



Cost Efficiency of U.S. School Districts: Stochastic Frontier Evidence from the National Education Cost Model

Bruce D. Baker
University of Miami

I estimate the technical efficiency of U.S. school districts in producing middle-grades achievement with a stochastic frontier model built on the National Education Cost Model (NECM), which measures each district's spending relative to the predicted cost of national-average outcomes. Because outcomes are measured in standard-deviation units, efficiency is expressed in input terms: the share of a district's resources, relative to cost, that a frontier district would need to reach the same outcome. Using 128,713 district-years for 8,088 districts in 38 states (2009–2024), I find a concave frontier: a 10 percent increase in resources relative to cost raises frontier outcomes by 0.066 SD in the pooled model and 0.086 SD within states. States account for 20 percent of the variance in pooled efficiency but 1 percent once state frontier intercepts are included, and those intercepts track pooled state rankings almost exactly ($r = 0.96$). This pattern is consistent with cross-state measurement error in linked assessments and reported spending, which attenuates the pooled frontier and inflates state differences in efficiency. District rankings are otherwise stable across specifications (rank correlation 0.88). The share of district-years in the two lowest efficiency bands doubled between 2011 and 2024, from about 11 to 22 percent, and the widening occurs within states. Relative to their own state's averages, the least efficient districts are smaller, spend more and are more heavily staffed in every category, about 1 SD above the state mean in total staff per pupil. Differences in administrative spending, by contrast, are small in budget terms.

VERSION: October 2026

Suggested citation: Baker, Bruce D.. (2026). Cost Efficiency of U.S. School Districts: Stochastic Frontier Evidence from the National Education Cost Model. (EdWorkingPaper: 26-1626). Retrieved from Annenberg Institute at Brown University:
<https://doi.org/10.26300/v2th-4f94>

Cost Efficiency of U.S. School Districts: Stochastic Frontier

Evidence from the National Education Cost Model

Bruce D. Baker

University of Miami

bdb119@Miami.edu

Abstract

I estimate the technical efficiency of U.S. school districts in producing middle-grades achievement with a stochastic frontier model built on the National Education Cost Model (NECM), which measures each district's spending relative to the predicted cost of national-average outcomes. Because outcomes are measured in standard-deviation units, efficiency is expressed in input terms: the share of a district's resources, relative to cost, that a frontier district would need to reach the same outcome. Using 128,713 district-years for 8,088 districts in 38 states (2009–2024), I find a concave frontier: a 10 percent increase in resources relative to cost raises frontier outcomes by 0.066 SD in the pooled model and 0.086 SD within states. States account for 20 percent of the variance in pooled efficiency but 1 percent once state frontier intercepts are included, and those intercepts track pooled state rankings almost exactly ($r = 0.96$). This pattern is consistent with cross-state measurement error in linked assessments and reported spending, which attenuates the pooled frontier and inflates state differences in efficiency. District rankings are otherwise stable across specifications (rank correlation 0.88). The share of district-years in the two lowest efficiency bands doubled between 2011 and 2024, from about 11 to 22 percent, and the widening occurs within states. Relative to their own state's averages, the least efficient districts are smaller, spend more and are more heavily staffed in every category, about 1 SD above the state mean in total staff per pupil. Differences in administrative spending, by contrast, are small in budget terms.

Keywords: efficiency; stochastic frontier; education cost function; school finance; staffing; measurement error

JEL codes: I21, I22, H75, C23

1. Introduction

The question of whether money matters for schools has largely been settled. A decade of quasi-experimental evidence shows that increases in school spending, particularly those driven by state school finance reforms, improve student achievement and long-run attainment (Jackson, Johnson, & Persico, 2016; Lafortune, Rothstein, & Schanzenbach, 2018; Jackson & Mackevicius, 2024). The policy debate has shifted accordingly. Opponents of additional funding now argue less often that money does not matter and more often that districts waste what they already have. Claims about administrative growth, overstaffing and inefficient districts recur in legislative testimony and litigation, typically supported by simple comparisons of spending and test scores that ignore differences in the cost of achieving common outcomes.

Measuring efficiency credibly requires two things those comparisons lack. First, resources must be measured relative to what it costs a given district, with its students and its setting, to reach a common outcome target. A district spending \$18,000 per pupil may be generously or poorly funded depending on its student needs, scale and regional labor costs. Second, the analyst must separate shortfalls from best practice from noise and measurement error in both outcomes and spending. This second problem is acute in national work. State assessments differ in content, standards and testing populations, and states differ in how they record district expenditures. Errors of this kind are shared by every district in a state and will appear, in any national model, as state-level differences in efficiency.

This paper estimates the technical efficiency of U.S. school districts using a stochastic frontier production model built on the National Education Cost Model (NECM). The NECM provides, for each district and year, the predicted per-pupil cost of achieving national-average outcomes. I

model each district's outcome gap, in standard deviations of a nationally linked achievement index, as a concave function of the log of its spending relative to that cost, with a one-sided inefficiency term. Because outcomes in standard-deviation units can be negative, the conventional output-oriented measure is not meaningful. I instead express efficiency in input terms, as the share of a district's resources relative to cost that a frontier-efficient district would need to reach the same outcome.

The paper makes four contributions. First, it provides a national, district-level panel of efficiency estimates on a common scale across 14 years, which allows the distribution of efficiency to be tracked over time rather than standardized within each year. Second, it uses state fixed effects in the frontier to bracket the role of cross-state measurement error. A pooled model treats all state-level differences as district inefficiency; a state fixed effects model treats none of them as such. Comparing the two shows how much of what national comparisons call inefficiency is a state-level artifact. Third, it documents that the dispersion of efficiency has roughly doubled since 2011 and that this widening occurs within states. Fourth, it links efficiency to detailed staffing and spending data, comparing districts with their own state's averages rather than a national mean.

The main findings are as follows. The frontier is concave; a 10 percent increase in resources relative to cost raises frontier outcomes by 0.066 SD in the pooled model and 0.086 SD within states. The steeper within-state slope is what one would expect if state-level measurement error attenuates the pooled estimate. State intercepts track pooled state efficiency rankings almost exactly ($r = 0.96$), and the share of variance in efficiency between states falls from 19.5 to 1.1 percent once they are included. District rankings within states are otherwise stable (rank correlation 0.88), and no district-year moves between the most and least efficient bands. The

share of district-years in the two lowest efficiency bands rose from about 11 percent in 2011 to 22 percent in 2024 under both models. Finally, the least efficient districts are not underfunded or overfunded relative to cost; they are higher-cost, smaller districts that are more heavily staffed in every category than their state peers. Their excess administrative spending is real but small, roughly \$80 to \$100 per pupil.

Section 2 reviews related work. Section 3 describes the data and Section 4 the empirical approach. Section 5 presents results and Section 6 robustness checks. Section 7 discusses implications for school finance policy, and Section 8 concludes.

2. Related literature

2.1 Education cost functions and efficiency

Education cost functions model district spending as a function of outcomes, input prices, student needs and scale, and are the standard tool for estimating the cost of reaching a given outcome target (Duncombe, Ruggiero, & Yinger, 1996; Duncombe & Yinger, 2005; Imazeki & Reschovsky, 2004). Their treatment of efficiency is central to their credibility. Because observed spending includes spending that does not contribute to measured outcomes, cost functions either include efficiency controls, such as fiscal capacity and competition measures, or estimate a cost frontier directly (Gronberg, Jansen, & Taylor, 2011). Critics argue that efficiency controls cannot fully separate necessary cost from inefficiency, so cost estimates may embed waste (Costrell, Hanushek, & Loeb, 2008); defenders respond that the methods produce stable, validated estimates of relative cost (Duncombe & Yinger, 2011). The present paper takes the cost model as given and asks the complementary question: conditional on resources relative to estimated cost, how far does each district fall short of the best attainable outcome?

The NECM extends cost function methods to the national level, estimating district costs from a national panel and producing, for each district and year, the predicted spending needed to reach national-average outcomes (Baker, Weber, & Srikanth, 2021; Baker, 2025b). Appendix B documents the specification, estimates and validation tests. The funding gap, actual spending divided by this predicted cost minus one, is the resource measure used here. Normalizing resources by cost in this way addresses the first problem identified above: districts are compared on resources relative to need rather than raw dollars.

2.2 Frontier methods in education

Efficiency in education has been estimated with both nonparametric (data envelopment analysis) and parametric (stochastic frontier) methods (De Witte & López-Torres, 2017). Early district-level applications include Ruggiero (1996), who studied New York districts, and Grosskopf, Hayes, Taylor, and Weber (1997, 2001), who studied Texas districts and the role of competition and monitoring. Chakraborty, Biswas, and Lewis (2001) compared stochastic and nonstochastic approaches for Utah districts. Gronberg, Jansen, Karakaplan, and Taylor (2015) used a stochastic cost frontier to study scale and consolidation in Texas. These studies are almost all single-state, which sidesteps cross-state comparability but limits national inference.

The stochastic frontier model separates a one-sided inefficiency term from symmetric noise (Aigner, Lovell, & Schmidt, 1977; Meeusen & van den Broeck, 1977), with district-level efficiency recovered from the conditional distribution of inefficiency given the composed residual (Jondrow, Lovell, Materov, & Schmidt, 1982; Battese & Coelli, 1988). Identification rests on negative skewness of the composed residual; when skewness is weak, estimated inefficiency is small and imprecise (Waldman, 1982; Simar & Wilson, 2010). Greene (2005)

showed that, in panel settings, unobserved heterogeneity across units is easily mistaken for inefficiency, and proposed “true” fixed and random effects frontiers that separate the two. I apply that logic at the state level: state intercepts absorb heterogeneity, including measurement error common to a state, that a pooled model would attribute to districts. Kumbhakar and Tsionas (2006) discuss input-oriented efficiency in stochastic frontier production models, the measure adopted here.

2.3 Measurement across states

National comparisons of district achievement depend on linking state assessments to a common scale. The Stanford Education Data Archive (SEDA) does so using state NAEP results, and validation studies show the linked scores track district performance well on average, with error that varies by state (Reardon, Kalogrides, & Ho, 2021). Spending data present parallel problems: states differ in how they record state-paid pension contributions, payments to other agencies and similar items in the Census Annual Survey of School System Finances (F-33). Errors common to a state attenuate estimated relationships between resources and outcomes in pooled data, a standard errors-in-variables result. Within-state estimation removes error components shared by all districts in a state, although it can amplify the influence of district-specific error (Griliches & Hausman, 1986), which would bias a within-state slope toward zero rather than away from it. This paper uses that logic to interpret the gap between pooled and within-state frontier estimates.

3. Data

3.1 The NECM district panel

The analysis uses the 2024 release of the NECM district panel (Baker, Di Carlo, & Weber, 2024), covering fiscal years 2009 through 2024. For each district and year the panel provides three measures. The first is the predicted per-pupil current spending required to achieve national-average outcomes, given the district's student needs, scale, location and regional input prices (cost model 1; Appendix B). The second is actual current spending per pupil from the F-33 (U.S. Census Bureau, n.d.). The third is an outcome index in standard deviations relative to the national average.

The outcome index includes only state reading and mathematics assessments in grades 6 through 8, placed on a nationally normed scale by SEDA. Efficiency here therefore means efficiency in producing middle-grades achievement in two subjects, not in producing early-grade learning, high school completion or non-academic outcomes on which districts also spend.

The resource measure is the funding gap, actual spending divided by predicted cost minus one. A funding gap of zero means a district spends exactly what the NECM predicts it would need to reach national-average outcomes; a gap of -20 percent means it spends 20 percent less. The frontier models below use the log of the underlying spending-to-cost ratio.

3.2 Sample

In several states the outcome index is identical for a given district for three or more consecutive years, which indicates carried-forward or imputed values rather than measured outcomes. Eleven states in which this applies to nearly all districts are excluded: Alaska, Colorado, the District of

Columbia, Maine, New Hampshire, New Mexico, New York, North Dakota, Oregon, Pennsylvania and Washington. Elsewhere, individual districts with such runs are excluded. Hawaii is not in the NECM file, and Vermont’s single remaining district drops out. The spending-to-cost ratio is trimmed at the 0.1st and 99.9th percentiles. The analytic sample is 128,713 district-years for 8,088 districts in 38 states. Efficiency is averaged over trailing three-year windows, which yields 112,215 district-years with three-year estimates for 2011–2024.

3.3 Staffing and spending profiles

To characterize districts across the efficiency distribution, I link staffing counts from the NCES Common Core of Data (CCD; National Center for Education Statistics, n.d.) and functional spending detail from the F-33 for fiscal years 2018–2024 (56,005 linked district-years). Staffing is measured as full-time-equivalent staff per 100 pupils in ten categories, from classroom teachers to central office support staff; zero counts are treated as missing. Spending is measured as shares of current elementary–secondary spending in nine functional categories. Both are winsorized at the 1st and 99th percentiles.

Table 1 describes the analytic sample. The typical pupil attends a district slightly below the national average outcome (an enrollment-weighted mean outcome gap of -0.07 SD) that spends somewhat less than the NECM predicts it would need to reach that average: the enrollment-weighted median funding gap is -14 percent. Unweighted, the mean funding gap is positive ($+6$ percent) but the median is close to zero (-2 percent), reflecting a right tail of small, high-spending districts.

Table 1. Descriptive statistics for the analytic sample, 2009–2024

	Unweighted			Enrollment-weighted		
	Mean	SD	Median	Mean	SD	Median
Outcome gap (SD)	−0.03	0.38	−0.03	−0.07	0.37	−0.07
Current spending per pupil	\$11,833	\$3,928	\$10,797	\$11,355	\$3,626	\$10,356
NECM predicted cost per pupil	\$12,326	\$5,360	\$11,094	\$13,308	\$5,670	\$12,115
Funding gap (spending ÷ cost − 1)	+6.4%	43.5%	−1.9%	−5.2%	38.4%	−14.2%
ln(spending ÷ cost)	−0.011	0.377	−0.019	−0.123	0.365	−0.154
Enrollment	4,591	14,366	1,576	49,548	98,563	13,888

N = 128,713 district-years, 8,088 districts, 38 states, fiscal years 2009–2024. Spending and predicted cost are nominal dollars per pupil, pooled across years. Predicted cost is the NECM cost of national-average outcomes (cost model 1; Appendix B). Funding gap = spending ÷ predicted cost − 1. Enrollment-weighted statistics weight each district-year by enrollment and so describe the district attended by the average pupil.

4. Empirical strategy

4.1 The production frontier

The frontier relates district *i*’s outcome gap *y* in year *t* to the log of its spending relative to

NECM predicted cost:

$$y_{it} = \alpha_t + \gamma_s + \beta \ln \left(\frac{S_{it}}{C_{it}} \right) + v_{it} - u_{it}, v_{it} \sim N(0, \sigma_v^2), u_{it} \geq 0$$

Here α_t are year effects, *S* is current spending per pupil, *C* is NECM predicted cost, *v* is symmetric noise and *u* is inefficiency, the shortfall of a district’s outcome from the frontier. In the baseline model $\gamma_s = 0$. In the state fixed effects model γ_s is a full set of state intercepts, a “true fixed effects” frontier in the sense of Greene (2005) applied at the state level, so each state has its own frontier height and inefficiency is measured relative to it.

The log form captures diminishing returns. In preliminary specification tests it was preferred to linear, quadratic, cubic and spline alternatives; adding a squared log term here changes the frontier very little (Section 6). Inefficiency *u* is specified as exponential, with half-normal as an alternative. Models are estimated by enrollment-weighted pseudo-maximum likelihood, with

weights normalized to mean 1, so that estimates describe the distribution of pupils rather than of districts. Because conventional likelihood-based standard errors are not valid under weighting, I report robust standard errors clustered by district.

Cross-sectional frontier identification requires negatively skewed OLS residuals. In the baseline model, weighted skewness is -0.13 overall and -0.1 to -0.2 in most years, which identifies inefficiency but only modestly. With state fixed effects, overall skewness is $+0.04$. Inefficiency is then identified from a lower tail that is longer than the upper tail: the 1st percentile of standardized residuals is -2.94 , against $+2.77$ at the 99th. District estimates under state effects therefore lean more heavily on the exponential assumption, a point I return to in Section 6.

4.2 Input-oriented technical efficiency

Because y is measured in standard deviations and can be negative, the conventional output-oriented measure $\exp(-u)$ has no meaningful interpretation. I instead express efficiency in input terms. Given the log frontier, a district with inefficiency u could reach its current outcome with resources relative to cost reduced by the factor $\exp(-u/\beta)$. Technical efficiency (TE) is the conditional expectation of that factor given the composed residual $\varepsilon = v - u$:

$$TE_{it} = E \left[\exp \left(\frac{-u_{it}}{\beta} \right) \mid \varepsilon_{it} \right]$$

This is a Battese–Coelli (1988) type predictor evaluated at the input-oriented transformation. A TE of 0.85 means a frontier-efficient district would need 85 percent of the district's current resources, relative to cost, to achieve the same outcome. Because TE scales u by β , a steeper frontier implies higher TE for a given outcome shortfall: the same shortfall can be closed with fewer additional resources.

Annual estimates are noisy, so I average TE over a trailing three-year window ($t-2$ to t), requiring all three years. Annual three-year results therefore begin in 2011.

4.3 Efficiency bands

Three-year TE in the baseline model is concentrated between 0.80 and 0.93 with a long lower tail (Figure 1). Few districts exceed 0.93: with noise large relative to inefficiency, the predictor shrinks estimates toward the mean and places no district on the frontier. I therefore classify district-years into four bands with round cut points supported by the distribution (Table 2). The 0.80 and 0.70 cuts lie at natural breaks identified by one-dimensional k-means clustering (0.798 and 0.682). The 0.90 cut marks the top decile. Under state fixed effects TE levels are higher, so bands are defined to hold the same shares as the baseline bands, which places their cut points at 0.939, 0.880 and 0.816.

Table 2. Technical efficiency bands, baseline model

Band	Three-year TE	Share of district-years	Basis for cut point
Most efficient	≥ 0.90	10.9%	Top decile
Typical	0.80–0.90	73.0%	Central mass of the distribution
Less efficient	0.70–0.80	13.0%	1-D k-means break at 0.798
Least efficient	< 0.70	3.0%	1-D k-means break at 0.682

Shares of district-years with three-year TE, 2011–2024 ($N = 112,215$). State fixed effects bands hold the same shares, with cut points 0.939, 0.880 and 0.816.

4.4 State-standardized profiles

National comparisons of staffing and spending across bands mix district differences with state differences, since the low-efficiency bands are concentrated in a few states with distinctive staffing patterns. I therefore also express each staffing and spending measure x relative to the district's own state and year:

$$z_{ist} = \frac{x_{ist} - \mu_{st}}{\sigma_{st}}$$

where μ_{st} and σ_{st} are the enrollment-weighted mean and standard deviation of x across linked districts in state s and year t (264 state-years, each with at least five districts). Band means of z are enrollment-weighted, so $+0.5$ means the band's pupils attend districts half a standard deviation above their own state's average.

5. Results

5.1 The frontier

The frontier is concave and precisely estimated in both models (Table 3). In the baseline model the coefficient on log spending relative to cost is 0.691 (SE 0.015), so a 10 percent increase in resources relative to cost raises frontier outcomes by 0.066 SD. With state fixed effects the coefficient rises to 0.906 (SE 0.017), or 0.086 SD per 10 percent. The frontier improves on OLS in both models, by 345 pseudo-log-likelihood points in the baseline and 151 with state effects.

Table 3. Stochastic frontier estimates

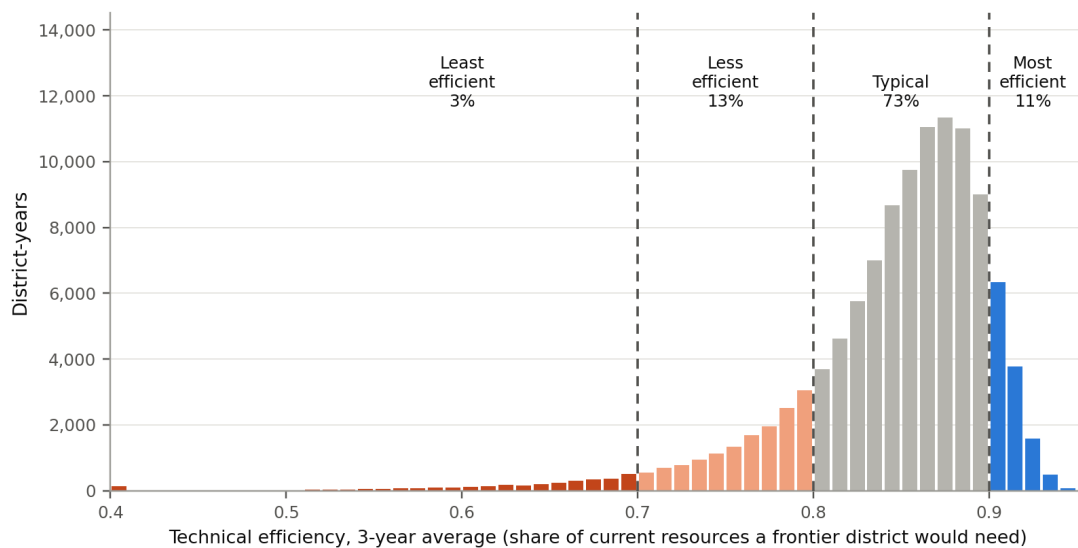
	Baseline (year effects)	State fixed effects
β , ln(spending/cost)	0.691 (0.015)	0.906 (0.017)
Frontier gain from +10% resources (SD)	0.066	0.086
σ_v	0.239	0.221
$E[u]$ (SD)	0.119	0.089
–Pseudo-LL, frontier	12,511	–1,750
–Pseudo-LL, OLS	12,856	–1,599
Skewness of OLS residuals (weighted)	–0.13	+0.04
Median three-year TE	0.859	0.914
Enrollment-weighted mean three-year TE	0.854	0.911
Share of variance in three-year TE between states	19.5%	1.1%

Exponential inefficiency; year fixed effects; N = 128,713 district-years; enrollment-weighted pseudo-likelihood (weights normalized to mean 1). Robust standard errors clustered by district in parentheses. Variance share is the R^2 from an enrollment-weighted regression of three-year TE on state indicators (N = 112,215).

The steeper within-state slope is the pattern expected if a substantial share of between-state variation in both variables is measurement error. Error in the spending-to-cost ratio common to a state attenuates the pooled coefficient, and state-level error in the outcome index adds noise unrelated to resources. Within a state, districts take the same assessments and report spending under the same rules, so these errors largely difference out.

Noise is large relative to inefficiency in both models (σ_v of 0.24 and 0.22 against mean inefficiency of 0.12 and 0.09 SD). As a result the frontier runs through the upper part of the district distribution rather than enveloping it (Figure 2), and district TE estimates are shrunk toward the mean. Three-year TE is concentrated between 0.80 and 0.93 in the baseline, with a long lower tail (Figure 1).

Figure 1. Distribution of three-year technical efficiency, baseline model, 2011–2024

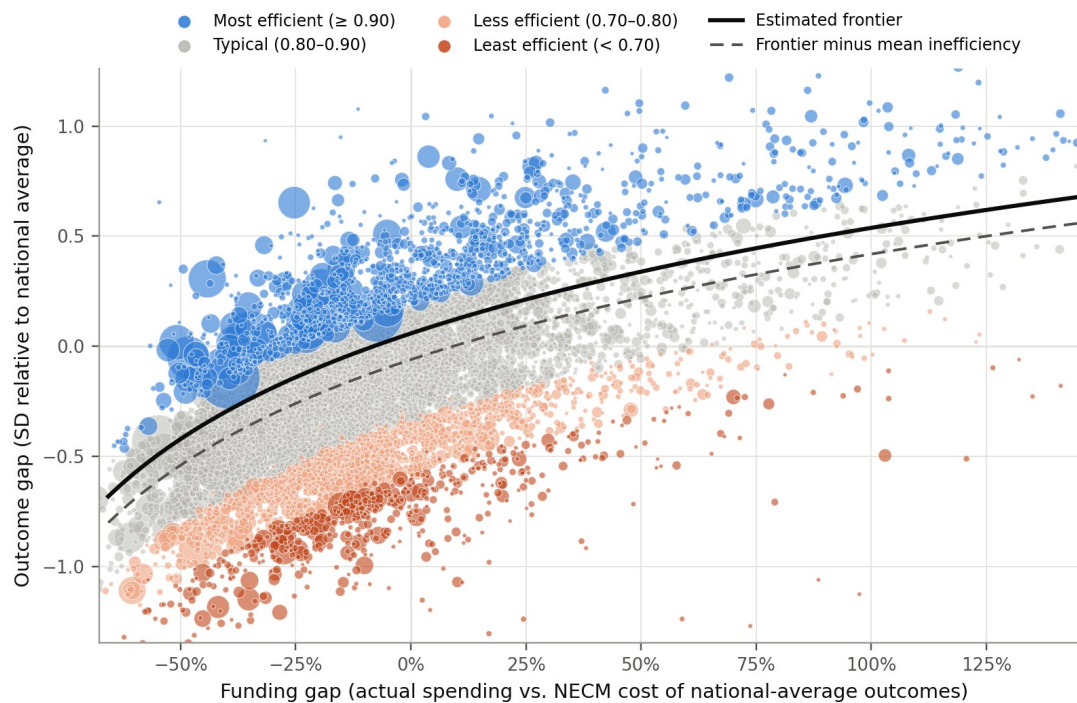


Dashed lines mark the band cut points. Values below 0.40 shown at 0.40.

Figure 2 plots district outcome gaps against funding gaps for 2022–2024. The frontier rises steeply among districts well below their cost targets and flattens among those well above them. Most efficient districts appear at every level of funding, including many that are substantially

underfunded relative to cost. Least efficient districts cluster near or somewhat below their cost targets, where outcomes fall well short of what similar resources yield elsewhere.

Figure 2. Outcome frontier and technical efficiency, 2022–2024 averages



7,891 districts with data in all three years; baseline model. Bubble area proportional to enrollment; colors show three-year TE as of 2024. Axes trimmed to show 99 percent of districts. Points above the frontier reflect positive noise, not efficiency above 1.

5.2 State concentration and measurement error

In the baseline model, efficiency is strongly concentrated by state (Table 4). Idaho and Utah have the largest shares of most efficient district-years (28 and 33 percent). Nearly three-quarters of West Virginia’s district-years and nearly half of Rhode Island’s and Connecticut’s fall below 0.80. California and New Jersey have sizable shares in both tails.

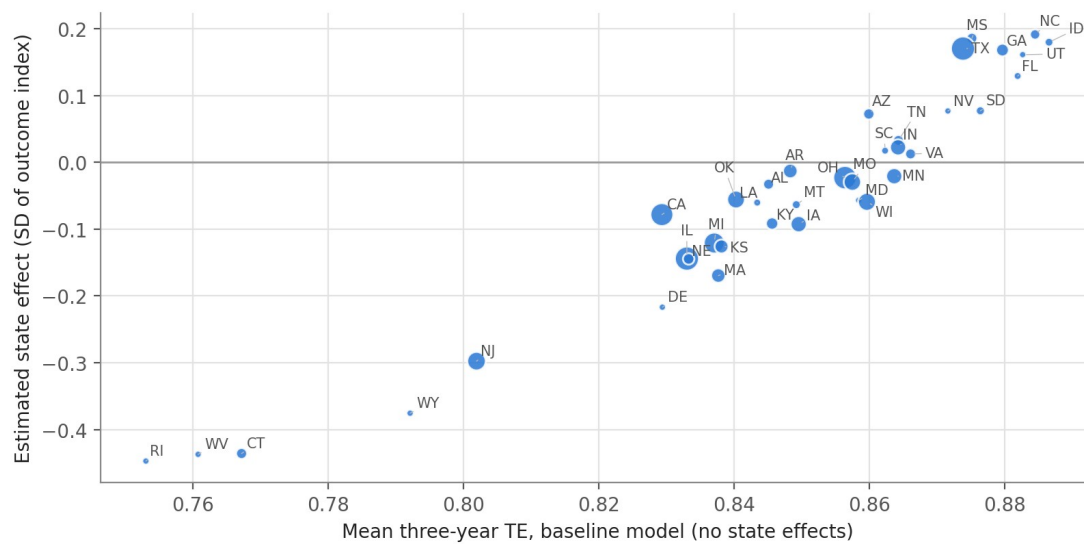
Table 4. States with the highest and lowest average technical efficiency, baseline model, 2011–2024

State	Mean TE	% ≥ 0.90	% < 0.80	State effect, FE model (SD)
Idaho	0.887	27.6	4.7	+0.18
North Carolina	0.884	13.2	3.9	+0.19
Utah	0.883	33.5	2.2	+0.16
Florida	0.882	11.6	3.6	+0.13
Georgia	0.880	12.1	4.7	+0.17
South Dakota	0.876	17.6	3.4	+0.08
California	0.829	16.7	28.3	−0.08
New Jersey	0.802	11.6	29.0	−0.30
Wyoming	0.792	1.4	39.1	−0.38
Connecticut	0.767	2.0	46.1	−0.44
West Virginia	0.761	0.0	73.7	−0.44
Rhode Island	0.753	4.8	48.4	−0.45

Mean TE enrollment-weighted; percentages are shares of district-years. State effect is the intercept from the state fixed effects frontier, centered on the enrollment-weighted mean. Delaware (13 districts) omitted for small numbers. All states in Appendix Tables A1 and A3.

The state fixed effects frontier absorbs almost all of this ranking. State intercepts range from +0.19 SD (North Carolina and Mississippi) to −0.45 SD (Rhode Island) and track baseline state mean TE almost exactly ($r = 0.96$; Figure 3). The share of variance in three-year TE between states falls from 19.5 to 1.1 percent.

Figure 3. State frontier effects and baseline state efficiency, 2011–2024



State intercepts from the state fixed effects frontier, centered on the enrollment-weighted mean across states. Marker size reflects number of districts; $r = 0.96$.

The model cannot say whether a state intercept reflects error in the outcome index, error in spending or cost, or genuine state-level efficiency in governance, policy or management. The baseline treats all three as district inefficiency; the state fixed effects model treats none of them that way. The two models bracket the truth. Given the known limits of cross-state test linking and of reported spending, and the steeper within-state slope, I regard much of the state-level variation as measurement error, and the state fixed effects estimates as the better guide to district efficiency.

Outside the states with large intercepts, district rankings are stable. The rank correlation between the two models' three-year TE is 0.88, 83 percent of district-years fall in the same band, and no district-year moves between the most and least efficient bands (Table 5). Movement concentrates in states with large state effects. The share of district-years below the typical band falls from 74 to 9 percent in West Virginia, 48 to 7 percent in Rhode Island and 46 to 15 percent in Connecticut. It rises from 13 to 33 percent in Texas, 4 to 19 percent in North Carolina and 5 to

21 percent in Georgia, states whose districts sit high on a common frontier but vary widely around their own.

Table 5. Band assignments with and without state fixed effects (percent of baseline band)

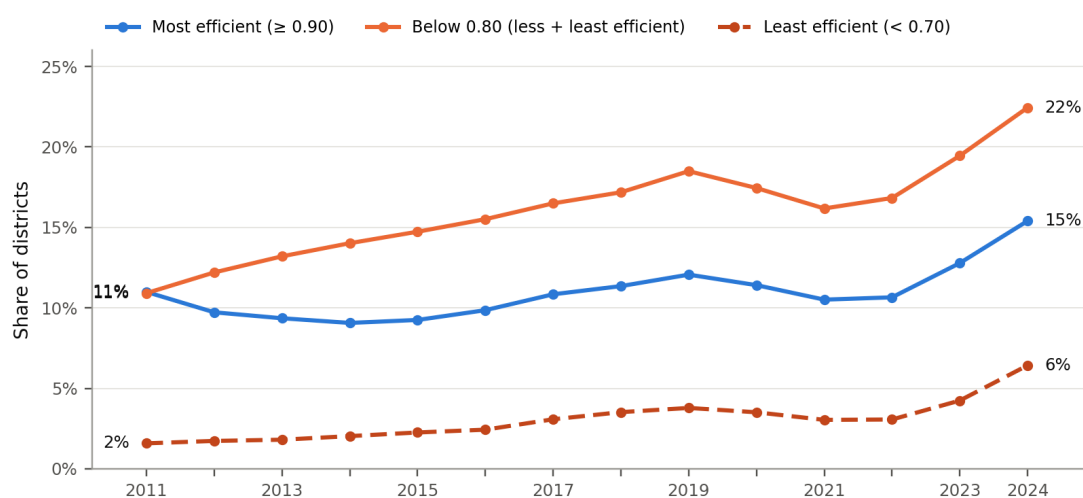
Baseline band	Most (FE)	Typical (FE)	Less (FE)	Least (FE)
Most efficient	72.5	27.5	0.0	0.0
Typical	4.1	89.5	6.3	0.0
Less efficient	0.0	34.4	57.2	8.4
Least efficient	0.0	4.9	32.1	63.0

Rows sum to 100. District-years with three-year TE, 2011–2024 (N = 112,215). State fixed effects bands hold the same shares as baseline bands.

5.3 Growing dispersion

Because TE is measured on a common scale across years, it can reveal changes in the spread of efficiency that within-year standardization would mask. Average enrollment-weighted TE barely changed (0.86 in 2011; 0.85 in 2024), but both tails grew (Figure 4). In the baseline model the share of district-years below 0.80 rose from 11 percent in 2011 to 18.5 percent in 2019, eased slightly in 2020–2022, and reached 22 percent in 2024. The share below 0.70 rose from under 2 percent to more than 6 percent. The share at or above 0.90 ranged between about 9 and 12 percent before 2022 (highest in 2019) and rose to 15 percent in 2024.

Figure 4. Share of districts by efficiency band, baseline model, 2011–2024



Three-year trailing averages; 2024 reflects 2022–2024.

The widening occurs within states. Under state fixed effects, the share of district-years in the two lower bands rises from 11.5 percent in 2011 to 22.3 percent in 2024, almost exactly the baseline change (10.9 to 22.4 percent). Districts with similar resources relative to cost, taking the same assessments and reporting under the same rules, are further apart in outcomes than a decade ago. The post-2022 portion is consistent with the documented post-pandemic divergence in achievement (Kuhfeld, Soland, & Lewis, 2022), though it may also reflect changes in assessments and reporting. The rise from 2011 to 2019 predates the pandemic.

5.4 Staffing and spending profiles

The least efficient districts are higher-cost, not underfunded or overfunded relative to need (Table 6). Under state fixed effects bands, per-pupil spending rises from \$12,682 in the most efficient band to \$16,701 in the least efficient, but predicted cost rises in parallel, from \$13,641 to \$18,664. Median funding gaps are similar across bands (+1 to –8 percent). Instruction falls from 61.8 to 56.6 percent of current spending, general administration rises from 1.62 to 2.09 percent, and total staff from 12.6 to 15.2 per 100 pupils.

Table 6. Profile of districts by efficiency band, state fixed effects model, FY2018–2024 (enrollment-weighted)

	Most efficient	Typical	Less efficient	Least efficient	All
District-years	6,960	38,987	7,928	2,130	56,005
Mean three-year TE	0.946	0.916	0.858	0.772	0.909
Current spending per pupil	\$12,682	\$12,782	\$15,147	\$16,701	\$13,145
NECM predicted cost	\$13,641	\$15,934	\$17,712	\$18,664	\$15,932
Median funding gap	+1%	–8%	–7%	–5%	–7%
Outcome gap (SD)	+0.38	–0.12	–0.49	–0.72	–0.12
Instruction (% of current spending)	61.76	60.31	58.19	56.58	60.15
General administration (%)	1.62	1.73	1.88	2.09	1.75
School administration (%)	5.43	5.49	5.54	5.27	5.48
Operation & maintenance (%)	9.03	9.37	9.60	10.40	9.38
Classroom teachers per 100 pupils	6.13	6.38	6.38	6.74	6.36
Instructional aides per 100	1.53	1.59	1.84	2.06	1.62
School administrators per 100	0.35	0.38	0.41	0.45	0.38
LEA administrators per 100	0.12	0.15	0.16	0.20	0.15
LEA admin. support staff per 100	0.36	0.38	0.48	0.59	0.39
Total staff per 100	12.61	13.23	14.00	15.19	13.29

Bands hold the baseline shares (cut points 0.939, 0.880, 0.816). Spending shares from Census F-33; staffing from CCD FTE, zero counts treated as missing; measures winsorized at the 1st and 99th percentiles. Funding gap = spending ÷ predicted cost – 1. Full category detail and baseline bands in Appendix Table A2.

These national averages mix district and state differences: in the baseline model the least efficient districts are concentrated in New Jersey, Connecticut, Rhode Island, California and Illinois, whose staffing patterns differ from the national average for reasons unrelated to district efficiency. Table 7 and Figure 5 therefore express each measure relative to the district’s own state and year.

Table 7. Staffing and budget allocation in SD from the enrollment-weighted state average, FY2018–2024

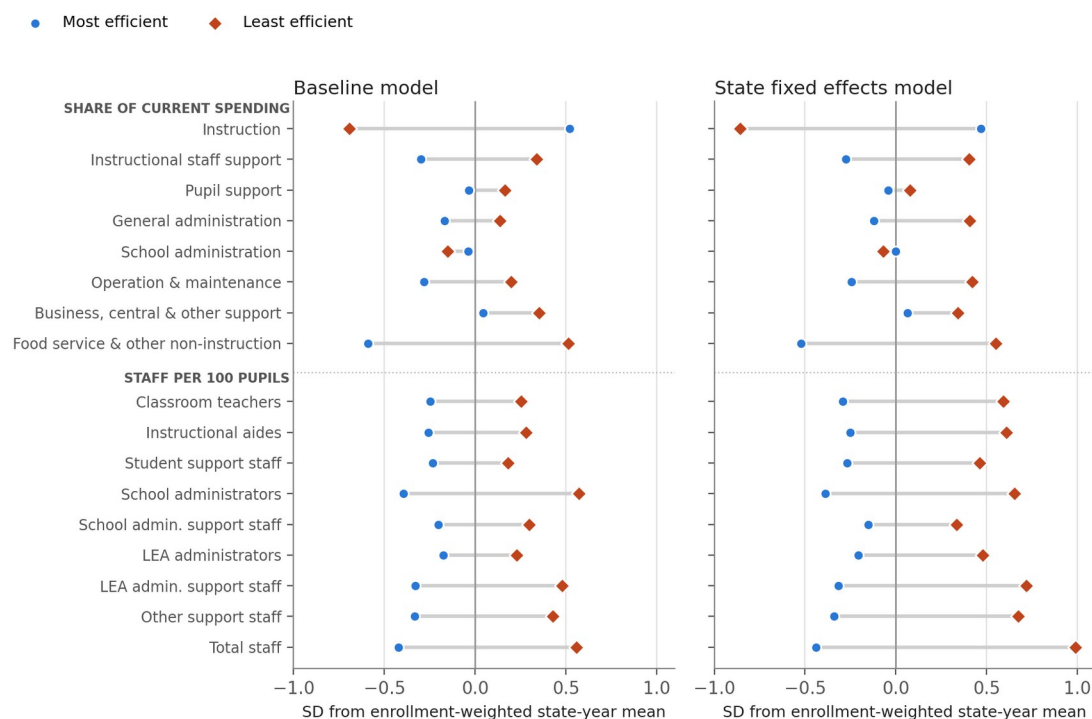
	Baseline: most	Baseline: least	State FE: most	State FE: least
Per-pupil current spending	−0.40	+0.82	−0.50	+1.19
NECM predicted cost	−0.50	+0.82	−0.51	+0.40
Funding gap	+0.39	−0.42	+0.34	+0.20
Outcome gap	+1.25	−1.43	+1.32	−1.42
Enrollment (log)	+0.20	−0.10	+0.09	−0.55
Instruction share	+0.52	−0.69	+0.47	−0.86
General administration share	−0.17	+0.14	−0.12	+0.41
School administration share	−0.04	−0.15	+0.00	−0.07
Operation & maintenance share	−0.28	+0.20	−0.25	+0.42
Business, central & other support share	+0.04	+0.35	+0.06	+0.34
Food service & other non-instruction share	−0.59	+0.51	−0.52	+0.55
Classroom teachers	−0.25	+0.25	−0.29	+0.59
Instructional aides	−0.26	+0.28	−0.25	+0.61
School administrators	−0.40	+0.57	−0.39	+0.65
LEA administrators	−0.18	+0.23	−0.21	+0.48
LEA admin. support staff	−0.33	+0.48	−0.31	+0.72
Other support staff	−0.33	+0.43	−0.34	+0.67
Total staff	−0.42	+0.56	−0.44	+0.99

Enrollment-weighted band means of $z = (x - \mu_{st})/\sigma_{st}$ across 264 state-years. Staffing per 100 pupils. Typical and less efficient bands, and all categories, in Appendix Table A4.

Relative to their own states, the least efficient districts are more heavily staffed in every category. Under state fixed effects bands they are +0.99 SD above their state average in total staff per pupil, including classroom teachers (+0.59), instructional aides (+0.61), school administrators (+0.65) and LEA administrative support staff (+0.72). The most efficient districts are below their state average in every staffing category. Removing state effects from the frontier sharpens the staffing contrast rather than eliminating it (+0.56 SD in total staff under baseline bands).

Budget shares move away from instruction (−0.86 SD), but not toward school administration (−0.07 SD), even though school administrator staffing is higher. The offsetting shares are general administration, operation and maintenance, instructional staff support, business and central services, and food service.

Figure 5. Staffing and budget allocation relative to state averages, most and least efficient districts, FY2018–2024



Enrollment-weighted mean z-scores (Table 7 and Appendix Table A4). Most and least efficient bands are baseline $TE \geq 0.90$ and < 0.70 ; state fixed effects bands are matched to the same shares.

Part of this contrast follows scale. The least efficient districts under state effects are smaller than their state peers (-0.55 SD in log enrollment; median enrollment 1,092 against 1,499 in the most efficient band), spend well above their state average ($+1.19$ SD) and face somewhat higher predicted costs ($+0.40$ SD). Small districts carry more staff per pupil for reasons the NECM treats as cost. What the comparison adds is that the least efficient districts do not have a thinner classroom workforce; they have more of everything, with the largest excess in central office and support staff.

In dollar terms the administrative difference is small. The general administration share in the least efficient band is 0.47 percentage points above the most efficient band, roughly \$80 per pupil. By the frontier estimate, \$80 to \$100 per pupil would be associated with an outcome

difference well under 0.01 SD, against an outcome gap of more than 1 SD between the two bands.

6. Robustness

Functional form and distributional assumptions. Table 8 compares the preferred baseline specification with a half-normal inefficiency distribution and with a squared log term. The squared term is small (0.026) and leaves the slope at the mean almost unchanged. The exponential distribution fits better than the half-normal. The two distributions imply different average inefficiency (0.119 versus 0.186 SD) and thus different TE levels, but identical district rankings. Cut points should therefore be read relative to the exponential specification; band membership, which is defined by distributional breaks and shares, is unaffected.

Table 8. Alternative frontier specifications, baseline model

Frontier	Inefficiency	β , $\ln(\text{spend}/\text{cost})$	β , $\ln^2$	σ_v	$E[u]$	–Pseudo-LL
Log	Exponential (preferred)	0.691		0.239	0.119	12,511
Log	Half-normal	0.691		0.227	0.186	12,621
Log + $\log^2$	Exponential	0.695	0.026	0.240	0.119	12,486
Log + $\log^2$	Half-normal	0.694	0.026	0.228	0.185	12,597

Year fixed effects; $N = 128,713$; enrollment-weighted pseudo-likelihood. District-clustered robust SE for the preferred specification in Table 3.

Comparison with a residual-based ranking. A simpler approach flags districts by the standardized residuals from an OLS regression of the outcome gap on log spending relative to cost. With a constant inefficiency distribution, frontier inefficiency is a monotone function of the composed residual, so the two approaches should rank districts almost identically. They do: the Spearman correlation between three-year TE and three-year standardized residuals is 0.997, 96 percent of district-years fall in the same band, and none moves between the top and bottom bands. The frontier’s value lies in two features the residual approach lacks: an interpretable

efficiency scale in resource terms, and a common metric across years. Within-year standardization of residuals would mask, by construction, the growth in dispersion documented in Section 5.3.

Identification under state effects. Within states the OLS residual distribution is close to symmetric (skewness +0.04), so inefficiency is identified by a lower tail that is longer than the upper tail. District TE under state effects therefore depends more heavily on the exponential assumption than in the baseline. Two features of the results limit this concern. Rankings are highly correlated across models (0.88), with no movement between extreme bands, and the time trend in the lower bands is nearly identical under both. Bootstrap intervals for district TE and a half-normal state fixed effects model are left for a future version.

Persistence. Inefficiency is treated as independent across years. Three-year averaging reduces the influence of transient noise, but a panel frontier separating persistent from transient inefficiency (Colombi, Kumbhakar, Martini, & Vittadini, 2014) would allow districts persistently in each band to be identified directly; I leave this extension to future work.

7. Discussion

National efficiency rankings are mostly state rankings, and state rankings are mostly measurement. One-fifth of the variance in pooled efficiency lies between states, and state frontier intercepts reproduce the pooled state ranking almost exactly. Some state differences may reflect real differences in policy or management. But the within-state frontier is steeper, as attenuation from state-level error would predict, and the states at the bottom of the pooled ranking are those where cross-state measurement problems are plausibly largest. Any analysis that ranks districts or states nationally on spending and linked test scores should be read with this

in mind. Claims that districts in particular states are wasteful, based on national comparisons, may say more about how those states test and account than about how their districts operate.

Resources relative to cost matter more than pooled estimates suggest. The within-state frontier implies that a 10 percent increase in resources relative to cost raises attainable outcomes by about 0.09 SD. This is a frontier relationship, not a causal effect of spending on the average district, and it is not directly comparable to quasi-experimental estimates, which measure the effects of spending increases sustained over several years of exposure (Lafortune, Rothstein, & Schanzenbach, 2018; Jackson & Mackevicius, 2024). It is nonetheless of the same order of magnitude as those estimates. It also implies that pooled national estimates of the spending–outcome relationship understate it.

Inefficiency is real and growing, but it is not where it is usually looked for. The share of district-years in the lowest two efficiency bands doubled between 2011 and 2024, within states, and the rise began well before the pandemic. The least efficient districts are not those with more money than they need: their funding gaps are similar to those of other districts. They are smaller, higher-cost districts with more staff of every kind. Administrative staffing and general administration spending are higher, but the differences are small in budget terms, around \$80 to \$100 per pupil, and by the frontier estimate could account for only a trivial share of the outcome gap. Policies that target administrative spending as a route to closing outcome gaps would therefore address a small part of the difference.

What efficiency estimates can and cannot do. Inefficiency here is a residual: any shortfall from the frontier not explained by resources relative to NECM cost, including unmeasured needs and cost factors. Estimates for individual districts are imprecise because noise is large relative to inefficiency, so band assignments are more reliable than point estimates. The staffing

comparisons are descriptive and not cost-adjusted within states; part of the excess staffing in small districts is a scale cost that the NECM already prices. The outcome measure is narrow, limited to grade 6–8 reading and mathematics. And excluding eleven states with carried-forward outcome data, including New York and Pennsylvania, improves measurement at the cost of national coverage. These limits argue for using frontier estimates to identify districts for closer study, not to set funding or impose sanctions.

8. Conclusion

This paper estimates the technical efficiency of U.S. school districts in producing middle-grades achievement, using a stochastic frontier built on the National Education Cost Model and an input-oriented efficiency measure suited to outcomes in standard-deviation units. Three results stand out. First, most of the between-state variation in national efficiency estimates is absorbed by state frontier intercepts, and the within-state frontier is steeper, consistent with cross-state measurement error in linked assessments and reported spending. Second, the dispersion of efficiency within states has roughly doubled since 2011. Third, the least efficient districts are smaller and higher-cost, with more staff of every kind relative to their state peers; their excess administrative spending is real but small.

The broader lesson is methodological as much as substantive. National efficiency comparisons are only as good as the comparability of the outcome and spending measures that feed them. Estimating frontiers within states, and comparing districts with their state peers, removes the largest source of spurious variation while preserving the district rankings that matter for identifying where resources are and are not translating into outcomes. Natural extensions include panel frontiers that separate persistent from transient inefficiency, models of inefficiency as a

function of district characteristics such as size, sparsity and governance, and staffing comparisons with districts of similar cost and scale within the same state.

Declarations

Data availability. The NECM district panel and School Finance Indicators Database (SFID) data are publicly available at <https://www.schoolfinancedata.org>. SEDA data are available from the Stanford Education Data Archive; CCD and F-33 data from NCES and the U.S. Census Bureau. Replication code is available from the author on request.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest. The author declares no competing interests.

Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work the author used Claude (Anthropic) to draft the manuscript text from the author's technical report. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

- Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), 21–37.
- Baker, B. D. (2025a). *Framework for evaluating & reforming education finance systems* (EdWorkingPaper No. 25-1127). Annenberg Institute at Brown University.
<https://doi.org/10.26300/pa0r-n548>
- Baker, B. D. (2025b). How and why racial isolation affects education costs and the provision of equal educational opportunity. *Education Policy Analysis Archives*, 33(55).
<https://doi.org/10.14507/epaa.33.8843>
- Baker, B. D., Di Carlo, M., & Weber, M. (2024). *School Finance Indicators Database and National Education Cost Model* [Data set]. Albert Shanker Institute.
<https://www.schoolfinancedata.org>
- Baker, B. D., Weber, M., & Srikanth, A. (2021). Informing federal school finance policy with empirical evidence. *Journal of Education Finance*, 47(1), 1–25.
<https://doi.org/10.1353/jef.2021.a806389>
- Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, 38(3), 387–399.
- Chakraborty, K., Biswas, B., & Lewis, W. C. (2001). Measurement of technical efficiency in public education: A stochastic and nonstochastic production function approach. *Southern Economic Journal*, 67(4), 889–905.

- Colombi, R., Kumbhakar, S. C., Martini, G., & Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis*, 42(2), 123–136.
- Costrell, R., Hanushek, E., & Loeb, S. (2008). What do cost functions tell us about the cost of an adequate education? *Peabody Journal of Education*, 83(2), 198–223.
- De Witte, K., & López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, 68(4), 339–363.
- Duncombe, W., Ruggiero, J., & Yinger, J. (1996). Alternative approaches to measuring the cost of education. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 327–356). Brookings Institution.
- Duncombe, W., & Yinger, J. (2005). How much more does a disadvantaged student cost? *Economics of Education Review*, 24(5), 513–532.
- Duncombe, W., & Yinger, J. (2011). Are education cost functions ready for prime time? An examination of their validity and reliability. *Peabody Journal of Education*, 86(1), 28–57.
- Fahle, E. M., saliba, j., Kalogrides, D., Shear, B. R., Reardon, S. F., & Ho, A. D. (2024). *Stanford Education Data Archive: Technical documentation* (Version 5.0).
<https://purl.stanford.edu/cs829jn7849>
- Geverdt, D. (2018). *Education Demographic and Geographic Estimates Program (EDGE): School neighborhood poverty estimates—Documentation* (NCES 2018-027). U.S. Department of Education, National Center for Education Statistics.

- Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2), 269–303.
- Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31(1), 93–118.
- Gronberg, T. J., Jansen, D. W., Karakaplan, M. U., & Taylor, L. L. (2015). School district consolidation: Market concentration and the scale-efficiency tradeoff. *Southern Economic Journal*, 82(2), 580–597.
- Gronberg, T. J., Jansen, D. W., & Taylor, L. L. (2011). The adequacy of educational cost functions: Lessons from Texas. *Peabody Journal of Education*, 86(1), 3–27.
- Grosskopf, S., Hayes, K. J., Taylor, L. L., & Weber, W. L. (1997). Budget-constrained frontier measures of fiscal equality and efficiency in schooling. *Review of Economics and Statistics*, 79(1), 116–124.
- Grosskopf, S., Hayes, K. J., Taylor, L. L., & Weber, W. L. (2001). On the determinants of school district efficiency: Competition and monitoring. *Journal of Urban Economics*, 49(3), 453–478.
- Imazeki, J., & Reschovsky, A. (2004). Is No Child Left Behind an un- (or under-) funded federal mandate? Evidence from Texas. *National Tax Journal*, 57(3), 571–588.
- Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *Quarterly Journal of Economics*, 131(1), 157–218.

- Jackson, C. K., & Mackevicius, C. L. (2024). What impacts can we expect from school spending policy? Evidence from evaluations in the United States. *American Economic Journal: Applied Economics*, 16(1), 412–446.
- Jondrow, J., Lovell, C. A. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2–3), 233–238.
- Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher*, 51(7), 500–506.
- Kumbhakar, S. C., & Tsionas, E. G. (2006). Estimation of stochastic frontier production functions with input-oriented technical efficiency. *Journal of Econometrics*, 133(1), 71–96.
- Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, 10(2), 1–26.
- Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2), 435–444.
- National Center for Education Statistics. (n.d.). *Common Core of Data: Local education agency universe and staff data, 2017–18 through 2023–24* [Data set]. U.S. Department of Education. <https://nces.ed.gov/ccd/>

- Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., & saliba, j. (2024). *Stanford Education Data Archive* (Version 5.0) [Data set].
<https://purl.stanford.edu/cs829jn7849>
- Reardon, S. F., Kalogrides, D., & Ho, A. D. (2021). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Journal of Educational and Behavioral Statistics*, 46(2), 138–167.
- Ruggiero, J. (1996). Efficiency of educational production: An analysis of New York school districts. *Review of Economics and Statistics*, 78(3), 499–509.
- Simar, L., & Wilson, P. W. (2010). Inferences from cross-sectional, stochastic frontier models. *Econometric Reviews*, 29(1), 62–98.
- Taylor, L. L., & Fowler, W. J., Jr. (2006). *A comparable wage approach to geographic cost adjustment* (NCES 2006-321). U.S. Department of Education, National Center for Education Statistics.
- U.S. Census Bureau. (n.d.). *Annual Survey of School System Finances, fiscal years 2009–2024* [Data set]. <https://www.census.gov/programs-surveys/school-finances.html>
- Waldman, D. M. (1982). A stationary point for the stochastic frontier likelihood. *Journal of Econometrics*, 18(2), 275–279.

Appendix A. Supplementary tables

Table A1. Technical efficiency by state, 2011–2024 (ordered by mean TE)

State	Districts	Mean TE (enroll.-wt.)	% Most efficient	% Typical	% Less efficient	% Least efficient
ID	82	0.887	27.6	67.7	2.3	2.4
NC	115	0.884	13.2	83.0	3.6	0.2
UT	32	0.883	33.5	64.3	2.2	0.0
FL	66	0.882	11.6	84.8	3.3	0.3
GA	177	0.880	12.1	83.2	4.4	0.3
SD	87	0.876	17.6	79.0	2.8	0.7
MS	128	0.875	15.1	81.4	3.5	0.0
TX	831	0.874	12.8	74.2	10.5	2.5
NV	11	0.872	0.0	75.3	16.2	8.4
VA	129	0.866	2.2	87.3	9.9	0.7
TN	127	0.864	13.2	83.2	3.6	0.1
IN	284	0.864	15.6	77.3	6.3	0.8
MN	283	0.864	10.8	78.4	8.5	2.4
SC	66	0.862	4.7	83.2	9.9	2.2
AZ	137	0.860	13.2	62.5	15.6	8.7
WI	350	0.860	13.1	80.4	5.8	0.7
MD	24	0.858	2.1	80.7	15.5	1.8
MO	331	0.857	10.2	83.4	6.0	0.4
OH	597	0.856	15.3	75.1	8.7	0.9
IA	285	0.850	4.7	87.8	7.2	0.4
MT	85	0.849	17.5	69.7	8.2	4.7
AR	231	0.848	5.7	74.3	17.1	2.9
KY	166	0.846	7.8	83.1	8.5	0.7
AL	129	0.845	8.7	70.4	19.9	1.0
LA	67	0.843	1.2	76.6	18.1	4.2
OK	342	0.840	8.4	74.8	14.4	2.4
KS	212	0.838	8.1	79.9	10.6	1.4
MA	226	0.838	18.3	60.8	17.1	3.9
MI	478	0.837	3.8	75.8	17.6	2.8
NE	145	0.833	6.4	73.4	16.7	3.5
IL	636	0.833	8.1	68.0	19.5	4.3
DE	13	0.829	0.0	72.5	25.3	2.2
CA	575	0.829	16.7	55.0	21.6	6.8
NJ	377	0.802	11.6	59.4	21.7	7.4
WY	42	0.792	1.4	59.5	30.8	8.3
CT	132	0.767	2.0	51.9	29.2	16.9
WV	54	0.761	0.0	26.3	56.0	17.7
RI	31	0.753	4.8	46.8	35.5	12.9

Shares of district-years with three-year TE in each band. Mean TE is enrollment-weighted. Restricted sample; Hawaii is not in the NECM file, and Vermont's single district drops out of the sample.

Table A2. Profile of districts by technical efficiency band, baseline model, FY2018–2024 (enrollment-weighted averages)

	Most efficient	Typical	Less efficient	Least efficient	All
District-years	6,730	39,038	8,034	2,203	56,005
Mean three-year TE	0.913	0.863	0.760	0.637	0.852
Per-pupil current spending	\$11,858	\$12,712	\$15,956	\$18,237	\$13,145
NECM predicted cost	\$13,676	\$15,785	\$18,642	\$20,417	\$15,932
Median funding gap	+2%	–8%	–6%	–4%	–7%
Outcome gap (SD)	+0.32	–0.13	–0.50	–0.75	–0.12
Percent of current spending					
Instruction	62.11	60.11	58.35	57.52	60.15
Pupil support	5.79	5.73	6.62	7.08	5.89
Instructional staff support	4.37	4.85	5.02	4.76	4.79
General administration	1.44	1.79	1.86	1.92	1.75
School administration	5.49	5.50	5.44	5.06	5.48
Operation & maintenance	9.00	9.45	9.44	9.52	9.38
Student transportation	3.62	3.91	3.58	3.66	3.81
Business, central & other support	4.36	4.13	5.21	5.89	4.36
Food service & other non-instruction	3.77	4.44	4.35	4.37	4.32
Staff per 100 pupils					
Classroom teachers	6.13	6.40	6.40	6.59	6.36
Instructional aides	1.43	1.60	1.87	2.05	1.62
Guidance counselors	0.23	0.24	0.25	0.24	0.24
Librarians/media specialists	0.09	0.11	0.09	0.10	0.10
Student support staff	0.74	0.80	0.89	0.94	0.81
School administrators	0.35	0.39	0.42	0.42	0.38
School admin. support staff	0.51	0.54	0.60	0.62	0.55
LEA administrators	0.12	0.15	0.17	0.19	0.15
LEA admin. support staff	0.33	0.38	0.50	0.60	0.39
Other support staff	2.47	2.57	2.68	2.80	2.58
Total staff	12.54	13.28	13.98	14.72	13.29

Baseline bands (Table 2 cut points). Same data and definitions as Table 6.

Table A3. State effects and band shares with and without state fixed effects, 2011–2024 (ordered by state effect)

State	State effect (SD)	Mean TE, state FE	% most efficient: baseline	% most efficient: state FE	% below typical: baseline	% below typical: state FE
NC	+0.19	0.914	13.2	1.0	3.9	18.9
MS	+0.19	0.912	15.1	8.0	3.5	12.6
ID	+0.18	0.913	27.6	11.6	4.7	13.6
TX	+0.17	0.908	12.8	4.4	12.9	33.2
GA	+0.17	0.913	12.1	3.8	4.7	21.3
UT	+0.16	0.912	33.5	11.8	2.2	7.6
FL	+0.13	0.917	11.6	2.3	3.6	13.4
SD	+0.08	0.912	17.6	7.1	3.4	11.9

State	State effect (SD)	Mean TE, state FE	% most efficient: baseline	% most efficient: state FE	% below typical: baseline	% below typical: state FE
NV	+0.08	0.918	0.0	0.0	24.7	50.0
AZ	+0.07	0.908	13.2	10.4	24.4	29.1
TN	+0.03	0.913	13.2	11.3	3.7	8.0
IN	+0.02	0.911	15.6	10.2	7.1	12.6
SC	+0.02	0.915	4.7	4.8	12.1	11.8
VA	+0.01	0.913	2.2	1.7	10.5	16.2
AR	−0.01	0.911	5.7	6.6	20.0	18.5
MN	−0.02	0.911	10.8	6.9	10.9	13.6
OH	−0.02	0.910	15.3	10.6	9.6	13.8
MO	−0.03	0.913	10.2	14.1	6.4	7.0
AL	−0.03	0.914	8.7	11.4	20.9	11.4
OK	−0.06	0.910	8.4	14.5	16.8	13.5
MD	−0.06	0.915	2.1	0.9	17.3	20.8
WI	−0.06	0.912	13.1	11.3	6.5	9.6
LA	−0.06	0.913	1.2	4.8	22.2	15.4
MT	−0.06	0.903	17.5	18.7	12.9	14.2
CA	−0.08	0.907	16.7	20.3	28.3	21.0
KY	−0.09	0.912	7.8	15.0	9.1	6.6
IA	−0.09	0.913	4.7	7.7	7.6	6.7
MI	−0.12	0.912	3.8	6.6	20.4	11.3
KS	−0.13	0.911	8.1	14.1	12.0	8.5
IL	−0.14	0.909	8.1	13.6	23.9	17.5
NE	−0.14	0.909	6.4	11.9	20.2	16.4
MA	−0.17	0.908	18.3	17.1	21.0	20.1
DE	−0.22	0.915	0.0	2.7	27.5	6.6
NJ	−0.30	0.905	11.6	20.2	29.0	13.1
WY	−0.38	0.906	1.4	18.4	39.1	11.7
CT	−0.44	0.908	2.0	14.3	46.1	15.0
WV	−0.44	0.915	0.0	5.4	73.7	8.9
RI	−0.45	0.911	4.8	21.7	48.4	7.1

State effect: intercept from the state fixed effects frontier, centered on the enrollment-weighted mean, in SD of the outcome index. Mean TE is enrollment-weighted. “Below typical” = less or least efficient bands. State fixed effects bands hold the same national shares as the baseline bands.

Table A4. Staffing and budget allocation in standard deviations from the enrollment-weighted state average, all bands, FY2018–2024

	Baseline bands				State fixed effects bands			
	Most	Typ.	Less	Least	Most	Typ.	Less	Least
Context								
Per-pupil current spending	−0.40	−0.05	+0.58	+0.82	−0.50	−0.06	+0.66	+1.19
NECM predicted cost	−0.50	−0.02	+0.58	+0.82	−0.51	+0.01	+0.38	+0.40
Funding gap	+0.39	−0.02	−0.33	−0.42	+0.34	−0.05	−0.06	+0.20
Outcome gap	+1.25	−0.06	−0.97	−1.43	+1.32	−0.02	−0.95	−1.42
Enrollment (log)	+0.20	−0.06	+0.10	−0.10	+0.09	+0.02	−0.11	−0.55

	Baseline bands				State fixed effects bands			
Percent of current spending								
Instruction	+0.52	-0.00	-0.52	-0.69	+0.47	+0.04	-0.57	-0.86
Pupil support	-0.03	-0.02	+0.13	+0.16	-0.04	-0.02	+0.16	+0.08
Instructional staff support	-0.30	-0.00	+0.32	+0.34	-0.28	-0.02	+0.33	+0.40
General administration	-0.17	+0.02	+0.07	+0.14	-0.12	-0.02	+0.17	+0.41
School administration	-0.04	+0.00	+0.06	-0.15	+0.00	-0.00	+0.02	-0.07
Operation & maintenance	-0.28	+0.04	+0.13	+0.20	-0.25	-0.01	+0.20	+0.42
Student transportation	-0.05	-0.01	+0.06	+0.17	-0.09	+0.00	+0.06	-0.01
Business, central & other support	+0.04	-0.05	+0.15	+0.35	+0.06	-0.05	+0.15	+0.34
Food service & other non-instruction	-0.59	+0.06	+0.32	+0.51	-0.52	+0.02	+0.31	+0.55
Staff per 100 pupils								
Classroom teachers	-0.25	-0.00	+0.29	+0.25	-0.29	-0.04	+0.41	+0.59
Instructional aides	-0.26	+0.01	+0.21	+0.28	-0.25	-0.03	+0.36	+0.61
Guidance counselors	-0.08	-0.03	+0.27	+0.08	-0.10	-0.04	+0.31	+0.20
Librarians/media specialists	-0.13	+0.04	-0.03	-0.07	-0.06	-0.00	+0.08	+0.07
Student support staff	-0.23	+0.00	+0.27	+0.18	-0.27	-0.02	+0.32	+0.46
School administrators	-0.40	-0.01	+0.44	+0.57	-0.39	-0.02	+0.42	+0.65
School admin. support staff	-0.20	+0.01	+0.16	+0.30	-0.15	-0.01	+0.18	+0.33
LEA administrators	-0.18	+0.00	+0.18	+0.23	-0.21	-0.03	+0.28	+0.48
LEA admin. support staff	-0.33	+0.01	+0.32	+0.48	-0.31	-0.02	+0.31	+0.72
Other support staff	-0.33	+0.02	+0.23	+0.43	-0.34	-0.01	+0.27	+0.67
Total staff	-0.42	+0.00	+0.40	+0.56	-0.44	-0.04	+0.54	+0.99

Entries are enrollment-weighted band means of $z = (x - \mu_{st}) / \sigma_{st}$, with μ and σ the enrollment-weighted mean and SD across linked districts in the same state and fiscal year (264 state-years). Typ. = typical. Baseline bands use the Table 2 cut points; state fixed effects bands hold the same shares.

Variable definitions

- Outcome gap: NECM outcome index, SD units relative to the national average; based only on SEDA nationally normed state reading and mathematics assessments, grades 6–8.
- Funding gap: actual current spending per pupil ÷ NECM predicted cost of national-average outcomes (cost model 1; Appendix B) – 1.
- Technical efficiency: $E[\exp(-u/\beta) \mid \varepsilon]$ from the exponential stochastic frontier; three-year trailing average.
- Staffing ratios: CCD full-time-equivalent staff per 100 enrolled pupils in the ten categories listed in Table A2.
- Spending shares: F-33 current spending in each of the nine functional categories listed in Table A2, as a percentage of total current elementary–secondary spending.

Appendix B. The National Education Cost Model

The NECM is a district-level education cost function estimated by two-stage least squares on a national panel. Predicted cost is the per-pupil current spending the model implies for reaching national-average outcomes, given each district's student needs, grade composition, scale, population density and labor costs (Baker, Weber, & Srikanth, 2021; Baker, 2025b). The specification follows the cost function tradition of Duncombe and Yinger (2005, 2011); its theoretical basis is set out in Baker (2025a). Estimates below are from the most recent published version (Baker, 2025b, Table 1).

B.1 Data

- **Spending:** current spending per pupil from the School Finance Indicators Database (Baker, Di Carlo, & Weber, 2024), built from the Census Annual Survey of School System Finances (F-33).
- **Outcomes:** the SEDA v5.0 overall outcome index, nationally normed and mean-centered, for reading and mathematics in grades 3–8 (Reardon et al., 2024; Fahle et al., 2024).
- **Student needs and context:** NCES EDGE data, the Census child poverty rate, and the NCES Neighborhood Poverty Index (Geverdt, 2018).
- **Labor costs:** an extended version of the Education Comparable Wage Index (Taylor & Fowler, 2006).
- **Sample:** 147,186 district-year observations in the published estimates.

B.2 Specification

The cost function models log current spending per pupil E for district i in year t as:

$$\ln E_{it} = \theta_0 + \theta_1 O_{it} + \theta_2 W_{it} + N_{it}' \delta + G_{it}' \lambda + S_{it}' \kappa + \theta_3 \ln D_{it} + Z_{it}' \phi + \tau t + e_{it}$$

The terms are defined as follows:

- **O** is the outcome index, in SD units relative to the national mean.
- **W** is the labor cost index (Education Comparable Wage Index).
- **N** is student need: the child poverty rate, the disability rate (centered on the state mean), the share of English learners, the share of Black enrollment, and Black share $\times$ population density.
- **G** is grade composition: the shares of enrollment in pre-kindergarten and in grades 9–12.
- **S** is scale: indicators for enrollment under 100, 101–300, 301–600, 601–1,200, 1,201–1,500 and 1,501–2,000, with districts over 2,000 pupils as the reference group.
- **D** is population per square mile.
- **Z** is efficiency controls. The share of the population aged 5–17 captures public monitoring, the housing value ratio captures fiscal capacity, and a Herfindahl index captures competition.
- **t** is a linear time trend.

Outcomes are treated as endogenous, because districts choose spending and outcomes jointly. O is instrumented with the demographic characteristics of surrounding districts. When the Black share enters the second stage, the percent Hispanic of surrounding districts serves as an instrument; when it does not, the percent Black or Hispanic of surrounding districts does.

Standard errors are robust.

Specification used here. This paper uses cost model 1 of the 2024 NECM release, the race-sensitive Black $\times$ population density specification that Baker (2025b) identifies as the preferred model.

B.3 Estimates

Table B1. NECM cost function estimates (dependent variable: ln current spending per pupil)

	Black $\times$ density (preferred)	Race-neutral
Outcome index (instrumented)	1.783*** (0.106)	1.956*** (0.127)
Labor cost index	0.499*** (0.034)	0.573*** (0.036)
Child poverty rate	2.219*** (0.155)	3.312*** (0.227)
Disability rate (state-centered)	2.929*** (0.141)	2.957*** (0.157)
% English learners	1.846*** (0.126)	1.700*** (0.133)
% Black	0.399*** (0.069)	
% Black $\times$ population density	0.115*** (0.012)	
% Pre-K	-0.007 (0.140)	0.173 (0.153)
% Grades 9–12	0.491*** (0.040)	0.499*** (0.044)
Enrollment < 100	0.557*** (0.079)	0.545*** (0.082)
Enrollment 101–300	0.371*** (0.021)	0.367*** (0.022)
Enrollment 301–600	0.213*** (0.016)	0.199*** (0.016)
Enrollment 601–1,200	0.140*** (0.012)	0.119*** (0.013)
Enrollment 1,201–1,500	0.112*** (0.014)	0.090*** (0.015)
Enrollment 1,501–2,000	0.104*** (0.012)	0.083*** (0.013)
Population density (ln)	-0.070*** (0.007)	-0.036*** (0.006)
% Population aged 5–17	-0.975*** (0.160)	-0.444*** (0.122)
Housing value ratio	-0.329*** (0.028)	-0.409*** (0.034)
Herfindahl index	-0.161 (0.462)	0.363 (0.506)
Time trend	0.015*** (0.001)	0.016*** (0.001)
Constant	8.558*** (0.063)	8.233*** (0.063)
N	147,186	147,186
Partial F, excluded instruments	191.96	160.12
Hansen J, p-value	0.795	0.339

Source: Baker (2025b), Table 1. Robust standard errors in parentheses. *** $p < 0.01$. The race-neutral model is shown for comparison.

The coefficients have direct cost interpretations. A 1 SD higher outcome target raises predicted cost by about 178 log points. A 10 percentage point increase in child poverty raises cost by about

22 log points. Districts with under 100 pupils face costs about 56 log points higher than districts over 2,000.

B.4 Predicted cost of national-average outcomes

Predicted cost C_{it} is obtained from the second-stage estimates with three adjustments:

- The outcome index is set to the national mean ($O = 0$).
- The efficiency controls Z are held at their national means, so spending explained by fiscal capacity, monitoring or competition does not count as cost.
- Each district's need, grade, scale, density and wage measures are set to their actual values.

The funding gap used in this paper is $S_{it} / C_{it} - 1$. Because the efficiency controls are held at national means, spending that the cost model attributes to them shows up in the funding gap rather than in cost.

B.5 Model selection and validation

Baker (2025b) compares four specifications on instrument diagnostics, out-of-sample prediction (80/20 split), the strength of the funding-gap–outcome-gap relationship, and residual bias by racial composition (Table B2). The Black $\times$ density specification performs best or near-best on every criterion.

Table B2. Model selection diagnostics for alternative NECM specifications

Criterion	Black	Race-neutral	Poverty × density	Black × density
Partial F, excluded instruments	177.54	160.12	157.19	191.96
Hansen J, p-value	0.625	0.339	0.041	0.795
Out-of-sample r with actual spending	0.298	0.271	0.291	0.339
Out-of-sample MAPE	36.2%	38.3%	37.2%	35.1%
Out-of-sample MPE	14.0%	17.6%	18.9%	15.2%
Funding gap × outcome gap R ²	0.450	0.383	0.391	0.453
Residual bias on % Black	0.055***	−0.417***	−0.394***	0.092***

Source: Baker (2025b), Table 1. Out-of-sample tests fit models to a random 80% training sample and predict spending in the remaining 20%.

B.6 Implications for the frontier analysis

Shared outcome measure. The cost model and the frontier use the same family of SEDA outcomes. The cost model is estimated to explain spending, not outcomes, but its validity check is the very relationship the frontier models: funding gaps against outcome gaps. The frontier therefore measures efficiency conditional on the NECM. Shortfalls from causes the cost model does not price, such as unmeasured needs, are attributed to inefficiency. As a consistency check, the inverse of the cost model’s outcome coefficient ($1/1.783 = 0.56$ SD per log unit of spending) is close to the pooled frontier slope of 0.69.

Efficiency controls. Holding fiscal capacity, monitoring and competition at national means removes their association with spending from cost, not from spending. A high-capacity district that spends above its predicted cost is therefore credited with a positive funding gap, and any shortfall in its outcomes counts toward inefficiency in the frontier.

Scale. The cost model prices substantial diseconomies of small scale. Part of the higher staffing in small, least efficient districts (Section 5.4) is therefore already treated as cost. The state-standardized staffing comparisons are not cost-adjusted, which is why they overstate the excess staffing attributable to inefficiency.

No state effects in the cost model. The NECM omits state fixed effects by design, to remain a generalizable national model (Baker, 2025b). State-level differences in spending accounting therefore pass into predicted cost and the funding gap. This is part of the motivation for the state fixed effects frontier in Section 5.2.

Outcome grade range. The cost model in Baker (2025b) is estimated on a grades 3–8 outcome index; the frontier outcome uses grades 6–8.