NBER WORKING PAPER SERIES

EDUCATION REFORM IN GENERAL EQUILIBRIUM: EVIDENCE FROM CALIFORNIA'S CLASS SIZE REDUCTION

Michael Gilraine
Hugh Macartney
Robert McMillan

Working Paper 24191
http://www.nber.org/papers/w24191

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2018

We would like to thank Pat Bayer and Gregorio Caetano for helpful discussions, and workshop participants at the University of Bristol (CPMO) and the University of Toronto for additional comments. Financial support is gratefully acknowledged from the IES, SSHRC, Duke University, and the University of Toronto Mississauga. All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Education Reform in General Equilibrium: Evidence from California's Class Size Reduction
Michael Gilraine, Hugh Macartney, and Robert McMillan
NBER Working Paper No. 24191
January 2018
JEL No. H40,I21,I22

ABSTRACT

This paper sheds new light on general equilibrium responses to major education reforms, focusing on a sorting mechanism likely to operate whenever a reform improves public school quality significantly. It does so in the context of California's statewide class size reduction program of the late-1990s, and makes two main contributions. First, using a transparent differencing strategy that exploits the grade-specific roll-out of the reform, we show evidence of general equilibrium sorting effects: Improvements in public school quality caused marked reductions in local private school shares, consequent changes in public school demographics, and significant increases in local house prices -- the latter indicative of the reform's full impact. Second, using a generalization of the differencing approach, we provide credible estimates of the direct and indirect impacts of the reform on a common scale. These reveal a large pure class size effect of 0.11 SD (in terms of mathematics scores), and an even larger indirect effect of 0.16 SD via induced changes in school demographics. Further, we show that both effects persist positively, giving rise to an overall policy impact estimated to be 0.4 SD higher after four years of treatment (relative to none). The analysis draws attention, more broadly, to conditions under which the indirect sorting effects of major reforms are likely to be first order.

Michael Gilraine
New York University
Department of Economics
19 West 4th Street
New York, NY 10012
mike.gilraine@nyu.edu

Hugh Macartney
Duke University
Department of Economics
239 Social Sciences Building
Box 90097
Durham, NC 27708
and NBER
hugh.macartney@duke.edu

Robert McMillan
University of Toronto
Department of Economics
150 St. George Street
Toronto, ON M5S 3G7
CANADA
and NBER
mcmillan@chass.utoronto.ca

1 Introduction

Empirical policy analysis often focuses on the direct, intended effects of policies, "holding all else equal." While measuring such direct effects accurately is an important ingredient in policy making, it is well appreciated that large-scale reforms may also give rise to indirect general equilibrium effects that work in offsetting or reinforcing ways. Because of their potential to alter the policy-making calculus significantly, estimating the scale of any indirect effects is of considerable interest; yet doing so presents challenges for empirical research, given they constitute additional sources of endogeneity that can be difficult to identify. As a consequence, the literature seeking to gauge indirect general equilibrium effects of policy is relatively undeveloped.[1]

To help fill the gap, this paper aims to measure the indirect effects of large-scale reforms in a credible way and place them alongside the direct benefits on a common scale. We focus on an education context and a type of general equilibrium response that is likely to matter whenever a reform improves public school quality significantly – the basic goal of most education reforms – and where private school options are popular pre-reform.[2] In this common configuration of circumstances, some households will tend to re-sort by switching out of private schools, potentially changing the mix of students in public schools. In turn, to the extent that compositional changes influence education production, so they should affect measured outcomes, though to a degree that has not been well established in prior work.

The reform we analyze – California's class size reduction (CSR) program of the late-1990s – was very large indeed, being the largest state-led education reform ever implemented in the United States up to that point.[3] Inspired by Project STAR in Tennessee – a well-known experimental evaluation in education and the subject of a number of influential studies[4] – the

[1]Notable existing contributions include papers by Jepsen and Rivkin (2009) and Dinerstein and Smith (2016), discussed below. Bianchi (2017) provides a thorough analysis of indirect effects arising from an Italian university reform.

[2]Low public school quality gives rise both to greater demand for private schools *and* greater pressure for public school reform, making such a combination of circumstances common in practice.

[3]This assessment comes from the 1998/99 report of the CSR Research Consortium.

[4]Prominent among these are Krueger (1999), Krueger and Whitmore (2001), and Chetty et al. (2011).

California legislature under the leadership of Governor Wilson sought to replicate Project STAR's publicized experimental benefits at an altogether larger scale. To that end, the CSR program targeted kindergarten through third grade (as did Project STAR), and cut class sizes in these early grades by a substantial amount throughout the state.

The timing of implementation was quite specific, and is useful for identifying policy impacts. Smaller classes in first grade were phased in during the 1996-97 school year, with schools having to hire enough teachers in that grade to lower class sizes below 20 in order to be eligible for CSR funding – a significant amount ($650 per student). In 1997-98, classes in second grade became eligible, and schools could then seek to reduce class sizes in either kindergarten and third grade the following year (and the remaining grade the year after).

Given the combination of the strong financial incentives to implement the reform according to this timetable, the substantial reductions in class size it produced – a 20 percent reduction (on average) in elementary schools across the state – and the sheer scale of California's education system, one would expect CSR to have broader effects. Major impacts on the teacher labor market have already been highlighted in compelling work by Jepsen and Rivkin (2009), who show that there was a sudden, significant increased need for new teacher hires, which dampened CSR's benefits in the short term. Further, the effects of the policy were non-uniform, with some schools experiencing reductions in class size without any discernible reduction in teacher quality, becoming potentially more attractive to parents as a result.

These changes in public school quality make it likely that sociodemographic sorting from private to public schools occurred in response to CSR. Our first contribution is to document significant general equilibrium sorting effects using a transparent differencing approach. Exploiting the unique roll-out of CSR, grade-by-grade, we show that improvements in public school quality caused sizable reductions in local private school share of 1.8 percentage points for relevant elementary grades (equivalent to 15 percent of the pre-CSR K-3 average private school share of 11.7 percent and a fifth of its standard deviation),[5] and pronounced changes in the sociodemographic composition of public schools using difference-in-differences and triple-differences research designs. These results are consistent with an inflow of higher socioeconomic status students into public schools.[6]

Drawing on similar sources of variation, we also estimate the total value of the reform in terms of local house prices, finding that households were willing to pay a significant premium for a house in an area where the policy had been implemented.[7] Specifically, a one standard deviation increase in CSR implementation, captured by an intuitive index that we propose, is associated with approximately a 2.6 percent increase in house prices.[8] Such overall benefits of the program combine reductions in class size, changes in teacher quality, and student re-sorting. We show, further, that the magnitude of these benefits is cut by one quarter once we control for observable measures of teacher quality and particularly school sociodemographics, suggestive of the importance of indirect general equilibrium effects.

Our second contribution is to develop an estimable framework that allows us to measure the relative importance of the policy's direct and indirect benefits on a common scale, using test scores. For this part of the analysis, we parametrize the education production function and focus on test scores (rather than the hedonic function and house prices), given we measure test scores well at the school-grade-year level, and because our test scores are available for given grades and years within-school. This allows us to disentangle relevant direct and indirect channels using our differencing approach – something that cannot be done in a precise way with house prices, given they do not have grade-level granularity.

[5] In the same vein as the first part of the current paper, Estevan (2015) uses an education finance reform in Brazil to highlight reductions in private school enrollment in response to increased public education expenditure.

[6] Researchers studying California's CSR reform face the data limitation that individual students cannot be followed over time (see Section 2). Our estimation approach here makes use of more aggregated data along with the grade-specific timing of the reform.

[7] Several papers have sought to estimate parents' willingness-to-pay for smaller classes based on hedonic estimation strategies, examples including Clark and Herrin (2000) and Rohlfs and Zilora (2014).

[8] To be precise, a one standard deviation increase in the CSR implementation measure (0.15) corresponds to an increase in house prices of $5,145. This increase is 2.6 percent of the 1995 pre-CSR (weighted) mean of $195,000.

In line with influential recent research (see Chetty et al. 2011 and Chetty, Friedman and Rockoff 2014), the framework allows the direct and indirect effects to carry over into the future, persisting at different rates. To the extent that persistence is important, we show that a simple difference-in-differences approach for estimating even the direct effect of the reform will not be valid.

Using a generalization of the differencing approach, we estimate the parameters of this framework in a credible way, showing how each observed grade-year score can be separated into a pure class size effect (comparable to experimental estimates) and the effect arising from the general equilibrium change in student quality (subsuming the effects of a student's own ability and peers). Our main focus is on these two channels, while controlling for induced changes in teacher quality.

The key parameters of interest are identified based on two assumptions. First, we appeal to the plausible assumption of common trends across grades, so that any differences in scores between grades arising independently of CSR are time-invariant. This assumption implies that pre-treatment grade-year combinations provide suitable controls for unobserved differences in post-treatment test scores that are unrelated to CSR; event study-type graphs indicate that parallel trends hold pre-treatment. Second, we assume that teacher effects are determined according to variation in observable teacher characteristics, such as experience and credentials (as in Jepsen and Rivkin 2009).

Our estimates indicate that both direct and indirect effects are significant and positive,[9] with the indirect sorting effect being greater in magnitude (0.16σ in terms of mathematics scores) than the direct effect (0.11σ); lending credence to our approach, the estimate of the direct effect is in keeping with leading estimates in the prior literature.[10] Further, we find

[9]Across a variety of settings, studies in the class size literature find both positive effects on achievement (Finn and Achilles 1990; Krueger 1999; Molnar et al. 1999, Angrist and Lavy 1999, Krueger and Whitmore 2001, Cho, Glewwe and Whitler 2012, Gilraine 2017) and no effects (Hoxby 2000; Dobbelsteen, Levin and Oosterbeek 2002; Asadullah 2005; Gary-Bobo and Mahjoub 2006; Leuven, Oosterbeek and Rønning 2008; Battistin, Angrist and Vuri 2017). Research studying the Californian context has delivered similar mixed results to the broader literature (see Bohrnstedt and Stecher 2002; Stecher, McCaffrey and Bugliari 2003 and Funkhouser 2009 for non-effects, and Unlu 2005 and Jepsen and Rivkin 2009 for positive test score effects).

[10]For instance, our estimate is nearly identical to the experimental estimates from the Krueger and

that both the direct and indirect effects persist into subsequent years, at approximately the same rate in each case. In turn, estimates of the direct effect using a standard difference-in-differences strategy are biased, favoring the multiple-differencing estimation approach we follow. Based on the parameter estimates, our framework indicates an overall policy impact of CSR among fourth grade students who have been treated for the first four years of public schooling to be 0.4σ (relative to receiving no treatment).

Our analysis complements recent research by Dinerstein and Smith (2016), who provide persuasive evidence that increased funding for public schools in New York City drew private school students into the public system, both by choice and through the forced closure of (typically small) private schools. The authors also analyze the achievement effects of higher public school quality, comparing incumbent public school students with students who switch from private to public school.[11] Our school-focused approach allows us to decompose the direct effect of CSR that is analogous to experimental estimates and the indirect general equilibrium effect, which combines the 'own' effect of private-to-public switchers along with their spillover effect on incumbent peers. Here, a bounding exercise indicates that peer spillovers are likely to be a significant fraction of the indirect sorting effect.

The magnitudes of our direct and indirect estimates in the context of CSR support the view that researchers should treat household sorting as a primary factor when assessing the impact of large-scale reforms that alter public school quality. This is especially likely to be the case when private school enrollment is high pre-reform, and a large number of households are on the margin of switching. Based on our findings, analyses that ignore sorting effects are likely to understate – to a high degree in some circumstances – the overall impact of major education reforms in other contexts. We develop these implications below.

The rest of the paper is organized as follows: Section 2 sets out the institutional background to CSR in California and describes the data used in our analysis. Section 3 then

Whitmore (2001) analysis of Project STAR.

[11] The incumbent quality effect includes any peer effect from incoming private school students. The effect for switchers is the difference in value added between the private and public schools they attended.

presents the strategies that we use to explore the general equilbrium effects of the reform on school enrollments and house prices respectively, along with the corresponding causal estimates. Those motivate an estimation framework for separately identifying the direct and indirect general equilibrium effects of the reform on the basis of test scores, which we develop in Section 4. Estimates of the framework are presented and interpreted in Section 5, and Section 6 concludes.

2 Background and Data

Given we analyze the general equilibrium effects of major education policies in the context of California's state-sponsored class size reduction (CSR) program, we discuss in this section the policy context and relevant institutional background to the CSR reform, along with the rich data set we have assembled.

2.1 Institutional and Policy Background

California's CSR program was put in place in the spring of 1996 – up to that time, the largest of its kind implemented in the United States.[12] Impetus for the reform arose in the wake of disappointing national test score rankings four years earlier, when National Assessment of Educational Progress (NAEP) scores became available on a state-by-state basis for the first time. These revealed California to be among the worst-performing states in both mathematics and reading. Furthermore, it became clear that the low performance issue was persistent.[13]

California lawmakers, motivated in part by Project STAR,[14] enacted the class size reduc-

[12]After full implementation, California's CSR program cost about $1.5 billion each year. Following California, several large-scale CSR programs were implemented, with the federal government spending $1.2 billion a year from 1999 to 2001 on class size reduction, and Florida instituting a CSR program in 2002 that cost over $2 billion a year.

[13]For instance, the 1994 NAEP results showed California to be the very bottom state (along with Louisiana) in fourth grade reading, and in 1996, it tied with Tennessee at the bottom of the eighth grade mathematics rankings.

[14]See, for example, a report from the associated legislative discussions, available at `http://files.eric.`

tion reform to address these problems in July 1996. While the policy was widely supported by both parents and teachers, fierce disagreement between the Republican Governor Pete Wilson and the California Teachers' Association over education policy meant that its implementation did not arise in a consensual way, with the Governor adamant that extra funding available from the state's budget surplus in the mid-1990s – which by a narrowly-passed 1988 constitutional initiative had to be spent on education – would not be used as discretionary funding that could flow into higher teacher salaries. To ensure this, Governor Wilson avoided funding the union-dominated education boards by arranging to give the money directly to schools that had class sizes below a certain threshold.

The reform provided targeted incentives to reduce class sizes in lower grades from a statewide average of 28.5 down to 20.[15] For the first year of operation – the school year 1996-97 – the program applied only to first graders. Second grade classes then became subject to the program incentives in the following year (1997-98), and schools were able to choose to implement CSR in either kindergarten or third grade beginning in 1998-99.[16] Although participation was voluntary, substantial financial payments of $650 per pupil enrolled in a class of 20 or fewer students (relative to average 1995-96 per-pupil expenditures of $6,068) led to nearly universal adoption by districts and high levels of adoption by schools.[17]

Table 1 shows the policy coverage (and also data availability, discussed below), making clear the nature of the roll-out. Table 2 highlights the timing of CSR implementation that we exploit, showing average class sizes in grades K-5 for school years 1997-98 (when first and second grade were already affected) through 2001-02. Despite participation in CSR being nearly universal, some districts and schools still chose not to implement CSR. Districts did

`ed.gov/fulltext/ED407699.pdf`.

[15]This subsection draws on the lively account of the background to CSR in Schrag (2006). As detailed there, an unidentified staffer for the Governor stated that the class size goal of 20 was set based on affordability, rather than with any specific policy rationale in mind.

[16]We exploit this differential timing of implementation by grade in our identification strategies below, when studying changes in private school share, sociodemographic compositions and test scores.

[17]For districts to participate in CSR, they only needed to opt into the program, whereas schools only received CSR funding if they reduced class sizes in the relevant grade. We use the school adoption decision in our school-level house price design and our triple-differences approach, both described below.

not implement CSR either because (i) they had class sizes just above twenty, and did not think it was worth seeking the extra funding to hire a new teacher, or (ii) they already had many class sizes below twenty and did not realize they were eligible.[18] At the school level, schools often delayed their implementation of CSR due to a lack of space: in a survey by the CSR Research Consortium, eighty percent of principals who had not implemented CSR stated that space issues were the main reason.[19]

The initial announcement and roll-out of the actual policy was both sudden and unanticipated, generating headlines such as "Sacramento Surprise – Extra Funds / Governor wants to use money to cut class size" in the *San Francisco Chronicle* (Lucas 1996). As a consequence, no systematic program evaluation method was put in place.[20] Several other factors make studying CSR challenging. In terms of student performance, student testing did not begin until the 1997-98 school year, when the Standardized Testing and Reporting Program – another initiative of the Republican Governor – began. Thus, researchers do not have access to a comparable pre-reform test.[21] Additional data limitations included a lack of student or classroom-level data and an inability to track teachers over time.[22]

2.2 Data

Several useful data sources are available for our empirical analysis (previously discussed limitations notwithstanding), which we now describe. For a more detailed description, see Appendix Table A.1.

The California Department of Education (CDE) provides three types of data used in the study. The first consists of the student enrollment for all public schools and districts at the

[18] See `http://www.lao.ca.gov/1997/021297_class_size/class_size_297.html.`

[19] See `http://www.classize.org/summary/97-98/summaryrpt.pdf`

[20] The legislature did create the CSR Research Consortium to conduct a four-year comprehensive study to evaluate the implementation and impact of CSR, though it had to confront the same data limitations that we highlight below. The Consortium was composed of five research institutions: the American Institutes for Research, RAND, Policy Analysis for California Education (PACE), WestEd, and EdSource.

[21] Earlier tests in the state – for instance the CLAS test – were discontinued in the face of budget cuts and union resistance. Appendix D offers a quick primer on California statewide testing.

[22] California's teacher identifiers were scrambled each year to prevent following the same teacher over time. They continue to be scrambled in the statewide files to the present.

grade level, from the 1990-91 through the 2012-13 school years. We augment the enrollment data to incorporate demographic characteristics, including race, 'English as a Second Language' (ESL) status, Free or Reduced-Price Meal status[23] and CSR implementation status for all schools in the 1998-99 to 2003-04 school years inclusive.

Second, the CDE also provides grade-level enrollment data for private schools from 1990-91 to 2012-13 inclusive. No demographics are available beyond overall private school enrollments from this source. Together, these two data sets allow us to study the effects of CSR on public school compositions and the local private school share, starting well before CSR's introduction.[24]

The third data source comprises test score data from California's Standardized Testing and Reporting Program (STAR) for second grade and higher. All students in second through eleventh grade (with some minor exceptions[25]) took the Stanford Achievement Test in both mathematics and English near the end of the academic year.[26] The Stanford Achievement Test was a national norm-referenced multiple-choice test introduced in the 1997-98 school year. Given that the policy was in place for first grade since the 1996-97 school year and included second grade beginning in 1997-98, we do not observe a purely pre-reform period in terms of scores. Thus, identifying the effect of CSR on test scores necessarily involves exploiting differences in treatment over time. Our estimation strategy is designed to use that variation.[27]

To keep test scores similar over time, we use the percentile ranking as our test score

[23]This serves as a measure of the poverty rate of the student body since only students whose household income is below a threshold based on a percentage of the poverty line are eligible.

[24]The test score data, described next, do not have this desirable feature.

[25]Students were exempted if they were special education students or if a parent or guardian submitted a written request for exemption. Test taking rates were high nonetheless. For example, in 1998-99, over ninety-three percent of students in grades 2-11 took the relevant test.

[26]Testing dates were generally between March 15 through May 25 of a given academic year.

[27]We are limited in terms of the years we can use in some of our analyses. In the 2002-03 school year, California's STAR program was reauthorized and the State Board of Education issued a request for potential contractors to submit proposals for administering STAR. The contract was won by CTB/McGraw-Hill, and led to the test being changed from the Stanford Achievement Test (run by Harcourt Educational Measurement) to the California Achievement Tests. Because the monotonicity of scores by grade is no longer preserved for the 2002-03 academic year and onward due to the test change, we must limit some analyses to the academic years 1997-98 through 2001-02, even though test scores are reported until 2012-13.

measure. This ranking reflects the percentage of students in a nationally-representative sample of students, in the same grade, tested at a comparable time of the school year, who fall below the test score for the mean student in a given school-grade-year. For example, if the average student in a school-grade-year scored at the 60th percentile on the standardized test, this would mean that they did as well as or better than 60 percent of the students in the national sample. The availability of these data alongside the CSR policy are represented in Table 1 (already referred to concerning the policy's rollout), while summary statistics for the test score data in mathematics are provided in Appendix Table A.2.

Since schools (and districts) chose whether or not to adopt CSR, and adoption – looking ahead – is used in our identification strategy to determine the impact of CSR on house prices, Table 3 reports district and school characteristics for schools that implemented CSR alongside those that did not. The table shows that 'High-CSR' districts (those in the top three quartiles in terms of CSR implementation) and CSR-implementing schools have a larger fraction of their student body that is white and a correspondingly lower fraction of their student body that is Hispanic, relative to 'Low-CSR' districts (in the bottom quartile in terms of CSR implementation) and schools that did not implement CSR. In terms of districts, Low-CSR implementing districts were also likely to be smaller relative to their High-CSR counterparts.

A fourth dataset originates from the DataQuick DataFile Service and contains housing price information for 90 percent of all sale and loan housing transactions covering all regions of California, from 1990 to 2012 inclusive. These data also include a rich set of housing characteristics, such as the number of bedrooms, lot size, and square footage. We map the house prices to school districts using the year 2000 California school district boundary files created by the U.S. Census Bureau. We also map them to specific school attendance zones using the 2009-10 boundary files for school attendance areas provided by The College of William and Mary and the Minnesota Population Center (2011).[28] The boundary files

[28]Appendix E provides a brief description of the process we use to match the housing transactions data to school districts and school attendance zones.

provide coverage for about 40 percent of elementary schools in California.[29]

Table 4 provides summary statistics for the main variables used in our analysis. Along with overall means, we break these down into the period preceding the introduction of the CSR reform in California (1990-91 through 1995-96), the period during which it was phased in across grades (1996-97 through 1998-99), and the period following its full implementation (2000-01 through 2012-13).

In terms of the school data (Panel A), the evolution of the student-teacher ratio over time indicates that the CSR reform had a dramatic effect: the ratio fell from over 25 to about 20, reflecting a 20 percent decline in class size.[30] Panel A also reveals that the private school share of enrollment at the state level declined during the period of interest, falling from 9.9 percent prior to CSR implementation to 8.4 percent afterward.[31] In addition, there was a marked change in the composition of public school students, with a reduction of about 10 percentage points in the share of white students and a corresponding increase in the fraction of Hispanic students.[32] In the house price data (Panel B), we observe a near doubling of house prices during the period of interest. There is also substantial heterogeneity, particularly during the collapse of the housing bubble in the post-CSR period.

3 Evidence of General Equilibrium Responses

In this section, we present causal evidence shedding light on general equilibrium responses to the reform. We study three outcomes of interest: private school share, public school com-

[29]It is worth noting that the school attendance zone boundary files provide disproportionate coverage in urban areas. As a consequence, the school-level house price design we employ below may not be representative of the effect of CSR on California as a whole.

[30]The student-teacher ratio is used as a proxy for class size as we do not observe teacher assignment data prior to the introduction of CSR. Data on the number of elementary school teachers are drawn from the National Center for Education Statistics (https://nces.ed.gov).

[31]This cannot be taken of proof that CSR caused a re-sorting of students between private and public schools, as there was a similar national trend of declining private school shares during the time period (Buddin 2012). For this reason, we adopt a grade-by-grade research design that relies on local intensity measures, described below, to test whether CSR had an impact over and above the national trend.

[32]Again, we will not take this as direct evidence of re-sorting. For that, we will need to draw on variation in school-level implementation and across CSR and non-CSR grades.

position, and house prices. The first two relate to sorting responses, while the latter allows us to speak to the full effect of the reform (incorporating the direct effect of smaller classes, sorting, and changes in teacher quality). In each instance, we have assembled consistent data from well before the reform was introduced (not possible for test scores), and exploit differences in CSR adoption across time, grades, schools and districts. Thus, for each outcome in turn, we describe the differencing strategy we use, then present results, consisting of both visual evidence and regression estimates.

3.1 Private School Share

To explore the effect of CSR on private school shares, we take advantage of the reform's differential impact on kindergarten through third grade, made clear in Table 2. For each period t, we define the treatment group as any school-grade that implements CSR and the control group as any grade that does not.[33]

We begin with a simple difference-in-differences approach, which compares treatment and control grades before and after the reform came into effect. The analysis uses the following regression (weighted by district-grade-year enrollment[34]):

$$share_{dgt} = \beta_0 + \beta_1 post_{gt} + \beta_2 treat_g + \beta_3 (post_{gt} * treat_g) + \eta_d + \theta_t + \phi X_{dgt} + \epsilon_{dgt} \,, \tag{3.1}$$

where $share_{dgt}$ is the private school share for district d in grade g at time t,[35] $post_{gt}$ indicates whether (or not) CSR had been implemented for grade g, $treat_g$ indicates whether grade g was ever subject to the CSR reform, X_{dgt} is a set of district-grade-year covariates (percent ESL, race and enrollment), and η_d and θ_t are district and time fixed effects, respectively.

The difference-in-differences coefficient of interest is β_3. It is identified under the as-

[33]Later on, we also use differences in CSR adoption across districts to create an additional layer of contrast.

[34]Weighting is used to account for smaller districts that do not contain any private schools. Alternatively, the regression can be restricted to only those school districts with a private school option. We present results for the 'weighting' method, as the sample restriction produces similar estimates and so is omitted.

[35]Formally, $share_{dgt}$ is defined as the enrollment in private schools for d-g-t divided by the total enrollment for d-g-t.

sumption that CSR and non-CSR grades would have experienced the same change in private school share in the absence of the reform. While this 'parallel trends' assumption is not directly testable, the lack of differential pre-trends favors the validity of this assumption in the results that follow.

We also allow for the possibility that trends are not parallel by introducing an additional layer of differencing. Specifically, we augment the difference-in-differences analysis by estimating a triple-differences specification that further differences according to a measure of the local intensity of CSR. This intensity measure is created using the share of CSR-eligible students in a district who are in a CSR school-grade. It takes advantage of the fact that while most districts opted into CSR,[36] the school-grade level implementation was uneven across them. Given that school-level CSR participation data are only available for the 1998-99 through 2003-04 school years, we define our local intensity measure (CSR_d) as the percentage of K-3 students in a CSR participating school-grade within a district for the 1998-99 school year.[37] Formally,

$$CSR_d = \frac{\sum_{s \in d} \sum_{g=0}^{3} \mathbb{1}\{CSR_{sg}\} * (enroll_{sg})}{\sum_{s \in d} \sum_{g=0}^{3} enroll_{sg}}, \tag{3.2}$$

where $enroll_{sg}$ is the enrollment of grade g students in school s and district d (kindergarten is defined as $g = 0$), and $\mathbb{1}\{CSR_{sg}\}$ indicates whether the school implemented CSR for the particular grade in the 1998-99 school year.

Using this measure, we implement the triple-differences approach by estimating the following weighted[38] regression:

[36] In the first year of CSR, only 56 of 895 districts in California did not opt in. In the following year, twenty districts remained non-participating districts. For every year thereafter in our sample period, the number of non-participating districts was about ten.

[37] Results are similar if this variable is averaged over the 1998-99 through 2003-04 school years.

[38] Again, we weight by district-grade-year enrollment.

$$share_{dgt} = \beta_0 + \beta_1 post_{gt} + \beta_2 treat_g + \beta_3 CSR_d + \beta_4 (post_{gt} * treat_g) + \beta_5 (post_{gt} * CSR_d)$$
$$+ \beta_6 (treat_g * CSR_d) + \beta_7 (post_{gt} * treat_g * CSR_d) + \theta_t + \phi X_{dgt} + \epsilon_{dgt} \, , \tag{3.3}$$

where all variables other than the intensity measure CSR_d are identical to those in equation (3.1). The triple-differences coefficient of interest is β_7. Identification of the parameter depends on a less restrictive variant of the parallel trends assumption – that the difference in the way that the private school share evolves between CSR and non-CSR grades would have been the same for low- and high-share CSR districts in the absence of the reform.

3.1.1 Private School Results

Private School Share: Our difference-in-differences approach exploits variation in the time when different grades became subject to CSR. Our triple-differences identification strategy adds an extra layer (as noted) in the form of the local intensity of adoption.

With respect to the former type of variation, Figure 1 plots the change in private school enrollment share overall and in CSR grades over time. The visual evidence is clear: when CSR is first implemented in the public system for a particular grade, the corresponding private share for that grade declines noticeably relative to other grades, suggesting that the reform attracts private school students into the public system. For example, the share of students in private schools in the entire state in first grade is flat in the two academic years preceding 1995-96. Then, by the start of 1996-97 (the first year that CSR affects public school class sizes in first grade), there is a pronounced dip down in first grade while the shares for other grades remain steady, consistent with there being a switch into public schools for that grade. Similarly, when second grade becomes eligible for CSR in public schools at the start of 1997-98, we also see a pronounced decline in the private school share relative to the previous academic year (and relative to other grades). The same is true for kindergarten and third grade in the first year when those grades became eligible (1998-99).

Given the patterns in Figure 1, we start by reporting the most basic contrast in Table 5 – private school shares 'before' and 'after' CSR's introduction among the treated and untreated grades. We find that treated grades experienced a precisely-estimated 1.4 percentage point decline in private school share relative to untreated grades following implementation of the policy. This corresponds to the estimate from the difference-in-differences estimator specified by equation (3.1) without including any controls. We estimate that equation and report the results in Table 6, though now including various controls. Our preferred specification with all controls included is similar to its unconditional counterpart: treated grades experience a 1.8 percentage point decline in private school share relative to untreated grades as a result of CSR. This decline is equivalent to 15 percent of the pre-CSR K-3 average private school share of 11.7 percent – a significant amount – and 21 percent of its standard deviation.

As for our measure of the local intensity of CSR given in equation (3.2), Figure A.2 provides a map for the state, showing significant variation.[39] Using district-level CSR participation intensity as an additional dimension of differencing, our preferred triple-differences analysis from equation (3.3) yields qualitatively similar findings, although noting that the difference-in-differences and triple-differences estimates are not directly comparable since almost all districts have some level of CSR implementation.[40] With all controls, column (4) of Table 7 shows that CSR is associated with a 1.5 percentage point decline in private school share. Thus, the evidence indicates that private school shares experienced a substantial reduction as a result of the reform, concentrated precisely in the grades that were treated.

We provide support for the 'parallel trends' assumption that underlies these results by plotting difference-in-differences estimates by year, using the treatment of grades (CSR versus non-CSR) and district-level CSR participation intensity as the two dimensions of differencing. Figure 2 shows that there is no effect on private school share prior to the implementation of the reform, followed by a clear decline after.

[39]Figure A.1 provides a map of the state showing substantial variation in our outcome variable: the change in private school share from the 1995-96 to 1999-2000 school years by school district.

[40]Thus, the triple-differences coefficient cannot be interpreted as the effect of CSR relative to a non-CSR baseline, as such a comparison extends beyond the support of the data.

The steep decline in private school share induced by CSR makes it likely that the extensive margin would also be affected. Figure 3 plots the number of private schools per 1000 school-aged children in California and the rest of the country.[41] As expected, there is a steep reduction in private schools per capita after the 1996-97 reform in California relative to the rest of the country. These can be further broken into private school entry and exit responses. After the 1996-97 CSR reform, Figures A.3(a) and A.3(b) show a sharp increase in private school exit rates and a decline in entry rates in California relative to the rest of the country.

Table A.3 estimates these extensive margin effects of CSR in difference-in-differences and triple-differences frameworks.[42] The point estimate from the triple-differences specification indicates that the CSR reform caused a decline of 0.065 private schools per 1000 school-aged children, a 23 percent decline off California's pre-CSR mean number of private schools per 1000 school-aged children.

Persistence after Switching: Given the evidence relating to the initial impact of the reform, we would like to know whether sorting is transitory or not − relevant when separating the reform's direct and indirect effects, as this depends on the degree to which students return to private school over time. Three main possibilities come to mind: students previously in the private school system might return to the private system after completing third grade; they might return after completing all grades offered by the public school that they switched into (say, after completing fifth grade in a K-5 school); or they might remain in the public system for the duration of their primary and secondary education. In Appendix B, we present a regression discontinuity approach showing that approximately two-thirds of the CSR 'treatment effect' on private school share disappears when making the transition to middle school, consistent with a significant share of students transitioning back into the private system.

[41] Data come from the Private School Universe Survey, run by the National Center for Education Statistics. It is available at `https://nces.ed.gov/surveys/pss/pssdata.asp`.

[42] The latter uses time (pre- vs. post-CSR), state (California vs. the rest of the country) and whether the private school served CSR grades as the three layers of differencing.

3.2 Public School Composition

The previous set of results indicates that CSR induced students in relevant grades to switch from private school into the public system. Our public school data allow us to explore how this influx of new students from private schools affected public school sociodemographic compositions at the school-grade-year level.

To do so, we analyze the effect of CSR on public school student demographics by exploiting the degree of private school presence locally.[43] Our econometric approach involves a triple-differences design, using as the third dimension of differencing whether a private school is nearby. The weighted regression is then:

$$\begin{aligned} demo_{sgt} = \beta_0 + \beta_1 post_{gt} + \beta_4(post_{gt} * treat_g) + \beta_5(post_{gt} * \mathbb{1}\{Buffer < x\ km\}_s) \\ + \beta_6(treat_g * \mathbb{1}\{Buffer < x\ km\}_s) + \beta_7(post_{gt} * treat_g * \mathbb{1}\{Buffer < x\ km\}_s) \\ + \eta_s + \theta_t + \phi X_{sgt} + \epsilon_{sgt}\ , \end{aligned} \tag{3.4}$$

where $demo_{sgt}$ is a school-grade-year demographic variable, $\mathbb{1}\{Buffer < x\ km\}_s)$ indicates if a private school is within a radius of $x\ km$ of school s and all other variables are defined as before.[44] The triple-differences estimate β_7 has a causal interpretation under the assumption that the difference in the change in demographic share between CSR and non-CSR grades would have been the same in the absence of the reform for public schools within $x\ km$ of a private school and those further away.

3.2.1 Public School Composition Results

We are interested in measuring whether public school demographics shifted as a result of the CSR-induced decline in private school share. Given that the proportion of white students in private school is initially about fifteen percent higher and the proportion of Hispanic students about twenty three percent lower compared to their public counterparts (see Table

[43] Appendix Section C reports additional evidence taking advantage of school-level differences in the reform's implementation, corroborating the results presented here.

[44] Only private schools with ten or more students in kindergarten through third grade are included.

8), inflows to the public system are likely to consist mainly of these two groups.[45] Although public-private demographic disparities in the proportion of Asian and black students are smaller, one might also expect an influx of Asian students and a decline in the proportion of black students in the public system. This is indeed what we find.

Figure 4 represents the variation used in our approach, which compares student demographics for public schools that have a nearby private school with those that do not. We see a drop over time in the proportion of Hispanic and black students following the reform along with a rise in the fraction of white students.

We then incorporate the comparison between CSR and non-CSR grades described in equation (3.4). Table 8 reports estimates for the distance design, using three different 'closeness' measures: 1.5, 3 and 5 kilometers.[46] Relative to public schools that have no nearby private competitors, the results indicate that CSR led to a significant increase in the fraction of white students and a decline in the fraction of Hispanic and black students in public schools with nearby private alternatives.

As a validity check, we compute difference-in-differences estimates by year, using the treatment of grades (CSR versus non-CSR) and the distance to the nearest private school competitor (within 3 kilometres versus more than 3 kilometres) as the two dimensions of differencing. Figures 5(a) and 5(b) plot these estimates for the fraction of white and Hispanic students, respectively, showing an increase in the fraction of whites and a reduction for Hispanics. For white students, there is evidence of a small pre-trend in the pre-reform years, although the magnitude of the post-CSR treatment effects is about four times that of the pre-reform estimates. For Hispanic students, the point estimates are indistinguishable from zero in the pre-reform years and become negative and statistically significant once CSR is implemented.

[45] While we do not have access to very detailed private school demographic data, the NCES provides school-level demographics for the 1997-98 school year and every two years thereafter. The public-private demographic disparities we report are thus one year after CSR began in 1996-97.

[46] As is apparent, the choice of the 'closeness' measure has very little impact on the results.

3.3 House Prices

We have already shown evidence of reform-induced student sorting between the private and public school systems. Along with the direct effect of the reform on class size, these changes are likely to alter public school quality. In turn, to the extent that this amenity matters to households, equilibrium house prices should adjust accordingly.

In this subsection, we are interested in measuring the full impact of the reform on house prices, combining both the direct effect of CSR on public school quality along with indirect demographic changes. We will also provide suggestive estimates of the relative importance of the direct and indirect effects of the reform using a straightforward regression approach: the framework we propose in the following section is designed to measure the direct and indirect effects of the reform in a credible way, based on a common test score metric.

To quantify how far the effects of the reform are capitalized in the housing market, we rely on a difference-in-differences design that exploits variation in house prices over time (before and after CSR) and the local district treatment intensity measure (CSR_d) defined in Section 3.1.[47] Accordingly, we estimate the following regression:

$$price_{dt} = \beta_0 + \beta_1 Post_t + \beta_2 CSR_d + \beta_3 (Post_t * CSR_d) + \phi X_{dt} + \epsilon_{dt} \,, \tag{3.5}$$

where $price_{dt}$ is the average house price in district d at time t,[48] $Post_t$ is a CSR implementation indicator, X_{dt} is a set of controls consisting of house-level characteristics (number of bedrooms, lot size and square feet), student characteristics (percent ESL, race, percent free or reduced-price meal, enrollment, and enrollment squared), and teacher characteristics (average teacher experience, proportion of teachers without a bachelor degree, and proportion of teachers with a graduate degree). The coefficient of interest is β_3, which has a causal in-

[47]We are limited to two dimensions of differencing, given that differences between CSR and non-CSR grades cannot be exploited when house prices are the outcome of interest. Appendix Section C exploits school-level differences in the reform's implementation to highlight the effect of CSR on house prices, corroborating the results presented here.

[48]To account for large house prices increases in California during the early 2000s, we deflate all house prices to 1995 dollars using California's house price index (available at `https://fred.stlouisfed.org/series/CASTHPI`).

terpretation under the parallel trends assumption that high- and low-intensity CSR districts would have experienced the same change in house prices in the (counterfactual) absence of CSR.

3.3.1 House Price Results

We identify the effect of CSR on house prices (as discussed) using district-level variation by comparing post-reform house prices to their pre-reform baseline across high- and low-share CSR districts. The district variation that we draw on is relatively evenly distributed across the state, as Figure A.2 shows. Figure 6(a) displays trends among high- and low-share CSR districts (specifically, the top three-quarters versus the bottom quartile of CSR-implementing districts in 1996-97), showing that there do not appear to be any differential trends in house prices prior to the reform's implementation. Once CSR comes into effect, however, house prices show a significant increase in high-share CSR districts relative to their low-share counterparts.

This visual evidence maps directly into the econometric specifications given by equation (3.5), which are reported in Table 9. It is important to differentiate between columns (1)-(3) and columns (4)-(6), since the former do not control for indirect general equilibrium effects of the reform in terms of adjustments in school demographics and teacher quality, which may themselves be highly valued by parents.

Our preferred estimate to capture the *full effect* of the reform is given in column (3), which controls for house characteristics and district fixed effects, but does not control for any indirect general equilibrium effects of the reform via adjustments to teacher or peer quality. The point estimate of $34,300 implies that a one standard deviation increase in CSR implementation, measured in equation 3.2, is associated with around a 2.6 percent increase in house prices – a substantial amount.[49]

[49]Based on the $34,300 point estimate, a one standard deviation increase in the CSR implementation measure (0.15) corresponds to an increase in house prices of $5,145. This increase is 2.6 percent of the 1995 pre-CSR (weighted) mean of $195,000.

While column (3) captures the full capitalization of the reform into house prices, it is informative to look (in a suggestive way) at the effect of the reform on house prices once the indirect general equilibrium effects on teacher and demographic characteristics are accounted for. First, column (4) controls for student characteristics, which we find are capitalized into house prices: once student characteristics are controlled for, a one standard deviation increase in CSR implementation is now associated with a 2.0 percent increase in house prices (the difference between column (3) and (4) is statistically significant at the one percent level). Column (6) controls for teacher characteristics such as education and experience, with column (5) re-analyzing column (4) with data restricted to years where we have teacher characteristics data to provide a benchmark. We do not find that the general equilibrium teacher effects studied in Jepsen and Rivkin (2009) are capitalized in house prices.[50] Controlling for the indirect general equilibrium effects on peer quality thus decreases the capitalization of the reform into house prices by about one-quarter.

Regarding the validity of these results, we report the effect of district CSR treatment intensity on house prices by year in Figure 6(b). The effect of CSR on house prices is indistinguishable from zero in the pre-reform period, suggesting negligible differences in pre-trends between high- and low-CSR treatment intensity districts. The effect becomes positive upon implementation of the reform, however, and continues to grow even after the reform is fully implemented, which may reflect local housing markets being slow to reach new equilibria.

The evidence we have presented relating to house prices is suggestive, both of the significant size of the overall impact of the reform and the importance of indirect effects, especially related to sociodemographic sorting. Next, we develop a more formal approach for separating out the constituent effects of the reform. As noted in the Introduction, we focus on the education production function (and test score outcomes) rather than the hedonic function (and house prices), both because test scores are well measured at the school-grade-year level,

[50]This is consistent with the result in Imberman and Lovenheim (2016) that house prices do not respond to teacher quality, *per se*.

and – given our test scores are available for particular grades and years within-school – we are able to disentangle relevant channels using a differencing approach: in contrast, house prices do not have the requisite granularity.[51]

4 A Framework for Separating Direct and Indirect Effects

This section sets out an estimable framework for measuring the direct and indirect general equilibrium effects of a major education reform on a common footing. The framework makes clear how we will parameterize the direct and indirect effects of the reform. When estimating these effects, empirical challenges arise; building on the variation we have already exploited, we outline a multiple differencing strategy that allows us to address these challenges.

In what follows, we present the technology, our main estimating equation, and then explain how suitable differencing allows us to recover the technological parameters, contrasting this with a naive difference-in-differences approach. We close the section by discussing the sources of identification for the framework's key parameters.

Technology

Consider an environment in which an outcome y depends on a policy variable and a set of other relevant inputs. We will think of outcomes primarily being test scores, given an education setting that features a major policy change. In that setting, the direct and indirect effects of the policy associated with changing the policy variable can be understood in terms of an education production technology in which educational inputs jointly affect measured test score performance.

Our goal will be to uncover the parameters of the technology. To that end, we will

[51] We also have a more precise handle on the inputs to the education production process in this context than to the various local 'house price' processes.

make explicit the assumptions that allow us to apply the multiple differencing approach developed below. We focus on a specification in which output is affected by three main inputs $\{R, Q, X^S\}$, where R measures school resources (under the direct control of the policy maker), Q represents teacher quality, and student characteristics in the school are given by X^S. There is also a further noise component, ϵ, reflecting unobservable random influences on contemporaneous test scores.

Assumption 1: linear technology

For tractability, we will use a linear approximation to the true education production function, following the bulk of the education literature, with additive inputs (including an additive error). This structure will be necessary in order to apply our differencing strategy.

Assumption 2: cumulative technology

In light of compelling evidence from the recent empirical literature (see Rivkin, Hanushek and Kain (2005) and Chetty, Friedman and Rockoff (2014), for example) that underscores the cumulative nature of education production, we allow current inputs in one period to have persistent effects in subsequent periods as students acquire (and retain) knowledge and skills.

Based on these two assumptions, we write the linear technology to account for the persistent effects of both class size and student sociodemographics on scores in period t as

$$y_t = \gamma + \gamma_R \sum_{\tau=0}^{L} \delta_R^{\tau} R_{t-\tau} + \gamma_S \sum_{\tau=0}^{L} \delta_S^{\tau} X_{t-\tau}^{S} + h(Q_t) + \epsilon_t \tag{4.1}$$

The proposed structure is parsimonious, chosen in light of what we are able to identify using our estimation approach. The introduction of the reform is assumed to have two main effects: (1) the direct effect of reducing class size on student learning, γ_R; and (2) the indirect general equilibrium effect from changes in student composition γ_S, which arises as a result of sorting between the private and private school systems – this includes the 'own' effect and a spillover effect. As will be seen below, we will use a control technique for teacher quality,

Q (captured by $h(Q_t)$ in 4.1), and so do not impose parametric assumptions regarding the impact of teacher quality in the equation.[52]

In terms of persistence, the effect of past 'resources' (smaller classes) on current test scores is parametrized by δ_R – a parameter of interest in prior research (see, for example, Krueger and Whitmore 2001 and Ding and Lehrer 2010). Specifically, δ_R measures the persistent effect on test scores of a one unit increase in resources one period ago into the present. Further, resources from at most L periods ago are allowed to influence current test scores, following a geometric decay. Similarly, δ_S captures the persistent effects of past school demographic compositions on current test scores, also following a geometric decay over the same number of periods.

According to the above parameterization, the introduction of the policy can be represented by the change in school resources (ΔR) associated with the policy intervention, comparing before and after. The *direct*, contemporaneous impact of the policy change in terms of test scores is given by the product of the γ_R parameter and the change in class size (determined by school resources).

For large-scale reforms, policy makers would like to have a sense of induced effects that may reinforce or counteract the direct effect of the policy on outcomes. Our main focus is on the induced effect of more resources for public schools on the demographic composition of students in public school, as students switch from local private schools (at least where private schools have some prior presence), and in turn, the impact on test score outcomes. Changing the mix of students (ΔX^S) may affect outcomes through two channels. First, if the incoming students are of higher ability than students already enrolled in public school and score more highly themselves, then outcomes will improve through what we term the 'own' effect. Second, the change in the demographic composition of public schools may result in *spillover* benefits to incumbent public school students, perhaps via positive peer influences in the classroom.

[52] Consequently, we will omit teacher quality from the discussion of the framework for clarity, returning to it when describing identification.

As we are limited in our capacity to separate these two channels out by the aggregated nature of the data, for the most part, we will combine them together. Thus, given the simple technology above, the combined indirect sorting effect on contemporaneous scores can be written $\gamma_S \frac{\Delta X^S}{\Delta R}$.[53]

Adding persistence implies that direct and indirect effects from earlier periods can now accumulate over time: the structure allows expressions for each effect to be written down. For example, a shock to class sizes l periods ago will give rise to an indirect sorting effect (in terms of test scores) in the current period of $\gamma_S \delta_S^l \frac{\partial X_l^S}{\partial R_l}$, where $\frac{\partial X_l^S}{\partial R_l}$ measures the induced within-period sorting response l periods ago. In turn, taking a forward-looking perspective, a class size shock at t will have a total indirect sorting effect on scores that propagates into the future according to $\gamma_S \frac{\partial X_t^S}{\partial R_t} \sum_{\tau=0}^{L} \delta_S^\tau$. We will use the estimates and this simple structure to compute a measure of the overall test score benefits of CSR below.

Estimating Equation

To uncover the parameters in equation (4.1), we need to locate sufficient sources of independent variation. In a setting where a policy shock occurs, a before/after contrast may be used to identify the direct effect of school resources, as in several prior studies. Yet given our interest in wider general equilibrium responses to major education reforms, we wish to identify more than just the direct effect. Further, we need to account for persistence, which compounds the difficulty of disentangling any direct and indirect effects. As shown below, the persistence of the direct and indirect effects in terms of test scores potentially invalidates a simpler reduced-form strategy that uses untreated grades as controls, as test scores for those grades could include past treatments received in earlier treated grades.

Using the specification of the technology in (4.1), our estimation approach is intended to address these issues. To uncover the causal impact of grade-specific reductions in class size, including any induced changes in school demographics, we will draw notional contrasts

[53] We will assume that indirect effects are induced within the same period as the resource shock. The effects on scores may be longer-lasting.

between observed scores at the school-grade-year level and counterfactual scores that would have prevailed had the reform not been enacted.

The technology above allows us to write those contrasts down, forming the basis of our estimating equation. To help clarify the identification argument, we track both grades and time, using the τ index to increment both successive grades ($g \in \{0, 1, 2, 3, 4, ...\}$) and academic years ($t \in \{1996\text{-}97, 1997\text{-}98, 1998\text{-}99, 1999\text{-}00\}$). Adapting (4.1) from above (with additional subscripts) to reflect the available data, the school-grade-year (s-g-t) test score y_{sgt} can be written as a function of current and past student, school and teacher inputs:

$$y_{sgt} = \gamma + \gamma_R \sum_{\tau=0}^{L} \delta_R^{\tau} R_{s,g-\tau,t-\tau} + \gamma_S \sum_{\tau=0}^{L} \delta_S^{\tau} X^S_{s,g-\tau,t-\tau} + \epsilon_{sgt} \, , \tag{4.2}$$

suppressing the teacher quality effect for clarity.[54] To capture the cumulative nature of the student learning technology in line with the above discussion, this equation allows inputs from L periods prior to current time t to affect current scores.

The comparisons we will make involve school averages, where the total number of schools is given by N_s. Thus we define $\Delta y_{gt} \equiv y_{gt} - y^u_{gt} \equiv \frac{1}{N_s}\sum_s (y_{sgt} - y^u_{sgt})$ as the difference between the actual average test score in grade g at time t, averaging over all schools serving that grade and the unobserved (superscripted by 'u') average score that would arise in a counterfactual setting where the reform had never been implemented. We can define ΔR_{gt} and ΔX^S_{gt} analogously. Further, given our emphasis on persistent effects, we will allow lags of these inputs to affect test scores contemporaneously. Forming first differences along the above lines, we obtain the following estimating equation:

$$\Delta y_{gt} = \gamma_R \sum_{\tau=0}^{L} \delta_R^{\tau} \Delta R_{g-\tau,t-\tau} + \gamma_S \sum_{\tau=0}^{L} \delta_S^{\tau} \Delta X^S_{g-\tau,t-\tau} + \Delta\epsilon_{gt} \, , \tag{4.3}$$

where $\Delta R_{g-\tau,t-\tau}$ and $\Delta X^S_{g-\tau,t-\tau}$ represent the change in school resources and the mix of

[54] We describe how teacher quality is included in the estimation approach at the end of the section, and in a more detailed appendix.

students arising from CSR (relative to the counterfactual baseline) for students in grade $g-\tau$ and academic year $t-\tau$.

Treated grade-year combinations satisfy $t \geq$ 1997-98 and $2 \leq g \leq 2+t-$ 1997-98: they, along with 'control' grade-year combinations, are represented in Table 1. For those treated grade-years, $\Delta y_{gt} \neq 0$, given $\Delta R_{gt} \neq 0$ and $\Delta X^S_{gt} \neq 0$: for all other control combinations, we have that $\Delta R_{gt} = \Delta X^S_{gt} = 0$, so $\Delta y_{gt} = 0$ for untreated grade-year combinations. The above equation implies, for instance, that $\Delta y_{gt} = 0$ for untreated pre-reform grade-year combinations such as third grade and above in 1997-98, since the reform had yet to be implemented.[55]

4.1 Estimation Approach

A practical challenge in taking this estimating equation to the data is that Δy_{gt} is not observed for treated grade-year combinations, as it depends on counterfactual test scores. The natural approach is to use scores from other school-grade-year combinations as controls for the counterfactual scores in treated grades. Identification from observed test scores is then obtained if we impose the plausible assumption of common trends across grades g and g' and time periods t and t': in our notation, $y^u_{gt} - y^u_{g't} = y^u_{gt'} - y^u_{g't'}$.

Difference-in-Differences: One estimation approach would involve a simple difference-in-differences ('D-in-D') specification, making a before-after comparison of the average scores of students in a grade that became subject to CSR in a given year with the average before-after scores of students in a control grade. Under standard assumptions – a linear technology, no spillovers, and no persistence of inputs – this would recover the direct causal impact of class size reduction.

Because of the evidence in the prior literature that class size reductions *do* have persistent effects (and large-scale reforms are likely to have spillovers), a difference-in-differences mea-

[55]Note that second grade cannot be used to identify any parameters, as no pre-reform observations exist to construct $y^u_{2,t}$ (recalling that 1997-98 is the first year for which we have test score data).

surement approach applied to test scores will not be valid. It is instructive to show where the simple D-in-D approach goes awry, in the process pin-pointing where we get identification of the direct and indirect effects of CSR using our approach.

Referring to back to Table 1, third grade became subject to the reform in the 1999-2000 academic year (and remained treated subsequently). In prior years, third grade class sizes were unaffected. Thus, one could construct the first difference when applying D-in-D by deducting the statewide average test score in third grade in 1998-99 from the statewide average for third grade in 1999-2000. The table also suggests numerous suitable control grades: those that, over the same time span, were never subject to CSR. Take, for instance, fourth grade in 1998-99 and 1999-2000, and construct the analogous difference in average scores.

From the structure above, we can see when the simple D-in-D recovers the direct causal impact of the policy – that is, $\Delta y_{3,99\text{-}00} - \Delta y_{4,99\text{-}00} = \gamma_R \Delta R_{3,99\text{-}00}$. Specifically, equation 4.3 allows us to write student achievement in terms of the changes associated with the reform for a given cohort in a given year, without needing to specify how the counterfactual is constructed in practice. For example, the differences between observed and counterfactual average test scores can be expressed in terms of the parameters for third grade students in 1999-00 as:

$$\begin{aligned}\Delta y_{3,99\text{-}00} = {} & \delta_R^2 \gamma_R \Delta R_{1,97\text{-}98} + \delta_R \gamma_R \Delta R_{2,98\text{-}99} + \gamma_R \Delta R_{3,99\text{-}00} \\ & + \delta_S^2 \gamma_S \Delta X^S_{1,97\text{-}98} + \delta_S \gamma_S \Delta X^S_{2,98\text{-}99} + \gamma_S \Delta X^S_{3,99\text{-}00} + \Delta \epsilon_{3,99\text{-}00} \, . \end{aligned} \tag{4.4}$$

Similarly, considering fourth grade students in the 1999-00 school year (who were subject to the CSR reform in first grade (1996-97), second grade (1997-98) and third grade (1998-99)), the differences between observed and counterfactual average test scores can be expressed as:

$$\Delta y_{4,99\text{-}00} = \delta_R^3 \gamma_R \Delta R_{1,96\text{-}97} + \delta_R^2 \gamma_R \Delta R_{2,97\text{-}98} + \delta_R \gamma_R \Delta R_{3,98\text{-}99}$$
$$+ \delta_S^3 \gamma_S \Delta X^S_{1,96\text{-}97} + \delta_S^2 \gamma_S \Delta X^S_{2,97\text{-}98} + \delta_S \gamma_S \Delta X^S_{3,98\text{-}99} + \gamma_S \Delta X^S_{4,99\text{-}00} + \Delta\epsilon_{4,99\text{-}00}\,, \quad (4.5)$$

where there is no effect of the reform in kindergarten ('grade 0') as CSR was not yet implemented for this cohort during that grade (i.e., $\Delta R_{0,95\text{-}96} = 0$).

Comparing students in CSR and non-CSR grades amounts to subtracting equation 4.5 from 4.4, which yields:

$$\Delta y_{3,99\text{-}00} - \Delta y_{4,99\text{-}00} = \gamma_R \Delta R_{3,99\text{-}00} - \delta_R^3 \gamma_R \Delta R_{1,96\text{-}97} - \delta_S^3 \gamma_S \Delta X^S_{1,96\text{-}97}\,, \quad (4.6)$$

where the equality follows because CSR affects all grades that have implemented CSR equivalently (i.e., $\Delta R_{g,t} = \Delta R_{g',t} \;\; \forall g, g'$).[56] It is clear in the above equation that the direct effect of class size, $\gamma_R \Delta R_{3,99-00}$, which is analogous to experimental estimates (for instance, from Project STAR), is not identified if there is persistence in the student learning technology and spillovers (i.e., $\delta_R \neq 0$, or $\delta_S \neq 0$ and there are sorting effects).

Our Strategy: The preceding argument underscores the need for an approach that can account for the possible effects of persistence and sorting. With that aim in mind, we take advantage of the way the policy was rolled out. In successive years, recall that schools were able to reduce class sizes in first grade (in 1996-97), then in second grade the next year, followed by both kindergarten and third grade in 1998-99 and 1999-2000 (depending on whether schools opted for kindergarten or third grade first), leading to differential exposure to the reform. Further, within each year, adoption was not uniform. A multiple differencing approach allows us to exploit these various contrasts.

To explain the essence of the approach, we focus on separating out the direct effects

[56]See Table 2, where the CSR grades have similar class sizes once CSR is implemented.

of the reform from the indirect sorting effect, abstracting from changes in teacher quality. Consider the incoming first grade cohort for the 1998-99 school year and the impact of the reform on their test scores by fourth grade (in 2001-02). This cohort was subject to the reform for grades 1-3 though not while in kindergarten, as CSR had yet to be implemented there in 1997-98. In addition, the cohort was not subject to CSR in fourth grade (given that fourth grade was never part of the reform), although any change to the student composition engendered by the reform remained, given that the students attracted into the public system by the reform had yet to return to the private sector (see the regression discontinuity results in Appendix B). The achievement of that cohort is given by:

$$\begin{aligned}\Delta y_{4,01\text{-}02} &= \delta_R^3 \gamma_R \Delta R_{1,98\text{-}99} + \delta_R^2 \gamma_R \Delta R_{2,99\text{-}00} + \delta_R \gamma_R \Delta R_{3,00\text{-}01} \\ &\quad + \delta_S^3 \gamma_S \Delta X^S_{1,98\text{-}99} + \delta_S^2 \gamma_S \Delta X^S_{2,99\text{-}00} + \delta_S \gamma_S \Delta X^S_{3,00\text{-}01} + \gamma_S \Delta X^S_{4,01\text{-}02} + \Delta\epsilon_{4,01\text{-}02}\,, \quad (4.7)\end{aligned}$$

since the students are not subject to the class size reform in fourth grade (i.e., $\Delta R_{4,02\text{-}03} = 0$), but student sorting still affects their achievement in fourth grade. The third grade cohort for the 2002-03 school year was subject to the policy for four consecutive years (given the policy was in effect when they were in kindergarten during 1999-00) and so their achievement is given by:

$$\begin{aligned}\Delta y_{3,01\text{-}02} &= \delta_R^3 \gamma_R \Delta R_{0,98\text{-}99} + \delta_R^2 \gamma_R \Delta R_{1,99\text{-}00} + \delta_R \gamma_R \Delta R_{2,00\text{-}01} + \gamma_R \Delta R_{3,01\text{-}02} \\ &\quad + \delta_S^3 \gamma_S \Delta X^S_{0,98\text{-}99} + \delta_S^2 \gamma_S \Delta X^S_{1,99\text{-}00} + \delta_S \gamma_S \Delta X^S_{2,00\text{-}01} + \gamma_S \Delta X^S_{3,01\text{-}02} + \Delta\epsilon_{3,01\text{-}02}\,. \quad (4.8)\end{aligned}$$

The CSR reform reduced class sizes to twenty or below in each treated grade, making the size of the reform identical in each grade and year, so we have that $\Delta R_{0,t} = \Delta R_{1,t} = \Delta R_{2,t} = \Delta R_{3,t}$, $\forall t$. Therefore, if we take the difference between equations 4.7 and 4.8, we are left with the term $\gamma_R \Delta R_{3,t}$, which represents the direct, contemporaneous effect of the change in school resources (here, given by the reduction in class size) on student achievement.

In turn, the framework can be used to recover γ_S – the effect of the influx of private school students into the public system. To do so, we draw on the regression discontinuity evidence in Appendix B, which shows that a high proportion of students (approximately two-thirds) left the public system when forced to transition to middle school. In light of that evidence, the achievement of sixth grade students in year t is given by:

$$\begin{aligned}\Delta y_{6,t} = \delta_R^5\gamma_R\Delta R_{1,t} + \delta_R^4\gamma_R\Delta R_{2,t} + \delta_R^3\gamma_R\Delta R_{3,t} + \delta_S^5\gamma_S\Delta X_{1,t}^S + \delta_S^4\gamma_S\Delta X_{2,t}^S \\ + \delta_S^3\gamma_S\Delta X_{3,t}^S + \delta_S^2\gamma_S\Delta X_{4,t}^S + \delta_S\gamma_S\Delta X_{5,t}^S + \gamma_S\Delta X_{6,t}^S + \Delta\epsilon_{6,t}.\end{aligned} \tag{4.9}$$

In California, schools are divided among K-5 and K-6 configurations relatively evenly (as shown in Table A.7). If a proportion ψ (estimable from the regression discontinuity analysis) leaves the public system when students move to a new school, then we have that $\Delta X_{6,t,K6}^S = \psi\Delta X_{6,t,K5}^S$, where the extra subscripts 'K5' and 'K6' represent students who are in schools with a K-5 and a K-6 configuration, respectively. Taking the difference in achievement between sixth grade students in a K-6 configuration and those in a K-5 configuration gives:

$$\begin{aligned}\Delta y_{6,t,K6} - \Delta y_{6,t,K5} &= \gamma_S\Delta X_{6,t,K6}^S - \gamma_S\Delta X_{6,t,K5}^S + \Delta\epsilon_{6,t,K6} - \Delta\epsilon_{6,t,K5} \\ &= (1-\psi)\gamma_S\Delta X_{6,t,K6}^S\,,\end{aligned} \tag{4.10}$$

allowing us to solve for the γ_S parameter, given that we know ψ (already recovered from the regression discontinuity analysis).

Using a similar structure, we also uncover the fade-out rate of the reform, δ_R, and the fade-out rate of the effect of the change in student demographics, δ_S.[57] To do so, we take the parameters γ_R and γ_S to be known and construct the following two differences: (i) between fourth grade and third grade test scores in the 2000-01 school year, and (ii) between fourth and fifth grade test scores in the 2000-01 school year. These two equations (which can be found in Appendix G.1) form a system of two non-linear equations with two unknowns (δ_R,

[57]We are restricted to identifying the fade-out parameters only, rather than the non-parametric effect of the reform in each period, due to the change in the test format for the 2003-04 school year.

δ_S), which we then solve for.

4.2 Identification

Before presenting the main estimates in the next section, we first discuss the identification of the key parameters.

The Direct CSR Effect (γ_R): Identification of γ_R requires a third and fourth grade cohort within the same year (to avoid year effects), whereby the fourth grade cohort has been treated in grades 1-3 and the third grade cohort has been treated in grades K-3. Due to the timing of CSR implementation across grades, this unusual sequence of treatments occurs in our context in the 2001-02 school year.

Specifically, the 2001-02 third grade cohort received class size reduction in that year and was subject to the class size reform for the previous three years: kindergarten (1998-99), grade 1 (1999-2000), and grade 2 (2000-01). The 2001-02 fourth grade cohort, in contrast, did not receive class size reduction in the current year (since fourth grade is not eligible), but did so in the prior three years, namely first grade (1998-99), second grade (1999-2000), and third grade (2000-01). Note that this cohort did not receive class size reduction in kindergarten (1997-98). The average achievement level of these two grades in 2001-02 can therefore be differenced, just as in the model equations (4.7) and (4.8), to solve for γ_R. Observed achievement differences between these two cohorts are therefore attributable to the fact that the third grade cohort received class size reduction contemporaneously, while the fourth grade cohort did not.

Test scores for the third grade cohort may differ from the fourth grade cohort even had there been no class size reform. To account for these differences, we impose the assumption of common trends across grades and use observed test scores in the pre-CSR time period (1997-98) as the counterfactual difference in test scores between third and fourth grade in the absence of the class size reform.

The Indirect Sorting Effect (γ_S): We identify γ_S with the 2001-02 fifth and sixth grade cohorts to add another layer of differencing to control for differences between students in schools with different grade configurations.[58] The 2001-02 grade cohort was subject to the policy since first grade (1997-98) and so faced three years of reduced class sizes (1997-98, 1998-99, 1999-00) and five years of altered demographics within the public system (1997-98, 1998-99, 1999-00, 2000-01, 2001-02). Differencing the achievement of students in a school with a K-6 configuration from those in a K-5 configuration therefore gives us the estimate of γ_S in equation (4.10).

Even without the reform, observed achievement differences among sixth grade students may arise because of innate differences across students in schools with different configurations. To account for those, we first control for such differences between students in K-5 and K-6 schools by differencing out fifth grade achievement in each school type from their sixth grade achievement. Then, as for γ_R, we assume common trends across grades and use observed test scores in the pre-CSR time period (1997-98) as the counterfactual difference in test scores between fifth and sixth grade in the absence of the class size reform.

Persistence (δ_R, δ_S): The persistence parameters are identified by solving the nonlinear system of equations found in Appendix G.1 (specifically, equations G.7 and G.8). Intuitively, we pin down the fade-out in the direct and indirect effects by using the 2000-01 third, fourth, and fifth grade cohorts. All three cohorts were affected through the direct class size reduction channel for three years, but were affected by the indirect channel for different lengths of time (three, four, and five years for third, fourth, and fifth grade, respectively). This allows us to separate the persistence of the indirect effect, δ_S, which affected some grades more than others, from the persistence of the direct effect that influenced all grades equally (although it was applied at different points in time).

Accounting for Teacher Quality

[58] Any fifth and sixth grade cohort after 2001-02 could, in principle, be used to identify γ_S. In practice, the change to test scores in the 2002-03 school year prevents us from using these cohorts.

We further extend our identification strategy by augmenting the equations to include indirect teacher quality effects. Given the number of new parameters introduced, we do not have sufficient degrees of freedom to identify them using variation in test scores alone. To overcome that obstacle, we follow Jepsen and Rivkin (2009) by appealing to variation in observable teacher experience as a proxy for teacher quality. Those authors document a pronounced increase in the overall proportion of inexperienced teachers upon the introduction of the California CSR reform, and a subsequent decline to pre-CSR levels after a few years. Given that our structural framework relies on variation across CSR and non-CSR grades over time, we draw on evidence about the way in which teacher inexperience evolved by grade, presented in Table A.8.[59] Since these changes are observable in each year, we can be non-parametric in our treatment of these indirect teacher quality effects, rather than following the 'geometric decay' treatment used for class size and student sorting.[60]

5 Estimates of the Direct and Indirect Effects

This section presents results from implementing our estimation approach, followed by a detailed discussion of how we interpret the main findings.

Table 10 provides estimates of the contemporaneous partial equilibrium effect of CSR (γ_R) and the general equilibrium effect of CSR on student composition (γ_S). To explain the table layout, the estimate for the latter parameter is calculated using two different assumptions about the proportion, ψ, of students who return to private school after completing all grades offered by the public school that they switched to initially. In columns (1) and (2), we follow the regression discontinuity evidence in Table A.4 and treat $\psi = \frac{2}{3}$, using the fact that two-thirds of the students are estimated to return to the private system when the middle school transition occurs. A lower bound estimate for γ_S is provided in columns (3) and (4) by assuming that *all* students who were drawn into the public system by CSR return to the

[59]Jepsen and Rivkin (2009) control implicitly for teacher observables that evolve by grade, using school-grade-year controls and grade-year fixed effects, and so do not document patterns at the grade-year level.

[60]Appendix F provides further details about our estimation approach when accounting for teacher quality.

private system during the middle school transition ($\psi = 1$). In keeping with the evidence in Jepsen and Rivkin (2009), we find that controlling for teacher quality is important and so only report structural estimates that include teacher quality controls.[61]

All estimates are highly significant. Focusing on column (2), which uses the estimated share $\psi = \frac{2}{3}$ and includes county fixed effects, the direct impact of CSR accounts for a 2.2 unit increase in the mean percentile rank of students, which corresponds to a 0.11σ increase in the school-grade test score distribution. The magnitude of this estimate is in line with experimental estimates: for instance, Krueger and Whitmore (2001) find that Project STAR raised test scores by around 0.1 standard deviations.

By way of contrast, results from a simple difference-in-differences specification[62] are shown in Appendix Table A.5 (with Appendix Figure A.5 indicating that pre-trends hold for test scores). These imply treatment effects in the region of 0.07σ for mathematics scores. We will see that these estimates underestimate our preferred estimates of the class size effect that account for persistence by over a third.[63]

The general equilibrium sorting effect accounts for a 3.3 unit increase in the mean percentile rank of students, which is equivalent to a 0.16σ increase in the school-grade test score distribution. This is precisely estimated, and larger in magnitude than the direct effect. We discuss the interpretation of this effect below – specifically, whether it is plausible to think that spillovers from incoming to existing public school students might be important.

We can compare estimates of the direct effect using our framework with reduced form estimates that compare sixth grade versus fifth grade in K6 schools and K5 schools. Doing so yields estimates very close to the general equilibrium sorting effects we identify based on our

[61] The teacher quality estimates themselves are given in Table A.9.

[62] Specifically, we run the following event study regression: $y_{sgt} = \beta_0 + \beta_1 post_{sgt} + \beta_2 treat_{sg} + \beta_3 post_{sgt} * treat_{sg} + \eta_s + \theta_t + \delta_g + \phi X_{sgt} + \epsilon_{sgt}$, where y_{sgt} is the average test score in school s in grade g at time t, $post_{sgt} \equiv \mathbb{1}\{CSR_3 \geq 0\}$ is an indicator variable that school s has implemented CSR in third grade, $treat_{sg}$ is a third grade dummy, and η_s, θ_t and δ_g represent school, year and grade fixed effects, respectively. β_3 is our coefficient of interest.

[63] As an aside, we note here that the D-in-D results align nicely with our estimates if the structural δs and γs are inserted into equation (4.3). This suggests that the additivity assumption provides a reasonable approximation.

estimation framework. The reduced-form approach involves a triple-differences regression,[64] which is reported in Appendix Table A.6. The results we find here range between 0.10-0.13σ, similar to those we obtain from our estimating framework, helping to corroborate those results since they use a similar source of variation to the reduced-form analysis.

Turning to the persistence parameters, δ_R and δ_S, we find that both the direct and indirect effects fade out in the range of 45-70 percent each year. These estimates are in line with much of the literature on fade-out, which finds that the class size test score gain is "reduced approximately to half to one quarter of its previous magnitude" (Krueger and Whitmore 2001, page 11), although such test score gains then reappear later in the labor market (Chetty et al. 2011). These estimates are also consistent with fade-out estimates in the teacher effects literature (see Jacob, Lefgren and Sims 2010 and Kinsler 2012).

To summarize, the evidence indicates that general equilibrium student sorting in response to a reform that improves school quality is first order: it is at least as great as the large direct partial equilibrium effect we find. Thus, the indications are that focusing only on the direct channel may substantially underestimate the overall effect of CSR. More generally, to the extent that CSR is representative of other major reforms intended to improve school quality, estimates of those overall effects that abstract from induced sorting are likely to suffer from considerable omitted variables bias, points we develop next.

[64] Specifically, we restrict our data to grades g and $g-1$ and schools with K-5 or K-6 configurations. We then use the following regression:

$$\begin{aligned} y_{sgt} &= \alpha + \phi_g G_g + \phi_k K5_s + \phi_t post_t + \zeta_{gk} G_g * K5_s + \zeta_{gt} G_g * post_t + \zeta_{kt} K5_s * post_t \\ &\quad + \Phi_{K6-K5,g-(g-1),post-pre} G_g * K5_s * post_t + \phi X_{sgt} + \epsilon_{sgt} , \end{aligned} \tag{5.1}$$

where y_{sgt} is the test score in school s in grade g at time t, G_g is an indicator for grade g, $K5_s$ is an indicator for the K-5 grade span configuration (i.e. $K5 = 1$ denotes the K-5 configuration), $post_t$ refers to the 2001-02 school year and later, and X_{sgt} represent school-grade-year characteristics. Our coefficient of interest is $\Phi_{K6-K5,g-(g-1),post-pre}$, which compares g and $g-1$ grade scores between K-5 and K-6 schools before and after the 2001-02 school year (when the first CSR cohort entered sixth grade). Since 2001-02 represents the first sixth grade cohort that experienced CSR, we expect that $\Phi_{K6-K5,6-5,post-pre}$ will be positive and (roughly) similar in magnitude to γ_S. All other triple-differences between adjacent grades are placebo tests.

5.1 Interpretation

We interpret our main estimates in three parts. First, we discuss the likely extent of spillovers experienced by existing public school students arising from the estimated indirect sorting effect. Then we consider the magnitude and policy implications of the estimated indirect sorting effect, followed by its potential relevance in other contexts.

Spillovers

To interpret the indirect effect identified by γ_S, we wish to shed more light on the relative magnitude of its two components – the compositional ('own') effect and the spillover effect – noting the data limitations in a Californian context. The compositional effect occurs mechanically because students who would have enrolled in a private school in the absence of the reform would be expected to score higher on standardized tests (on average) than their public school counterparts. The spillover effect occurs because public school students might receive benefits from their new classmates, most likely through peer effects.

To gauge the relative impacts of the two, we consider two contrasting scenarios. In the first, we assume that there are no spillovers, so the indirect effect is due entirely to the compositional channel. In the second, we use California's pre-CSR private-public test score gap[65] combined with an assumption that the marginal private school student is high up the private school test score distribution. This assumption is motivated by models such as that of Epple and Romano (1998), in which high-ability, low-income students are likely to be the most responsive to an increase in public school quality.

We see from the estimates in Column (4) of Table 6 that there is a 1.8 percent increase in the proportion of students who, in the absence of the reform, would have entered the private rather than the public system. An average public school in the sample has K-3 enrollment of approximately fifty students per grade, indicating that it receives around one marginal

[65]Specifically, we use California's private-public fourth grade test score gap of 0.54σ from the 1996 NAEP (available at `https://nces.ed.gov/nationsreportcard//pdf/main1996/97488.pdf`).

private school student. In the 'no spillovers' scenario, the entire 0.16 school-grade standard deviation increase in test scores that is attributed to the indirect effect of the reform would be caused by the compositional change. For this indirect effect to be due *solely* to that change, the students induced into the public system by the reform would have to score, on average, 2.7σ higher than students in the public system – an implausibly large test score gap.[66] With California's public-private school test score gap being around 0.54σ, it seems highly unlikely that the entire indirect effect could be attributed solely to the change in the composition of students in the public system.

Under the second scenario, we assume that the marginal students brought into the public system are relatively high-ability private school students (consistent with Epple and Romano (1998)). For concreteness, we suppose that the marginal private school student entering the public system due to CSR is at the 75th percentile of the private school test score distribution and thus scores 1.56σ higher than the average public school student. Under this scenario, the composition effect leads to a 0.09σ (that is, $3 * \frac{1.56}{50}$) increase in the school-grade test score distribution. Peer effects must therefore account for the remaining 0.07σ of our indirect test score effect, implying a social multiplier of 1.75.[67] This social multiplier is very similar to the social multiplier estimated in Graham (2008), who uses Project STAR data to estimate a linear-in-means peer effects model and finds a social multiplier of 1.9. We therefore find that about forty percent of the indirect effect come from positive spillovers onto public school students arising from peer effects.

[66] An extra student in a school-grade of fifty students who scored 2.7σ higher would increase the average student-level standard deviation of test scores by 0.054. Since class-level standard deviations are much smaller than individual-level standard deviations, we multiply the 0.054 increase in the student-level standard deviation by three (0.054*3=0.162) to place that test score change in the distribution of school-grade test scores. (See Finn and Achilles 1990 where effect sizes are three-fold in the distribution of class means relative to individual means, for instance.)

[67] The social multiplier is given by $1 + \frac{0.07}{3 * \frac{1.56}{50}}$.

Policy Implications of the Sorting Effect

Our application has focused on class size reduction – a type of education policy that is controversial, given long-standing debates about its benefits and the clear acknowledgement that it can be very expensive.

In the absence of substantial positive general equilibrium effects, CSR is often viewed as a relatively unattractive policy, given its enormous costs. Brewer, Krop, Gill and Reichardt (1999) estimate that reducing class sizes to 18 (from 24) for students in first through third grade in the United States would require hiring an additional 100,000 teachers, at a cost of \$5-6 billion per year. As discussed in Hanushek (1999), such numbers make it unclear whether CSR policies would pass a sensible cost-benefit test, particularly since Brewer et al. (1999) did not account for the additional costs of the five states (including California) that had previously implemented CSR. In a meta-analysis, Hattie (2005) finds that reducing class sizes from 25 to 15 improves student achievement by about 0.10-0.20σ; yet class size reduction ranks well-down – fortieth out of forty-six possible interventions – intended to serve the same end.

Our analysis sheds new light on the benefits of class size reduction once general equilibrium sorting is accounted for, with an indirect sorting effect at least as great as the direct effect that has been the focus of much of the prior literature. Further, the indirect effects we have estimated will be magnified, given the evidence of positive persistence we uncover. In this regard, our results accord with the convincing studies that document longer-term benefits of class size reduction, focusing on Project STAR – see Krueger and Whitmore (2001) and Chetty et al. (2011).

Against the considerable costs, we can construct use our framework to construct a measure of the overall benefit of CSR. After one year of exposure, we expect that students in the public school system should score 0.21σ higher, with 0.11σ of that coming from the direct effect, 0.16σ from the indirect effect and the general equilibrium effect to teacher quality causing about a 0.06σ *decline* in test scores.[68]

In addition, we can use the linear technology from equation 4.1 combined with our estimated persistence parameters to estimate the full effects of CSR on cohorts that experienced the full four years of the program. To do so, we assume that teacher quality effects persist at the same rate as the direct effect (with similar persistence to Chetty, Friedman and Rockoff 2014) and find that fourth grade students who experienced CSR in grades K-3 score 0.39σ higher[69] than they would have in the absence of CSR.[70]

Noting the magnification of the benefits in light of the positive sorting (and spillover) effects we have found, it is important to emphasize that policy makers should place CSR alongside feasible reforms that target incentives – for schools, teachers, students, and even parents – as a means of improving education outcomes. Such a comparison is beyond the scope of this paper.

Indirect Sorting Effects: General Relevance

The size of estimates of the indirect sorting effect we obtained from California's CSR reform are likely to carry over to other settings when certain preconditions hold, each serving to increase its size. In particular, the reform-related shock to public school quality needs to be large; there should be a non-trivial share of households with children in private school pre-reform; and the characteristics of students in private versus public school have to differ in order to generate changes in peer quality post-reform (distinct from changes in scale).[71]

[68] This estimate comes from subtracting the CSR teacher quality from the non-CSR teacher quality in the final year we have data (2001-02) in Table A.9. Our general equilibrium teacher effect is significantly larger than the -0.01σ found in Jepsen and Rivkin (2009), although our higher general equilibrium teacher quality effect only acts to *dampen* the total impact of CSR.

[69] We find the overall fourth grade test score effect by plugging in the parameter estimates from Tables 10 and A.9 (in s.d. units) into equation (4.1). Formally, we plug our parameter estimates ($\hat{\gamma}_R = 0.111$, $\hat{\gamma}_S = 0.163$, $\hat{\gamma}_Q = -0.061$, $\hat{\delta}_R = 0.448$, $\hat{\delta}_S = 0.565$, δ_Q(assumed)$= \hat{\delta}_R$) into: $\hat{y}_4 = \sum_{\tau=1}^{4} \hat{\delta}_R^\tau \hat{\gamma}_R + \sum_{\tau=0}^{4} \hat{\delta}_S^\tau \hat{\gamma}_S + \sum_{\tau=1}^{4} \delta_Q^\tau \hat{\gamma}_Q$.

[70] This overall estimate is roughly in line with findings from Unlu (2005), who compared California NAEP scores to other states before and after CSR and found that four years of exposure to CSR raised fourth grade mathematics test scores by $0.2\text{-}0.3\sigma$.

[71] Further, student peer characteristics need to be relevant in the production of student achievement, as they indeed are.

The evidence indicates that all these pre-conditions hold in a Californian context. More broadly, the conditions in California are quite similar to those in other states. At the time of CSR, it ranked 20th (out of 50 states) in its private school enrollment rate (Yun and Reardon 2005). It also has a large private-public test score gap as in most other states – Altonji, Elder and Taber (2005) find a national eighth grade private-public test score gap of 0.4σ, for instance. Thus we expect similar sorting responses to large reform-related shocks to public school quality elsewhere in the United States.

Once the reform has occurred, two additional factors make the sorting effect larger – namely, that the private school students are responsive to relative changes in quality between public and private schools, placing a large share on the margin of switching;[72] and private schools need to be relatively passive to the reform.

To keep track of such features from a policy prediction perspective, one could envisage a model of public and private school behavior and 'consumer' choice, though estimating that would require more disaggregated data covering the relevant economic agents than we currently have access to.[73] Nevertheless, in a Californian context, with more aggregated data, we have shown that private school students differ from their public school counterparts (as one would expect). At the same time, there is clear evidence of adjustment on the part of private schools, which serves to mitigate the size of the sorting effect we have estimated. Specifically, following CSR, fewer private schools enter in the state, relative to trend, and more private schools exit, consistent with the evidence in New York City presented in Dinerstein and Smith (2016). On the quality margin, we also see suggestive evidence that private schools in the state responded to the boost in public school quality associated with CSR by *lowering* their own class sizes.[74]

[72] To the extent that more students are marginal, so the contrasts (under the third precondition) are less likely to be pronounced.

[73] For example, Bayer, Ferreira and McMillan (2004) uses an equilibrium sorting model to gauge the reinforcing effect of improvements to public school quality.

[74] These results are available on request.

6 Conclusion

In this paper, we have presented a transparent new empirical approach for gauging the extent of general equilibrium sorting responses to major policies. These are typically difficult to pin down and so have not been a prime focus of empirical research in a policy context. Yet we show them to be large – at least as important (in our education setting) as the direct effects of policy that have been carefully measured in the prior literature.

Our application used data from a major education reform in the late-1990s – California's CSR program. Without any doubt, CSR is immensely costly. That said, parents and teachers routinely and actively lobby for smaller classes, pressuring politicians to implement class size reduction initiatives. It is clear that econometric estimates of the direct, contemporaneous effects of class size reduction – the main focus in the literature – are not the only metric with which to measure the success of CSR programs. Influential recent research – see Chetty et al. (2011) – has drawn attention to longer term effects of class size reduction policies in the context of Project STAR. In a setting where CSR programs are large in scale, general equilibrium responses may also arise that can both dampen the effects of CSR, most notably through the need to hire new teachers (as in Jepsen and Rivkin 2009), and also magnify the benefits, possibly through student sorting. Our understanding of the latter induced responses is limited, however, in no small part because of the difficulty of finding independent variation needed to separate the indirect from the direct effects.

This paper has provided the first credible evidence in the literature relating to the magnitude of the indirect general equilibrium sorting channel, likely to matter in well-defined and often-encountered circumstances. Using the grade-specific timing of the reform, we documented important general equilibrium sorting responses, beginning with a significant decrease in private school share following California's CSR program, which in turn led to marked compositional changes in the public school system. Further, we showed that the combination of smaller classes and general equilibrium sorting was valued highly by parents, who were willing to pay substantially more to live in a region that had implemented CSR.

We then set out a framework that allowed us to estimate the direct effect of CSR and the indirect sorting effect on a common footing for the first time. Here we used similar sources of across-grade/over time variation, while controlling for changes in observable teacher quality. We found the direct effect to be 0.11σ (in terms of mathematics scores), and the indirect effect to be even larger (0.16σ). Our estimation approach also allowed us to recover the persistence of the direct and indirect effects of a major reform. Once these are accounted for, we show that the benefit calculus changes markedly.

Beyond a class size reduction setting, our approach and estimates are relevant when assessing the effects of major reforms in other contexts. For example, alternative reforms with different cost implications – for instance, incentive-based policies – that boost public school quality are likely to change the mix of students across public and private systems, with consequences for education production. We leave assessing the indirect general equilibrium effects of such policies to future work.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy*, 113(1): 151–184.

Angrist, Joshua D, and Victor Lavy. 1999. "Using Maimonides' rule to estimate the effect of class size on scholastic achievement." *Quarterly Journal of Economics*, 114(2): 533–575.

Asadullah, M Niaz. 2005. "The effect of class size on student achievement: Evidence from Bangladesh." *Applied Economics Letters*, 12(4): 217–221.

Battistin, E., Joshua D. Angrist, and Daniela Vuri. 2017. "In a Small Moment: Class Size and Moral Hazard in the Italian Mezzogiorno." *American Economic Journal: Applied Economics*, 9(4): 216–249.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2004. "Tiebout Sorting, Social Multipliers and the Demand for School Quality." National Bureau of Economic Research Working Paper 10871.

Bianchi, Nicola. 2017. "The Indirect Effects of Educational Expansions: Evidence from a Large Enrollment Increase in STEM Majors."

Bohrnstedt, George W., and Brian M. Stecher. 2002. "What We Have Learned about Class Size Reduction in California. Capstone Report."

Brewer, Dominic J., Cathy Krop, Brian P. Gill, and Robert Reichardt. 1999. "Estimating the Cost of National Class Size Reductions under Different Policy Alternatives." *Educational Evaluation and Policy Analysis*, 21(2): pp. 179–192.

Buddin, Richard. 2012. "The impact of charter schools on public and private school enrollments." *Cato Institute Policy Analysis*, 707.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review*, 104(9): 2633–2679.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *Quarterly Journal of Economics*, 126(4): 1593–1660.

Cho, Hyunkuk, Paul Glewwe, and Melissa Whitler. 2012. "Do reductions in class size raise students' test scores? Evidence from population variation in Minnesota's elementary schools." *Economics of Education Review*, 31(3): 77–95.

Clark, David E., and William E. Herrin. 2000. "The impact of public school attributes on home sale prices in California." *Growth and Change*, 31(3): 385–407.

Dinerstein, Michael, and Troy Smith. 2016. "Quantifying the Supply Response of Private Schools to Public Policies."

Ding, Weili, and Steven F. Lehrer. 2010. "Estimating treatment effects from contaminated multiperiod education experiments: The dynamic impacts of class size reductions." *Review of Economics and Statistics*, 92(1): 31–42.

Dobbelsteen, Simone, Jesse Levin, and Hessel Oosterbeek. 2002. "The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition." *Oxford Bulletin of Economics and Statistics*, 64(1): 17–38.

Epple, Dennis, and Richard E. Romano. 1998. "Competition between private and public schools, vouchers, and peer-group effects." *American Economic Review*, 33–62.

Estevan, Fernanda. 2015. "Public education expenditures and private school enrollment." *Canadian Journal of Economics/Revue canadienne d'économique*, 48(2): 561–584.

Finn, Jeremy D, and Charles M Achilles. 1990. "Answers and questions about class size: A statewide experiment." *American Educational Research Journal*, 27(3): 557–577.

Funkhouser, Edward. 2009. "The effect of kindergarten classroom size reduction on second grade student achievement: Evidence from California." *Economics of Education Review*, 28(3): 403–414.

Gary-Bobo, Robert J., and Mohamed Badrane Mahjoub. 2006. "Estimation of Class-Size Effects, Using 'Maimonides' Rule': The Case of French Junior High Schools." CEPR Discussion Papers.

Gilraine, Michael. 2017. "Identifying Multiple Treatments from a Single Discontinuity with an Application to Class Size Caps." Unpublished manuscript.

Graham, Bryan S. 2008. "Identifying social interactions through conditional variance restrictions." *Econometrica*, 76(3): 643–660.

Hanushek, Eric A. 1999. "Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects." *Educational Evaluation and Policy Analysis*, 21(2): 143–163.

Hattie, John. 2005. "The paradox of reducing class size and improving learning outcomes." *International Journal of Educational Research*, 43(6): 387 – 425.

Hoxby, Caroline M. 2000. "The effects of class size on student achievement: New evidence from population variation." *Quarterly Journal of Economics*, 115(4): 1239–1285.

Imberman, Scott A., and Michael F. Lovenheim. 2016. "Does the market value value-added? Evidence from housing prices after a public release of school and teacher value-added." *Journal of Urban Economics*, 91: 104–121.

Jacob, Brian A., Lars Lefgren, and David P. Sims. 2010. "The persistence of teacher-induced learning." *Journal of Human Resources*, 45(4): 915–943.

Jepsen, Christopher, and Steven Rivkin. 2009. "Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size." *Journal of Human Resources*, 44(1): 223–250.

Kinsler, Josh. 2012. "Beyond levels and growth estimating teacher value-added and its persistence." *Journal of Human Resources*, 47(3): 722–753.

Krueger, Alan B. 1999. "Experimental estimates of education production functions." *Quarterly Journal of Economics*, 114(2): 497–532.

Krueger, Alan B., and Diane M. Whitmore. 2001. "The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR." *Economic Journal*, 111(468): 1–28.

Leuven, Edwin, Hessel Oosterbeek, and Marte Rønning. 2008. "Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway." *Scandinavian Journal of Economics*, 110(4): 663–693.

Lucas, Greg. 1996. "Sacramento Surprise – Extra Funds / Governor wants to use money to cut class size." *San Francisco Chronicle.*

Molnar, Alex, Philip Smith, John Zahorik, Amanda Palmer, Anke Halbach, and Karen Ehrle. 1999. "Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin." *Educational Evaluation and Policy Analysis*, 21(2): 165–177.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica*, 73(2): 417–458.

Rohlfs, Chris, and Melanie Zilora. 2014. "Estimating Parents' Valuations of Class Size Reductions Using Attrition in the Tennessee STAR Experiment." *BE Journal of Economic Analysis & Policy*, 1–36.

Schrag, Peter. 2006. "Policy from the Hip: Class size reduction in California." *Brookings Papers on Education Policy*, 2006(1): 229–243.

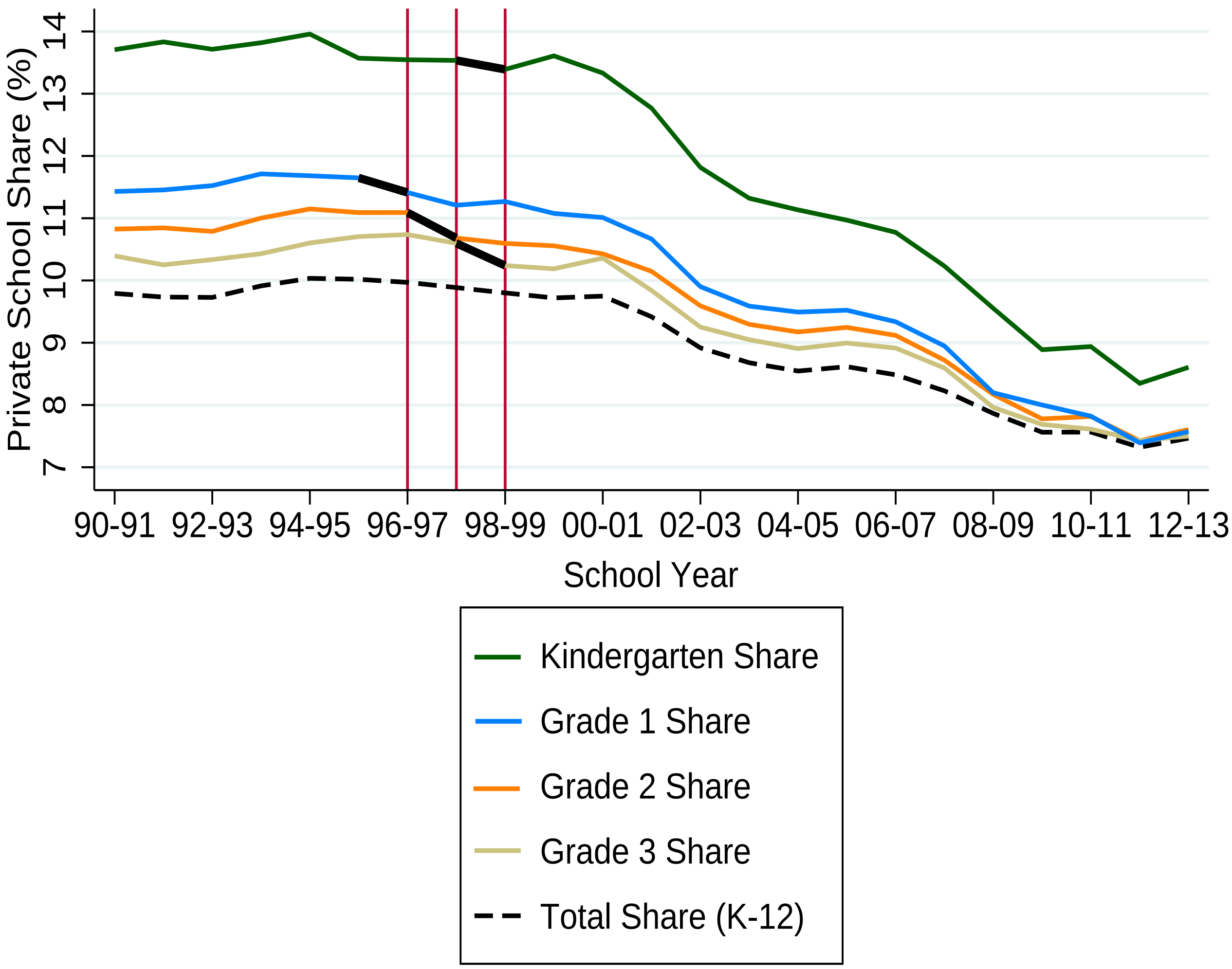
Stecher, Brian M., Daniel McCaffrey, and Delia Bugliari. 2003. "The relationship between exposure to class size reduction and student achievement in California." *Education Policy Analysis Archives*, 11(40): 1–27.

The College of William and Mary and the Minnesota Population Center. 2011. " School Attendance Boundary Information System (SABINS): Version 1.0. Minneapolis, MN: University of Minnesota."

Unlu, Fatih. 2005. "California class size reduction reform: New findings from the NAEP."

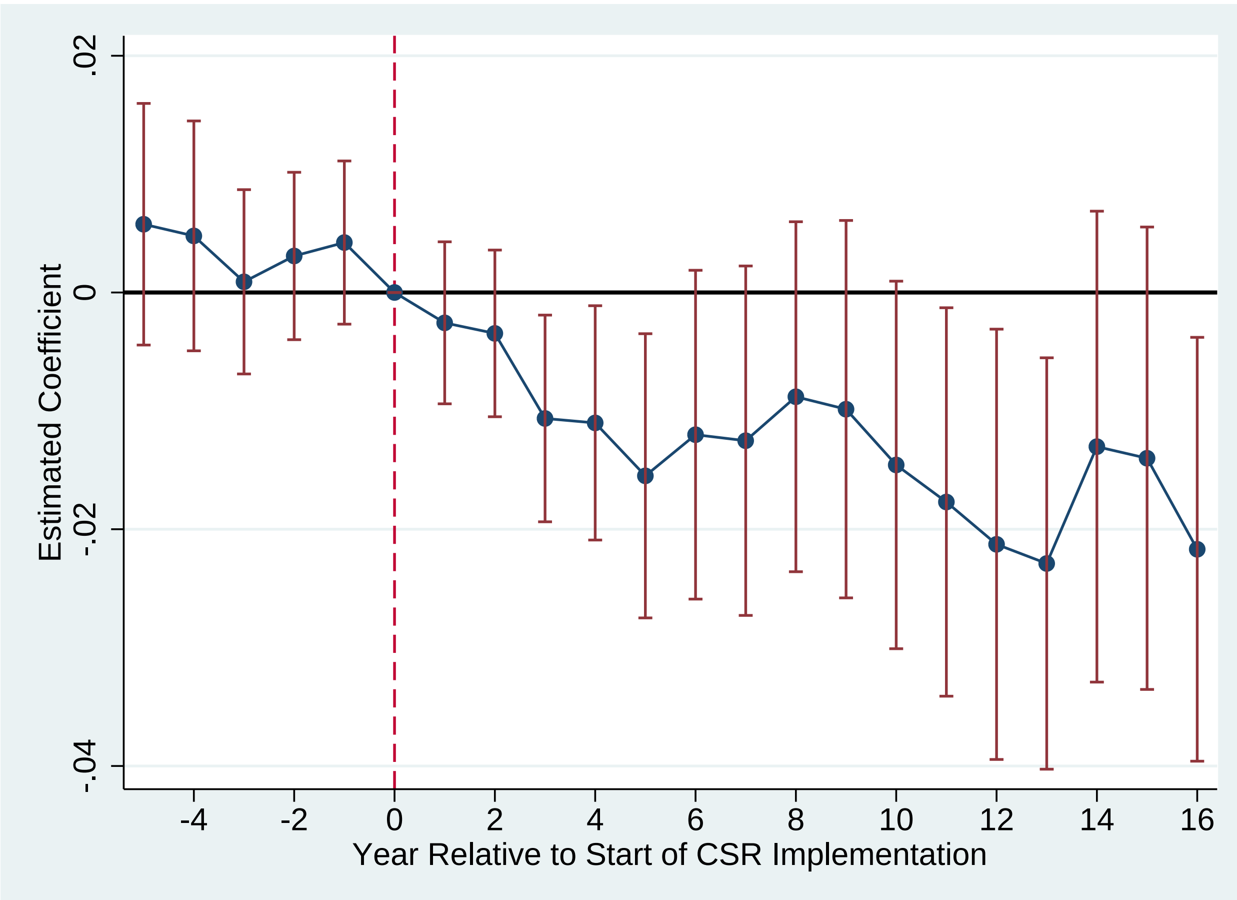
Yun, John T., and Sean F. Reardon. 2005. "Private school racial enrollments and segregation." *School choice and diversity: What the evidence says*, 42–58.

Figure 1: Private School Share Trends by Grade



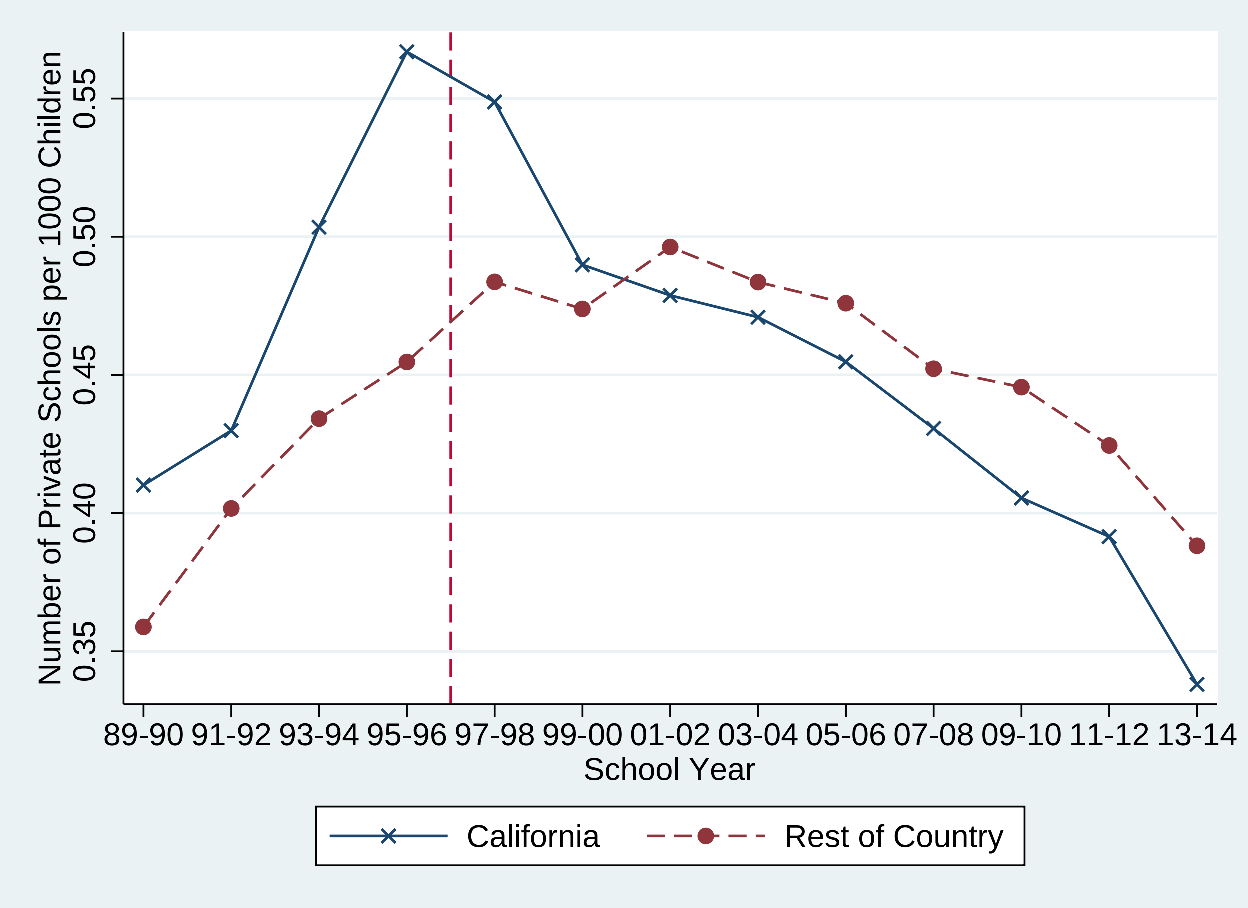
Notes: This figure shows aggregate private school share trends by grade over the two decades surrounding CSR. 'Private School Share' is defined as the aggregate number of students in private school in each grade in the state divided by the total number of public and private school students in that grade. Each year 'School Year' label corresponds to the start of the respective academic year. The vertical lines represent the start of school years 1996-97, 1997-98 and 1998-99 respectively, when different grades became eligible for CSR. Specifically, first grade became eligible for the 1996-97 school year, second grade for the 1997-98 school year, and third grade and kindergarten for the 1998-99 school year. The darkened thick line segments indicate the effect of CSR on the grade-level private school share when CSR was first implemented for that particular grade.

Figure 2: The Effect of CSR on Private School Share for Years Before, During, and After Implementation



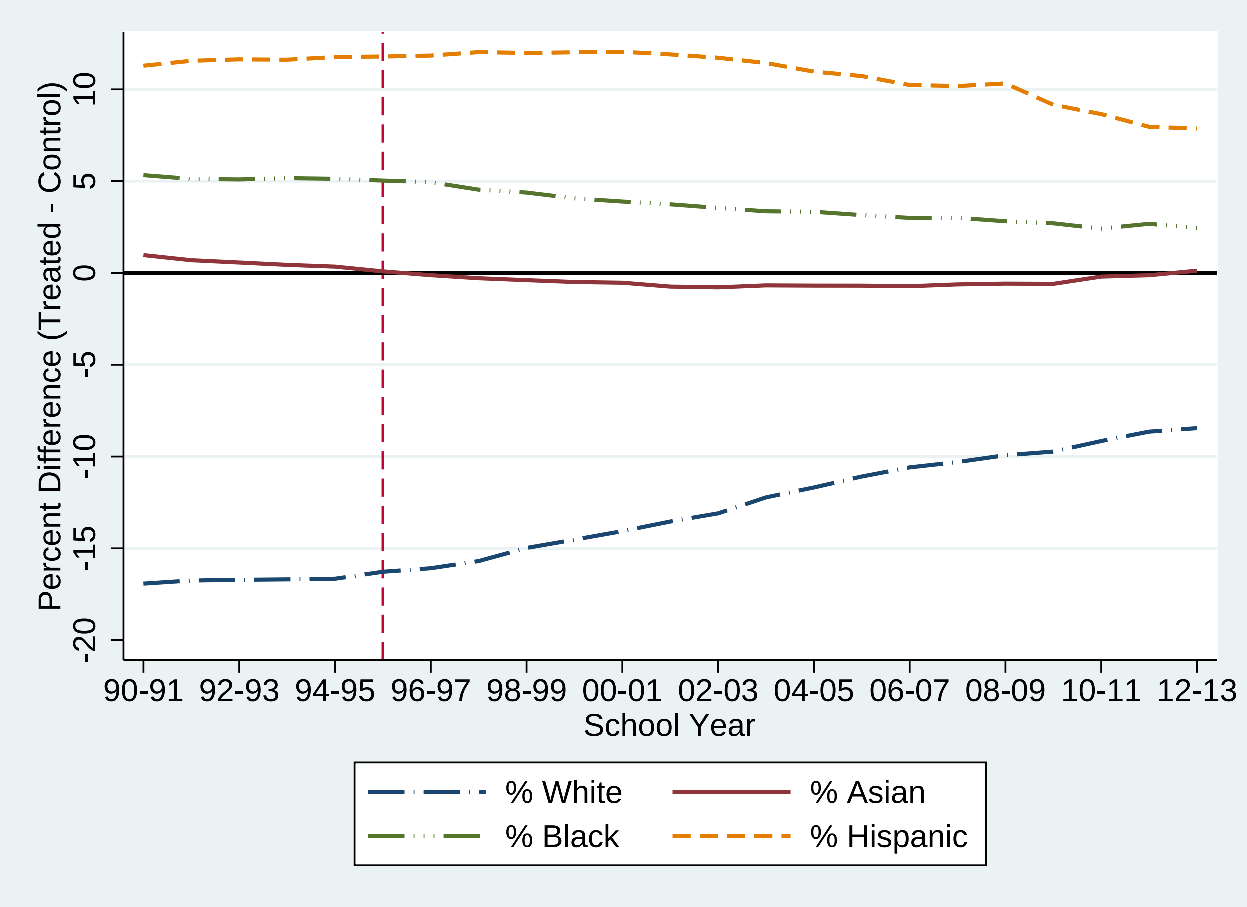
Notes: The figure shows the estimated effects of CSR on private school share by year relative to when CSR was implemented for a given grade. The figure uses the treatment of grades (CSR versus non-CSR) and district-level CSR participation intensity as the two dimensions of differencing. The dashed vertical line represents the start of CSR implementation in a CSR grade: for first grade, the vertical line represents the 1996-97 school year, for second grade, the 1997-98 school year, and for kindergarten and third grade, the 1998-99 school year. The horizontal line indicates an estimate of zero. The estimate at the start of CSR implementation is normalized to zero. Vertical bands represent 95% confidence intervals for each point estimate. Covariates and grade, year and district fixed effects are included. Standard errors are clustered at the district level.

Figure 3: Number of Private Schools per 1000 School-Aged Children by Year



Notes: The dashed vertical line indicates the 1996-97 introduction of the CSR reform. Data are available only every two years. The figure only includes private schools that primarily serve CSR grades. A private school is determined to serve CSR grades if, on average, the school consists of twenty percent or more students in K-3 in the 1989-90 through 2013-14 school years. The population of children are defined as all individuals aged 5-17 living in a state.

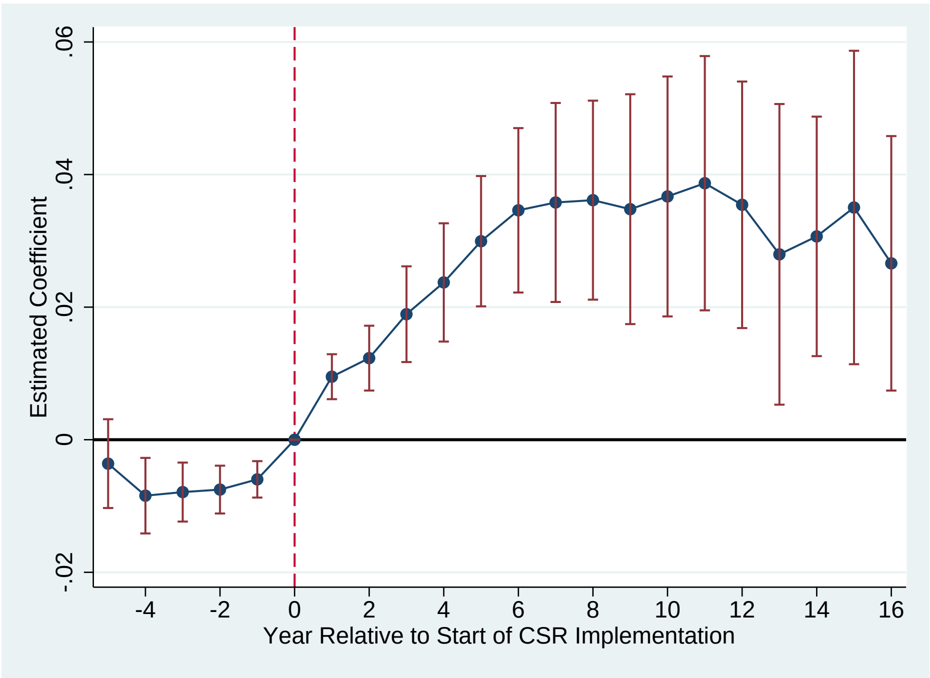
Figure 4: Demographic Trends by Public Schools: Treated (Private School within $3km$) minus Untreated



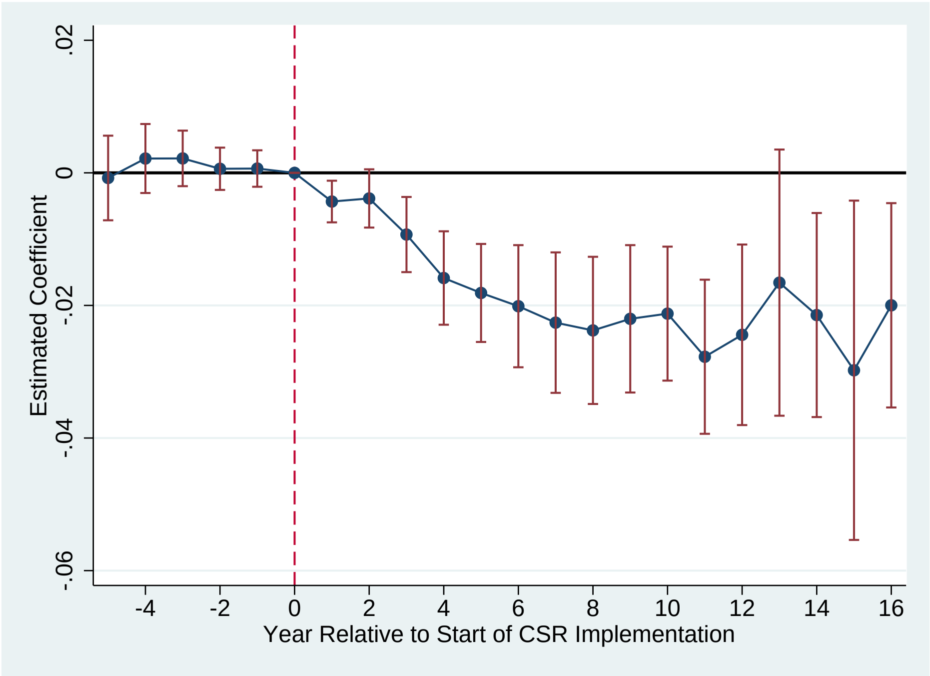
Notes: The figure shows the percent difference in demographics between public schools with a private school within $3km$ versus those that did not from 1990-91 through 2012-13. The data generating the figures are weighted by school K-3 enrollment. Each year label refers to the start of the respective academic year. The dashed vertical line represents the 1995-96 school year so that all periods thereafter incorporate the effects of CSR. The horizontal 'zero' line represents no difference between treated and control schools.

Figure 5: The Effect of CSR on Public School Demographics for Years Before, During, and After Implementation

(a) Percent White



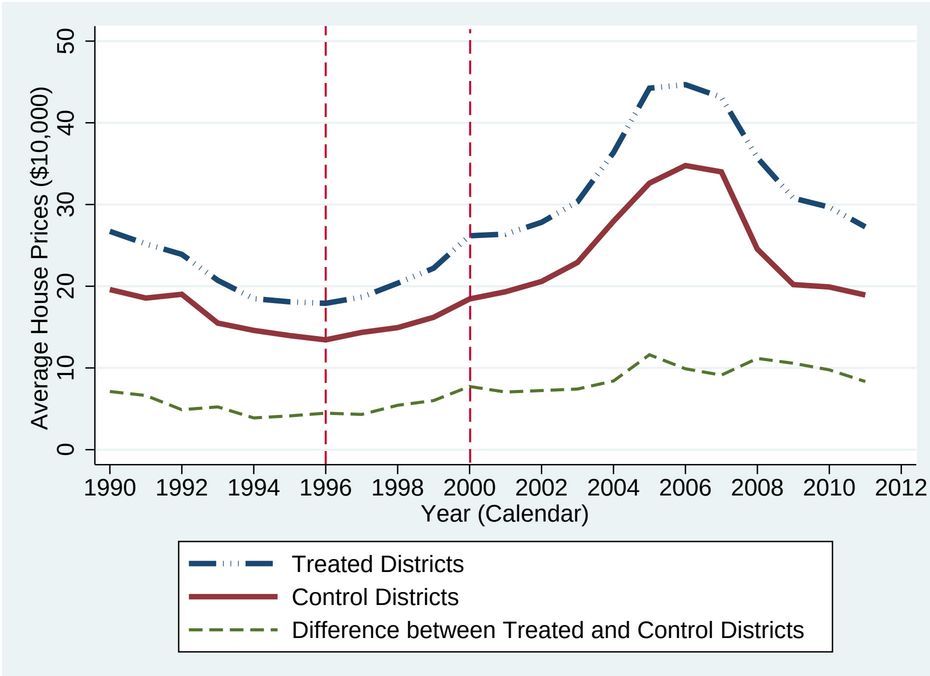
(b) Percent Hispanic



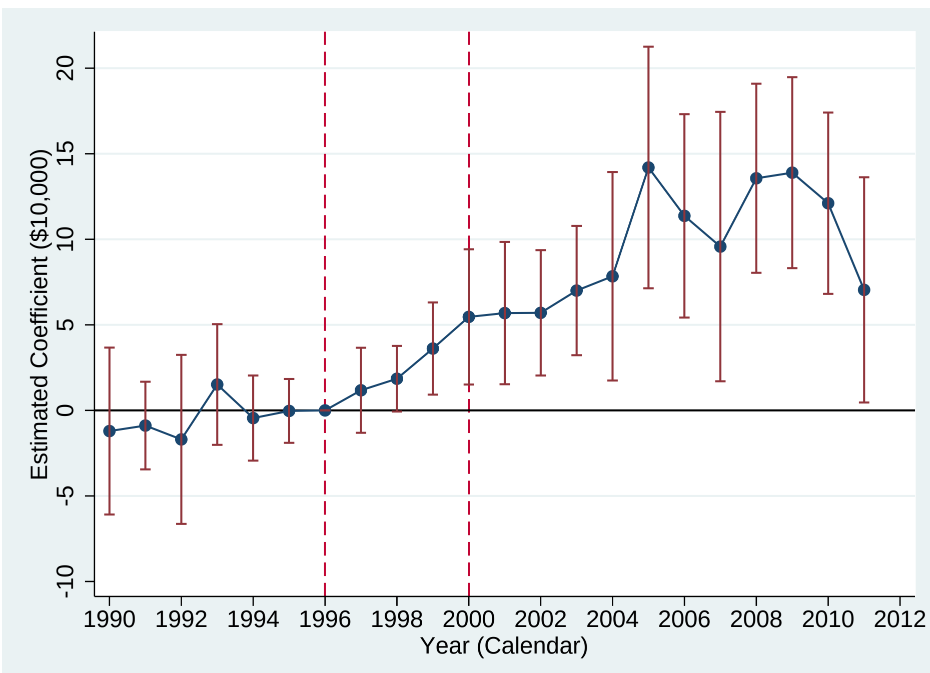
Notes: The above figures show the estimated effects of CSR on public school white and Hispanic demographics by year relative to when CSR was implemented for a given grade. The figures use the treatment of grades (CSR versus non-CSR) and whether a school is close to a private school (within $3km$) as the two dimensions of differencing. The dashed vertical line represents the start of CSR implementation in a CSR grade: for first grade, the vertical line represents the 1996-97 school year, for second grade, the 1997-98 school year, and for kindergarten and third grade, the 1998-99 school year. The horizontal line represents an estimate of zero. The estimate at the start of CSR implementation is normalized to zero. Vertical bands represent 95% confidence intervals for each point estimate. Grade, year and school fixed effects are included. Standard errors are clustered at the district level.

Figure 6: Visual Evidence of CSR on House Prices

(a) House Prices by Treatment Status



(b) Estimated Effects of CSR on House Prices by Year



Notes: Figure 6(a) shows average house prices (in 1995 dollars) in 'treated' districts (top three-quarters of CSR implementing districts in 1996-97) and 'control' districts (bottom quartile of CSR implementing districts in 1996-97). Figure 6(b) reports the effect of district CSR treatment intensity on house prices (in 1995 dollars) by year. The estimate at the start of implementation (in 1996) is normalized to zero. Vertical bands represent 95% confidence intervals for each point estimate. Demographic covariates are omitted, though house characteristics and district fixed effects are included. Standard errors are clustered at the district-level. Each year label refers to the calendar year. Dashed vertical lines represent the start of CSR implementation in the 1996-97 school year and first year (2000-01 school year) when CSR was fully implemented in all grades K-3, respectively. Horizontal lines indicate an estimate of zero.

Table 1: Coverage of Policy and Data Availability

	Year							
Grade	**1996-97**	**1997-98**	**1999-00**	**2000-01**	**2001-02**	**2002-03**	**2003-04**	**2004-05**
K	·	·	×	×	×	×	×	×
1	×	×	×	×	×	×	×	×
2	·	×	×	×	×	×	×	×
3	·	·	×	×	×	×	×	×
4	·	·	·	·	·	·	·	·
5	·	·	·	·	·	·	·	·
6	·	·	·	·	·	·	·	·

Notes: The reform is in effect for a particular grade-year if the corresponding cell contains a × symbol and it is not if it contains a · symbol. While the earliest grade of implementation is kindergarten (K), test score data are only available for grades two and above and from 1997-98 onward. The × symbols in the first two rows (and first column) are in a lighter shading to reflect this.

Table 2: Average Class Size by Grade and Year

Grade	School Year 1997-98	1998-99	1999-2000	2000-01	2001-02
			Average Class Size		
Kindergarten	24.2	**21.0**	**19.9**	**19.6**	**19.5**
First Grade	**19.2**	**19.2**	**19.2**	**19.2**	**19.2**
Second Grade	**19.4**	**19.2**	**19.1**	**19.0**	**19.0**
Third Grade	22.4	**20.1**	**19.6**	**19.4**	**19.3**
Fourth Grade	29.1	28.9	28.9	28.7	28.5
Fifth Grade	29.4	29.3	29.2	29.3	29.0

Notes: The numbers in the table represent average class sizes by grade and year. Grade-year combinations that were affected by CSR are in bold font. Since grade level class sizes are not observed before 1997-98, first grade and second grade have no pre-CSR comparison because those grades implemented CSR during the 1996-97 and 1997-98 school years, respectively. Some pre-kindergarten classes are included in the kindergarten average class size calculation.

Table 3: CSR versus non-CSR Implementing Schools and Districts

Student Demographics (in 1997-98)

	High-CSR Districts (1)	Low-CSR Districts (2)	CSR Schools (3)	Non-CSR Schools (4)
Percent White	40.4	32.9	37.8	31.9
	(26.0)	(23.8)	(29.6)	(26.3)
Percent Hispanic	37.6	49.6	42.2	47.8
	(23.0)	(25.3)	(29.6)	(28.1)
Percent Black	9.2	7.2	8.7	10.9
	(9.1)	(8.7)	(12.7)	(13.0)
Percent Asian	8.8	7.1	7.6	6.1
	(9.8)	(8.7)	(11.4)	(9.4)
Percent ESL	32.8	38.0	40.7	46.2
	(19.1)	(21.5)	(27.6)	(28.6)
Percent FRPM	46.2	56.0	55.3	59.4
	(22.6)	(22.5)	(30.6)	(31.6)
Enrollment	10,134	1,626	585	522
	(18,817)	(1,157)	(280)	(532)
Observations	7,384	2,809	4,791	526

Notes: All demographics are for the 1997-98 school year. High-CSR districts are in the top three quartiles of CSR implementation, while Low-CSR districts are in the bottom quartile of CSR implementation, meaning that less than 85 percent of their (enrollment-weighted) schools implemented CSR. CSR schools are defined as schools that had implemented CSR in kindergarten or third grade in the 1998-99 school year, while non-CSR schools had not implemented CSR in neither kindergarten or third grade in the 1998-99 school year. All summary statistics are enrollment-weighted. District demographics are for all public school students (K-12) within a district.

Table 4: Descriptive Statistics

	Mean (1990-91 to 2012-13)	Pre-CSR (90-91 to 95-96)	CSR (96-97 to 98-99)	Post-CSR (99-00 to 12-13)
A. School Data				
Elementary Student-Teacher ratio[1]	22.3	25.5	23.9	20.7
Private School Share (%)	9.0	9.9	9.9	8.4
(enrollment weighted)	(8.4)	(8.5)	(8.5)	(8.3)
CSR Intensity[2]	84.7	85.4	84.5	84.5
	(28.7)	(28.0)	(29.1)	(29.0)
% English Learner[3]	27.7	24.8	26.2	29.2
	(25.6)	(25.4)	(25.8)	(25.5)
% White	52.2	61.0	57.1	47.6
	(29.1)	(27.6)	(28.5)	(28.8)
% Hispanic	33.3	27.2	30.0	36.4
	(27.6)	(25.2)	(26.4)	(28.3)
% Black	4.0	3.9	4.2	4.0
	(7.5)	(7.8)	(7.9)	(7.3)
% Asian	4.5	4.1	4.4	4.7
	(8.2)	(7.1)	(7.8)	(8.7)
enrollment	582	533	572	604
	(2246)	(2135)	(2249)	(2288)
% Free and Reduced Price Meals[4]	40.5	37.0	43.1	43.1
	(25.7)	(24.2)	(25.9)	(26.7)
Observations (District-Grade-Year)	253,056	63,983	32,761	156,312
B. House Price Data	1990-2012	1990-96	1997-99	2000-12
Transfer Price (*10,000)	28.8	17.4	18.2	36.1
	(25.8)	(12.1)	(14.2)	(29.3)
Lot Size (/1000)	42.0	41.4	44.3	41.5
	(81.8)	(79.2)	(89.4)	(80.3)
Bedrooms	2.9	2.9	2.9	2.9
	(0.8)	(0.8)	(0.8)	(0.8)
Square Feet (/1000)	1.7	1.7	1.7	1.7
	(0.4)	(0.4)	(0.4)	(0.4)
Observations (District-Year)	11,321	2,508	2,009	6,804

[1] Elementary Student-Teacher ratio is calculated as the number of elementary school teachers divided by the number of K-6 students.
[2] 'CSR Intensity' measures the proportion of K-3 students in CSR school-grades in the 1998-99 school year. The measure varies slightly year-to-year due to district closures and missing data for some districts in some years (87% of observations are for districts with at least 20 years of data).
[3] Data only include public school students. Some observations are missing values for this variable. There are a total of 237,468 observations with non-missing values.
[4] This variable is only available at the district-year level.

Table 5: Effect of CSR on Private School Share

Dependent Variable: Private School Share (%)			
Variable	Untreated Grades (Grades 4-12)	Treated Grades (Grades K-3)	Difference (Untreated-Treated)
Before	8.88	11.76	- 2.87
	(8.80)	(7.62)	(0.25)
After	8.16	9.63	-1.47
	(8.71)	(7.31)	(0.28)
Change	0.73	2.13	-1.41
(Before-After)	(0.15)	(0.25)	(0.17)
Observations	165,950	87,106	253,056

Notes: Observations are at the district-grade-year level, and cover 1990-91 through 2012-13 school years. Means are weighted by district-grade-year enrollment. Standard errors for the difference-in-means cells are clustered at the district level.

Table 6: Difference-in-Differences Estimates of CSR on Private School Share

Outcome Variable: Private School Share (%)				
	(1)	(2)	(3)	(4)
Treatment*Post	-1.41***	-1.35***	-1.45***	-1.78***
	(0.17)	(0.18)	(0.28)	(0.26)
Treatment	2.87***	2.82***	3.95***	5.29***
	(0.25)	(0.52)	(0.50)	(0.51)
Post	-0.73***	0.26*	0.15	0.44**
	(0.15)	(0.15)	(0.16)	(0.18)
Year/Grade FE	No	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes
District FE	No	No	No	Yes
Number of Observations	253,056	253,056	215,139	215,139

Notes: Observations are at the district-grade-year level and cover the 1990-91 through 2012-13 school years. Demographic controls include student race, gender, English second language, enrollment and enrollment squared. All regressions are weighted by district-grade-year enrollment. Standard errors are clustered at the district level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

Table 7: Triple-Differences Estimates of CSR on Private School Share

Outcome Variable: Private School Share (%)

	(1)	(2)	(3)	(4)
Treatment*Post*CSR	-1.45**	-1.47**	-1.42*	-1.53**
	(0.61)	(0.61)	(0.75)	(0.66)
Treatment*Post	-0.00	0.11	-0.21	-0.39
	(0.53)	(0.53)	(0.60)	(0.51)
Treatment*CSR	2.44**	2.46**	2.25*	2.80***
	(1.09)	(1.09)	(1.32)	(0.99)
Post*CSR	1.91***	1.86***	1.43**	1.01
	(0.66)	(0.66)	(0.71)	(0.57)
Treatment	0.00	0.13	3.32**	2.84***
	(1.00)	(1.12)	(1.35)	(1.04)
Post	-2.54***	-1.42**	-1.08*	-0.42
	(0.58)	(0.63)	(0.65)	(0.50)
CSR	5.67**	5.66**	2.15	-
	(2.25)	(2.25)	(1.96)	
Year/Grade FE	No	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes
District FE	No	No	No	Yes
Number of Observations	233,466	233,466	200,568	200,568

Notes: Observations are at the district-grade-year level and cover the 1990-91 through 2012-13 school years. Demographic controls include student race, gender, English second language, enrollment and enrollment squared. All regressions are weighted by district-grade-year enrollment. Standard errors are clustered at the district level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

Table 8: Triple-Differences Estimates of Compositional Changes

Outcome Variable: Student Demographics (%)	Percent White (1)	Percent Hispanic (2)	Percent Black (3)	Percent Asian (4)
Treatment*Post*$\mathbb{1}\{Buffer < 1.5\ km\}$	3.16***	-1.69***	-0.65***	0.04
	(0.73)	(0.48)	(0.19)	(0.24)
Treatment*Post*$\mathbb{1}\{Buffer < 3\ km\}$	3.13***	-1.64***	-0.70***	0.06
	(0.74)	(0.49)	(0.19)	(0.26)
Treatment*Post*$\mathbb{1}\{Buffer < 5\ km\}$	3.15***	-1.62***	-0.71***	0.05
	(0.74)	(0.49)	(0.19)	(0.26)
% Share in Private School (1997-98)	52.93	17.21	7.10	12.30
% Share in Public School (1997-98)	38.75	40.49	8.75	11.14
School/Grade/Year FE	Yes	Yes	Yes	Yes

Notes: Observations are at the school-grade-year level, and cover 1990-91 through 2012-13 school years. There are 1,147,271 observations. Enrollment and enrollment squared are included as controls. 'Post' is defined based on a 'before' and 'after' CSR implementation dummy. The table refers to the regression design described by equation (3.4). $\mathbb{1}\{Buffer < x\ km\}$ is the distance from a private school that a public school must be to be considered 'treated'. Three alternative buffers are provided for robustness. Private and public school demographic shares from the National Center for Education Statistics for the 1997-98 school year are provided for reference. All regressions are weighted by school-grade-year level enrollment and standard errors are clustered at the district level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

Table 9: Difference-in-Differences Estimates of Impact on House Prices

Outcome Variable: Average House Price in 1995 Dollars ($10,000s)

	All Years (1990-2012)				*1994-2012 only*	
	(1)	(2)	(3)	(4)	(5)	(6)
CSR_d*Post	1.62* (0.91)	3.11*** (1.10)	3.43*** (0.90)	2.63*** (0.80)	2.03** (0.83)	1.88** (0.82)
CSR_d	9.44*** (2.08)	6.73*** (1.89)	-	-	-	-
Post	-1.44 (0.95)	-1.23 (1.12)	0.11 (0.91)	4.40*** (1.40)	4.17*** (1.36)	2.68** (1.16)
House Characteristics	No	Yes	Yes	Yes	Yes	Yes
District FE	No	No	Yes	Yes	Yes	Yes
Demographic Controls	No	No	No	Yes	Yes	Yes
Teacher Controls	No	No	No	No	No	Yes
Observations	14,580	14,580	14,580	14,531	12,213	12,213

Notes: Observations are at the district-year level. Columns (1)-(4) cover the 1990 through 2012 calendar years while columns (5) and (6) are restricted to the 1994 through 2012 calendar years because teacher controls are only available from 1994 onward. House prices are deflated to 1995 dollars using California's house price index. All regressions include cubic controls for enrollment and year fixed effects. House characteristics consist of number of bedrooms and quadratic controls for square feet and lot size. Teacher controls include experience and education levels. Demographic controls include student race, gender, and free and reduced price meal eligibility. Standard errors are clustered at the district level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

Table 10: Structural Estimates

Outcome Variable: Mathematics Test Scores

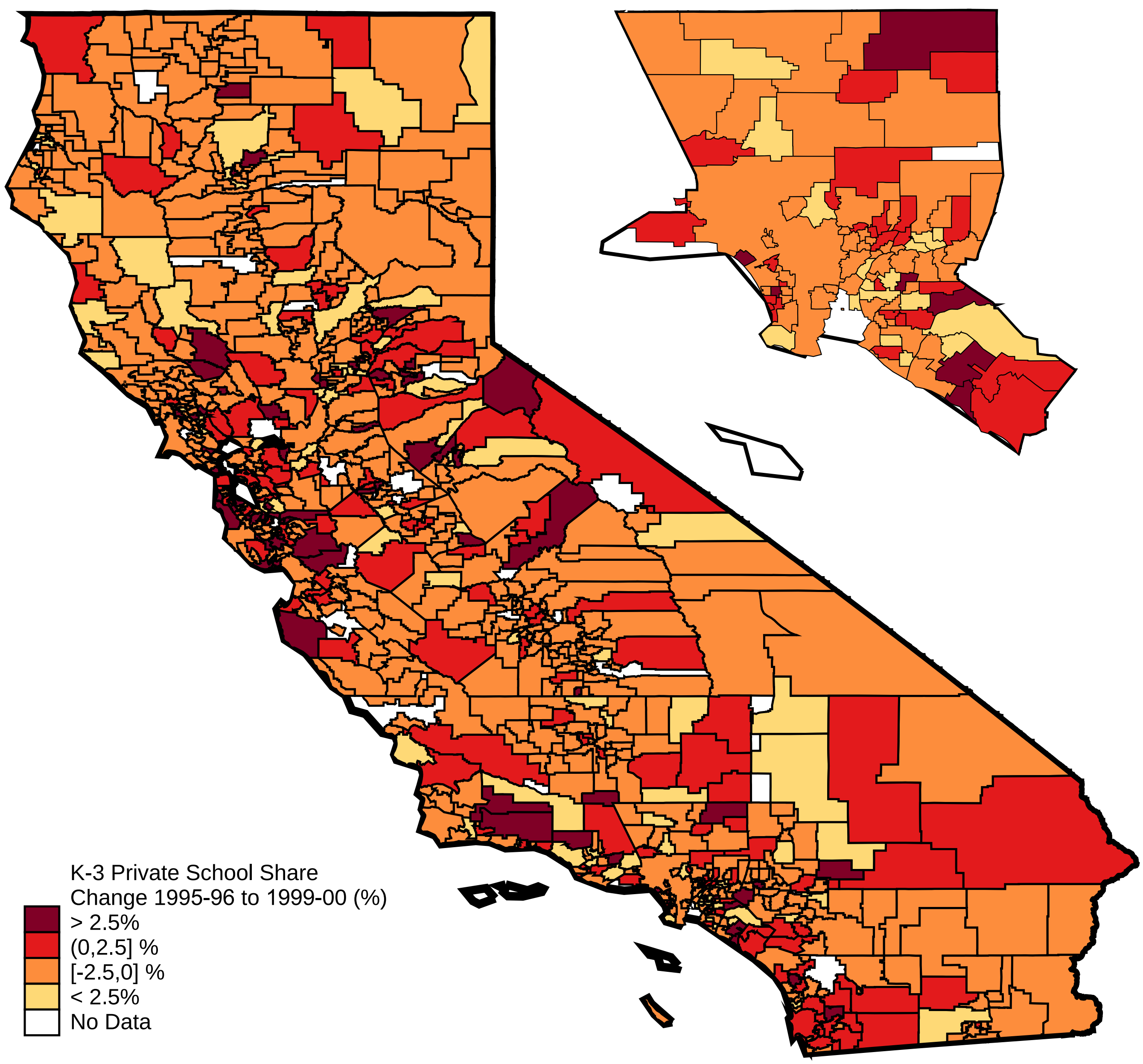
	With $\psi = \frac{2}{3}$		With $\psi = 1$	
	(1)	(2)	(3)	(4)
γ_R	2.10***	2.22***	2.10***	2.22***
	(0.20)	(0.20)	(0.20)	(0.20)
γ_S	2.27	3.26**	1.63	2.42**
	(1.69)	(1.54)	(1.24)	(1.13)
δ_R	0.49*	0.45**	0.50*	0.46**
	(0.26)	(0.21)	(0.26)	(0.21)
δ_S	0.64**	0.57**	0.70**	0.62**
	(0.30)	(0.27)	(0.31)	(0.30)
County FE	No	Yes	No	Yes
Observations	147,636	147,636	147,636	147,636

Notes: Observations are at the school-grade-year level, and cover the 1997-98 through 2003-04 school years. Mathematics test scores are shown in percentile ranks relative to a national norming sample, where one percentile rank roughly equates to 0.05σ in the distribution of school-grade level test scores. All parameter estimates include controls for teacher quality. Standard errors for γ_R and γ_S are computed using the delta method and are clustered at the school level. Standard errors for δ_R and δ_S are bootstrapped. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

A Appendix Figures and Tables

Figure A.1: Private School Share Change (1995-96 to 1999-2000)

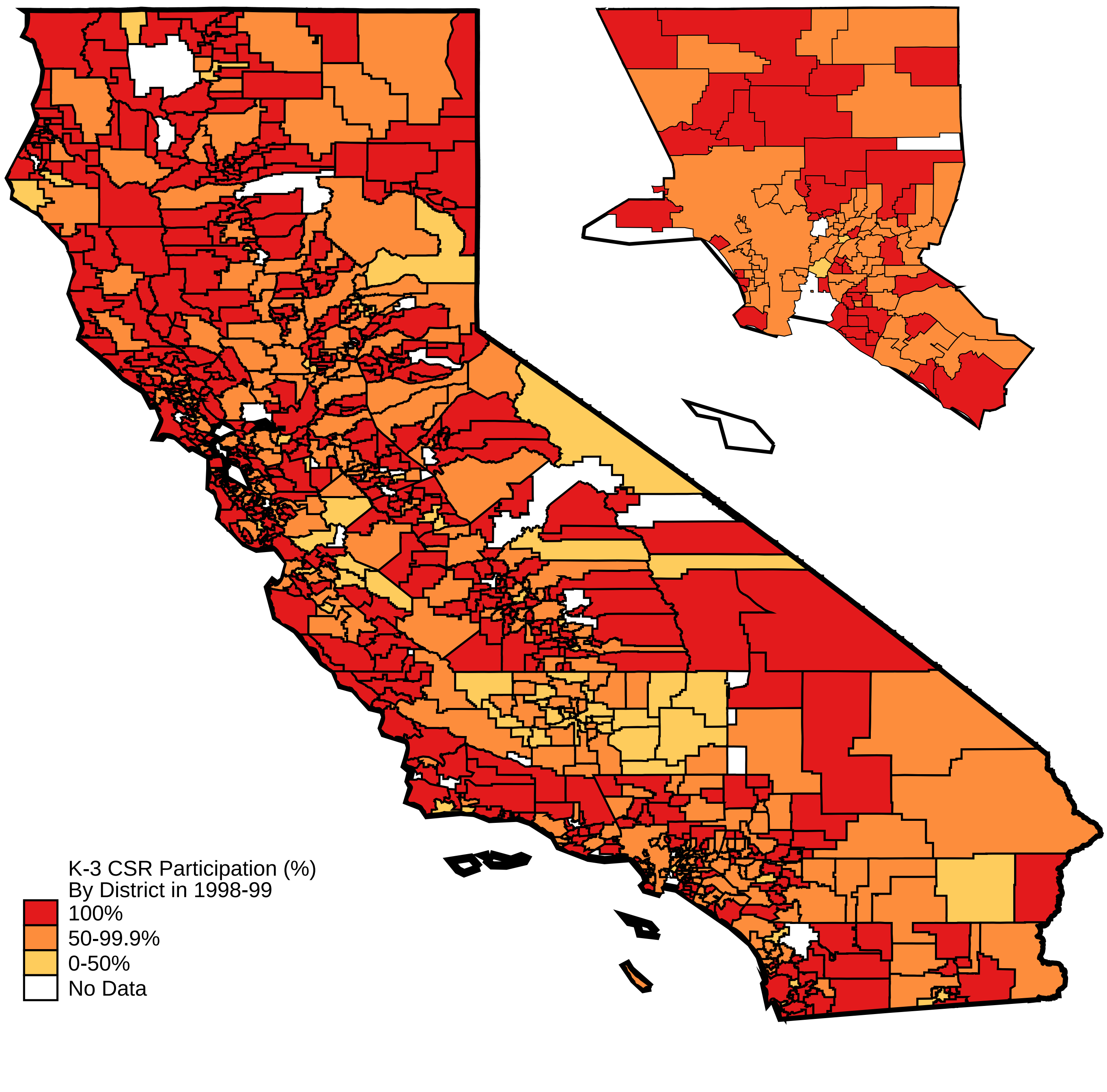
(a) California **(b)** Los Angeles and Orange Counties



Notes: The above figure shows the change in private school share for grades K-3 from 1995-96 to 1999-2000 school years for 876 school districts in California. Los Angeles and Orange Counties combined are shown separately for better visualization of that region. White areas denote regions that cannot be assigned to a school district.

Figure A.2: K-3 CSR Participation by District in 1998-99 ('CSR Intensity' Measure)

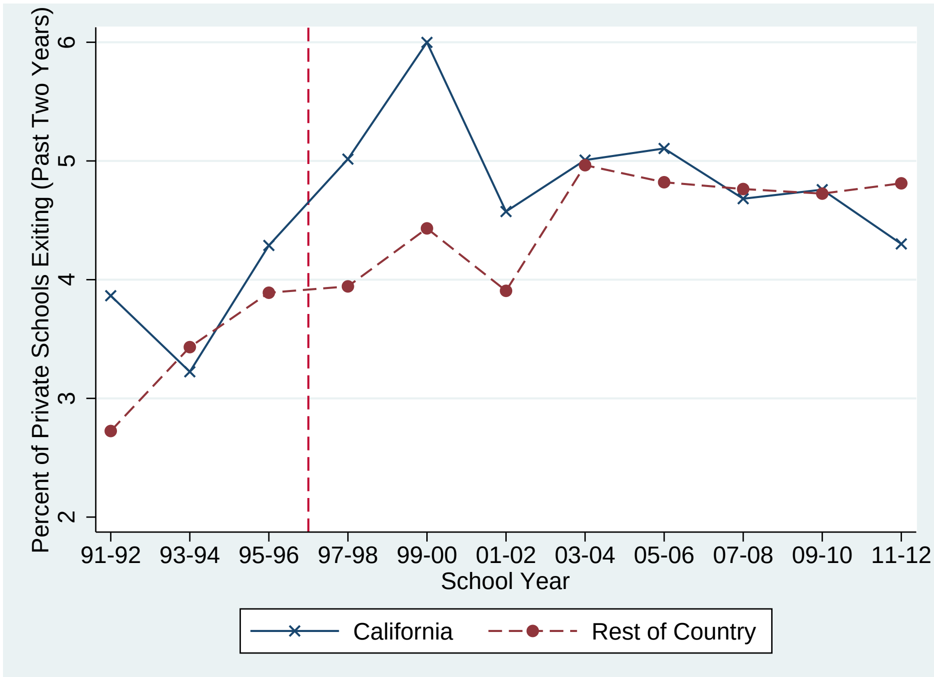
(a) California **(b)** Los Angeles and Orange Counties



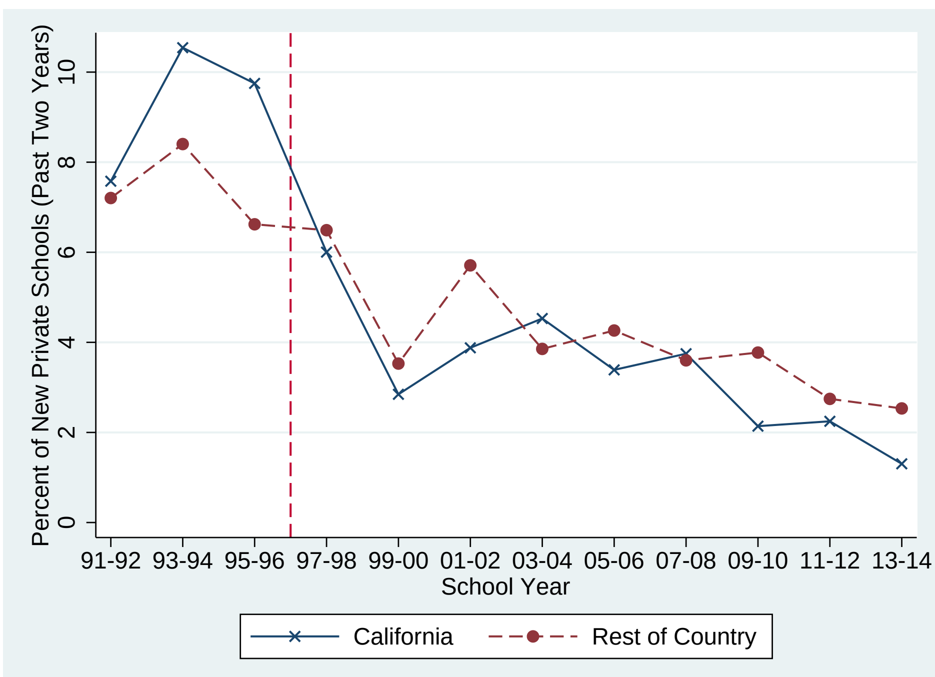
Notes: The above figure shows the percent of district-level K-3 enrollment in a CSR participating school-grade for the 1998-99 school year. Los Angeles and Orange Counties combined are shown separately for better visualization of that region. White areas denote regions that cannot be assigned to a school district.

Figure A.3: Biannual Private School Entry and Exit Rates

(a) Private School Exit Rates

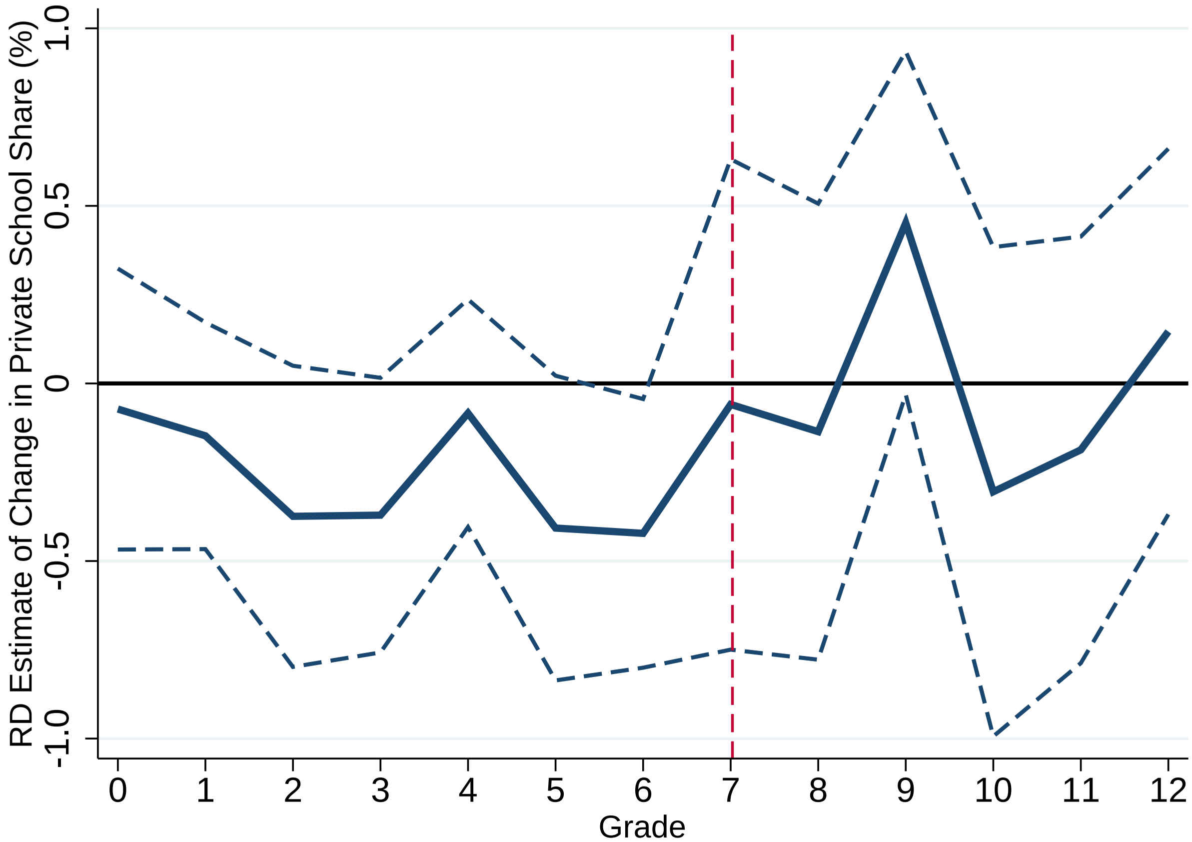


(b) Private School Entry Rates



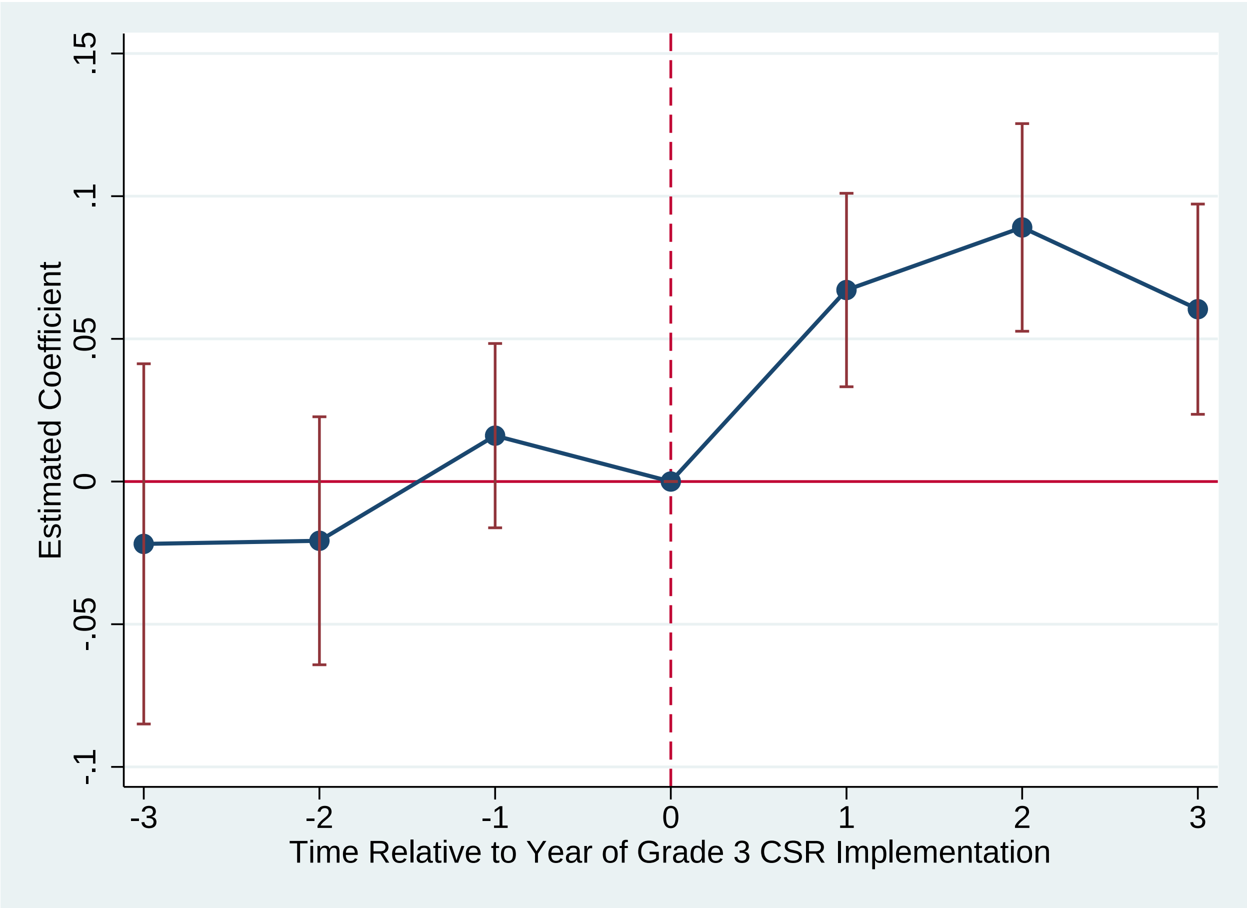
Notes: The dashed vertical line indicate the 1996-97 introduction of the CSR reform. Data are available only every two years. Figures only include private schools that primarily serve CSR grades. A private school is determined to serve CSR grades if, on average, the school consists of twenty percent or more students in K-3 in the 1989-90 through 2013-14 school years.

Figure A.4: The Effect of CSR on Private School Share by Grade



Notes: This figure shows the estimated effect of CSR on private school share for each grade using the RD design described in Appendix Section B. The vertical line at seventh grade indicates the first grade where (almost) all students have transitioned to middle school from their elementary school. The horizontal line represents an estimate of zero. The kindergarten effect here represents a placebo test as kindergarten was not a CSR grade for the cohorts around the discontinuity. The effect for each grade is estimated using a local linear regression allowing for a different functional form on either side of the cutoff. District fixed effects and demographic controls are included in all regressions. The bandwidth used is three. Standards errors are clustered at the district level.

Figure A.5: The Effects on Mathematics Test Scores by Event Time



Notes: The above figures show the estimated effects of CSR on public school mathematics test scores using a difference-in-differences design (see Section 5). The dashed vertical line represents the last year before CSR was implemented in the school in third grade and is normalized to zero. The horizontal line represents an estimate of zero. The estimate at the start of CSR implementation is normalized to zero. Vertical bands represent 95% confidence intervals for each point estimate. Demographic controls and grade, year and school fixed effects are included. Standard errors are clustered at the school level.

Table A.1: Data Availability and Sources

Data	Observation Level (1)	Years Covered (2)	Number of Observations (3)	Data Source (4)
Data Type: California Department of Education Data				
Public School Enrollment Data (includes race)	School-Grade-Year	1990-91 to 2012-13	1,147,271[a]	www.cde.ca.gov/ds/sd/sd/filesenr.asp
Private School Enrollment Data[b]	District-Grade-Year	1990-91 to 2012-13	316,069	www.cde.ca.gov/ds/si/ps/index.asp
Public School ESL Data[c]	School-Grade-Year	1990-91 to 2012-13	1,147,271	www.cde.ca.gov/ds/sd/sd/fileselsch.asp
Public School Free or Reduced-Price Meal Data	School-Year	1990-91 to 2012-13	200,848	www.cde.ca.gov/ds/sh/cw/filesafdc.asp
CSR Implementation Data	School-Grade-Year (grades K-3 only)	1998-99 to 2003-04	130,011	www.cde.ca.gov/ds/si/ps/index.asp
Standardized Testing and Reporting Data	School-Grade-Year (grades 2-11 only)	1997-98 to 2001-02	231,129[d]	star.cde.ca.gov
Teacher Assignment and Demographic Data	School-Grade-Year	1997-98 to 2001-02	222,626[d]	www.cde.ca.gov/ds/sd/df/filesassign.asp
Teacher Demographic and Experience Data	School-Year	1994-95 to 2010-11	166, 036	www.cde.ca.gov/ds/sd/df/filescertstaff.asp
Data Type: DataQuick House Price Data				
Housing Price Data (District-level)	District-Year	1990-2012	15,993	Private use data
Housing Price Data (School-level)	School-Year	1990-2012	50,139	Private use data
Data Type: Other Data				
Private School Universe Survey (State-level)	State-Year	1989-2014 (biannual)	663	nces.ed.gov/surveys/pss/
U.S. Population Data (State-level)	State-Year	1989-2014	1,326	seer.cancer.gov/popdata/download.html

[a] Only non-zero grade-level observations are included in this observation count.
[b] Private school enrollment data for 1990-91 through 1998-99 inclusive are not available on the CDE website. They were provided upon request by the CDE.
[c] California divides ESL students into English Learners and Fluent English Proficient. Since schools can alter students' ESL designations, we combine these two categories at the observation level into an ESL control to avoid picking up any endogenous responses in ESL designations following CSR.
[d] Data are available up to 2012-13, but we only use observations from 1997-98 to 2001-02 due to the switch from the Stanford Achievement Test to the California Achievement Test in the 2002-03 academic year.

Notes: All data can be aggregated to higher levels. Thus, 'school-grade-year' observations can be aggregated into 'district-grade-year' or 'school-year' observations.

Table A.2: Mathematics Test Score Summary Statistics

School Year	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
1997-98	44.6	43.6	41.4	43.3	50.6
	(19.2)	(19.5)	(19.1)	(19.7)	(19.0)
1998-99	44.6	43.6	41.4	43.4	50.6
	(19.2)	(19.5)	(19.1)	(19.7)	(18.9)
1999-00	58.5	58.2	52.4	52.2	58.8
	(18.6)	(18.1)	(18.6)	(19.4)	(18.2)
2000-01	59.8	61.1	55.4	55.8	61.4
	(18.0)	(17.5)	(18.2)	(18.9)	(17.8)
2001-02	62.6	63.5	58.1	58.2	63.1
	(16.9)	(16.8)	(17.5)	(18.1)	(17.3)
Total Observations (School-Grade-Year)	33,044	33,209	32,678	32,111	16,498

Notes: Test scores are for the Stanford 9 test and report the mean percentile ranking of students relative to a nationally representative reference group. The increase in test scores from the 1998-99 school year to the 1999-00 school year were due in part to a change in the norm-referencing group.

Table A.3: Triple-Differences on Number of Private Schools

Outcome Variable: Private Schools per 1000 School-Aged Children

	D-in-D (CSR Schools) (1)	D-in-D (non-CSR Schools) (2)	Triple-D (3)
Estimate	-0.078*** (0.015)	-0.013** (0.006)	0.065*** (0.013)
State and Year FE	Yes	Yes	Yes
Observations	663	663	1,326

Notes: Observations are at the state-by-biennial year level and cover 1989-90 through 2013-14 school years. The number of school-aged children by state is measured as the number of 5-17 year old children in the state according to data given to the National Cancer Institute by the U.S. Census Bureau (available at https://seer.cancer.gov/popdata/download.html). The D-in-D regression uses time (pre- vs. post-CSR) and state (CA vs. rest-of-country) as the two layers of differencing and restricts to private schools that primarily serve CSR grades in column (1) and private schools not primarily serving CSR grades in column (2). As an additional layer of differencing, the 'Triple-D' column adds whether the private school primarily serves CSR grades. Standard errors are clustered at the state level. *,** and *** denote significance at the 10%, 5% and 1% levels, respectively.

Table A.4: Regression-Discontinuity Estimates by Grade Span

Outcome Variable: Private School Share for Grade Span

	Kindergarten (Placebo) (1)	Elementary School CSR Grades (1-3) (2)	Elementary School non-CSR Grades (4-6) (3)	Middle School Grades (7-8) (4)	High School Grades (9-12) (5)
Average Effect	-0.07	-0.30**	-0.30**	-0.10	0.03
	(0.22)	(0.15)	(0.15)	(0.28)	(0.13)
Observations	2,874	8,825	9,251	6,390	11,680

Notes: Observations are at the district-cohort-grade level. The kindergarten effect here represents a placebo test as kindergarten was not a CSR grade for the cohorts around the discontinuity. To calculate average effects across grade spans, we estimate a separate local linear regression allowing for a different functional form on either side of the cutoff (see Equation B.1) for each grade. We then average these grade-level estimates to find the average effect over the grade span. The bandwidth used is three. Standards errors are calculated using the delta method and are clustered at the district level. Demographic controls are used in all regressions. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

Table A.5: Difference-in-Differences Estimates of CSR on Test Scores

Outcome Variable: Math Scores (σ)

	(1)	(2)	(3)
Treat*Post	0.105***	0.070***	0.067***
	(0.016)	(0.011)	(0.011)
Post	0.137***	0.022**	0.028***
	(0.032)	(0.010)	(0.010)
Treat	-0.094***	-0.065***	-0.066***
	(0.015)	(0.011)	(0.011)
Grade/Year/School FE	No	Yes	Yes
Demographic Controls	No	No	Yes
Number of Observations	207,926	207,926	207,523

Notes: Observations are at the school-grade-year level, and cover 1996-97 through 2003-04 school years. Test scores are normalized by grade-year to have mean zero and standard deviation one. Demographic controls include student race, enrollment and enrollment squared. Standard errors are clustered at the school level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

Table A.6:
Difference-in-Difference-in-Differences Estimates of CSR General Equilibrium Effects on Test Scores

Outcome Variable: Math Scores (σ)

	(1)	(2)	(3)
A. Grade 6 versus Grade 5 (Coefficient of Interest)			
$\Phi_{K6-K5,6-5,post-pre}$	0.128**	0.112**	0.099**
	(0.054)	(0.055)	(0.050)
A. Other Grade Differences (Placebo Tests)			
$\Phi_{K6-K5,7-6,post-pre}$	0.041	0.026	0.028
	(0.032)	(0.028)	(0.026)
$\Phi_{K6-K5,5-4,post-pre}$	-0.017	-0.017	-0.018
	(0.014)	(0.014)	(0.015)
$\Phi_{K6-K5,4-3,post-pre}$	-0.019	-0.018	-0.018
	(0.017)	(0.018)	(0.018)
$\Phi_{K6-K5,3-2,post-pre}$	-0.003	-0.003	-0.004
	(0.016)	(0.016)	(0.016)
Grade/Year/School FE	No	Yes	Yes
Demographic Controls	No	No	Yes

Notes: Observations are at the school-grade-year level, and cover 1996-97 through 2008-09 school years. Test scores are normalized by grade-year to have mean zero and standard deviation one. Demographic controls include student race, enrollment and enrollment squared. Standard errors are clustered at the district-level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

Table A.7: School Statistics by Grade Span Configuration

Grade Span	Number of Schools (1)	% Implementing CSR in First Year (2)
K-5	2183	95.9
K-6	1954	92.5
K-8	455	90.1
K-12	49	48.3
Other	289	90.3

Notes: Number of Schools refers to number of schools of that grade span serving second grade in the 1998-99 school year.

Table A.8: Percentage of Inexperienced Teachers

	Year					
Grade	1997-98	1998-99	1999-00	2000-01	2001-02	2002-03
2	27.3	26.7	21.6	18.8	17.0	15.2
3	26.8	26.3	22.0	17.9	16.4	14.0
4	26.9	33.1	32.9	30.1	27.6	24.5
5	24.0	27.8	28.8	28.4	25.7	22.5
6	23.5	26.7	27.4	26.5	27.0	23.2

Notes: Percent experiences is defined as the fraction of full time equivalent teacher with less than three years of experience teaching in the state of California.

Table A.9: Estimates of Teacher Quality

Outcome Variable: Mathematics Test Scores		
	CSR (1)	non-CSR (2)
$Q_{CSR/non,01-02}$	1.123*** (0.057)	-0.089*** (0.004)
$Q_{CSR/non,00-01}$	0.929*** (0.047)	-0.361*** (0.018)
$Q_{CSR/non,99-00}$	0.520*** (0.026)	-0.643*** (0.032)
$Q_{CSR/non,98-99}$	0.041*** (0.002)	-0.678*** (0.034)
$Q_{CSR/non,99-00,K5}$	0.997*** (0.078)	-1.132*** (0.089)
$Q_{CSR/non,99-00,K6}$	0.632*** (0.050)	-0.673*** (0.053)

Notes: This table shows estimates of teacher quality. Observations are at the school-grade-year level, and cover the 1997-98 through 2001-02 school years. Mathematics test scores are shown in percentile ranks relative to a national norming sample, where one percentile rank roughly equates to 0.05σ in the distribution of school-grade level test scores. Standard errors are computed using the delta method and are clustered at the school level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

B Persistence

This Appendix section examines how long the initial impact of the reform persisted in terms of students switching from private schools. To shed light on this issue, and noting that we do not observe individual switching behaviour, we implement the following regression discontinuity design, which exploits the differential exposure of cohorts to the reform:

$$\begin{aligned} grade\text{`}i\text{'}share_{dc} &= \beta_0 + \beta_1 D_{dc} + \beta_2 f(cohort_{dc}) + \beta_3 D_{dc} * f(cohort_{dc}) + \eta_d + \epsilon_{dc} \\ for \ \ -b &\leq cohort_{dc} \leq b \,, \end{aligned} \tag{B.1}$$

where $grade\text{`}i\text{'}share_{dc}$ is the private school share for a student in grade i belonging to cohort c in district d, D_{dc} indicates whether cohort c was exposed to CSR, $f(\cdot)$ is a flexible polynomial function, $cohort_{dc}$ is the cohort number (defined by the year that the student enters kindergarten and normalized by that year's relation to the reform's year of introduction),[75] η_d is a district fixed effect, and b is some bandwidth. This regression discontinuity design identifies the CSR effect on the private school share for each grade, the idea being to see whether pronounced changes in private school share line up with elementary school grade spans. Our coefficient of interest is β_1, which represents the effect of CSR on the private school share of cohorts in grade i. Given that the most common grade configurations in California are K-5 and K-6, the second hypothesis would imply that β_1 increases from elementary school non-CSR grades (4-6) to the middle school grades (7-8), while the first hypothesis would imply no such increase.

In terms of persistence results, Figure A.4 plots the estimated effect of CSR on private school share for each grade, and in Appendix Table A.4, we report average effects by grade span (elementary, middle, and high school) to increase power. The effect for kindergarten should be considered as a placebo, as the first CSR cohort was exposed in first grade only. This is borne out by an estimate that is statistically indistinguishable from zero. Estimates for subsequent grade spans indicate that the CSR reform induced private school students to enter the public school system and that they remained there until completion of the elementary grades.[76] Approximately two-thirds of the CSR 'treatment effect' on private school share disappears when making the transition to middle school, consistent with students transitioning back into the private system. (Here, the lack of individual data prevents us from tracking switching behavior with precision.)

[75] The cohort entering kindergarten in 1995-96 is designated 'cohort zero' as it is the first cohort to be exposed to CSR in first grade. Since the cohort variable is discrete, we add 0.5 to each value so that zero is the midpoint between the first treated and untreated cohorts.

[76] Table A.7 reports the number of elementary schools by grade configuration in California. Elementary schools are divided approximately equally between K-5 and K-6 configurations.

C School-Level Implementation Evidence: Public School Compositions and House Prices

This Appendix section provides evidence regarding CSR-induced changes to public school student composition and house prices. It exploits variation induced by schools choosing whether or not to implement CSR. This additional evidence is relevant to the effects of CSR on both public school composition and house prices from Section 3: the results herecorroborate the evidence in Section 3 above.

Public School Composition: The main text focuses on proximity to a private school in providing evidence of compositional change induced by CSR. To use school-level CSR implementation as our source of variation, we follow a similar triple-differences methodology, whereby we compare demographic characteristics in school-grades that implemented CSR with those that did not. As CSR implementation is not observed until the 1998-99 school year, we define any school that had implemented CSR in kindergarten or third grade in that year as a CSR-implementing school.[77] The weighted estimating equation is:

$$\begin{aligned} demo_{sgt} = \beta_0 + \beta_1 post_{gt} + \beta_2 treat_{gt} + \beta_4(post_{gt} * treat_{gt}) + \beta_5(post_{gt} * CSR_s) \\ + \beta_6(treat_{gt} * CSR_s) + \beta_7(post_{gt} * treat_{gt} * CSR_s) + \eta_s + \theta_t + \delta_g + \epsilon_{dgt}\,, \end{aligned} \tag{C.1}$$

where $demo_{sgt}$ is the demographic share of interest for grade g student in school s at time t, $post_{gt}$ indicates whether CSR had been implemented for grade g (as before), $treat_{gt}$ indicates whether grade g was subject to the CSR reform in year t, CSR_s indicates whether school s implemented CSR, X_{sgt} is a set of school-grade-year covariates, and η_s, θ_t and δ_g are school, time and grade fixed effects, respectively. The triple-differences coefficient of interest is β_7. To identify it, we assume that the change in the demographic share between CSR and non-CSR grades would have been the same for CSR and non-CSR schools in the absence of the reform.

School Level Composition Results: Figure C.1 provides a visual representation of the change in demographics in the schools that had implemented CSR relative to those that did not. The figure reveals a substantial relative increase in the fraction of Asian students in CSR schools, and a decline in the fraction of Hispanic students. While suggestive, these differences do not include a comparison between CSR and non-CSR grades (as described in equation (C.1)).

[77] This definition of treatment is motivated by the fact that any school that implemented CSR in the first possible year (the 1996-97 school year) would begin doing so for first grade, followed by second grade in the 1997-98 school year, followed by either kindergarten or third grade in the 1998-99 school year. Therefore, any school that had not implemented CSR for these grades in 1998-99 would also not have implemented CSR in the 1996-97 school year, making it a non-CSR-implementing school. According to this definition, around sixty percent of all schools implemented CSR by the 1998-99 school year.

This leads us to estimate triple-differences specifications in Panel A of Table C.1 to incorporate variation across CSR and non-CSR grades. Relative to schools that did not implement the reform, we find that CSR led to a reduction of almost 2 percent in the share of Hispanic students and an increase of almost 1.5 percent in the share of Asian students within schools that did implement the program. Further, point estimates are in line with the proportion of white and black students rising and declining, respectively, though neither is statistically significant.

Compared to the 'distance to private school' design in the main text, the results are similar for the decline in the share of Hispanic students, although the design does not capture the increase in the proportion white students or the decline in the proportion black students and estimates a large increase in the proportion Asian. Regardless, both designs point to changes in public school demographics whereby students who are over-represented in private schools relative to public schools (white and Asian students) are drawn into the public system.

House Prices: The main text focuses on district-level variation to identify the effect of CSR on house prices, due in part to the fact that housing transactions cannot be matched to school attendance zones statewide since our school attendance zone boundary dataset does not cover the entire state, but rather is focused on urban population centers. A school-level analysis may be illuminating, however, since aggregating to the district level obscures important variation. Figure C.2 highlights variation in the CSR implementation in four densely populated school districts for which we have data.[78] Visually, we see that non-CSR implementing schools tend to be clustered within certain regions of the state, with Sacramento Unified accounting for a disproportionate number of non-CSR implementing schools.

To conduct our analysis at the *school* level, we use a school-level CSR implementation variable (CSR_s) to construct a difference-in-differences regression. In that vein, we estimate the following regression:

$$price_{st} = \beta_0 + \beta_1 post_t + \beta_3(post_t * CSR_s) + \phi X_{st} + \eta_s + \theta_t + \epsilon_{st}\,, \tag{C.2}$$

where $price_{st}$ is the average house price in the attendance zone for school s at time t, CSR_s indicates whether the school implemented CSR in 1996, X_{st} is a set of non-demographic controls (enrollment, number of bedrooms, lot size and square feet), and η_s and θ_t are school and time fixed effects, respectively. As before, in a difference-in-differences context, the coefficient of interest is β_3, which can be interpreted causally under the assumption that CSR and non-CSR schools would have experienced the same change in house prices in the

[78] These four large districts (Los Angeles, Orange, Sacramento and Riverside) account for 52 percent of our total observations.

absence of CSR.

School-Level House Price Results: Figure C.3(a) displays trends among treated and control groups based on the school-level variation. Visually, there do not appear to be any differential trends in house prices prior to the reform's implementation. Once CSR comes into effect, however, house prices experience a significant increase in treated schools relative to their control school counterparts.

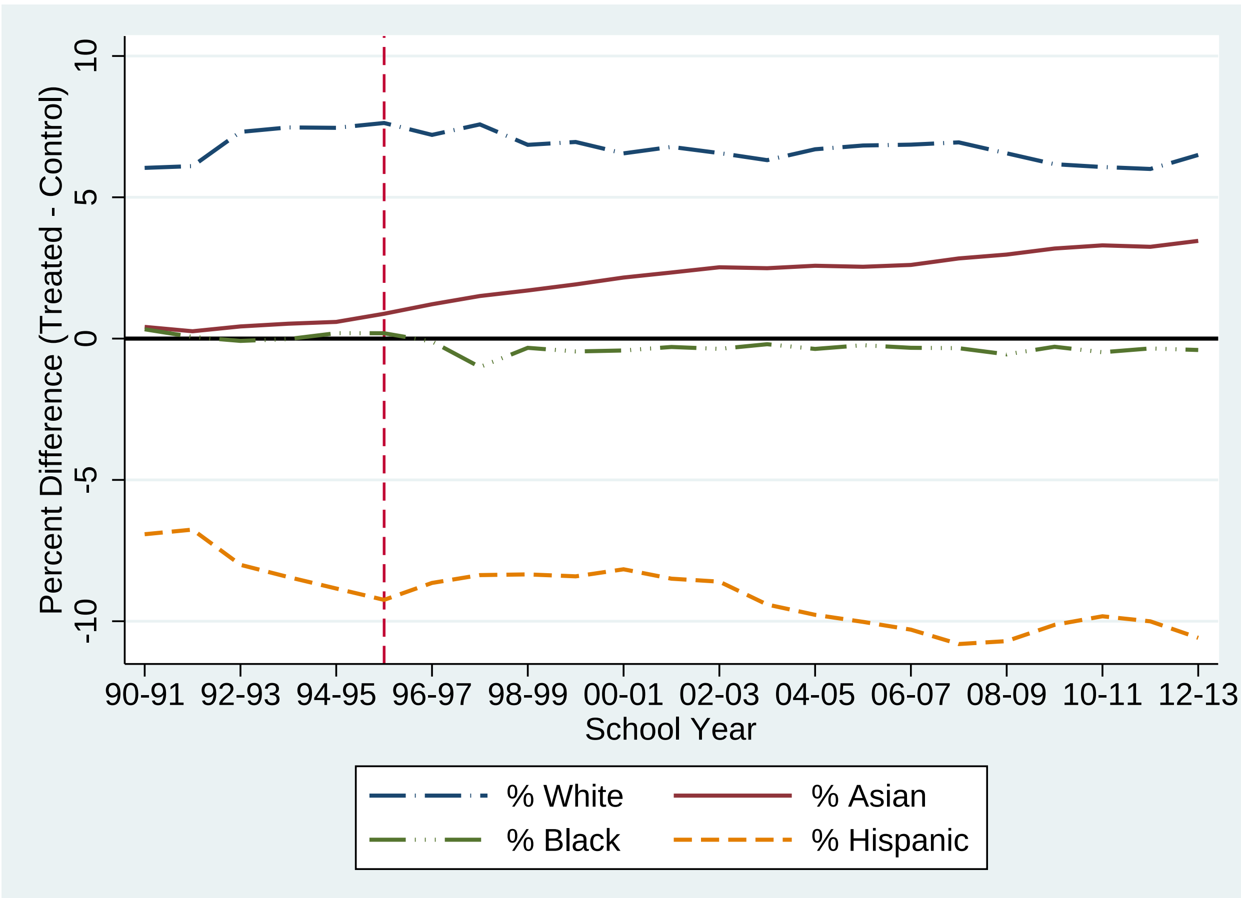
This evidence maps directly into the econometric specifications given by equation (C.2), which Table C.2 reports. As with the district level design in the main text, columns (1)-(3) should be differentiated from columns (4)-(6), since the former do not control for indirect general equilibrium effects of the reform, which may themselves be highly valued by parents. Column (3), which controls for house characteristics and district fixed effects but does not control for any indirect general equilibrium effects of the reform on teacher or peer quality, gives our preferred estimate of the *full effect* of the reform.

The point estimate of $8,200 implies that schools implementing CSR saw around a 4.2 percent increase in house prices.[79] While comparing with the district-based estimate is not straightforward since nearly every district implemented CSR to at least some extent, the 4.2 percent increase we find here is consistent with the district-based finding of 2.6 percent for a one standard deviation increase in CSR implementation. Similar to the district-based design, it is informative to look at the partial equilibrium impacts of the reform once the indirect general equilibrium effects on teacher and demographic characteristics are accounted for. As in the district-based design, we find that controlling for student demographics dramatically decreases the capitalization of the reform into house prices.

Regarding the validity of these results, we report the effect of district CSR treatment intensity and school CSR implementation on house prices by year in Figure C.3(b). The effect of CSR on house prices is indistinguishable from zero in the pre-reform period, while the effect becomes positive upon implementation of the reform. The effect also continues to grow even after the reform is fully implemented, which likely reflects local housing markets being slow to reach new equilibria.

[79] The point estimate is $8,200, which represents a 4.2 percent increase in house prices from the 1995 pre-CSR (weighted) mean of $195,000.

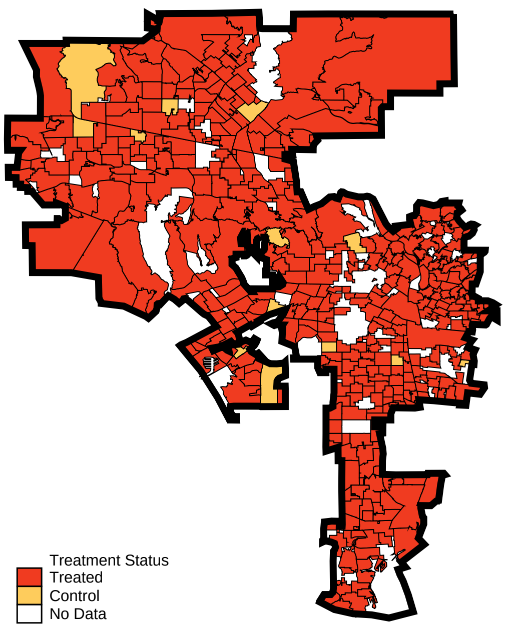
Figure C.1: Demographic Trends by Public Schools: Treated minus Untreated (CSR Implementation Design)



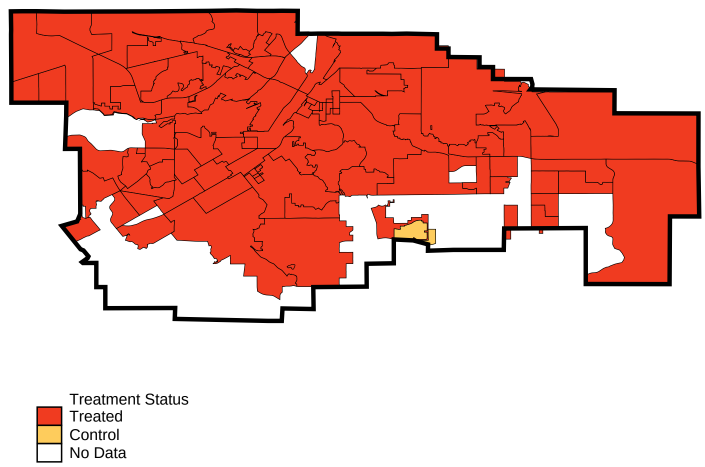
Notes: Figure C.1 shows the percent difference in demographics between schools that had implemented CSR versus those that did not from 1990-91 through 2012-13. The data generating the figures are weighted by school K-3 enrollment. Each year label refers to the start of the respective academic year. The dashed vertical line represents the 1995-96 school year so that all periods thereafter incorporate the effects of CSR. The horizontal 'zero' line represents no difference between treated and control schools.

Figure C.2: School-Level Treatment For Selected Districts

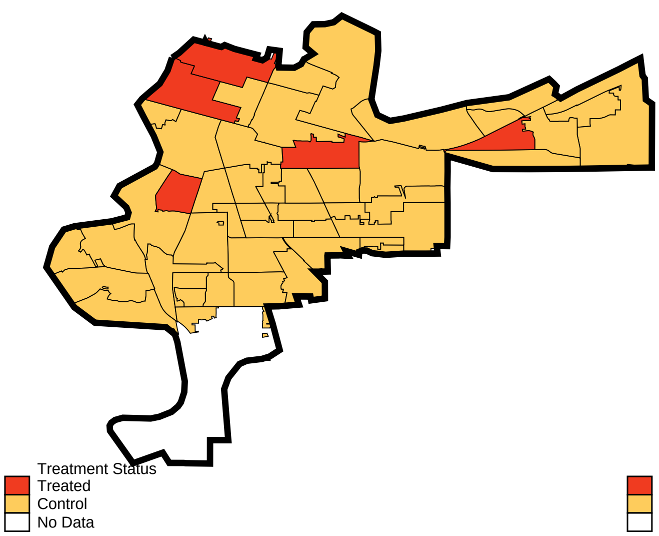
(a) Los Angeles Unified District



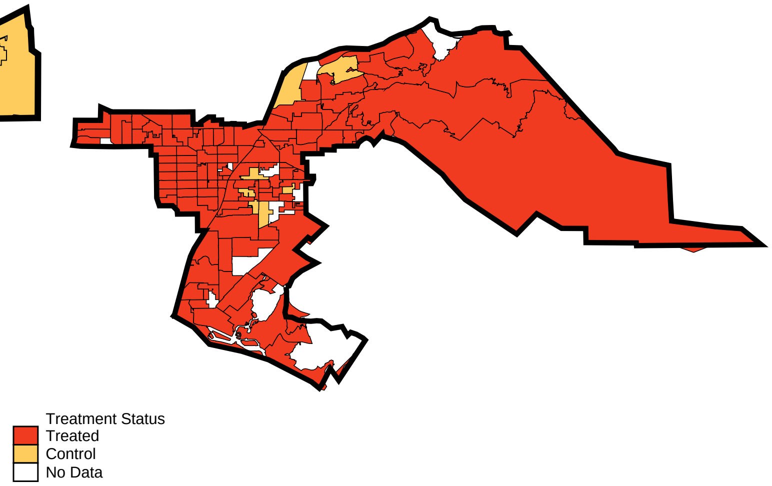
(b) Riverside County Region



(c) Sacramento Unified District



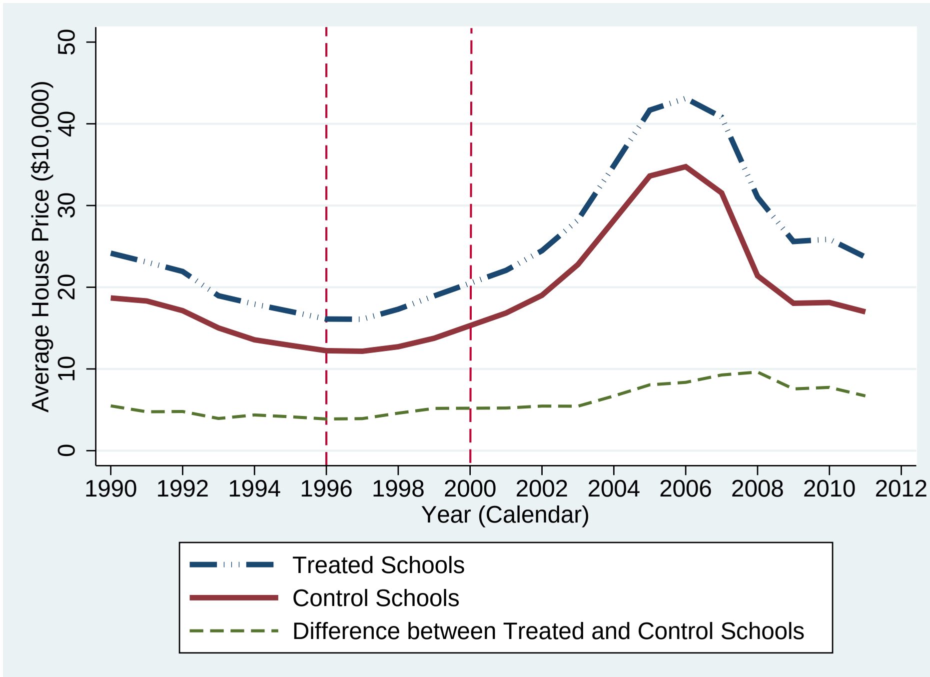
(d) Orange County Region



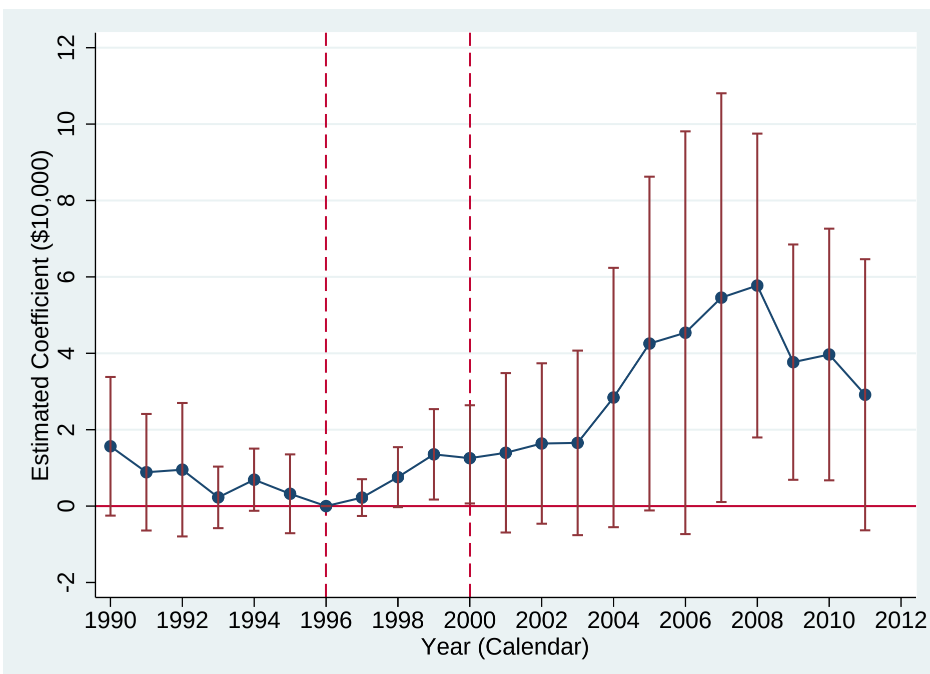
Notes: The above figure shows treated and control schools for two districts and two regions in California. The Riverside County region includes Riverside Unified, Alvord Unified, Jurupa Unified, and Moreno Valley school districts, while the Orange country region includes Santa Ana, Orange Unified, Garden Grove, and Newport school districts. Treated schools had CSR implemented in at least third grade or kindergarten in 1998, while control schools had no kindergarten or third grade CSR implementation in 1998. These two counties and regions cover 53.2% of all schools used in the school-level analysis.

Figure C.3: Visual Evidence of CSR on House Prices (School Level)

(a) House Prices by Treatment Status



(b) Estimated Effects of CSR on House Prices by Year



Notes: Figure C.3(a) shows average house prices (in 1995 dollars) in schools that implemented CSR ('treated' schools) and those that did not ('control schools'). Figure C.3(b) reports the effect of CSR school level implementation on house prices (in 1995 dollars) by year. The estimate at the start of implementation (in year 1996) is normalized to zero. Vertical bands represent 95% confidence intervals for each point estimate. Demographic covariates are omitted, though house characteristics and school fixed effects are included. Standard errors are clustered at the district-level. Each year label refers to the calendar year. Dashed vertical lines represent the start of CSR implementation in the 1996-97 school year and first year (2000-01 school year) when CSR was fully implemented in all grades K-3, respectively. Horizontal lines indicate an estimate of zero.

Table C.1: Triple-Differences Estimates of Compositional Changes (School-Level Implementation Design)

Outcome Variable: Student Demographics (%)				
	Percent White	Percent Hispanic	Percent Black	Percent Asian
	(1)	(2)	(3)	(4)
A. School Implementation Design				
Treatment*Post*CSR_s	0.90	-2.33**	-0.39	1.62***
	(1.43)	(1.11)	(0.48)	(0.57)
% Share in Private School (1997-98)	52.93	17.21	7.10	12.30
% Share in Public School (1997-98)	38.75	40.49	8.75	11.14
School/Grade/Year FE	Yes	Yes	Yes	Yes

Notes: Observations are at the school-grade-year level, and cover 1990-91 through 2012-13 school years. There are 1,017,865 observations. Enrollment and enrollment squared are included as controls. 'Post' is defined based on a 'before' and 'after' CSR implementation dummy. The table refers to the school level implementation design described by equation (C.1). Private and public school demographic shares from the National Center for Education Statistics for the 1997-98 school year are provided for reference. All regressions are weighted by school-grade-year level enrollment and standard errors are clustered at the district level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

Table C.2: Difference-in-Differences Estimates of Impact on House Prices (School CSR Implementation Design)

Outcome Variable: Average House Price in 1995 Dollars ($10,000s)

	All Years (1990-2012)				*1994-2012 only*	
	(1)	(2)	(3)	(4)	(5)	(6)
CSR_s*Post	0.29	0.43	0.82**	0.21	-0.14	-0.16
	(0.39)	(0.39)	(0.40)	(0.66)	(0.59)	(0.56)
House Characteristics	No	Yes	Yes	Yes	Yes	Yes
District/School FE	No	No	Yes	Yes	Yes	Yes
Teacher Controls	No	No	No	Yes	No	Yes
Demographic Controls	No	No	No	No	Yes	Yes
Observations	41,112	41,112	41,112	41,102	34,089	34,089

Notes: Observations are at the school-year level. Columns (1)-(4) cover the 1990 through 2012 calendar years while columns (5) and (6) are restricted to the 1994 through 2012 calendar years because teacher controls are only available from 1994 onward. House prices are deflated to 1995 dollars using California's house price index. All regressions include cubic controls for enrollment and year fixed effects. House characteristics consist of number of bedrooms, square feet and lot size. Teacher controls include experience and education levels. Demographic controls include student race, gender, and free and reduced price meal eligibility. Standard errors are clustered at the school level. ***,** and * denote significance at the 1%, 5% and 10% levels, respectively.

D California State Testing – a Quick Primer

Statewide testing in California started in 1961 for mathematics, reading and writing in grades 5, 8 and 10. In 1972, the California Assessment Program was created, which tested reading in grades 2 and 3 and mathematics, reading and writing in grades 6 and 12. It lasted (with a few test additions) until 1991, when it was replace by the California Learning Assessment System (CLAS), which covered reading, writing and mathematics in grades 4, 5, 8 and 10.

In 1994, under public pressure from civil rights groups that the CLAS was inaccurate and intruded upon students' privacy (due to numerous race-based questions on the test), Governor Pete Wilson vetoed a Senate bill to extend CLAS.[80] Therefore, there were no statewide tests for the 1994-95 and 1995-96 school years, although districts often did conduct standardized tests during this time; the state even provided funding for this through the Pupil Testing Incentive Program.

In the 1996-97 school year, the Standardized Testing and Reporting program (STAR) – an initiative of the Governor – was implemented, which tested reading, writing and math in grades 2-8 and reading, writing, mathematics, history, and science in grades 9-11. These are the tests we use in this study.

E Housing Data Linkage: a Description

This Appendix section describes how we map the housing transactions data obtained from DataQuick into school districts and school attendance zones. First, the DataQuick dataset reports the latitude and longitude of the housing sale in some instances, while in others only reports a physical address. If the latitude and longitude were given, we used those values directly. When only the physical address was given, we mapped the physical address to a latitude and longitude using the US address locator available at http://gis.ats.ucla.edu/.

With the house transactions now identified with a specific latitude and longitude, we then mapped each housing transaction into a given school district or school attendance zone using ArcGIS. To identify the coverage of school districts in California, we used the Elementary and Unified District 2000 Census shapefiles.[81] For school attendance zone coverage, we used the 2009 kindergarten school attendance zone shapefile from The College of William and

[80]The Governor stated that his veto was due to the fact that it did not give teachers and parents individual student achievement scores (scores were available at the school level only).

[81]These are available from `http://geocommons.com/overlays/206` and `http://geocommons.com/overlays/1252`, respectively.

Mary and the Minnesota Population Center (2011).[82] The geocoded housing data were then spatially joined to a specific school district or school attendance zone. School attendance zones were available for slightly over 10 percent of school districts in California, although they cover over 25 percent of the schools in California as these were urban school districts.

For the school-level house price, we needed to ascertain whether all public schools were within xkm of a private school. To do this, we mapped the address of all California public and private schools using the California Public School Directory[83] and the California Private School Directory[84] to latitudes and longitudes using the US address locator. A dummy variable was then created indicating whether a given public school was within xkm of a private school.

F Estimating Teacher Quality

This Appendix section explains in detail how we incorporate the estimation of teacher quality into our approach.

Let $Q^l_{CSR,t}$ and $Q^l_{non,t}$ denote the effect of teacher quality in year t for students in a CSR and non-CSR grade, respectively. We allow these effects to persist by using the l superscript, which represents the effect of being treated to a CSR or non-CSR teacher $l \geq 0$ periods ago (where 0 is the contemporaneous effect). Note that we do not look at teacher quality at the grade level, but rather distinguish between CSR and non-CSR grades since CSR should affect teachers across all CSR grades equally.

Our data begin in 1997-98, following the initial increase in the share of inexperienced teachers. The proportion of inexperienced teachers is similar across CSR (second and third) and non-CSR (fourth) grades for that first year.[85] An interesting pattern emerges over the next three years once the CSR program expands to kindergarten and third grade: teacher inexperience falls substantially for CSR grades and rises for non-CSR grades. Inexperience then falls for all grades thereafter.[86]

We incorporate variation in teacher inexperience into our estimation strategy. We estimate the teacher quality parameters $Q^l_{CSR,t}$ and $Q^l_{non,t}$ for each lag l according to the

[82] Since virtually all California schools cover grades K-3, it makes no difference which K-3 grade attendance zone shapefile is used.

[83] Available from http://www.cde.ca.gov/ds/si/ds/pubschls.asp

[84] http://www.cde.ca.gov/ds/si/ps/

[85] Inexperience in fifth and sixth grades is close to but slightly lower than for second through fourth grades.

[86] It may seem puzzling why schools would maintain teacher quality for CSR grades at the expense of non-CSR grades, since formal incentives under the 1999 Public Schools Accountability Act were not provided differentially by grade. Schools perhaps believed policymakers were paying closer attention to CSR grades or schools may have worked to ensure the success of a promising reform.

following two-step procedure. First, we regress test scores in 1997-98 + l ($y_{s,g,\text{1997-98}+l}$) on the share of teacher inexperience in 1997-98 ($X_{s,g,\text{1997-98}}$), including grade fixed effects (ϕ_g):

$$y_{s,g,\text{1997-98}+l} = \gamma_l X_{s,g,\text{1997-98}} + \phi_g + \epsilon_{s,g,\text{1997-98}} \,.$$

Second, we use the resulting estimate $\hat{\gamma}_l$ to compute teacher quality relative to the 1997-98 baseline:[87]

$$Q^l_{CSR,t} = \hat{\gamma}_l \times (X_{3,t} - X_{3,\text{1997-98}})$$
$$Q^l_{non,t} = \hat{\gamma}_l \times (X_{4,t} - X_{4,\text{1997-98}}) \,,$$

where CSR and non-CSR values of Q use variation in third- and fourth-grade inexperience, respectively. Thus, the relevant parameters to compute $\hat{\gamma}_R$, are $Q^0_{CSR,\text{2001-02}} = \hat{\gamma}_0 \times (X_{3,\text{2001-02}} - X_{3,\text{1997-98}})$, $Q^0_{non,\text{2001-02}} = \hat{\gamma}_0 \times (X_{4,\text{2001-02}} - X_{4,\text{1997-98}})$ and $Q^4_{CSR,\text{1997-98}} = \hat{\gamma}_4 \times (X_{4,\text{1997-98}} - X_{4,\text{1997-98}}) = 0$. The necessary parameters to compute $\hat{\gamma}_S$ are estimated analogously.[88]

[87] Defining teacher quality relative to 1997-98 controls for preexisting differences between grades that are unrelated to the implementation of the CSR program. Using 1997-98 as a baseline is justified given that CSR had yet to apply to third grade in that year. Indeed, Table A.8 shows that the share of teacher inexperience is essentially identical across third and fourth grades in 1997-98.

[88] We estimate the parameters $Q^2_{CSR,\text{2000-01},K6}$, $Q^2_{non,\text{2000-01},K6}$, $Q^2_{CSR,\text{2000-01},K5}$ and $Q^2_{non,\text{2000-01},K5}$. Due to a lack of test score data in 1996-97, the parameters $Q^5_{CSR,\text{1996-97},K5/K6}$ and $Q^5_{non,\text{1996-97},K5/K6}$ cannot be estimated. However, as with $Q^4_{CSR,1997-98}$, we can assume that they are negligible since teacher quality across grades is likely to be similar in 1996-97 and 1997-98.

G Estimating Equations

This Appendix section discusses the identification of the main parameters without, then with, teacher effects.

G.1 Without Teacher Effects

$\boldsymbol{\gamma_R}$**:** To identify γ_R, we subtract $\Delta y_{4,01-02}$ from $\Delta y_{3,01-02}$:

$$\begin{aligned}\Delta y_{3,01-02} - \Delta y_{4,01-02} = {} & \delta_R^3 \gamma_R \Delta R_{0,98-99} + \delta_R^2 \gamma_R \Delta R_{1,99-00} + \delta_R \gamma_R \Delta R_{2,00-01} + \gamma_R \Delta R_{3,01-02} \\ & + \delta_S^3 \gamma_S \Delta X^S_{0,97-98} + \delta_S^2 \gamma_S \Delta X^S_{1,98-99} + \delta_S \gamma_S \Delta X^S_{2,99-00} + \gamma_S \Delta X^S_{3,00-01} + \Delta\epsilon_{3,01-02} \\ & - (\gamma + \delta_R^3 \gamma_R \Delta R_{1,98-99} + \delta_R^2 \gamma_R \Delta R_{2,99-00} + \delta_R \gamma_R \Delta R_{3,00-01} \\ & + \delta_S^3 \gamma_S \Delta X^S_{1,98-99} + \delta_S^2 \gamma_S \Delta X^S_{2,99-00} + \delta_S \gamma_S \Delta X^S_{3,00-01} + \gamma_S \Delta X^S_{4,2001-02} + \Delta\epsilon_{4,2001-02}) \\ & = \gamma_R \Delta R_{2001-02}\,, \end{aligned} \tag{G.1}$$

where the final equality comes the fact that CSR affected all grades equally once it was implemented, so that $\Delta R_{gt} = \Delta R_{g't}$ and $\Delta X^S_{gt} = \Delta X^S_{g't}$ $\forall g, g'$.

Since we do not observe the counterfactual test scores in the absence of the reform, $\Delta y_{3,01-02}$ and $\Delta y_{4,01-02}$ are not observed. Therefore, we use the pre-reform test scores $y_{3,97-98}$ and $y_{4,97-98}$ as counterfactuals for $y_{3,01-02}$ and $y_{4,01-02}$, respectively. Thus, we have:

$$y_{3,01-02} - y_{4,01-02} - (y_{3,97-98} - y_{4,97-98}) = \gamma_R \Delta R_{01-02}\,. \tag{G.2}$$

$\boldsymbol{\gamma_S}$**:** Similarly, for γ_S, we subtract sixth grade test scores for K-6 schools ($\Delta y_{6,01-02,K6}$) from K-5 schools ($\Delta y_{6,01-02,K5}$):

$$\begin{aligned} & \Delta y_{6,01-02,K6} - \Delta y_{6,01-02,K5} = \delta_R^5 \gamma_R \Delta R_{1,96-97,K6} + \delta_R^4 \gamma_R \Delta R_{2,97-98,K6} + \delta_R^3 \gamma_R \Delta R_{3,98-99,K6} \\ & + \delta_S^5 \gamma_S \Delta X^S_{1,96-97,K6} + \delta_S^4 \gamma_S \Delta X^S_{2,97-98,K6} + \delta_S^3 \gamma_S \Delta X^S_{3,98-99,K6} \\ & + \delta_S^2 \gamma_S \Delta X^S_{4,99-00,K6} + \delta_S \gamma_S \Delta X^S_{5,00-01,K6} + \gamma_S \Delta X^S_{6,01-02,K6} + \Delta\epsilon_{6,01-02,K6} \\ & - (\gamma + \delta_R^5 \gamma_R \Delta R_{1,96-97,K5} + \delta_R^4 \gamma_R \Delta R_{2,97-98,K5} + \delta_R^3 \gamma_R \Delta R_{3,98-99,K5} \\ & + \delta_S^5 \gamma_S \Delta X^S_{1,96-97,K5} + \delta_S^4 \gamma_S \Delta X^S_{2,97-98,K5} + \delta_S^3 \gamma_S \Delta X^S_{3,98-99,K5} \\ & + \delta_S^2 \gamma_S \Delta X^S_{4,99-00,K5} + \delta_S \gamma_S \Delta X^S_{5,00-01,K5} + (1-\psi) \gamma_S \Delta X^S_{6,01-02,K5} + \Delta\epsilon_{6,01-02,K5}) \\ & = \psi \gamma_S \Delta X^S_{6,01-02}\,, \end{aligned} \tag{G.3}$$

where the final equality comes the fact that CSR affected K-5 and K-6 schools equally (until the switch back into the private system in sixth grade) so that $\Delta R_{g,t,K6} = \Delta R_{g,t,K5}$ and

$\Delta X^S_{g,t,K6} = \Delta X^S_{g,t,K5}\ \forall g$.

Since we do not observe the counterfactual test scores in the absence of the reform, $\Delta y_{6,01-02,K6}$ and $\Delta y_{6,01-02,K5}$ are not observed. In this case, we use two levels of differencing to act as the counterfactual. First, to account for systematic differences between K-6 and K-5 schools, we use fifth grade test scores in K-5 ($y_{5,01-02,K6}$) and K-6 schools ($y_{5,01-02,K5}$) as are first level of differencing. Then, we use the pre-reform test scores for both fifth and sixth grades, $y_{5,97-98}$ and $y_{6,97-98}$, in K-5 and K-6 schools as counterfactuals for the observed test scores in fifth and sixth grades in the 2001-02 school year. Therefore, we have:[89]

$$\begin{aligned}\psi\gamma_S\Delta X^S_{6,01-02} = {} & [y_{6,01-02,K6} - y_{5,01-02,K6} - (y_{6,97-98,K6} - y_{5,97-98,K6})] \\ & - [y_{6,01-02,K5} - y_{5,01-02,K5} - (y_{6,97-98,K5} - y_{5,97-98,K5})]\,. \end{aligned} \tag{G.4}$$

$\boldsymbol{(\delta_R, \delta_S)}$**:** Identification of δ_R and δ_S, takes the parameters γ_R and γ_S to be known and differences the test scores in fourth grade and third grade in the 2000-01 school year, which yields:[90]

$$\begin{aligned}\Delta y_{4,00-01} - \Delta y_{3,00-01} = {} & \delta_R^3\gamma_R\Delta R_{1,97-98} + \delta_R^2\gamma_R\Delta R_{2,98-99} + \delta_R\gamma_R\Delta R_{3,99-00} \\ & + \delta_S^3\gamma_S\Delta X^S_{1,97-98} + \delta_S^2\gamma_S\Delta X^S_{2,98-99} + \delta_S\gamma_S\Delta X^S_{3,99-00} + \gamma_S\Delta X^S_{4,00-01} + \Delta\epsilon_{4,01-02} \\ & - (\gamma + \delta_R^2\gamma_R\Delta R_{1,98-99} + \delta_R\gamma_R\Delta R_{2,99-00} + \gamma_R\Delta R_{3,00-01} \\ & + \delta_S^2\gamma_S\Delta X^S_{1,98-99} + \delta_S\gamma_S\Delta X^S_{2,99-00} + \gamma_S\Delta X^S_{3,00-01} + \Delta\epsilon_{3,01-02}) \\ = {} & \delta_R^3\gamma_R\Delta R_{1,97-98} - \gamma_R\Delta R_{3,00-01} + \delta_S^3\gamma_S\Delta X^S_{1,97-98}\,. \end{aligned} \tag{G.5}$$

Similarly, comparing test scores between fourth and fifth grade in the 2000-01 school year

[89] Here, we are overidentified since we could use 1997-98, 1998-99 and 1999-00 as counteractuals since those cohorts in fifth and sixth grades were not subject to CSR in those years. In practice, we use all three and take an average of the estimates, although estimates are quantitavely similar regardless which counterfactual year we use.

[90] $\Delta y_{3,99-00} - \Delta y_{4,99-00}$ yields the same structural equation as $\Delta y_{3,00-01} - \Delta y_{4,00-01}$ and $\Delta y_{5,01-02} - \Delta y_{4,01-02}$ yields the same structural equation as $\Delta y_{5,00-01} - \Delta y_{4,00-01}$. This equation is therefore overidentified. Once again, we use all both equations and take an average of the estimates, although estimates are quantitatively similar regardless which structural equation is used.

yields:

$$
\begin{aligned}
\Delta y_{5,00-01} - \Delta y_{4,00-01} = {} & \delta_R^4 \gamma_R \Delta R_{1,96-97} + \delta_R^3 \gamma_R \Delta R_{2,97-98} + \delta_R^2 \gamma_R \Delta R_{3,98-99} + \delta_S^4 \gamma_S \Delta X^S_{1,96-97} \\
& + \delta_S^3 \gamma_S \Delta X^S_{2,97-98} + \delta_S^2 \gamma_S \Delta X^S_{3,98-99} + \delta_S \gamma_S \Delta X^S_{4,99-00} + \gamma_S \Delta X^S_{5,00-01} + \Delta \epsilon_{5,00-01} \\
& - (\gamma + \delta_R^3 \gamma_R \Delta R_{1,97-98} + \delta_R^2 \gamma_R \Delta R_{2,98-99} + \delta_R \gamma_R \Delta R_{3,99-00} \\
& + \delta_S^3 \gamma_S \Delta X^S_{1,97-98} + \delta_S^2 \gamma_S \Delta X^S_{2,98-99} + \delta_S \gamma_S \Delta X^S_{3,99-00} + \gamma_S \Delta X^S_{4,00-01} + \Delta \epsilon_{4,00-01}) \\
& = \delta_R^4 \Delta R_{1,96-97} - \delta_R \Delta R_{3,99-00} + \delta_S^4 \gamma_S \Delta X^S_{1,96-97} \, . \qquad \text{(G.6)}
\end{aligned}
$$

Since CSR affected all grades equally, we have that $\Delta R_{1,96-97} = \Delta R_{1,97-98} = \Delta R_{3,99-00} = \Delta R_{3,00-01}$ and $\Delta X^S_{1,96-97} = \Delta X^S_{3,97-98}$. Suppressing the grade and year notation on the ΔR_{gt} and ΔX^S_{gt} variables yields the following two equations with two unknowns (δ_R, δ_S):

$$
\begin{aligned}
y_{4,00-01} - y_{3,00-01} - (y_{4,97-98} - y_{3,97-98}) &= \gamma_R \Delta R(\delta_R^3 - 1) + \delta_S^3 \gamma_S \Delta X^S \qquad \text{(G.7)} \\
y_{5,00-01} - y_{4,00-01} - (y_{5,97-98} - y_{4,97-98}) &= \delta_R \Delta R(\delta_R^3 - 1) + \delta_S^4 \gamma_S \Delta X^S \, . \qquad \text{(G.8)}
\end{aligned}
$$

G.2 With Teacher Effects

We incorporate general equilibrium teacher effects by controlling for differences in observed teacher quality proxies. To incorporate the teacher effects (as defined in the main text), we express differences between observed and counterfactual test scores with differences in teacher quality by whether students were in a CSR or non-CSR grade. For example, we now express the difference between observed and counterfactual third grade test scores in 2001-02 as:

$$
\begin{aligned}
\Delta y_{3,01-02} = {} & \delta_R^3 \gamma_R \Delta R_{0,98-99} + \delta_R^2 \gamma_R \Delta R_{1,99-00} + \delta_R \gamma_R \Delta R_{2,00-01} + \gamma_R \Delta R_{3,01-02} \\
& + \delta_S^3 \gamma_S \Delta X^S_{0,98-99} + \delta_S^2 \gamma_S \Delta X^S_{1,99-00} + \delta_S \gamma_S \Delta X^S_{2,00-01} + \gamma_S \Delta X^S_{3,01-02} \\
& + \gamma_Q (Q^3_{CSR,98-99} + Q^2_{CSR,99-00} + Q^1_{CSR,00-01} + Q^0_{CSR,01-02}) + \Delta \epsilon_{3,01-02} . \qquad \text{(G.9)}
\end{aligned}
$$

$\boldsymbol{\gamma_R}$**:** Incorporating general equilibrium teacher effects, the differences between observed and counterfactual test scores that yield γ_R can be expressed in terms of the parameters in the following way:

$$
\begin{aligned}
y_{3,01-02} - y_{4,01-02} - (y_{3,97-98} - y_{4,97-98}) &= \gamma_R \Delta R_{01-02} \\
&+ \gamma_Q (Q^3_{CSR,98-99} + Q^2_{CSR,99-00} + Q^1_{CSR,00-01} + Q^0_{CSR,01-02}) \\
&- \gamma_Q (Q^4_{CSR,97-98} + Q^3_{CSR,98-99} + Q^2_{CSR,99-00} + Q^1_{CSR,00-01} + Q^0_{non,01-02}) \\
&= \gamma_R \Delta R_{01-02} + \gamma_Q (Q^4_{CSR,97-98} + Q^0_{CSR,01-02} - Q^0_{non,01-02}) \, . \qquad \text{(G.10)}
\end{aligned}
$$

$\boldsymbol{\gamma_S}$**:** Similarly, the differences between observed and counterfactual test scores that yield γ_S can be expressed in terms of the parameters in the following way:

$$
\begin{aligned}
&[y_{6,01-02,K6} - y_{5,01-02,K6} - (y_{6,97-98,K6} - y_{5,97-98,K6})] \\
&- [y_{6,01-02,K5} - y_{5,01-02,K5} - (y_{6,97-98,K5} - y_{5,97-98,K5})] = \psi \gamma_S \Delta X^S_{6,01-02} \\
&+ \gamma_Q (Q^5_{CSR,96-97,K6} + Q^4_{CSR,97-98,K6} + Q^3_{CSR,98-99,K6} + Q^2_{non,99-00,K6} + Q^1_{non,00-01,K6} + Q^0_{non,01-02,K6}) \\
&- \gamma_Q (Q^4_{CSR,97-98,K6} + Q^3_{CSR,98-99,K6} + Q^2_{CSR,99-00,K6} + Q^1_{non,00-01,K6} + Q^0_{non,01-02,K6}) \\
&- [\gamma_Q (Q^5_{CSR,96-97,K5} + Q^4_{CSR,97-98,K5} + Q^3_{CSR,98-99,K5} + Q^2_{non,99-00,K5} + Q^1_{non,00-01,K5} + Q^0_{non,01-02,K5}) \\
&- \gamma_Q (Q^4_{CSR,97-98,K5} + Q^3_{CSR,98-99,K5} + Q^2_{CSR,99-00,K5} + Q^1_{non,00-01,K5} + Q^0_{non,01-02,K5})] \\
&= \psi \gamma_S \Delta X^S_{6,01-02} + \gamma_Q (Q^5_{CSR,96-97,K6} + Q^2_{non,99-00,K6} - Q^2_{CSR,99-00,K6}) \\
&- [\gamma_Q (Q^5_{CSR,96-97,K5} + Q^2_{non,99-00,K5} - Q^2_{CSR,99-00,K5})] \, . \qquad \text{(G.11)}
\end{aligned}
$$

$\boldsymbol{(\delta_R, \delta_S)}$**:** Finally, to solve for δ_R and δ_S, we incorporate teacher effects into the final two regressions:

$$
\begin{aligned}
y_{4,00-01} - y_{3,00-01} - (y_{4,97-98} - y_{3,97-98}) &= \gamma_R \Delta R_{00-01} (\delta^3_R - 1) + \delta^3_S \gamma_S \Delta X^S_{02-03} \\
&+ \gamma_Q (Q^4_{non,96-97} + Q^3_{CSR,97-98} + Q^2_{CSR,98-99} + Q^1_{CSR,99-00} + Q^0_{non,00-01}) \\
&- \gamma_Q (Q^3_{non,97-98} + Q^2_{CSR,98-99} + Q^1_{CSR,99-00} + Q^0_{CSR,00-01}) \\
&= \gamma_R \Delta R_{00-01} (\delta^3_R - 1) + \delta^3_S \gamma_S \Delta X^S_{02-03} \\
&+ \gamma_Q (Q^4_{non,96-97} + Q^3_{CSR,97-98} - Q^3_{non,97-98} + Q^0_{non,00-01} - Q^0_{CSR,00-01}) \, . \qquad \text{(G.12)}
\end{aligned}
$$

$$
\begin{aligned}
y_{5,00-01} - y_{4,00-01} - (y_{5,97-98} - y_{4,97-98}) &= \delta_R \gamma_R \Delta R_{00-01} (\delta^3_R - 1) + \delta^4_S \gamma_S \Delta X^S_{02-03} \\
&+ \gamma_Q (Q^4_{CSR,96-97} + Q^3_{CSR,97-98} + Q^2_{CSR,98-99} + Q^1_{non,99-00} + Q^0_{non,00-01}) \\
&- \gamma_Q (Q^4_{non,96-97} + Q^3_{CSR,97-98} + Q^2_{CSR,98-99} + Q^1_{CSR,99-00} + Q^0_{non,00-01}) \\
&= \delta_R \gamma_R \Delta R_{00-01} (\delta^3_R - 1) + \delta^4_S \gamma_S \Delta X^S_{02-03} \\
&+ \gamma_Q (Q^4_{CSR,96-97} - Q^4_{non,96-97} + Q^1_{non,99-00} - Q^1_{CSR,99-00}) \, . \qquad \text{(G.13)}
\end{aligned}
$$