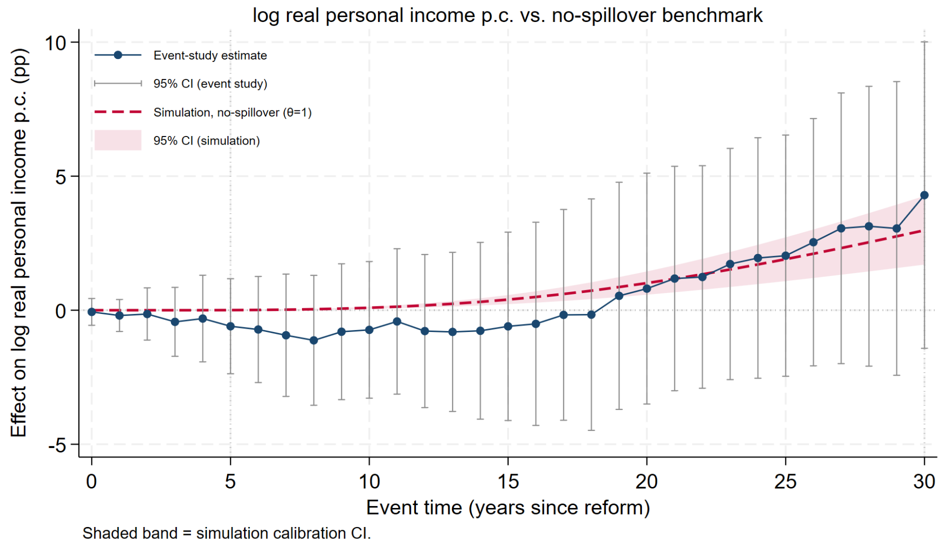
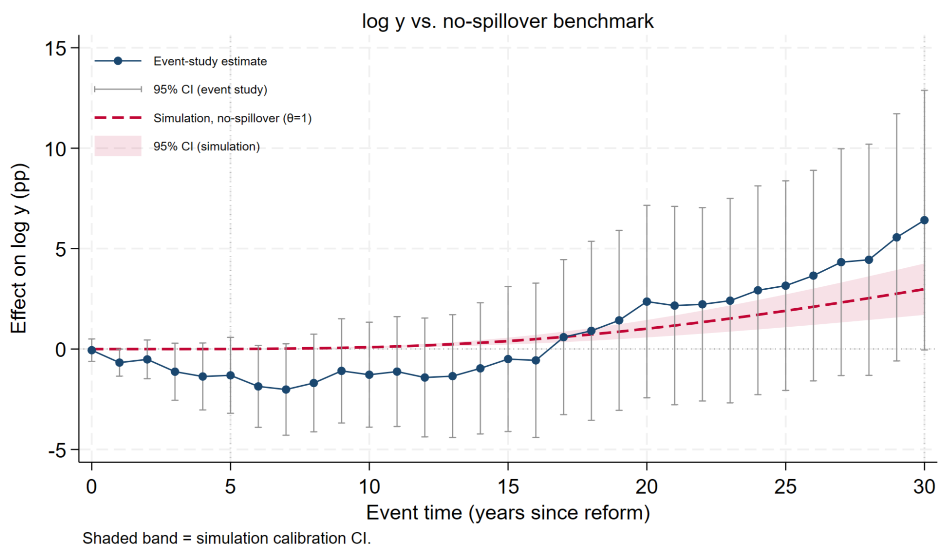


Figure 4. Illustrative Transition Dynamics Under the Benchmark Model
Notes: The dashed line shows the simulated transition path of aggregate labor earnings implied by the microeconomic estimates. The solid line illustrates one possible path for GDP under the benchmark framework after accounting for a temporary transition wedge. The transition wedge is purely illustrative and represents short-run forces discussed in the text—including fiscal adjustment, temporary resource reallocation, implementation costs, delayed complementary investment, and other transitional frictions—that may temporarily reduce output relative to labor earnings following the reform. By construction, the transition wedge dissipates over time, so the illustrative GDP path converges to the simulated labor-earnings path. The figure is intended only to illustrate the benchmark framework and is not an estimated decomposition of the observed GDP response.



(a) Log personal income per-capita



(b) Log real GDP per-capita

Figure 5. Personal income and GDP per-capita relative to the micro-based benchmark

Notes: The solid lines with markers report event-study estimates, and the vertical bars report 95 percent confidence intervals. The dashed lines report the aggregate labor-earnings path implied by the micro estimates, with the shaded regions showing the corresponding simulation confidence intervals. Panel (a) reports effects on log personal income per-capita, and Panel (b) reports effects on log real GDP per-capita. The coefficients therefore represent approximately proportional changes relative to the omitted pre-reform period. Note that the full event study estimates including the pre-period are presented in Appendix Figures A3 and A4.

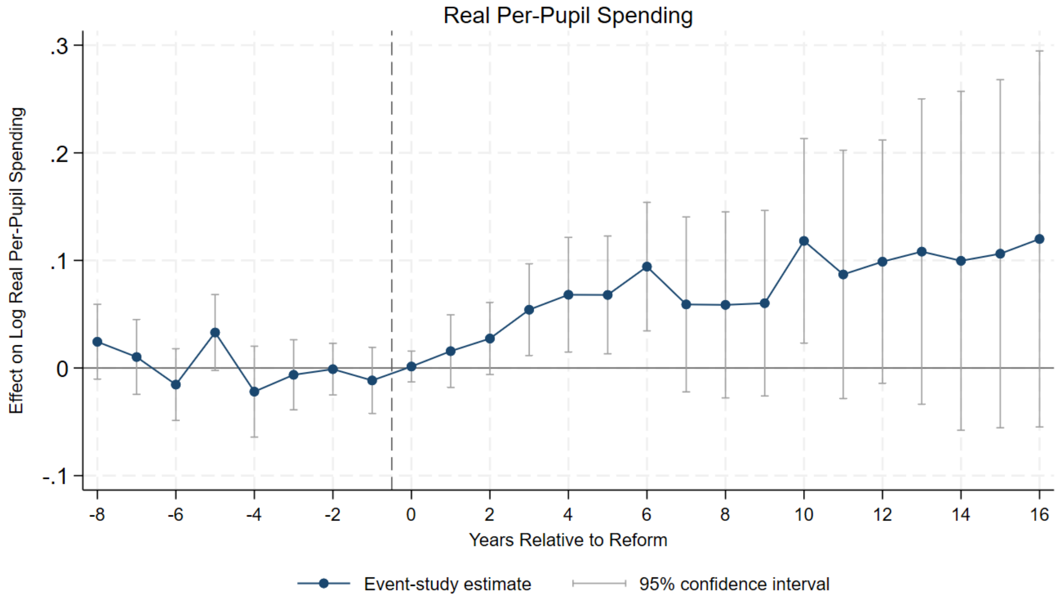
Appendix

Table A1. School Finance Reforms: Coding and Sources

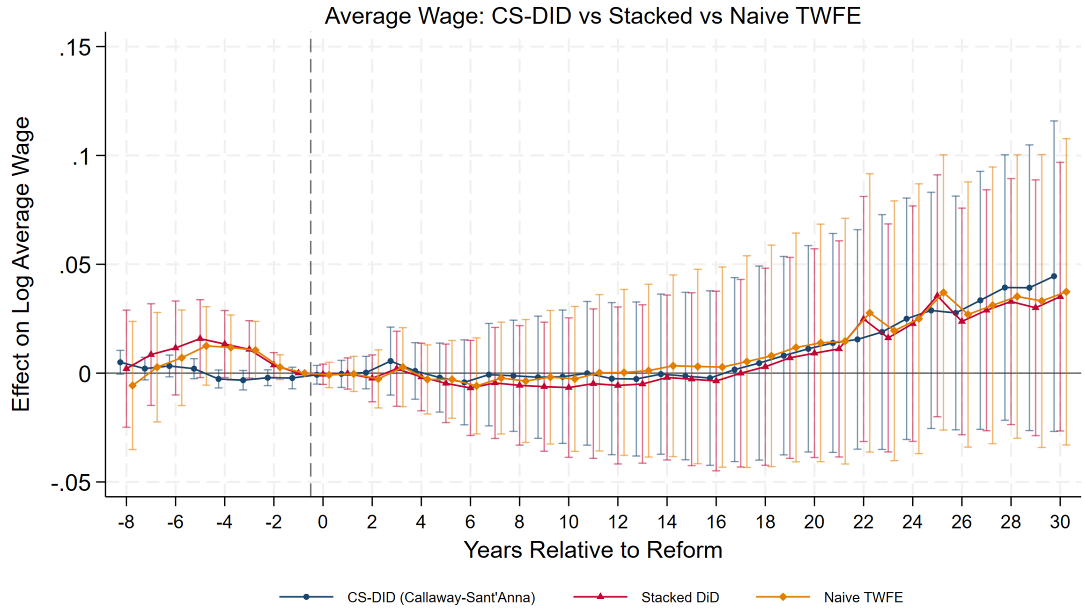
State	Abbr.	Reform Year	Source	Notes
Michigan	MI	1973	JJP (2016)	*Milliken v. Green*; legislative overhaul following court order
New Jersey	NJ	1973	JJP (2016)	*Robinson v. Cahill*; first of multiple NJ court orders
California	CA	1976	JJP (2016)	*Serrano v. Priest*; Prop. 13 (1978) complicated implementation
Connecticut	CT	1977	JJP (2016)	*Horton v. Meskill*
Kansas	KS	1976	JJP (2016)	Legislative reform
Washington	WA	1978	JJP (2016)	*Seattle School District v. State*
Wyoming	WY	1980	JJP (2016)	*Washakie County v. Herschler*
West Virginia	WV	1982	JJP (2016)	*Pauley v. Kelly*
Arkansas	AR	1983	JJP (2016)	*Dupree v. Alma School District*
Oklahoma	OK	1987	JJP (2016)	Legislative reform;
Kentucky	KY	1989	JJP (2016)	*Rose v. Council for Better Education*; comprehensive reform
Montana	MT	1989	JJP (2016)	*Helena Elementary v. State*;
Texas	TX	1989	JJP (2016)	*Edgewood v. Kirby*;
Oregon	OR	1991	JJP (2016)	*Measure 5* property tax limitation triggered reform
Alabama	AL	1993	LRS (2018)	*Alabama Coalition for Equity v. Hunt*
Massachusetts	MA	1993	JJP (2016)	Education Reform Act; legislative reform
Minnesota	MN	1993	JJP (2016)	Legislative reform
Missouri	MO	1993	LRS (2018)	Legislative reform
New Hampshire	NH	1993	LRS (2018)	*Claremont School District v. Governor*
Tennessee	TN	1993	JJP (2016)	*Tennessee Small School Systems v. McWherter*
Arizona	AZ	1994	JJP (2016)	*Roosevelt Elementary v. Bishop*;
North Dakota	ND	1994	LRS (2018)	Legislative reform
Florida	FL	1996	LRS (2018)	Legislative reform
Illinois	IL	1996	LRS (2018)	Legislative reform
Maryland	MD	1996	LRS (2018)	*Bradford v. Maryland State Board*
North Carolina	NC	1997	LRS (2018)	*Leandro v. State*
Ohio	OH	1997	LRS (2018)	*DeRolph v. State*
Vermont	VT	1997	LRS (2018)	*Brigham v. State*; Act 60
New Mexico	NM	1999	LRS (2018)	Legislative reform
Pennsylvania	PA	1999	LRS (2018)	Legislative reform
South Carolina	SC	1999	LRS (2018)	*Abbeville County School District v. State*
Colorado	CO	2000	LRS (2018)	*Lobato v. State*; legislative precursor
New York	NY	2003	LRS (2018)	*Campaign for Fiscal Equity v. State*
Georgia	GA	2005	LRS (2018)	Legislative reform (A-Plus Education Reform Act)
Idaho	ID	2005	LRS (2018)	Legislative reform

Notes: Reform years correspond to the first year of implementation of a court-ordered or legislative school finance reform that materially altered the distribution of per-pupil spending across districts. JJP (2016) refers to Jackson et al. (2016). LRS (2018) refers to Lafortune et al. (2018). Never-treated states used as controls (not listed) are: AK, DC, DE, HI, IA, IN, LA, ME, MS, NE, NV, RI, SD, UT, VA, and WI.

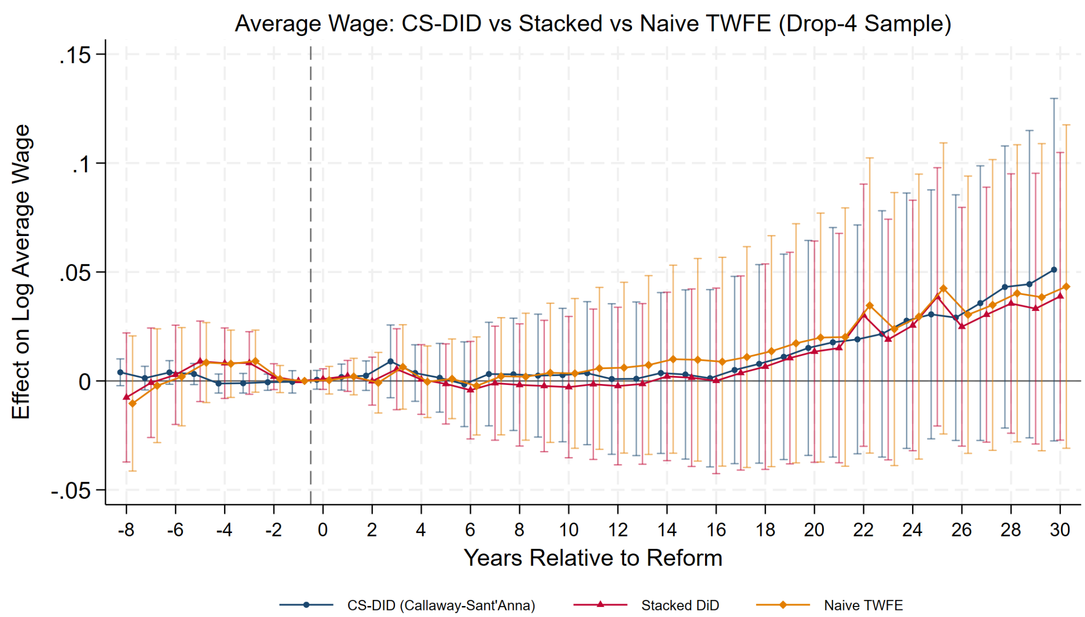
Figure A1. Effect of School Finance Reforms on Real Per-Pupil Spending



Notes: This figure presents event-study estimates of the effect of school finance reforms on log real per-pupil education spending. District-level spending and enrollment data are aggregated to the state-year level. State-level per-pupil spending is calculated by dividing total education spending across districts by total enrollment and is expressed in 2012 dollars. The sample is restricted to reforms implemented between 1976 and 2002, providing sufficient pre-reform data to assess differential trends and sufficient post-reform data to estimate medium-run effects. Estimates are obtained using the doubly robust inverse probability weighting estimator implemented by `csdid`. Because the estimation sample contains states treated at different dates, treatment effects are estimated relative to states that have not yet implemented a reform. Standard errors are clustered at the state level. The omitted reference period is the period immediately preceding the reform. Vertical bars report 95 percent confidence intervals. The vertical dashed line denotes the onset of the reform.

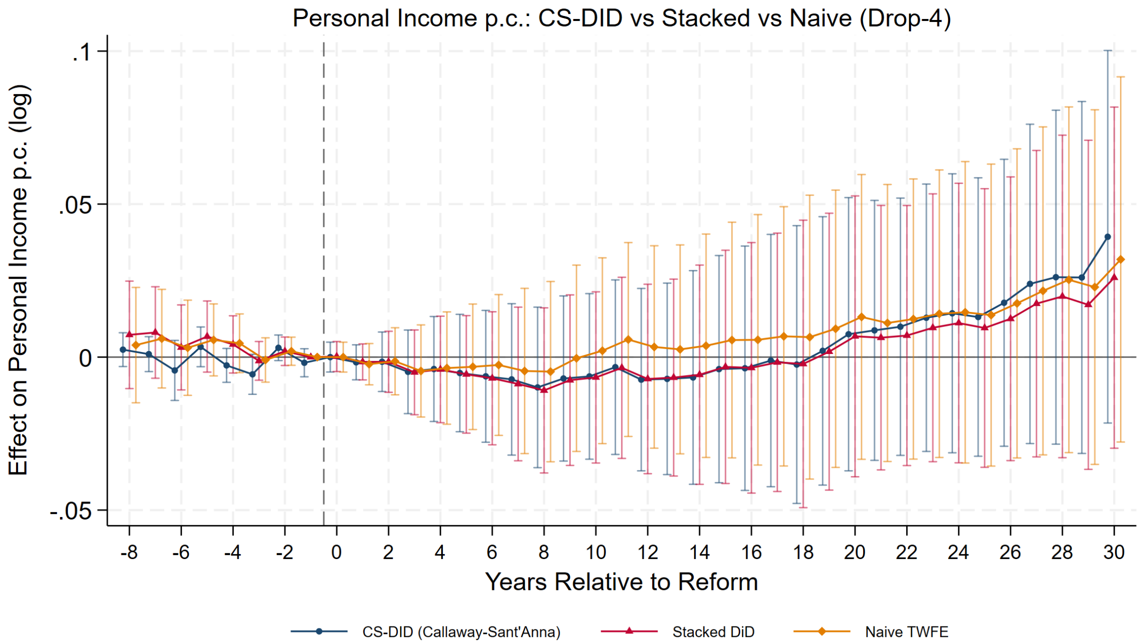


(a) Full sample

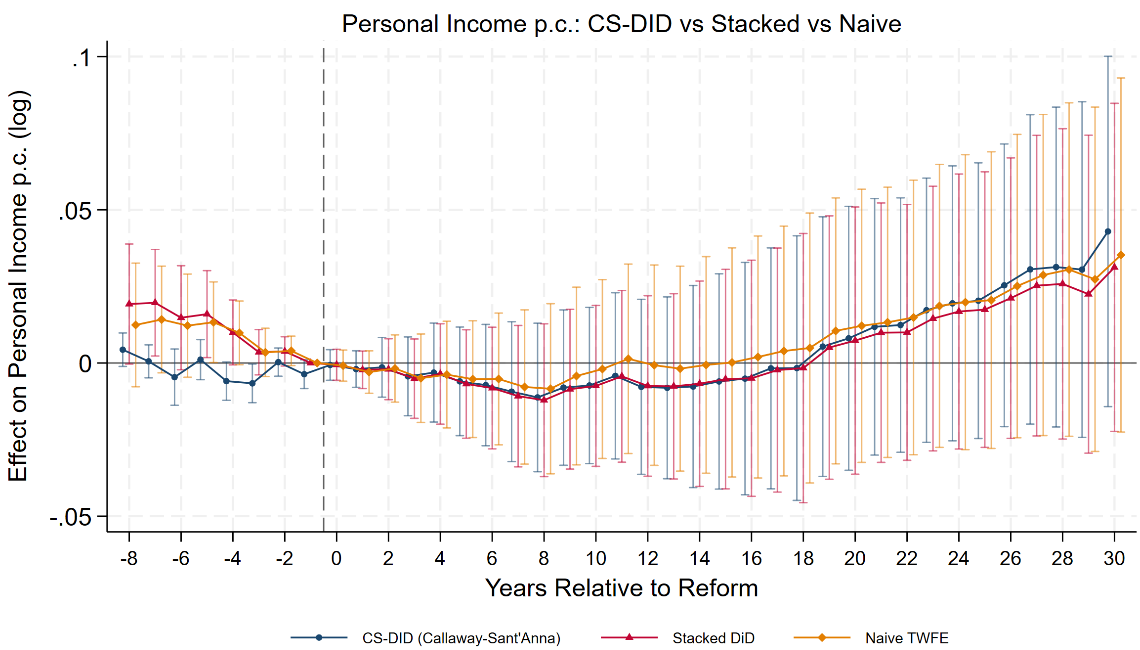


(b) Excluding Arizona, Montana, Oklahoma, and Texas

Figure A2. Comparison of event-study estimators for average wages. The figure compares dynamic treatment effect estimates obtained using the Callaway–Sant'Anna estimator, a stacked difference-in-differences estimator, and a conventional two-way fixed effects event study. Panel (a) reports estimates using the full sample of states. Panel (b) excludes Arizona, Montana, Oklahoma, and Texas, four energy and resource intensive states whose wage and income dynamics are driven in part by commodity price cycles unrelated to school finance reform, to confirm that the estimator comparison is not sensitive to their inclusion.

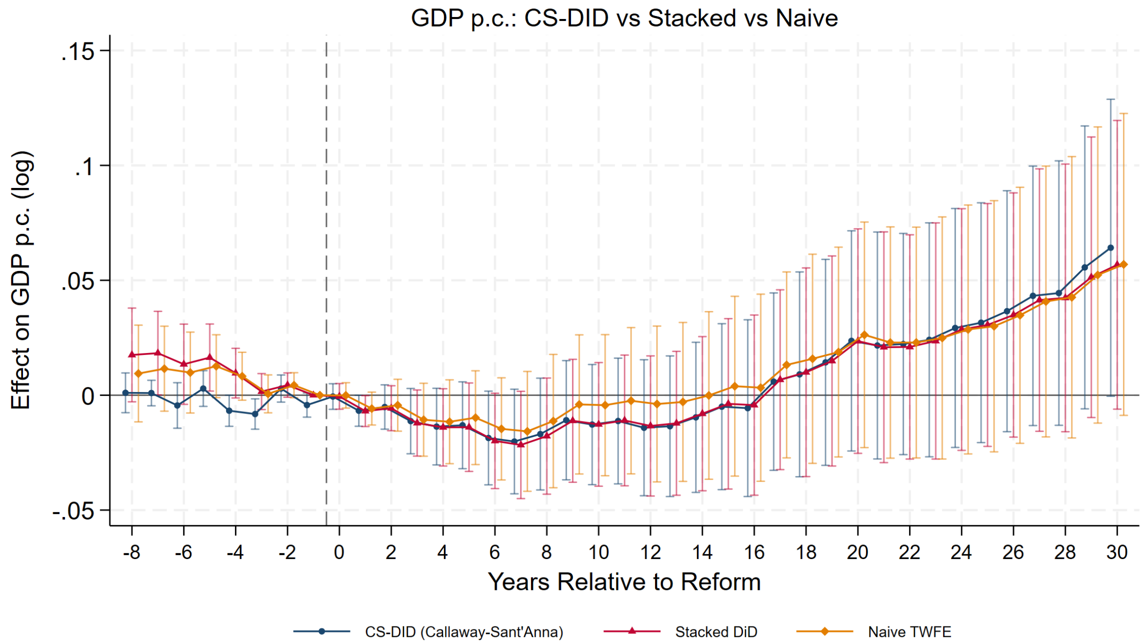


(a) Excluding energy and resource-intensive states

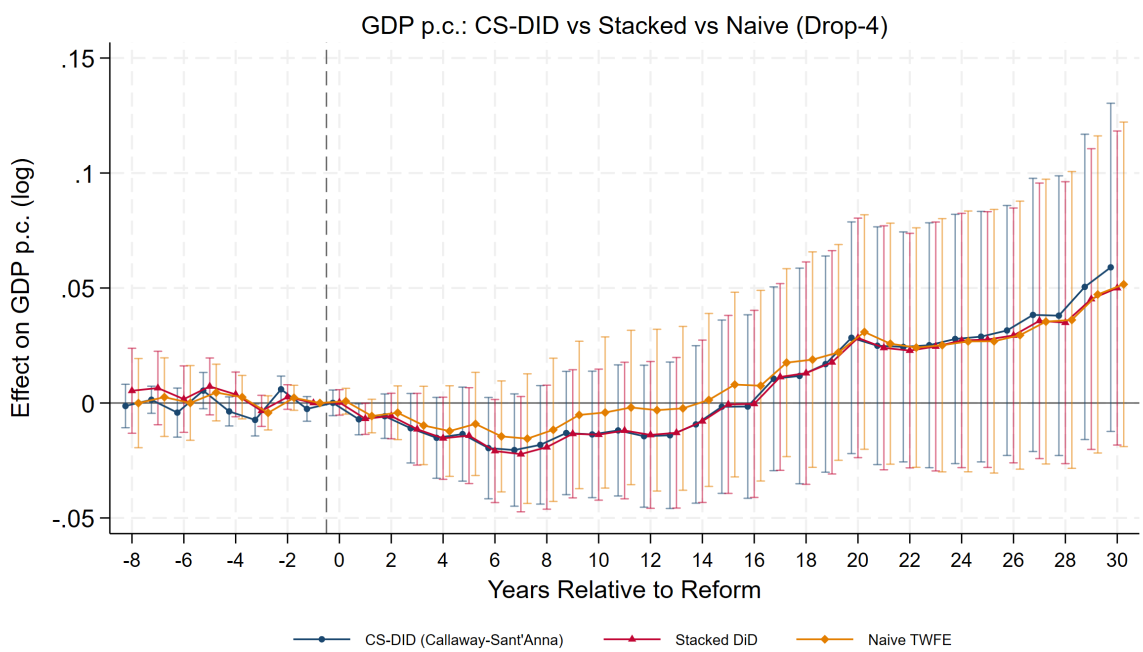


(b) Full sample

Figure A3. Comparison of event-study estimators for personal income per-capita. The figure compares dynamic treatment effect estimates obtained using the Callaway–Sant'Anna estimator, a stacked difference-in-differences estimator, and a conventional two-way fixed effects event study. Panel (a) excludes Arizona, Montana, Oklahoma, and Texas, four energy- and resource-intensive states whose wage and income dynamics are driven in part by commodity price cycles unrelated to school finance reform. Panel (b) reports estimates using the full sample of states. The similarity across panels confirms that the estimator comparison is not sensitive to the inclusion of these states.

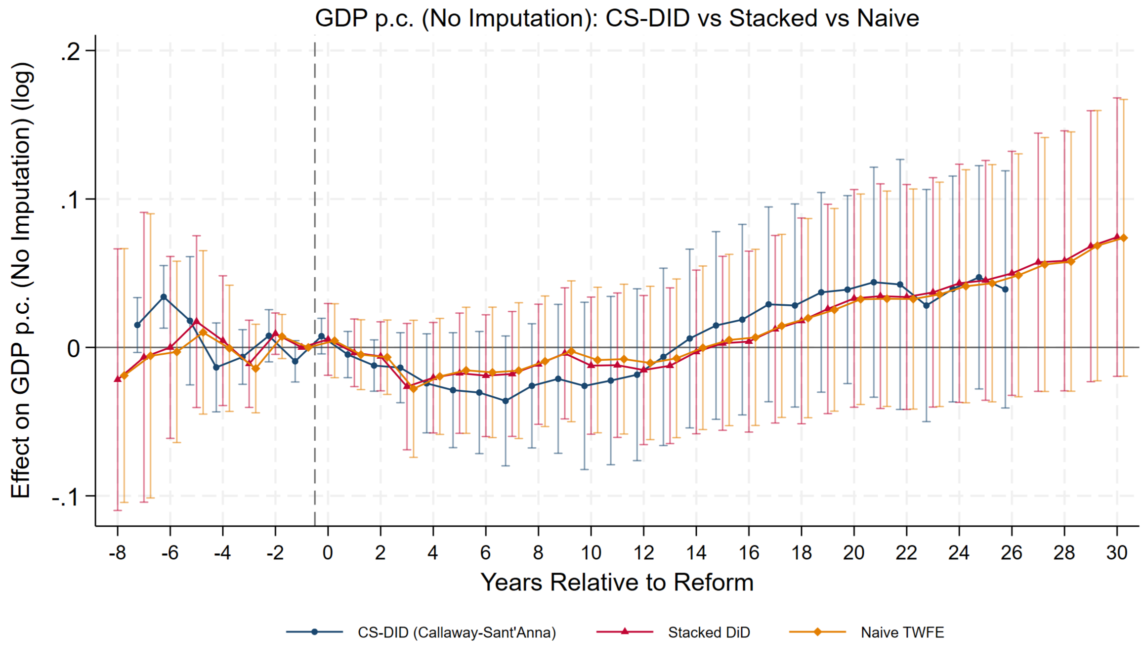


(a) Full sample

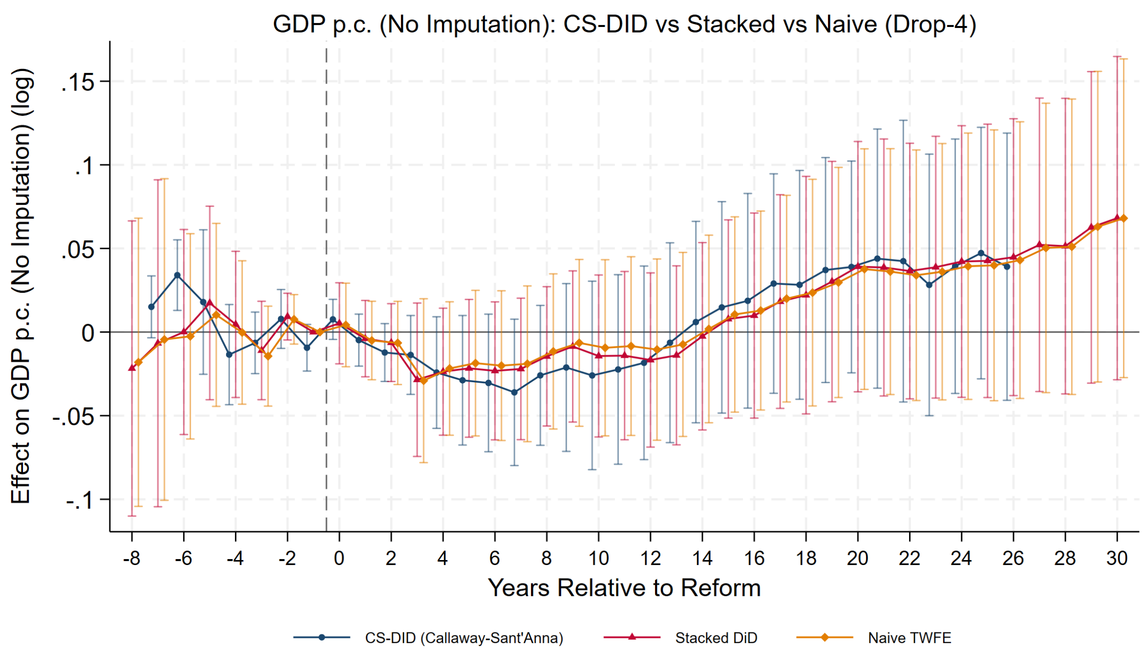


(b) Excluding energy and resource-intensive states

Figure A4. Comparison of event-study estimators for GDP per-capita. The figure compares dynamic treatment effect estimates obtained using the Callaway–Sant'Anna estimator, a stacked difference-in-differences estimator, and a conventional two-way fixed effects event study. Panel (a) reports estimates using the full sample of states. Panel (b) excludes Arizona, Montana, Oklahoma, and Texas, four energy- and resource-intensive states whose wage and income dynamics are driven in part by commodity price cycles unrelated to school finance reform. The similarity across panels confirms that the estimator comparison is not sensitive to the inclusion of these states.



(a) Full sample



(b) Excluding energy and resource-intensive states

Figure A5. Comparison of event-study estimators for GDP per-capita using observed GDP data. The figure compares dynamic treatment effect estimates obtained using the Callaway–Sant'Anna estimator, a stacked difference-in-differences estimator, and a conventional two-way fixed effects event study. Unlike the baseline GDP analysis, these estimates use only observed GDP data and do not rely on the backcast for years prior to 1997. Panel (a) reports estimates using the full sample of states. Panel (b) excludes Arizona, Montana, Oklahoma, and Texas, four energy- and resource-intensive states whose wage and income dynamics are driven in part by commodity price cycles unrelated to school finance reform. The similarity across panels and estimators confirms that the GDP results are not driven by the backcast, estimator choice, or the inclusion of these states.

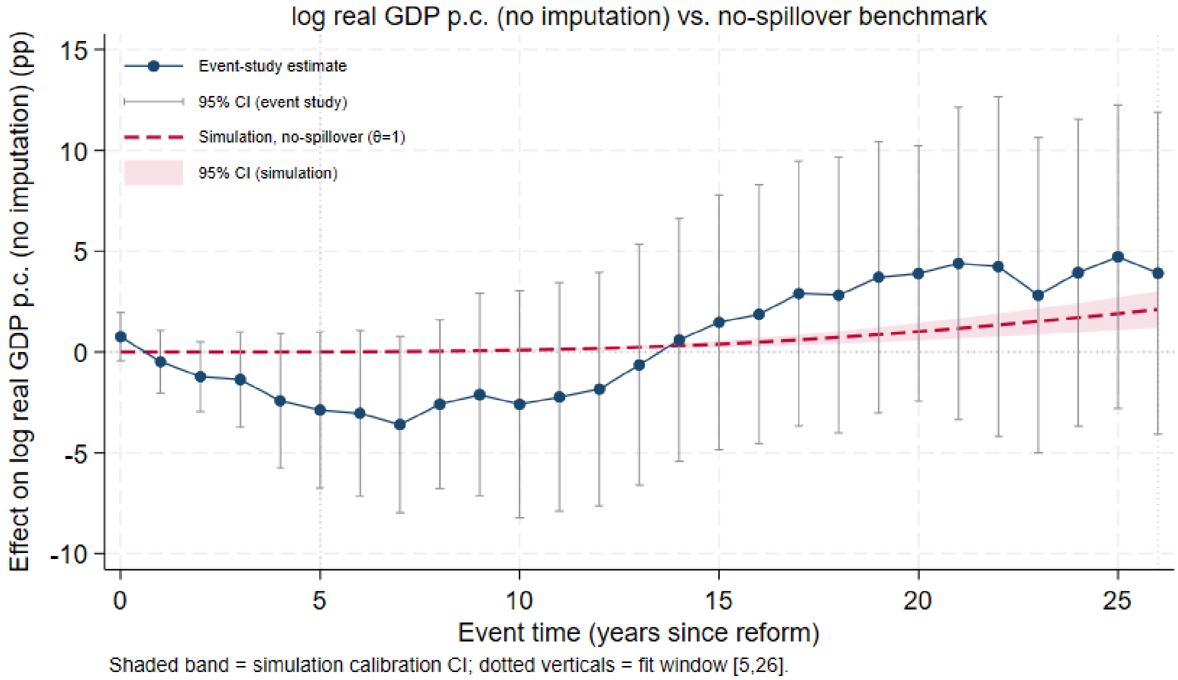
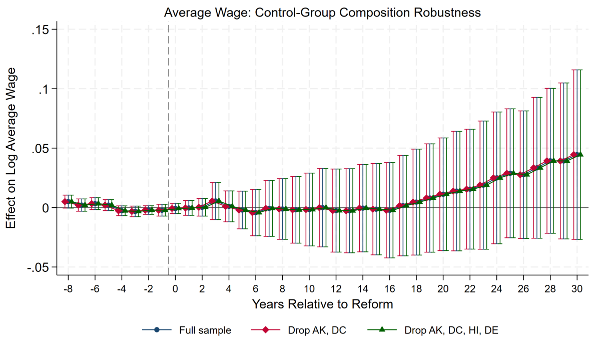


Figure A6. Estimated GDP Response (no imputation) and the No-Spillover Benchmark

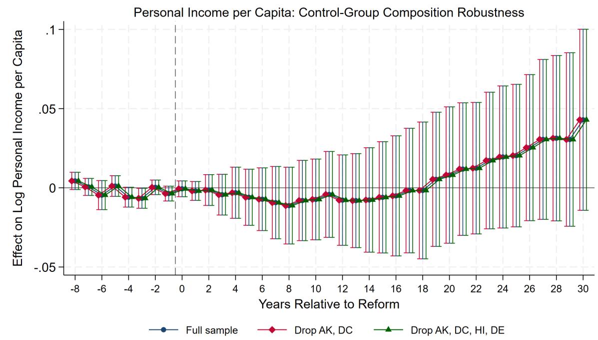
Notes: The solid line reports the event-study estimates of the effect of school finance reform on log GDP per-capita, with 95 percent confidence intervals shown by the vertical bars. The dashed line plots the simulated earnings path implied by the micro estimates under the no-spillover benchmark, $\theta = 1$, and the shaded region denotes the 95 percent confidence interval associated with uncertainty in the micro calibration.

Note: Because the GDP series begins in 1997 and the Callaway–Sant'Anna estimator requires at least one pre-treatment observation, this event-study analysis is necessarily restricted to reforms implemented after 1997. In addition, to estimate effects through event year 20—the point at which the estimated effects have largely stabilized—reforms must have been implemented by 2005. These restrictions leave seven reforms (New Mexico, Pennsylvania, South Carolina, Colorado, New York, Georgia, and Idaho), each of which has both pre-treatment observations and at least twenty years of post-treatment data. Consequently, every treated reform contributes to each event year in the estimation window, yielding a balanced event-study sample through year 20.

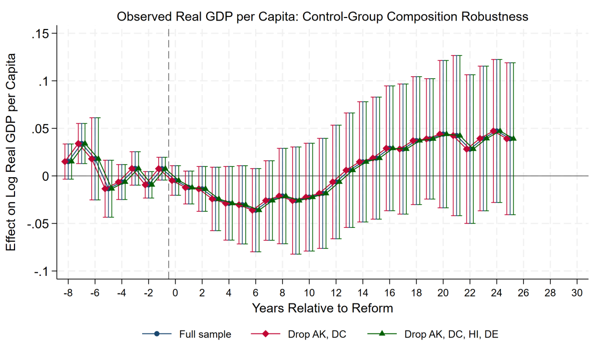
Figure A7. Robustness to Control-Group Composition



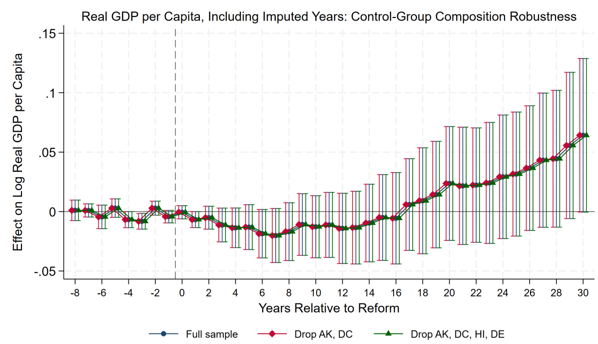
(a) Average wage



(b) Personal income per-capita



(c) Observed real GDP per-capita



(d) Real GDP per-capita, including imputed years

Notes: Each panel presents event-study estimates under alternative control-group composition restrictions. The full-sample specification is compared with specifications excluding Alaska and the District of Columbia, and excluding Alaska, the District of Columbia, Hawaii, and Delaware. Vertical bars report 95 percent confidence intervals.

Simulating the Aggregate Income Path Implied by Micro Estimates of School Finance Reform Effects

A key part of the analysis is to work out what the existing micro-level evidence on school finance reform implies for aggregate earnings. The logic is simple. The micro studies tell us how much a child's adult earnings rise after exposure to a reform when observed at certain ages. If earnings respond *only* through this childhood channel then the aggregate earnings effect in any year is a weighted average of these individual effects across the workers then in the labor force. This section constructs that weighted average and traces how it evolves over time, yielding a predicted time path for the aggregate, per-worker earnings effect of a reform.

The construction relies on three components: an *age profile* that describes how the earnings effect of childhood exposure evolves over the life cycle; an *exposure rule* that determines how fully each birth cohort was treated given the reform's timing; and a fixed *population age structure* that gives the age composition of the workforce over which individual effects are averaged into a population mean in each year. I calibrate the simulation to a benchmark based on two published micro estimates and propagate the sampling uncertainty in that benchmark into confidence intervals. Because the construction shuts down general equilibrium spillovers and any crowd-out of private investment, the resulting path is an empirically testable partial-equilibrium benchmark for the aggregate earnings response to a reform.

IX.A. Age profile

I model the earnings effect of childhood exposure as a smooth non-decreasing function of age, normalized to rise from zero to one between ages 20 and 45. The unscaled shape blends a linear ramp with a concave quadratic ramp,

$$g(a) = (1-w)x + w(2x - x^2), \qquad x = \frac{a-20}{25}, \tag{6}$$

for $20 < a < 45$, with $g(a) = 0$ for $a \leq 20$ and $g(a) = 1$ for $a \geq 45$. The weight $w \in [0,1]$ governs the concavity, and I set $w = 0.5$.[9] The realized profile scales this shape by a plateau P, as in $\varphi(a) = P\,g(a)$, so the effect reaches its maximum P at all ages beyond 45. The plateau is pinned down by the calibration below. While the assumed earnings profile has conceptual meaning, as I show below, the predicted aggregate path over the observable data horizon (roughly 0–30 years) is essentially invariant to the assumed life-cycle earnings profile.

[9] The linear component captures the steady accumulation of the earnings gap with experience, while the concave quadratic component allows the rate of accumulation to slow as workers approach their earnings plateau. Equal weighting yields a profile that rises briskly through the twenties before gradually approaching its plateau, a parsimonious approximation to the gradual emergence of earnings effects in the underlying studies. Results are largely similar under a simple linear profile.

IX.B. Population age structure

Let $s(a)$ denote the share of the working-age population at age a, taken from the Census age distribution and normalized to sum to one over ages 18 to 65. These shares are used to convert age-specific effects into population averages and supply the weights used to harmonize the two source estimates (which are estimated over different age ranges). Holding the age distribution fixed isolates the dynamics of interest, which are driven by the replacement of untreated cohorts by treated cohorts, from secondary variation in the population's age composition. Note that the simulated path is largely unchanged when using worker age weights instead of population age weights. See Appendix Figure A10.

IX.C. Source estimates, harmonization, and calibration

I anchor the scale of the earnings profile to two influential studies of school finance reforms. Jackson et al. (2016) report that school finance reforms increased per-pupil spending by approximately 10 percent, while Rothstein and Schanzenbach (2022) estimate increases of roughly 5 percent. As a validation exercise, Appendix Figure A1 presents an event-study analysis of the reforms in my sample. The estimates indicate that, following a gradual phase-in over approximately five years, the reforms increased real per-pupil spending by about 10 percent, with these gains remaining persistent for more than a decade. These results closely mirror the findings of the existing literature and provide additional evidence that school finance reforms generated sustained increases in educational spending. Note that I take the reduced form effect of the reforms on outcomes as reported in the papers rather than using percent of spending as the unit.

Turning to the effect on adult earnings, Jackson, Johnson, and Persico (2016) estimate that, for the full population, exposure to reform-induced school spending increases adult log earnings by approximately 0.078 (SE = 0.0196) for individuals exposed to the reform throughout all twelve school-age years, with earnings averaged over ages 20 to 45. Rothstein and Schanzenbach (2022) estimate an effect of approximately 0.0036 (SE = 0.0015) per additional year of exposure to a post-1990 school finance reform, with earnings averaged over ages 26 to 39. These estimates provide the empirical foundation for the calibration that follows.

I place both estimates on a common basis before combining them. First, I express each as the effect of thirteen years of exposure, consistent with the exposure rule below and assuming effects are linear in years of exposure.[10] Second, I harmonize the reporting age ranges. The two samples share essentially the same midpoint age (32.5), but the wider 20–45 window used by Jackson, Johnson, and Persico places more weight far from the midpoint, so under a concave profile its average lies slightly below that of the narrower 26–39 window used by Rothstein and Schanzenbach. I therefore multiply the Rothstein and Schanzenbach estimate by $K = 0.9555$ to express it over ages 20 to 45.[11]

The harmonized estimates are $0.078 \times (13/12) = 0.0845$ for Jackson, Johnson, and Persico

[10] I rescale the Jackson, Johnson, and Persico estimate by 13/12 and convert the Rothstein and Schanzenbach per-year estimate to full exposure by multiplying by 13. The same rescaling applies to the standard errors.

[11] The factor is the ratio of unscaled profile means over the two windows, $K = m(20, 45)/m(26, 39)$ with $m(\ell, h) = \sum_{a=\ell}^{h} s(a)g(a) / \sum_{a=\ell}^{h} s(a)$. Evaluated at the baseline profile and Census shares, $m(20, 45) = 0.585$ and $m(26, 39) = 0.612$, so $K = 0.9555$. With equal weights and a linear profile the windows share a midpoint and $K = 1$; the Census weights alone give 1.006, and the remaining 4.5 percent reflects the concavity of the profile.

and $0.0036 \times 13 \times 0.9555 = 0.0447$ for Rothstein and Schanzenbach. I combine them by a simple average, treating them as independent because the two studies exploit non-overlapping sets of reforms (pre-1990 and post-1990) and therefore distinct treated cohorts.[12] This gives a benchmark of $\bar{\beta} = 0.0646$ (SE 0.0141).

Pinning Down P

The published estimates are *average* earnings effects over ages 20 to 45, whereas the simulation requires an age-specific earnings profile, determined by the plateau P. Because the assumed profile is increasing over this range, the plateau (the effect for workers aged 45 and older) must exceed the observed average. I recover the implied plateau by choosing P so that the population-weighted average of $Pg(a)$ over ages 20 to 45 equals the benchmark,

$$P = \frac{\bar{\beta}}{m(20, 45)} = 0.1105. \tag{7}$$

An average effect of 6.5 percent over ages 20 to 45 thus implies a profile that is zero through age 20, rises over the early career, and reaches a maximum of about 11.05 percent by age 45.[13]

IX.D. Cohort exposure rule

A reform affects only cohorts whose school-age years overlap the post-reform period. I assign each cohort an exposure equal to the fraction of its thirteen school-age years (ages 5 to 17) that fell after the reform. Cohorts not yet in school at enactment receive full exposure, cohorts already finished receive none, and cohorts mid-schooling receive partial exposure. With effects being linear in years of exposure, a cohort exposed for x of thirteen years receives $x/13$ of the full effect.[14] Figure 1 shows the implied profiles for full, eight-year, and four-year exposure, illustrating both the gradual emergence of effects over the life cycle and the proportional scaling with exposure.

IX.E. Aggregating to the earnings path

With the simulated effect for each cohort, it is straightforward to construct the simulated aggregate earnings path for the economy assuming no spillovers or crowd-out. For each event year t, the aggregate log-earnings effect is the population-weighted sum of age-specific effects, each scaled by that age group's exposure,

$$A(t) = \sum_{a=18}^{65} s(a)\, e(a, t)\, \varphi(a). \tag{8}$$

[12]Writing the harmonized estimates and standard errors as β_j and SE_j, the benchmark and its standard error are $\bar{\beta} = (\beta_{\mathrm{JJP}} + \beta_{\mathrm{RS}})/2$ and $\mathrm{se}(\bar{\beta}) = \frac{1}{2}\sqrt{\mathrm{SE}_{\mathrm{JJP}}^2 + \mathrm{SE}_{\mathrm{RS}}^2}$. Because the two harmonized standard errors are similar, an inverse-variance weighted average yields nearly the same result.

[13]Because the calibration is linear, this is equivalent to computing the implied plateau from each study over its own age window and averaging the two plateaus.

[14]Formally, the exposure of workers aged a in event year t is $e(a, t) = \min\{1, \max(0, (18 + t - a)/13)\}$.

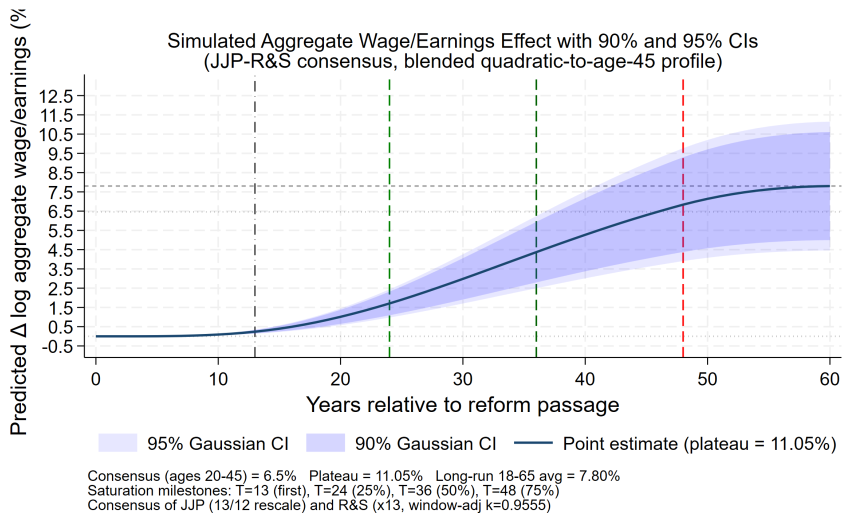
The age shares weight each group's contribution, the exposure fractions evolve as treated cohorts age into the workforce and untreated cohorts age out, and the age profile sets the effect for a fully exposed worker. I evaluate the path for event years 0 to 60. Because $A(t) = q(t)\bar{\beta}$ is linear in the benchmark, uncertainty propagates proportionally as $\mathrm{se}\{A(t)\} = q(t)\,\mathrm{se}(\bar{\beta})$, which permits analytical confidence intervals.[15]

Figure 1 presents the results, marking four saturation milestones. The first fully exposed cohort enters the workforce after 13 years, and 25, 50, and 75 percent of the working-age population is fully exposed after roughly 24, 36, and 48 years.

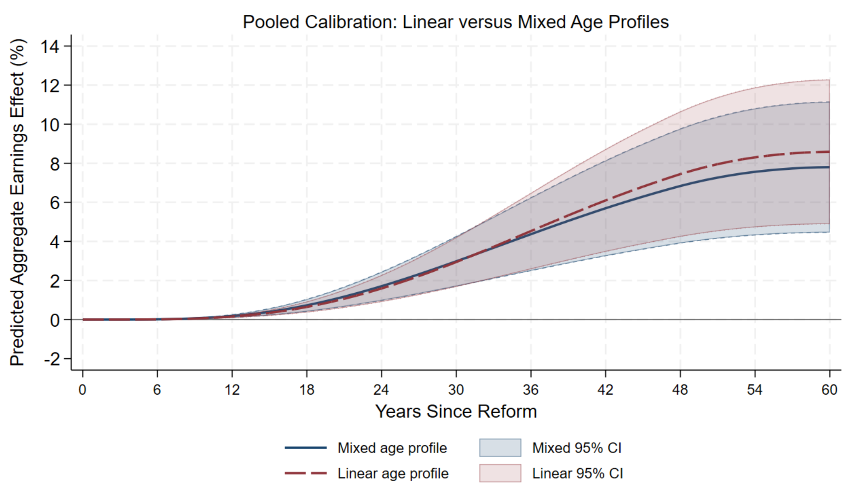
Given that the simulated path is a central object of interest in the analysis, Figure A8 also compares the baseline simulation, which assumes a mixed linear–quadratic age–earnings profile, with an alternative simulation based on a simple linear age–earnings profile. As the figure shows, the two specifications generate nearly identical aggregate earnings paths throughout the first 25–30 years after reform passage—the empirically relevant period examined in the event-study analysis. The paths begin to diverge only after approximately 35 years, when the linear profile reaches a somewhat higher long-run plateau. Even then, the differences remain small relative to the uncertainty implied by the sampling variation in the underlying micro estimates. This comparison demonstrates that the simulated aggregate earnings path is robust to reasonable alternative assumptions regarding the evolution of earnings over the life cycle.

Appendix Figure A9 presents analogous simulations calibrated separately to Jackson, Johnson, and Persico (2016) and to Rothstein and Schanzenbach (2022). The former implies a somewhat larger long-run earnings effect, while the latter implies a more modest one, but the timing and shape of the simulated paths are nearly identical. Because the paper identifies the timing and curvature of the response rather than its exact magnitude, the central conclusions are insensitive to which estimate is used. I therefore calibrate the baseline simulation to the simple average of the two estimates. This is a natural benchmark because the estimates are statistically consistent with one another, arise from independent reform eras (pre-1990 and post-1990), and the treated sample in this paper is composed of roughly equal numbers of earlier and later reforms. Combining the two estimates also yields substantially tighter confidence intervals than either study alone, providing a more informative calibration.

[15] I report 90 and 95 percent intervals. These reflect sampling uncertainty in the micro estimates only, not uncertainty in the age profile, exposure rule, or population age structure.



(a) Baseline simulated aggregate earnings path



(b) Comparison of mixed and linear age profiles

Figure A8. Predicted Aggregate Earnings Effect of School Finance Reform

Notes: Panel a plots the baseline simulated aggregate earnings effect of school finance reform by years since reform passage. The simulation is calibrated to the combined benchmark implied by Jackson, Johnson, and Persico (2016) and Rothstein and Schanzenbach (2022). The benchmark effect over ages 20–45 is 6.5 percent, implying an age-45 plateau of 11.05 percent and a long-run population-weighted average effect over ages 18–65 of 7.8 percent. The shaded regions denote 90 and 95 percent confidence intervals obtained by propagating the estimated sampling uncertainty in the combined benchmark. The vertical lines mark the entry of the first fully exposed cohort into the workforce and the horizons at which 25, 50, and 75 percent of the working-age population is fully exposed. Panel b assesses sensitivity to the assumed age profile of the earnings effect. The mixed profile, used in the baseline simulation, is an equal-weighted combination of a linear profile and a concave quadratic profile that rises from zero at age 20 to its plateau at age 45. The alternative profile increases linearly between ages 20 and 45 and remains constant thereafter. Each profile is separately calibrated so that its population-weighted average over ages 20–45 matches the combined benchmark. The shaded regions report 95 percent confidence intervals obtained by propagating uncertainty in the corresponding benchmark estimate. The similarity of the two paths indicates that the simulated aggregate earnings trajectory is not sensitive to using the mixed rather than the linear age profile.

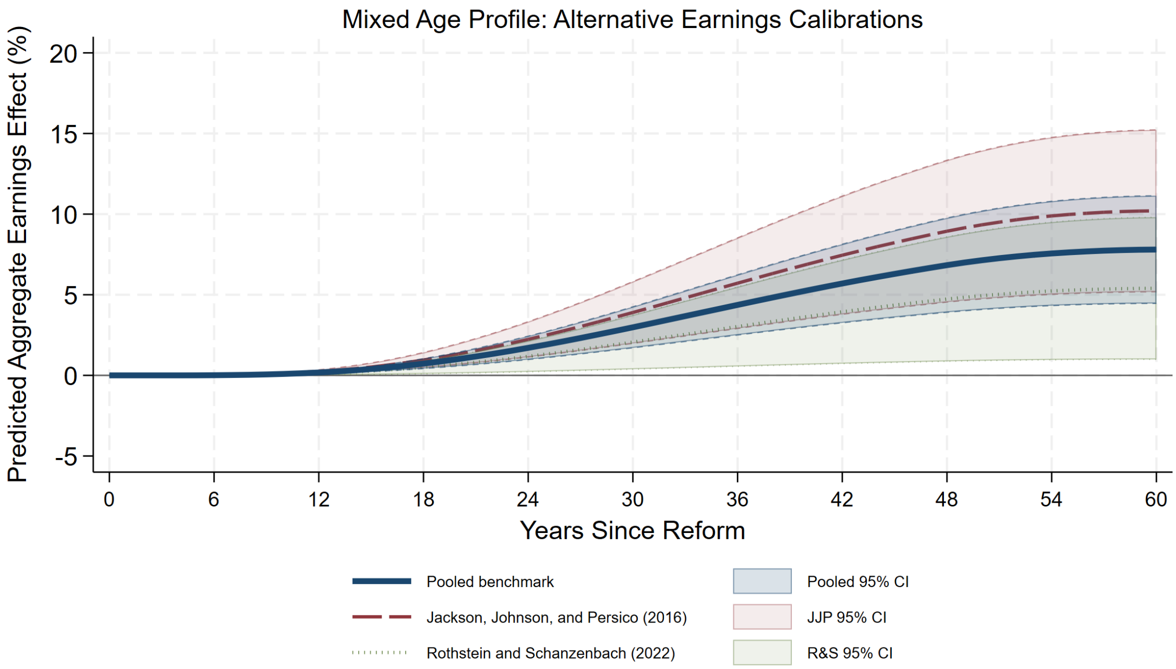


Figure A9. Alternative Earnings Calibrations Under the Mixed Age–Earnings Profile

Notes: The figure compares three alternative calibrations of the aggregate earnings simulation under the baseline mixed linear–quadratic age–earnings profile. The solid blue line denotes the baseline calibration, which averages the benchmark estimates reported by Jackson, Johnson, and Persico (2016) and Rothstein and Schanzenbach (2022). The dashed red line uses only the Jackson, Johnson, and Persico (2016) estimate, while the dotted green line uses only the Rothstein and Schanzenbach (2022) estimate. Shaded regions denote 95 percent confidence intervals obtained by propagating the sampling uncertainty from the underlying micro estimates through the simulation. Although the long-run magnitudes differ somewhat across calibrations, the predicted timing and curvature of the aggregate earnings response are similar. The pooled calibration yields substantially narrower confidence intervals than either study individually because it combines independent evidence from both reform eras.

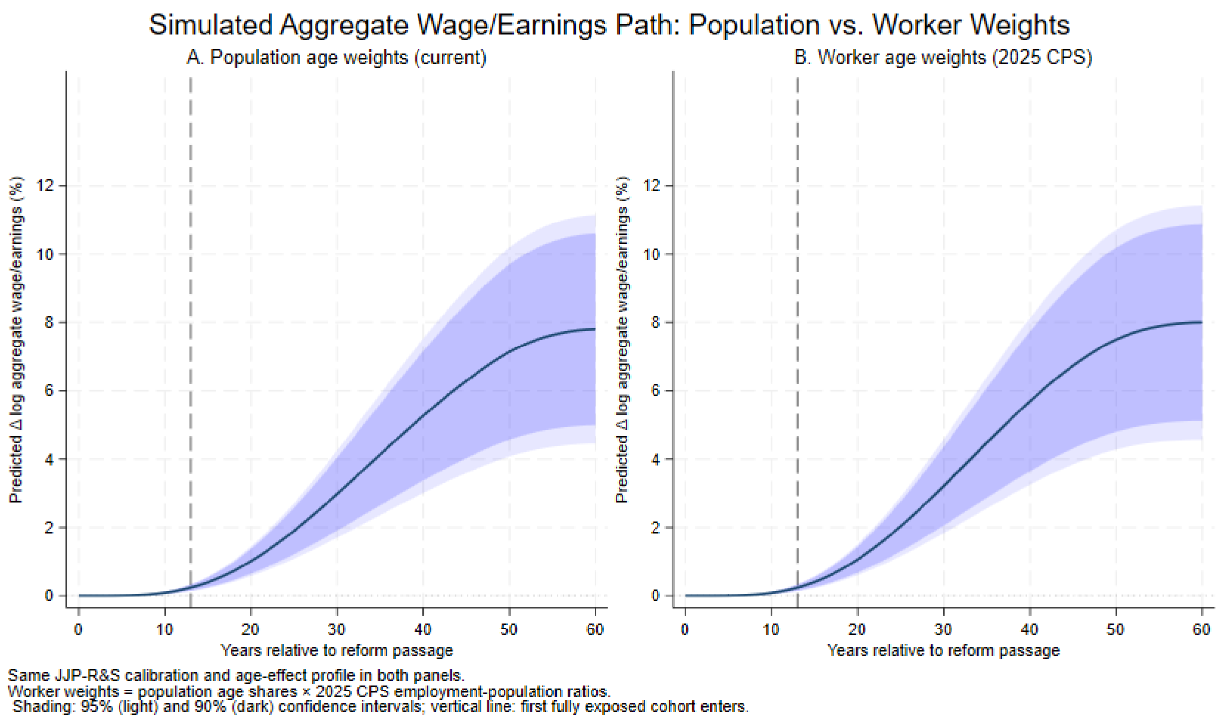


Figure A10. Simulated Aggregate Wage/Earnings Path Using Population and Worker Age Weights. Panel A presents the simulated aggregate wage/earnings path using the population age weights employed in the baseline analysis. Panel B instead uses worker age weights constructed by multiplying the population age distribution by age-specific employment-to-population ratios from the 2025 Current Population Survey and normalizing the resulting weights to sum to one over ages 18–65. The JJP–R&S calibration and age-specific earnings-effect profile are held fixed across panels, so the comparison isolates the effect of using worker rather than population weights in aggregating cohort-level earnings gains. Shaded regions denote 90 and 95 percent confidence intervals. The vertical dashed line indicates the entry of the first fully exposed cohort into the labor market.

X. Backcasting State GDP

Official state GDP data are available beginning in 1997. To extend the GDP series to earlier years, I use the relationship between real GDP per-capita and real personal income per-capita within each state during the period in which both are observed. I first convert personal income per-capita to real terms using the PCE price index and define

$$r_{st} = \log\left(GDP_{st}^{real}\right) - \log\left(PI_{st}^{real}\right). \tag{9}$$

I then calculate, separately for each state, the average value of this log GDP-to-PI ratio over the years for which GDP is observed:

$$\bar{r}_s = \frac{1}{T_s} \sum_{t \geq 1997} r_{st}. \tag{10}$$

For years prior to 1997, I impute log real GDP per-capita as

$$\log\left(\widehat{GDP_{st}^{real}}\right) = \log\left(PI_{st}^{real}\right) + \bar{r}_s. \tag{11}$$

Thus, the backcast assumes that each state's average GDP-to-personal-income ratio observed after 1997 also applies to the pre-1997 period, while allowing the imputed GDP series to evolve over time with that state's observed real personal income. Actual GDP is used for all years beginning in 1997; only pre-1997 observations are imputed.

Importantly, the main GDP results are similar when the analysis is restricted entirely to observed post-1997 GDP and no imputed observations are used (Appendix Figure A6).