Education Reform in General Equilibrium: Evidence from California’s Class Size Reduction*

Michael Gilraine, New York University
Hugh Macartney, Duke University and NBER
Robert McMillan, University of Toronto and NBER

September 5, 2017

Abstract

Understanding general equilibrium responses to major education reforms is important for policy, given their potential to swamp any direct reform effects. This paper focuses on a general equilibrium sorting response that is likely to matter whenever a reform improves public school quality significantly, leading some households to re-optimize by switching out of private schools. It does so in the context of the unique roll-out, grade-by-grade, of California’s statewide class size reduction program of the late-1990s – a measure inspired by the STAR Experiment, and up to that time, the most costly state education reform ever implemented in the United States. Using a transparent differencing strategy that exploits the grade-specific timing of the reform, we first show significant reductions in local private school share for relevant elementary grades, marked changes in public school sociodemographics, and large house price increases in areas where the policy had been implemented. These reduced-form findings motivate an estimable structural framework that allows us to gauge the policy’s direct and indirect effects on a common scale for the first time, as well as their respective persistence rates. Identifying the structural parameters using a generalization of the differencing approach, our estimates reveal a significant pure class size effect of 0.11σ (in terms of mathematics scores), and an even larger indirect effect via induced changes in school demographics, of 0.17σ: ignoring this indirect response would thus understate the reform’s full impact by well over half. We find both indirect and direct effects of the reform persist positively, invalidating a simple difference-in-differences approach; further, a bounding exercise suggests that the indirect sorting effect involves positive spillovers for students already in public school. The analysis draws attention, more broadly, to conditions in which the indirect sorting effects of large-scale education reforms in other contexts are likely to be first-order.

Keywords: Education Reform, General Equilibrium, Sorting, Class Size Reduction, Difference-in-Differences, Structural Model, Persistence

*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, and the University of Toronto Mississauga. All remaining errors are our own. Contact emails: Gilraine, mike.gilraine@nyu.edu; Macartney, hugh.macartney@duke.edu; McMillan, mcmillan@chass.utoronto.ca.
1 Introduction

Economic policy analysis often focuses on the direct, intended effects of policies, “holding all other things equal.” Measuring such effects is an important ingredient in any policy making, and our ability to estimate them accurately has improved with the development of appealing experimental or quasi-experimental methods that use transparent variation to uncover causal impacts. It is well appreciated, however, that large-scale reforms may give rise to indirect general equilibrium effects that reinforce or counteract the direct policy impact, sometimes in unintended ways. Understanding the scale of any indirect effects is of considerable interest, as they have the potential to alter the policy-making calculus significantly. Never-the-less, pinning down such indirect effects presents challenges for applied research, primarily because they represent additional sources of endogeneity that can be difficult to identify. As a consequence, the literature seeking to measure them is relatively undeveloped.\footnote{Notable existing contributions include papers by Jepsen and Rivkin (2009) and Dinerstein and Smith (2016), discussed in more detail in Section 2.}

The goal of this paper is to gauge the extent of the indirect effects of large-scale reforms, allowing them to be placed 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.\footnote{Such a combination of circumstances is common in practice for the reason that low public school quality gives rise both to greater demand for private schools and greater pressure for public school reform.}

In such a common setting, some households are likely 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, up to that point being the largest state-led education reform ever implemented in the United States.\footnote{This assessment comes from the 1998/99 report of the CSR Research Consortium.} Inspired by the notable success of Project STAR in...
Tennessee – perhaps the most famous experimental evaluation in education and the subject of a number of influential studies\(^4\) – the 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 identification: Smaller classes in grade 1 were phased in during the 1996-97 school year, with schools having to hire enough teachers in grade 1 to lower class sizes below 20 in order to be eligible for CSR funding – a substantial amount (over $600 per student). In 1997-98, classes in grade 2 became eligible, and schools could then seek to reduce class sizes in one of kindergarten and grade 3 in the following year, the other the year after.

Given the combination of the strong financial incentives to implement the reform according to this timetable, the substantial impact on class size it produced, and the sheer scale of California’s education system, one would expect CSR to have broader effects. Substantial impacts on teacher labor markets have already been documented by Jepsen and Rivkin (2009), who show that there was a sudden, significant increased need for new teacher hires, which dampened CSR’s effects in the short term. Further, the effects of the reform were non-uniform, with some schools experiencing reductions in class size without any discernible reduction in teacher quality, as a result becoming significantly more attractive to parents.

We focus on sociodemographic sorting 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 significant reductions in local private school share of 1.8 percentage points for relevant elementary grades, and marked changes in the sociodemographic composition of public schools using difference-in-differences and triple-differences research designs. These are consistent with an inflow of

\(^4\)Prominent among these are Krueger (1999) and Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan (2011).
higher socioeconomic status students into public schools.\textsuperscript{5}

Using similar sources of variation, we next estimate the total value of the reform using local house prices, finding that households were willing to pay a significant amount more for a house in an area where the policy had been implemented. Specifically, a one standard deviation increase in CSR implementation, captured by an intuitive index that we propose, is associated with around a 10 percent increase in house prices. 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 half once we control for observable measures of school sociodemographics and teacher quality, suggestive of the importance of indirect general equilibrium effects.

To separate these components out more precisely, accounting for correlated unobservables, our second contribution is to develop an estimable structural framework that allows us to gauge the relative importance of the policy’s direct and indirect benefits on a common scale for the first time, using test scores. In line with influential recent research (see Chetty et al. (2011) and Chetty, Friedman, and Rockoff (2014)), underscoring the importance of input persistence, the framework also 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 will not be valid.

As our third contribution, we use a generalization of the differencing approach to estimate the parameters of the structural framework, identifying direct and indirect effects on a common footing and the persistence of each. Given the parametrization of the production technology, we show how each observed grade-year score can be separated into a pure class size effect, comparable to experimentally-derived estimates, and the effect arising from the general equilibrium change in student quality, which subsumes the effect of own ability and peers. Our main focus is on these two of these channels, while controlling for induced changes in teacher quality.

\textsuperscript{5}Researchers studying California’s CSR reform face the data limitation that individual student information is not available, and students cannot be followed over time. See Section 4.
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. 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 structural estimates indicate that both effects are significant and positive, with the indirect sorting effect being at least as great in magnitude (0.17σ in terms of mathematics scores) as 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. Further, we find that both the direct and indirect effects persist into subsequent years, at approximately the same rate in each case.

The magnitudes of our direct and indirect estimates in the context of CSR suggest that researchers should view household sorting as a primary factor when assessing the impact of large-scale reforms that alter public school quality. Based on our findings, analyses that ignore re-sorting effects are likely to understate, perhaps to a high degree, the full impact of major education reforms in other contexts. We develop the implications of the analysis below.

The rest of the paper is organized as follows: The next section describes how the current study relates to prior work. Section 3 sets out a framework to clarify the direct and indirect effects of a major education reform, and highlight measurement difficulties that our approach seeks to overcome. Section 4 discusses the institutional background to CSR and the rich data we have assembled. Section 5 presents the two reduced-form strategies that we use to explore the effects of the reform on school enrollments and house prices respectively, along with the corresponding results. Those results motivate the structural framework for analyzing general

---

6 Event study-type graphs indicate that parallel trends hold pre-treatment.
equilibrium effects on the basis of test scores in Section 6. Estimates of the framework are give in Section 7, the results are interpreted in Section 8, and Section 9 concludes.

2 Relation to Prior Literature

This paper contributes to a substantial and important literature studying the direct effects of changes in class size. While the class size literature has yet to reach a consensus, as studies 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 et al., 2012; Gilraine, 2017) and no effects (Hoxby, 2000; Dobbelsteen et al., 2002; Asadullah, 2005; Gary-Bobo and Mahjoub, 2006; Leuven et al., 2008; Battistin et al., 2016), we view the experimentally-derived findings from Project STAR as the most relevant comparison set for our estimate of the direct effect, as the implementation details of the reform are quite similar and the associated prior estimates are highly credible. Our estimate is nearly identical to that from Krueger and Whitmore (2001)’s analysis of Project STAR.

Existing research studying the effect of CSR on student achievement in California has delivered mixed results. Bohrnstedt and Stecher (2002), Stecher et al. (2003) and Funkhouser (2009) find that CSR has minimal effects on student achievement, while Unlu (2005) and Jepsen and Rivkin (2009) indicate that CSR improved student test scores by 0.2-0.3 and 0.06-0.1 of a standard deviation, respectively. In attempting to explain these conflicting findings, researchers have focused on changes in the teacher labor market (Bohrnstedt and Stecher, 2002; Jepsen and Rivkin, 2009) and changes in the use of combination classes (Sims, 2008).

Our analysis complements that of Jepsen and Rivkin (2009), as noted, exploring indirect effects but with an emphasis on sociodemographic sorting. Our study makes use of the timing of the reform to identify the sorting effect: Bayer et al. (2004) follow a more structural approach using cross-sectional Census data to study the size of social multipliers in education:
an initial improvement in local public school quality leads to additional sorting by high-SES students given their higher valuations of school quality, which can further raise public school quality, as households care about the demographic composition of schools. In addition, Estevan (2015) use an education finance reform in Brazil to highlight reductions in private school enrollment in response to increase public education expenditure.

Recent innovative research by Dinerstein and Smith (2016) documents a similar unintended effect of improving public school quality in a New York City setting, following the city’s “Fair Student Funding” reform of 2007-08. Specifically, private school students are drawn into public schools either through choice or because of the forced closure of many small private schools. As some students were led to attend lower quality public schools than the private schools they left, this factor served to offset some of the aggregate achievement gains due to the funding reform. Our analysis complements that of Dinerstein and Smith, who contrast the effects of higher public school quality (measured in terms of value-added) on incumbent public school students, who also benefit from a change in peer quality, with the impact of the reform on students switching into public school. We measure the direct effect of the CSR reform on students in public school alongside an indirect general equilibrium effect consisting of both the benefits to former private school students switching into public school and benefits to incumbent public school students arising from improved peer quality. A suggestive bounding exercise indicates that the latter spillover gains constitute a significant share of the indirect effect.

An influential recent literature examines the persistent effects of inputs in education, notably Chetty et al. (2011) in the context of the long-run (labor market) impacts of Project STAR, and Chetty et al. (2014) in terms of the persistence of teacher inputs. Our research builds on that prior work, allowing the sociodemographic composition of schools (as well as smaller classes) to have persistent effects on student achievement.

Several papers have sought to estimate parents’ WTP for smaller classes based on hedonic estimation strategies. For example, Clark and Herrin (2000) find that reducing the district-
wide average class size by 1.6 students raises property values by 5.2% and Rohlfs and Zilora 
(2014) find that parents are willing to pay $2,000 to $18,000 per year in 2010 dollars for a 
class size of 15 relative to 24 students. Unlike those studies, we account for possible general 
equilibrium effects of large-scale CSR programs.

3 Framework for Policy Analysis

This section sets out a framework that forms the basis of the technology we take to the 
data in our empirical application. It clarifies the direct and full (general equilibrium) effects 
of the policy, the full effect including both the indirect sorting effect and changes in teacher 
quality. We also draw attention to challenges that arise when estimating both the direct 
and indirect effects. In Section 6 below, we will develop a multiple differencing strategy that 
allows us to address these challenges.

We start by considering a general environment in which some outcome $y$ depends on a 
policy parameter $\theta$ and a set of controls $X$. The relationship linking $y$ to $\theta$ and $X$ is captured 
by a function $f(\cdot)$. With time subscripts, at a time $t$ we have $y_t = f(X_t, \theta_t)$. Suppose in the 
simplest (binary) case that there is an initial period, $t = 0$, in which policy is characterized 
by $\theta = \theta_0$. This serves as a reference point for subsequent time differencing. In period 1, the 
policy maker implements a policy change, setting $\theta = \theta_1$. Interest centers on the impact of 
the policy change, $\Delta \theta \equiv \theta_1 - \theta_0$, on outcomes, the outcome vector in turn feeding into the 
policy maker’s objective.\(^7\)

In this general setting, we can think of the direct effect of the policy as the change 
in the outcome relative to the change in the policy, holding the control vector fixed, with 
$X_t = X_c, \forall t$. This can be written as $\frac{\Delta f}{\Delta \theta}|_{X=X_c} \equiv \frac{f(X_c, \theta_1) - f(X_c, \theta_0)}{\theta_1 - \theta_0}$. The full effect of the policy, in 
contrast, allows the control vector to change in response to the policy change. We write this 
as $\frac{\Delta f}{\Delta \theta} \equiv \frac{f(X_1, \theta_1) - f(X_0, \theta_0)}{\theta_1 - \theta_0}$, to make explicit the possible dependence of $X$ on the policy.

\(^7\)Our focus will be on the technology rather than decisions made by policy makers, the latter being taken 
as exogenous.
parameter $\theta$ in a given period. In turn, the difference between the full and direct effects of the policy can be divided into various indirect effects, though this requires more structure to be introduced.

In order to understand the direct, full, and indirect effects of the policy change on outcomes, a key task for applied research involves uncovering the function $f(.)$. That will be our primary focus. Given we will be considering an education setting that features a major policy change, we will think of outcomes being test scores – these could be at the individual student level or some higher level of aggregation. The function $f(.)$ then describes an education production function in which school resources, teacher quality and student characteristics jointly affect measured performance, captured by test scores.

For tractability, we will use a linear approximation to the true education production function, following the bulk of the education literature. We start with a simple specification that allows only contemporaneous inputs to affect output, focusing on $\{R_t, Q_t, X_t^S\}$, where $R_t$ measures resources at time $t$, $Q_t$ represents teacher quality, student characteristics in the school are given by $X_t^S$. We also allow for a further additive noise component, $\epsilon_t$, to reflect unobservable, random influences on contemporaneous test scores. Thus, we have:

$$y_t = \gamma + \gamma_R R_t + \gamma_{X} X_t^S + \gamma_Q Q_t + \epsilon_t.$$  

(3.1)

According to this parameterization, we will think of the policy change in the simplest instance as the product of the (fixed) parameter $\gamma_R$ multiplied by the change in school resources ($\Delta R \equiv R_1 - R_0$), comparing period 1 with base period 0. A substantial literature (discussed in the previous section) has sought to estimate the $\gamma_R$ parameter, particularly in the context of class size reduction. Multiplying the estimated parameter by the change in class size (determined by school resources) then yields the direct impact of the policy change in terms of test scores.

---

8See the classic discussion in Todd and Wolpin (2003).

9Below, as is realistic, we will allow for inputs to have cumulative effects.
For large-scale reforms, policy makers would like to have a sense of the full impact of the policy, including induced effects that may reinforce or counteract the direct effect of the policy on outcomes. Our goal will be to keep track of the indirect effects, and place these alongside the direct effect of the reform.

In a class size context, prior work has documented the way that changes in school resources across an entire education system lead to induced changes in teacher quality (\(\Delta Q\)) based on the same pre-post contrast, so \(\frac{\Delta Q}{\Delta R}\) is non-zero. These should be allowed for (via \(\Delta y = \gamma Q \Delta Q \Delta R\)), given the simple technology in (3.1)) when determining the full impact of the policy; in our application below, we will use a control strategy to account for class size changes affecting teacher quality (\(h(Q_t)\)) at the local level.

Our main focus is on the induced effect of more resources for public schools on the mix of students in public school, as students switch from local private schools (at least where private schools have some prior presence). Changing the mix of students (\(\Delta X^S\)) may affect outcomes if the incoming students are of higher ability than students already enrolled in public school and score more highly themselves (the ‘own’ effect) or through spillover benefits to incumbent students; given the simple technology above, this would be written \(\gamma_S \frac{\Delta X^S}{\Delta R}\).

The technology we seek to estimate below will be more general, in light of compelling evidence from the recent empirical literature (see Chetty et al. (2014)) that underscores the cumulative nature of education production. Accordingly, inputs in one period will be allowed to have persistent effects in subsequent periods as students acquire (and retain) knowledge and skills. We will write the technology to account for the persistent effects of both class size and student sociodemographics as

\[
y_t = \gamma + \gamma_R \sum_{\tau=0}^{L} \delta_R^t R_{t-\tau} + \gamma_S \sum_{\tau=0}^{L} \delta_S^t X^S_{t-\tau} + h(Q_t) + \epsilon_t
\]  (3.2)

The effect of past ‘resources’ (smaller classes) on current test scores is parametrized by \(\delta_R\) – a parameter of interest in prior research (see, for example, Krueger and Whitmore (2001).
and Ding and Lehrer (2010)). Specifically, $\delta_R$ measures the effect on test scores of a one unit increase in resources one period ago that persists into the present. Resources from at most $L$ periods ago are allowed to influence current test scores, following a geometric decay. Similarly, $\delta_S$ captures the persistent effects of past school demographic compositions on current test scores, also following a geometric decay over (at most) $L$ periods.

Adding persistence implies that direct and indirect effects from earlier periods can now accumulate over time: the richer structure (beyond the static case) allows expressions for each to be written down.\(^{10}\)

### 3.1 Estimation Challenges

The structural portion of the paper seeks to uncover the parameters in equation (3.2). Two main sets of challenges arise in doing so.

The first involves locating 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 the overall effect of major education reforms, we need to account for the fact that teacher quality and school compositions may adjust in response, in turn affecting outcomes through two separate (indirect) channels. Thus, there is a need to identify more than one main effect.

The second challenge involves accounting for persistence, which compounds the difficulty of disentangling direct and indirect effects. As we show below, persistence of the direct and indirect effects of the reform 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.

Our structural approach, using the relatively parsimonious specification of the technology in (3.2), is intended to address both issues. This approach makes use of institutional features

\(^{10}\)For example, a shock to class size $l$ periods ago will give rise to an indirect sorting effect (in terms of test scores) in the current period of $\gamma_S \delta_S \frac{\partial X^S_l'}{\partial R_l}$, where $\frac{\partial X^S_l'}{\partial R_l}$ measures the induced within-period sorting response $l$ periods ago.
associated with the CSR reform, which we describe next.

4 Institutional Background and Data

In this section, we discuss the policy context and relevant institutional background to the CSR reform, along with the rich data set we have assembled.

4.1 Institutional and Policy Background

California’s state-sponsored class size reduction program, put in place in the spring of 1996, was the largest of its kind implemented in the United States up to that time. 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.

California lawmakers, motivated in part by Project STAR, enacted the class size reduction 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.

---

11 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.

12 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.

13 See, for example, a report from the associated legislative discussions, available at http://files.eric.ed.gov/fulltext/ED407699.pdf.
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.\textsuperscript{14} 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.\textsuperscript{15} 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.\textsuperscript{16}

Table 1 shows the policy coverage (and also data availability, discussed below). Table 2 highlights the timing of CSR implementation that we exploit, also showing the proportion of students in a CSR-compliant class in grades K-3 for school years 1996-97 through 2001-02. Despite participation in CSR being nearly universal, some districts and schools still chose not to implement CSR. Districts did 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.\textsuperscript{17} 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 reason.\textsuperscript{18}

The initial announcement and roll-out of the actual policy was both sudden and unantic-

---

\textsuperscript{14}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.

\textsuperscript{15}We exploit this differential timing of implementation by grade in our identification strategies below, both when studying changes in private school share and in the structural analysis.

\textsuperscript{16}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 will use the school adoption decision in our school-level house price design and our triple-differences approach, both described below.

\textsuperscript{17}See http://www.lao.ca.gov/1997/021297_class_size/class_size_297.html.

\textsuperscript{18}See http://www.classize.org/summary/97-98/summaryrpt.pdf
ipated, 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.\footnote{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 in our paper. The Consortium was composed of five research institutions: the American Institutes for Research, RAND, Policy Analysis for California Education (PACE), WestEd, and EdSource.} This problem was compounded by issues related to testing and data collection. Student testing did not take place 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.\footnote{Earlier tests in the state – for instance the CLAS test – were discontinued in the face of budget cuts and union resistance. Appendix C offers a quick primer on California statewide testing.} Additional data limitations included a lack of student or classroom-level data and an inability to track teachers over time.\footnote{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.}

4.2 Data

In this subsection, we describe the data sources that we draw on in the empirical analysis. (For a more detailed description, see Appendix Table A.1.)

The California Department of Education (CDE) provides three types of data that we combine in our study.\footnote{The CDE data are accessible at http://www.education.ca.gov/ds/.} The first consists of the student enrolment of all public schools and districts at the grade level, from the 1990-91 through the 2012-13 school years. We augment the enrolment data to incorporate demographic characteristics, including race, ‘English as a Second Language’ (ESL) status,\footnote{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.} Free or Reduced-Price Meal status\footnote{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.} and CSR implementation status for all schools in the 1998-99 to 2003-04 school years inclusive.

Second, the CDE also provides grade-level enrolment data for private schools from 1990-91 to 2012-13 inclusive. No demographics are available beyond overall enrollments. Together,
these two data sets allow us to study the effects of CSR on public school compositions and local private school share.

The third data source comprises test score data from California’s Standardized Testing and Reporting Program (STAR) for grades two and higher. All students in grades 2 through 11 (with some minor exceptions\textsuperscript{25}) took the Stanford Achievement Test in both mathematics and English near the end of the academic year.\textsuperscript{26} The Stanford Achievement Test was a national norm-referenced multiple-choice test that was introduced in the 1997-98 school year. Given that the policy was in place for first grade since the 1996-97 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.\textsuperscript{27}

There is a further complication: in the 2002-03 school year, the 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. In order to make clear some of the constraints this introduced, we must limit some of our analysis to the academic years 1997-98 through 2001-02, even though test scores are reported until 2012-13. This is because the monotonicity of scores by grade during that time period is no longer preserved for the 2002-03 academic year and onward due to the test change.

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

\textsuperscript{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 none-the-less. For example, in 1998-99, over ninety-three percent of students in grades 2-11 wrote the relevant test.
\textsuperscript{26}Testing dates were generally between March 15 through May 25 of a given academic year.
\textsuperscript{27}Our structural strategy is designed to use that variation.
if students in a school-grade-year averaged at the 60th percentile on the standardized test, this would mean that, on average, they scored 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 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 they adopted CSR, and this – looking ahead – is used in our identification strategy to determine the impact of CSR on house prices, Table 3 gives 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 who are 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 relatively small.

A fourth data set 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 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).\textsuperscript{28} The boundary files provide coverage for about 40 percent of elementary schools in California.\textsuperscript{29}

Table 4 provides summary statistics for the main variables used in our analysis. Along

\textsuperscript{28} Appendix D provides a brief description of the process we use to match the housing transactions data to school districts and school attendance zones.

\textsuperscript{29} It is worth noting that the school attendance zone boundary files provide disproportionate coverage in urban areas. This implies that the school-level house price design we employ below may not be representative of the effect of CSR on California as a whole.
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.\textsuperscript{30} Panel A also reveals that the private school share of enrolment at the state level declined during the period of interest, falling from 9 percent prior to CSR implementation to 8.4 percent afterward.\textsuperscript{31} In addition, there was a dramatic 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.\textsuperscript{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.

5 Descriptive Evidence

In this section, we present descriptive evidence of general equilibrium responses to the reform. Here, we exploit differences in CSR adoption across time, grades, schools and districts to study three outcomes of interest: private school share, public school composition, and house prices. For each outcome in turn, we describe the differencing strategy we use, then present results, consisting of both visual evidence and regression estimates.

\textsuperscript{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).

\textsuperscript{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 overall trend.

\textsuperscript{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.
5.1 Private School Share

To explore the effect of CSR on private school shares and public school demographic compositions, 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.\textsuperscript{33}

We first specify a simple difference-in-differences approach, which compares treatment and control grades before and after the reform came into effect. Focusing on private school share, the analysis uses the following regression (weighted by district-grade-year enrollment):\textsuperscript{34}

\[
share_{dgt} = \beta_0 + \beta_1 post_{gt} + \beta_2 treat_g + \beta_3 (post_{gt} \ast treat_g) + \eta_d + \theta_t + \delta_g + \phi X_{dgt} + \epsilon_{dgt},
\]  

(5.1)

where $share_{dgt}$ is the private school share for district $d$ in grade $g$ at time $t$;\textsuperscript{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 $\eta_d$, $\theta_t$ and $\delta_g$ are district, time and grade fixed effects, respectively.

The difference-in-differences coefficient of interest is $\beta_3$. It is identified under the assumption 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, we provide suggestive evidence as to its validity in Subsection 5.1.1.

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

\textsuperscript{33}Later on, we also use differences in CSR adoption across districts to create an additional layer of differencing.

\textsuperscript{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 is therefore omitted.

\textsuperscript{35}Formally, $share_{dgt}$ is defined as the enrollment in private schools for $d$-$g$-$t$ divided by the total enrollment for $d$-$g$-$t$. 
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,\textsuperscript{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 percent of K-3 students in a CSR participating school-grade within a district for the 1998-99 school year.\textsuperscript{37} Formally,
\[
CSR_d = \sum_{s \in d} \sum_{g = 0}^{3} \frac{\mathbb{1}\{CSR_{sg}\} * (enroll_{sg})}{\sum_{s \in d} \sum_{g = 0}^{3} enroll_{sg}},
\]  
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\textsuperscript{38} regression:
\[
share_{dgt} = \beta_0 + \beta_1 post_{gt} + \beta_2 treat_{g} + \beta_3 (post_{gt} * treat_{g}) + \beta_4 (post_{gt} * CSR_d) + \beta_5 (post_{gt} * treat_{g} * CSR_d) + \gamma_d + \theta_t + \delta_g + \phi X_{dgt} + \epsilon_{dgt},
\]  
where all variables other than the intensity measure \(CSR_d\) are identical to those in equation (5.1). The triple-differences coefficient of interest is \(\beta_7\). 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.

\textsuperscript{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.

\textsuperscript{37}Results are similar if this variable is averaged over the 1998-99 through 2003-04 school years.

\textsuperscript{38}Again, we weight by district-grade-year enrollment.
5.1.1 Private School Share Results

**Private School Share:** Our triple-differences identification strategy (as specified in Subsection 5.1) exploits variation in both the time when different grades became subject to CSR and the local intensity of adoption.

With respect to the former type of variation, Figure 1 plots the change in private school enrollment share in CSR versus non-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 significantly 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 grade 1 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 grade 1), there is a pronounced dip down while the shares for other grades remain steady, consistent with there being a switch into public schools for that grade. Similarly, when grade 2 becomes eligible for CSR in public schools at the start of 1997-98, we 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 grade 3 in the first year when those grades became eligible (1998-99).

As for our measure of the local intensity of CSR given in equation (5.2), Figure A.1 provides a map of the change in private school share from the 1995-96 to 1999-2000 school years by school district and Figure A.2 provides an analogous map for our local intensity regressor. These reveal substantial variation, respectively, in our outcome variable and a key regressor of interest.

Given the patterns in Figure 1, we start by reporting the most basic contrast – private school shares ‘before’ and ‘after’ CSR’s introduction among the treated and untreated grades – in Table 5. We find that, relative to untreated grades, treated grades experience a precisely-estimated 1.4 percentage point decline in private school share following implementation of the policy. This corresponds to the estimate from the difference-in-differences estimator
specified by equation (5.1) without including any controls. We estimate that equation and report the results in Table 6, 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 18 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.

Using district-level CSR participation intensity as an additional dimension of differencing, our preferred triple-differences analysis from equation (5.3) 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. 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, private school share 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.

Number of Private Schools: With the steep decline in private school share induced by CSR, it is seems likely that the extensive private school margin is also affected. We thus briefly investigate private school supply responses on the extensive margin (exit and entry) to the public CSR reform. Using the Private School Universe Survey, Figures A.3(a) and A.3(b) private school entry and exit rates for California alongside those for the rest of the country. After the 1996-97 CSR reform, we see a steep increase in private school exit rates

---

39 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.

40 The Private School Universe Survey is run by the National Center for Education Statistics. It is available at https://nces.ed.gov/surveys/pss/pssdata.asp.
and a steep decline in entry rates in California relative to the rest of the country. Figure 3 then takes these differential entry and exit rates and maps them into the number of private schools per 1000 school-aged children in California and the rest of the country. As expected, there is a steep decline in private schools per capita after the 1996-97 reform 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, using time (pre- vs post-CSR), state (California vs. rest of the country) and whether the private school served CSR grades as the three layers of differencing. 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.

5.2 Persistence after Switching

Given the evidence relating to the initial impact of the reform, next we consider how long such changes persisted. This will be relevant in the structural section below when we wish to gauge the proportion of students re-entering private schools – say, after completing grade 5 in a K-5 school – in order to estimate whether sorting is transitory or not.

One hypothesis is that students previously in the private school system remain in the public system once they make the initial switch; another is that they return to the private system after completing all grades offered by the public school that they switched into initially. To ascertain which is more likely, 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:

$$ grade^i'share_{dc} = \beta_0 + \beta_1 D_{dc} + \beta_2 f(\text{cohort}_{dc}) + \beta_3 D_{dc} \times f(\text{cohort}_{dc}) + \eta_d + \epsilon_{dc} \quad \text{for} \quad -b \leq \text{cohort}_{dc} \leq b, \quad (5.4) $$

where $grade^i'share_{dc}$ 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, \( \text{cohort}_{dc} \) 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), \( \eta_d \) is a district fixed effect, and \( b \) is some bandwidth. This regression discontinuity design identifies the CSR effect on 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 \( \beta_1 \), 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 \( \beta_1 \) 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 Table 8, 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. \(^{42} \) 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.)

\(^{41} \)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 cohort.

\(^{42} \)Table A.4 reports the number of elementary schools by grade configuration in California. Elementary schools are divided approximately equally between K-5 and K-6 configurations.
5.3 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 the private system affects public school sociodemographic compositions. To do so, we analyze the effect of CSR on public school student demographics by exploiting the degree of private school presence locally. Appendix Section B reports additional evidence taking advantage of school-level differences in the reform’s implementation, corroborating the results presented here.

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:

\[
demo_{s,gt} = \beta_0 + \beta_1 post_{gt} + \beta_2 treat_{g} + \beta_3 1\{Buffer < x \ km\}_s + \beta_4 (post_{gt} * treat_{g}) \\
+ \beta_5 (post_{gt} * 1\{Buffer < x \ km\}_s) + \beta_6 (treat_{g} * 1\{Buffer < x \ km\}_s) \\
+ \beta_7 (post_{gt} * treat_{g} * 1\{Buffer < x \ km\}_s) + \eta_s + \theta_t + \delta_g + \phi X_{s,gt} + \epsilon_{s,gt},
\]

where \(1\{Buffer < x \ km\}_s\) indicates whether a private school is within a radius of \(x \ km\) of school \(s\) and all other variables are defined as before.\(^{43}\) The triple-differences estimate \(\beta_7\) 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 for public schools within \(x \ km\) of a private school and those further away in the absence of the reform.

5.3.1 Public School Composition Results

Here, we shed light on whether public school demographics shifted as a result of the CSR-induced decline in private school share. Given that the proportion of white students is about fifteen percent higher and the proportion of Hispanic students is about twenty three percent lower in private schools initially compared to their public counterparts (see Table 9),

\(^{43}\)Only private schools with ten or more students in kindergarten through third grade are included.
inflows to the public system are likely to consist mainly of the first two groups.\textsuperscript{44} Although public-private demographic disparities among the proportion of Asian and black students are smaller, one might also expect an influx of Asian students and a decline the black students in the public system. This is indeed what we find.

Figure 4 gives a visual representation of 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 considerable drop in the proportion of Hispanic and black students along with a substantial rise in the fraction of white students. We then incorporate the comparison between CSR and non-CSR grades described in equation (5.5).

Panel B of Table 9 reports estimates for the distance design, using three different ‘closeness’ measures: 1.5, 3 and 5 kilometres.\textsuperscript{45} Relative to public schools that have no nearby private competitors, the results suggest that CSR led to a significant increase in the fraction of white students and 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. 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 are about four times these pre-reform estimates. For Hispanic students, the point estimates are indistinguishable from zero in the pre-reform years and become statistically significant once CSR is implemented.

\textsuperscript{44}While we do not have access to very detailed private school demographics, 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.

\textsuperscript{45}As is apparent, the choice of the ‘closeness’ measure has very little impact on the results.
5.4 House Prices

We have already shown evidence of reform-induced student sorting between the private and public school systems. This is likely to alter public school quality. To the extent that such an amenity matters to households, equilibrium house prices should adjust accordingly. To quantify how far these changes 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 5.1.\(^{46}\) As in the prior subsection, Appendix Section B exploits school-level differences in the reform’s implementation to highlight the effect of CSR on house prices, corroborating the results presented here.

Accordingly, we estimate the following weighted regression:\(^{47}\)

\[
price_{dt} = \beta_0 + \beta_1 Post_t + \beta_2 CSR_d + \beta_3 (Post_t \ast CSR_d) + \phi X_{dt} + \eta_d + \theta_t + \epsilon_{dt}, \tag{5.6}
\]

where \( price_{dt} \) is the average house price in district \( d \) at time \( t \), \( Post_t \) is a CSR implementation indicator, \( X_{dt} \) is a set of controls that consist 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), and \( \eta_d \) and \( \theta_t \) are district and time fixed effects, respectively. The coefficient of interest is \( \beta_3 \), which has a causal interpretation 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.

---

\(^{46}\) We are limited to two dimensions of differencing, as differences between CSR and non-CSR grades cannot be exploited when house prices are the outcome of interest.

\(^{47}\) The regression is weighted by the number of housing transactions to account for the fact that some district-years are associated with very few transactions.
5.4.1 House Price Results

As discussed, we identify the effect of CSR on house prices using district-level variation by comparing post-reform house prices to their pre-reform baseline across high- and low-share CSR districts. Figure A.2 shows that the district variation that we draw upon is relatively evenly distributed across the state. Figure 6(a) displays trends among high- and low-share CSR districts: 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 (5.6), which are reported in Table 10. 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, which may themselves be highly valued by parents.

Our preferred estimate to capture the full effect of the reform – referring back to Section 3 – 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 on teacher or peer quality. The point estimate of $134,600 implies that a one standard deviation increase in CSR implementation, measured in equation 5.2, is associated with around a 10 percent increase in house prices – a substantial amount.48

While column (3) captures the full capitalization of the reform into house prices, it is informative to look at the partial equilibrium effect of the reform once the indirect general equilibrium effects on teacher and demographic characteristics are accounted for. First, column (4) controls for teacher characteristics such as education and experience. We do not find that the general equilibrium teacher effects studies in Jepsen and Rivkin (2009) are capitalized in house prices, which is consistent with Imberman and Lovenheim (2016) that house prices do not respond to teacher quality, per se.49 School student characteris-

---

48 A one standard deviation in the CSR implementation measure is 0.15 and the $134,600 point estimate represents a 69 percent increase in house prices from the 1995 pre-CSR (weighted) mean of $195,000.

49 Column (3) and (4) are not estimated off the same sample since teacher characteristics are not available.
tics are, however, capitalized into house prices: once student characteristics are controlled for in column (5), a one standard deviation increase in CSR implementation is now being associated with a 6 percent increase in house prices (the difference between column (3) and (5) is statistically significant at five percent). The partial equilibrium effect of the reform, controlling away the general equilibrium effects on teacher and peer quality, thus decreases the capitalization of the reform into house prices by over half.

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. 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.

The evidence we have presented relating to house prices is suggestive, of both the significant size of the full effect of the reform and the importance of indirect effects related to sociodemographic sort and teacher quality. Because the observable sociodemographic and teacher controls we include are likely to be correlated with unobservable determinants of school performance, however, we develop a more compelling approach to separating out the indirect general equilibrium effects of the reform in the next section.

6 Structural Framework

To understand how the overall effect of the California class size reform can be separated into direct and indirect channels affecting student achievement, we model test scores in a way that allows us to place each channel on a common footing.

Building on the discussion in Section 3 above, the introduction of the reform is assumed

before the 1994-95 school year. The sample selection has little effect; the difference-in-differences point estimate when column (3) is restricted to the column (4) sample is 12.32 (s.e. 4.21).
to have two main effects: (1) the direct effect of reducing class size on student learning, $\gamma_R$; and (2) the indirect general equilibrium effect from changes in student composition $\gamma_S$, which arises as a result of sorting between the private and private school systems. (This includes the ‘own’ effect and a spillover effect.) A third type of response involves the indirect general equilibrium effect from changes in teacher quality, which occurs through the teacher labor market (as in Jepsen and Rivkin 2009).

To identify the direct and indirect policy responses, we will draw contrasts with a counterfactual world in which the reform was not enacted. To do so, we describe the underlying technology, then show how a careful differencing approach allows us to uncover the causal effects of interest.

**Technology**

We consider an additive, linear technology, following the bulk of the literature. Further, we allow for the differential persistence of the direct and indirect sorting effects. Adapting equation 3.2 from above, we write the school-grade-year $(s,g,t)$ test score $y_{sgt}$ as a function of current and past student, school and teacher inputs:

$$y_{sgt} = \gamma + \gamma_R \sum_{\tau=0}^{L} (\delta_R)^{\tau} R_{sg,t-\tau} + \gamma_S \sum_{\tau=0}^{L} (\delta_S)^{\tau} X_{sg,t-\tau}^{S} + h(Q_{sgt}) + \epsilon_{sgt}. \quad (6.1)$$

To capture the cumulative nature of the student learning technology, this equation allows inputs from $L$ periods prior to current time $t$ to affect current scores, where the index $\tau$ increments both successive grades ($g \in \{0,1,2,3,4,...\}$) and academic years ($t \in \{1996-97, 1997-98, 1998-99, 1999-00\}$).

We assume that the class size effect and the sociodemographic composition effect persist over time at rates $\delta_R$ and $\delta_S$, respectively. Prior work (see Krueger and Whitmore 2001) indicates that class size effects are likely to have an impact on student achievement both contemporaneously and in future. Whether the same is true for student sociodemographics
is testable, as we will show in what follows.

The structure for measuring the impacts of class size and demographics is parsimonious, chosen in light of what we are able to identify using a multiple differencing approach (described in the next subsection). As will be seen below, we will use a control technique for teacher quality, $Q$, and so do not impose parametric assumptions in the equation. Consequently, we will omit teacher quality from the discussion for clarity and return to it when describing our estimation approach.

**Estimating Equation**

To understand the causal impact of grade-specific reductions in class size, including any induced changes in school demographics, we will draw notional contrasts between observed scores at the school-grade-year level and counterfactual scores that would have prevailed had the reform not been enacted. To make clear what these are, we first provide some notation. The comparisons involve school averages, where the total number of schools is given by $N_s$. 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 and the unobserved (superscripted by ‘$u$’) average score resulting from a counterfactual setting in which the reform was never implemented.

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 shown in Table 1. For those treated grade-years, we have that $y_{gt} \neq y_{u_{gt}}$. For all other control combinations, $y_{gt} = y_{u_{gt}}$, implying, for untreated grade-year combinations, that $\Delta y_{gt} = 0$. For instance, the above equation implies 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.

\footnote{Note that grade 2 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).}
This gives the following estimating equation:

\[ \Delta y_{gt} = \gamma_R \sum_{\tau=0}^{L} (\delta_R)^{\tau} \Delta R_{gt-\tau,t-\tau} + \gamma_S \sum_{\tau=0}^{L} (\delta_S)^{\tau} \Delta X_{gt-\tau,t-\tau} + \Delta \epsilon_{gt}, \]  

(6.2)
suppressing the teacher quality effect for clarity.\(^{51}\)

### 6.1 Estimation Approach

A practical challenge in taking this estimating equation to the data is that \( \Delta y_{gt} \) 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 \( t' \):

\[ y_{gt}^u - y_{g't}^u = y_{g't'}^u - y_{g't'}^u. \]

Because of the evidence in the prior literature that class size reductions have persistent effects (and large scale reforms are likely to have spillovers), a difference-in-differences measurement approach will not be valid, for reasons we now explain.

**Difference-in-Differences:** A simple difference-in-differences (‘D-in-D’) specification would make 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 several assumptions (linear technology, no spillovers, and no persistence of inputs), this recovers the direct causal impact of class size reduction.

It is instructive to work through an example in the context of our structural model and the institutional variation we use, both to show where the simple D-in-D approach goes awry, and to pin-point where we get identification of the direct and indirect effects of CSR from using our approach.

\(^{51}\)We describe how teacher quality measurements are included in the estimation approach at the end of the section, and in a more detailed appendix.
Referring to Table 1, grade 3 throughout the state became subject to the reform in the 1999-2000 academic year (and remained treated subsequently). In prior years, grade three class sizes were unaffected. Thus, one could construct the first difference by deducting the statewide average test score in grade 3 in 1998-99 from the statewide average for grade 3 in 1999-2000. The table suggests numerous suitable control grades: those that, over the same timespan, were never subject to CSR. Take, for instance, grade 4 in 1998-99 and 1999-2000, and construct the analogous difference in average scores.

The structure above makes clear when the simple D-in-D just rehearsed recovers the direct causal impact of the policy – that is, \( \Delta y_{3,99-00} - \Delta y_{4,99-00} = \gamma_R \Delta R_{3,99-00} \). Equation 6.2 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 grade 3 students in 1999-00 as:

\[
\Delta y_{3,99-00} = \delta^2_R \gamma_R \Delta R_{1,97-98} + \delta_R \gamma_R \Delta R_{2,98-99} + \gamma_R \Delta R_{3,99-00} + \delta^2_S \gamma_S \Delta X_{1,97-98}^S + \delta_S \gamma_S \Delta X_{2,98-99}^S + \gamma_S \Delta X_{3,99-00}^S + \Delta \epsilon_{3,99-00}.
\] (6.3)

Similarly, considering grade 4 students in the 1999-00 school year (who were subject to the CSR reform in grade 1 (1996-97), grade 2 (1997-98) and grade 3 (1998-99)), the differences between observed and counterfactual average test scores can be expressed as:

\[
\Delta y_{4,99-00} = \delta^3_R \gamma_R \Delta R_{1,96-97} + \delta^2_R \gamma_R \Delta R_{2,97-98} + \delta_R \gamma_R \Delta R_{3,98-99} + \delta^3_S \gamma_S \Delta X_{1,96-97}^S + \delta^2_S \gamma_S \Delta X_{2,97-98}^S + \delta_S \gamma_S \Delta X_{3,98-99}^S + \gamma_S \Delta X_{4,99-00}^S + \Delta \epsilon_{4,99-00}, \quad (6.4)
\]

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., \( \Delta R_{0,95-96} = 0 \)).

Comparing students in CSR and non-CSR grades amounts to subtracting equation 6.4
from 6.3, which yields:

\[
\Delta y_{3,99-00} - \Delta y_{4,99-00} = \gamma_R \Delta R_{3,99-00} - \delta_R^2 \gamma_R \Delta R_{1,96-97} - \delta_S^2 \gamma_S \Delta X_{1,96-97}^S ,
\]  

(6.5)

where the equality follows because CSR affects all grades that have implemented CSR equivalently (i.e., \( \Delta R_{g,t} = \Delta R_{g',t} \) \( \forall g, g' \)).\(^{52}\) It is clear in the above equation that the direct effect of class size, \( \gamma_R \Delta R \), which is analogous to experimentally-derived estimates (for instance, Project STAR), is not identified if there is persistence in the student learning technology (i.e., \( \delta_R \neq 0 \) or \( \delta_S \neq 0 \)).

**Our Strategy:** The preceding argument highlights the need for a more nuanced approach – one that takes advantage of the unique grade-by-grade roll-out of CSR in California.

Our strategy for gauging the relative sizes of the direct and indirect effects of more school resources (in the form of CSR) takes advantage of the way the policy was rolled out. In successive years, as already rehearsed, schools were able to reduce class sizes in grade 1 (in 1996-97), then in grade 2 the next year, followed by both kindergarten and grade 3 in 1998-99 and 1999-2000 (depending on whether schools opted for kindergarten or grade 3 first). Further, within each year, adoption was not uniform. A multiple differencing approach allows us to exploit these various contrasts.

In the next subsection, we show how the main parameters of interest can be identified using a combination of the rich data variation available and some minimal structure. To highlight the essence of the approach, we focus for now on separating out the direct effects of the reform from the indirect effect of student sorting. (We discuss accounting for the impact of the reform on teacher quality below.)

Consider the incoming grade 1 cohort for the 1998-99 school year and the impact of the reform on their test scores by grade 4 (in 2001-02). This cohort was subject to the reform for grades 1-3, but was not subject to the reform while in kindergarten as CSR had

\(^{52}\)See Table 2, where the CSR grades have similar class sizes once CSR is implemented.
yet to be implemented there in 1997-98. In addition, the cohort was not subject to CSR in grade 4 (given that grade 4 was never part of the reform), although any change to the student composition engendered by the reform remained. The achievement of that cohort is therefore given by:

$$\Delta y_{4,01-02} = \delta_{R}^{3}\gamma_{R}\Delta R_{1,98-99} + \delta_{R}^{2}\gamma_{R}\Delta R_{2,99-00} + \delta_{R}\gamma_{R}\Delta R_{3,00-01}$$

$$+ \delta_{S}^{3}\gamma_{S}\Delta X_{1,98-99}^{S} + \delta_{S}^{2}\gamma_{S}\Delta X_{2,99-00}^{S} + \delta_{S}\gamma_{S}\Delta X_{3,00-01}^{S} + \gamma_{S}\Delta X_{4,01-02}^{S} + \Delta \epsilon_{4,01-02} \quad (6.6)$$

since the students are not subject to the class size reform in grade 4 (i.e., $\Delta R_{4,02-03} = 0$), but student sorting still affects their achievement in grade 4 since the students induced into the public system by the reform have yet to return to the private section (see the RD results).

Similarly, the grade 3 cohort for the 2002-03 school year has been subject to the policy for four consecutive years (since they were subject to the policy when they were in kindergarten during 1999-00) and so their achievement is:

$$\Delta y_{3,01-02} = \gamma + \delta_{R}^{3}\gamma_{R}\Delta R_{0,98-99} + \delta_{R}^{2}\gamma_{R}\Delta R_{1,99-00} + \delta_{R}\gamma_{R}\Delta R_{2,00-01} + \gamma_{R}\Delta R_{3,01-02}$$

$$+ \delta_{S}^{3}\gamma_{S}\Delta X_{6,98-99}^{S} + \delta_{S}^{2}\gamma_{S}\Delta X_{1,99-00}^{S} + \delta_{S}\gamma_{S}\Delta X_{2,00-01}^{S} + \gamma_{S}\Delta X_{3,01-02}^{S} + \Delta \epsilon_{3,01-02} \quad (6.7)$$

The CSR reform reduced class sizes to twenty or below in each treated grade, making the size of the reform identical in each 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 6.6 and 6.7, we are left with the term $\gamma_{R}\Delta R_{3}$, which represents the contemporaneous effect of the change in school resources (here, given by the reduction in class size) on student achievement.

Next, we show how the structural framework can be used to uncover $\gamma_{S}$, the effect of the influx of private school students into the public system. To do so, we draw on the reduced-form evidence that students remain in the public system until the students are forced to transition to middle school. In light of that evidence, the achievement of grade 6 students
in year $t$ is given by:

$$
\Delta y_{6,t} = \gamma + \delta_R \gamma R \Delta R_{1,t} + \delta_R \gamma R \Delta R_{2,t} + \delta_R \gamma R \Delta R_{3,t}
$$

$$
+ \delta_S \gamma S \Delta X_{1,t}^S + \delta_S \gamma S \Delta X_{2,t}^S + \delta_S \gamma S \Delta X_{3,t}^S + \delta_S \gamma S \Delta X_{4,t}^S + \delta_S \gamma S \Delta X_{5,t}^S + \gamma_S \Delta X_{6,t}^S + \Delta \epsilon_{6,t}
$$

(6.8)

From the reduced-form analysis, we showed that a large proportion of student left the public system when students moved on to a grade in a new school. In California, schools are divided among K-5 and K-6 configurations relatively evenly (as shown in Table A.4). If a proportion $\psi$ (estimable from the reduced-form 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 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 grade 6 students in a K-6 configuration and those in a K-5 configuration gives:

$$
\Delta y_{6,t,K6} - \Delta y_{6,t,K5} = \gamma_S \Delta X_{6,t,K6}^S - \gamma_S \Delta X_{6,t,K5}^S + \Delta \epsilon_{6,t,K6} - \Delta \epsilon_{6,t,K5}
$$

$$
= (1 - \psi) \gamma_S \Delta X_{6,t,K6}^S ,
$$

(6.9)

allowing us to solve for the $\gamma_S$ parameter, given that we know $\psi$ (recovered from the reduced-form analysis).

Using a similar structure, we also uncover the fade-out rate of the reform, $\delta_R$, and the fade-out rate of the effect of the change in student demographics, $\delta_S$.\textsuperscript{53} To do so, we take the parameters $\gamma_R$ and $\gamma_S$ to be known and difference the test scores in grade 4 and grade 3

\textsuperscript{53}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.
in the 2000-01 school year:\textsuperscript{54}

\[
\begin{align*}
\Delta y_{4,00-01} - \Delta y_{3,00-01} &= \gamma + \delta_R \gamma_R \Delta R_{1,97-98} + \delta_R^2 \gamma_R \Delta R_{2,98-99} + \delta_R \gamma_R \Delta R_{3,99-00} \\
&\quad + \delta_S \gamma_S \Delta X_{1,97-98} + \delta_S^2 \gamma_S \Delta X_{2,98-99} + \delta_S \gamma_S \Delta X_{3,99-00} + \gamma_S \Delta X_{4,00-01} + \Delta \epsilon_{4,01-02} \\
&\quad - (\gamma + \delta_R^3 \gamma_R \Delta R_{1,98-99} + \delta_R \gamma_R \Delta R_{2,99-00} + \gamma_R \Delta R_{3,00-01} \\
&\quad + \delta_S \gamma_S \Delta X_{1,98-99} + \delta_S \gamma_S \Delta X_{2,99-00} + \gamma_S \Delta X_{3,00-01} + \Delta \epsilon_{3,01-02}) \\
&= \delta_R^3 \gamma_R \Delta R_{1,97-98} - \gamma_R \Delta R_{3,00-01} + \delta_S^2 \gamma_S \Delta X_{1,97-98}. \\
\end{align*}
\]

\text{Equation} 6.10

Similarly, comparing test scores between grades 4 and 5 in the 2000-01 school year yields:

\[
\begin{align*}
\Delta y_{5,00-01} - \Delta y_{4,00-01} &= \gamma + \delta_R^4 \gamma_R \Delta R_{1,96-97} + \delta_R^3 \gamma_R \Delta R_{2,97-98} + \delta_R^2 \gamma_R \Delta R_{3,98-99} + \delta_S^4 \gamma_S \Delta X_{1,96-97} \\
&\quad + \delta_S^3 \gamma_S \Delta X_{2,97-98} + \delta_S^2 \gamma_S \Delta X_{3,98-99} + \delta_S \gamma_S \Delta X_{4,00-01} + \gamma_S \Delta X_{5,00-01} + \Delta \epsilon_{5,00-01} \\
&\quad - (\gamma + \delta_R^3 \gamma_R \Delta R_{1,97-98} + \delta_R^2 \gamma_R \Delta R_{2,98-99} + \delta_R \gamma_R \Delta R_{3,99-00} \\
&\quad + \delta_S^3 \gamma_S \Delta X_{1,97-98} + \delta_S^2 \gamma_S \Delta X_{2,98-99} + \delta_S \gamma_S \Delta X_{3,99-00} + \gamma_S \Delta X_{4,00-01} + \Delta \epsilon_{4,00-01}) \\
&= \delta_R^4 \Delta R_{1,96-97} - \delta_R \Delta R_{3,99-00} + \delta_S^4 \gamma_S \Delta X_{1,96-97}. \\
\end{align*}
\]

\text{Equation} 6.11

Equations 6.10 and 6.11 form a system of two non-linear equations with two unknowns (\(\delta_R\), \(\delta_S\)), which we then solve for.

\section*{6.2 Identification}

In this subsection, we discuss the identification of key parameters, explaining how we solve for them by incorporating counterfactuals into the differencing strategy. We also highlight explicitly how the timed roll-out of the policy is used to identify the structural model described above.

\textsuperscript{54}The parameters \(\delta_R\) and \(\delta_S\) are over-identified, as a comparison between grades 3 and 4 in the 1999-00 school year will also yield an estimate of \(\delta_B\). The estimate of \(\delta_B\) is nearly identical in either case; we report the average of the two estimates in our results.

35
The Direct CSR Effect ($\gamma_R$)

Identification of $\gamma_R$ requires a grade 3 and grade 4 cohort within the same year (to avoid year effects), whereby the grade 4 cohort has been treated in grades 1-3 and the grade 3 cohort has been treated in grades K-3. In our CSR context, this occurs for the grade 3 and grade 4 cohorts in the 2001-02 school year. On the one hand, the 2001-02 grade 3 cohort receives class size reduction in that year and has been subject to the class size reform for the past three years: kindergarten (1998-99), grade 1 (1999-2000), and grade 2 (2000-01). The 2001-02 grade 4 cohort on the other hand does not receive class size reduction in the current year (since grade 4 is not eligible), but has done so in the prior three years: grade 1 (1998-99), grade 2 (1999-2000), and grade 3 (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 structural model equations (6.6) and (6.7), to solve for $\gamma_R$. Observed achievement differences between these two cohorts are therefore attributable to the fact that the grade 3 cohort receives class size reduction contemporaneously, while the grade 4 cohort does not.

Test scores for the grade 3 cohort may differ from the grade 4 cohort even in the absence of the class size reform. To account for these differences, we impose the plausible 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 grades 3 and 4 in the absence of the class size reform.

The Indirect Sorting Effect ($\gamma_S$)

While identification of $\gamma_S$ requires any grade 6 cohort, we choose the 2001-02 cohort to allow us to add another layer of differencing to control for differences between students in schools with different grade configurations. The 2001-02 grade cohort was subject to the policy since grade 1 (1997-98) and so has faced 3 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 $\gamma_S$ in equation (6.9).

Even without the reform, observed achievement differences among grade 6 students may arise because of innate differences across students in schools with different configurations. To account for those, we first control for innate differences between students in K-5 and K-6 schools by differencing out grade 5 achievement in each school type from their grade 6 achievement. Then, as for $\gamma_R$, we impose the plausible 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 grades 6 and 5 in the absence of the class size reform.

**Accounting for Teacher Quality: Estimation and Identification**

We further extend our identification strategy by augmenting the structural 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 this obstacle, we follow Jepsen and Rivkin (2009) by appealing to variation in observable teacher experience as a proxy for teacher quality. They 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 upon new evidence about the way in which teacher inexperience evolved by grade, presented in Table A.5.\textsuperscript{55} 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 we use for class size and student sorting.\textsuperscript{56}

\textsuperscript{55}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.

\textsuperscript{56}For further details about our estimation approach to account for teacher quality, please see Appendix E.
7 Structural Results

This section presents results from implementing our structural methodology, which places the direct and indirect general equilibrium effects of CSR on a common footing using test scores.

Table 11 provides estimates of the partial equilibrium effect of CSR ($\gamma_R$) and the general equilibrium effect of CSR on student composition ($\gamma_S$). To explain the table layout, the estimate for the latter parameter is calculated using two different assumptions about the proportion, $\psi$, 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 8 and treat $\psi = \frac{2}{3}$, using the fact that two-thirds of the students are estimated to return to the private system during the middle school transition. A lower bound estimate for $\gamma_S$ is provided in columns (3) and (4) by assuming that all students who were drawn into the public system by CSR return to the private system during the middle school transition ($\psi = 1$). In keeping with the findings of Jepsen and Rivkin (2009), we find that controlling for teacher quality is important and thus only report structural estimates that include controls for teacher quality.\footnote{The teacher quality estimates themselves are given in Table 12.}

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 effect of CSR accounts for a 2.2 unit increase in the mean percentile rank of students, which corresponds to a 0.11$\sigma$ increase in the school-grade test score distribution. The magnitude of this estimate is in line with experimentally-derived estimates: for instance, Krueger and Whitmore (2001) report Project STAR raised test scores by about 0.1 standard deviations.

The general equilibrium sorting effect accounts for a 3.3 unit increase in the mean percentile rank of students, which is equivalent to a 0.17$\sigma$ 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 in the next section – specifically, whether it is
plausible to think that spillovers from incoming to existing public school students might be important.

Turning to the fade-out parameters, $\delta_R$ and $\delta_S$, we find that both the direct and indirect effect fade-out at 45-70 percent each year – in line with much of the literature on fade-out which finds that class size test score gains are “reduced approximately to half to one quarter of its previous magnitude” (Krueger and Whitmore, 2001, p. 11), although these test score gains then reappear later in the labour market (Chetty et al., 2011). These estimates are also consistent with the fade-out of teacher effects (see Jacob et al. (2010)).

To summarize, the evidence indicates that general equilibrium student sorting in response to a reform that improves school quality is first order: either it is statistically indistinguishable from the direct partial equilibrium effect or it exceeds the partial equilibrium effect in an economically meaningful way. Thus, experiment-based analyzes appear to substantially underestimate the effect of CSR. More generally, to the extent that CSR is representative of other major reforms to improve school quality, estimates of those effects that abstract from induced sorting are likely to suffer from considerable omitted variables bias. We develop this point in the following section.

8 Interpretation

This section is in three parts. First, we discuss the likely extent of spillovers to existing public school students from the indirect sorting effect we estimated in the previous section. Second, we consider the policy implications of the estimated indirect sorting effect, and then we discuss its potential relevance in other contexts.

8.1 Spillovers

To help interpret the indirect effect identified by $\gamma_S$, we seek to gauge the relative magnitude of its two components: the composition effect and the spillover effect. The composition
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 some benefit from their new peers entering their classes, most likely through peer effects.

To gauge the relative impacts of the two – the composition and spillover effects – we consider two simple scenarios. In the first, we assume there are no spillovers, so the indirect effect is composed entirely of the composition effect. In the second, we assume that there are spillovers in the form of peer effects and use Graham (2008)’s estimate of the linear-in-means peer effect to find the proportion of the total sorting effect that is accounted for by spillovers. In the latter case, since we only assume spillovers in the form of peer effects, the proportion of the total indirect effect is accounted for by spillovers should be taken as a lower bound.

From the estimates in Column (4) of Table 6, there is a 1.8 percent increase in the proportion of students who, in the absence of the reform, would have entered the private system rather than the public system. An average school in the sample has K-3 enrollment of 52 students per grade, indicating that it receives 0.94 of the students who would have entered private school in the absence of the reform into each school-grade. In the ‘no spillovers’ scenario, the entire 0.17 school-grade standard deviation increase in test scores that is attributed to the indirect effect of the reform would be caused by the composition effect. For this indirect effect to be due solely to the change in student composition, the students induced into the public system by the reform would have to score, on average, $2.9\sigma$ higher than students in the public system – a huge effect.$^{58}$ With the literature generally finding a public-private school test score gap in the $0.4\sigma$ range (Altonji et al., 2005), it seems highly unlikely that the entire indirect effect could be attributed just to the change in the composition of students in the public system.

$^{58}$An extra student in a school-grade of fifty-two students that scores $2.9\sigma$ higher would increase the average student level standard deviation of test scores by 0.056. A 0.056 increase in the student-level standard deviation is increased roughly threefold ($0.056*3=0.168$) to place that test score change in the distribution of school-grade test scores (see Finn and Achilles (1990), for instance).
Next, we use the linear-in-means estimates from Graham (2008), who finds that a one standard deviation increase in mathematics test scores among peers leads to a $0.8\sigma$ increase in ‘own’ test scores. This estimate suggests that the students entering the public system from the private system in response to the reform score $1.6\sigma$ higher than their public school counterparts, and that the $0.17\sigma$ indirect test score effect is roughly evenly divided between the composition effect and spillovers to the public school students operating through peer effects. While the private-public school test score gap may seem large, the division into the composition and spillover effect should be seen as a upper bound on the proportion of the indirect effect caused by the composition component, since other types of spillovers (such as increased parental pressure) are also likely to occur in this context, and would provide further positive spillovers for students in the public system.

### 8.2 Policy Implications of the Sorting Effect

In the absence of substantial positive general equilibrium effects, all but the most optimistic estimates would suggest that CSR is a relatively unattractive policy, given its enormous costs.

Brewer et al. (1999) estimates that reducing class sizes to 18 (from 24) for students in grades one through three 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 (1999b), 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 $0.10-0.20\sigma$; yet class size reduction ranks well-down, fortieth out of forty-six possible interventions intended to serve the same end.

---

59If the students induced into the public system score on average $1.6\sigma$ higher, then the composition effect leads to a $0.09\sigma (3 \times 0.52) \text{ increase in the school-grade test score distribution and linear-in-means peer effects lead to a } 0.07\sigma (3 \times 0.8 \times 0.52) \text { increase in the school-grade test score distribution.}
Our analysis indicates that, as a consequence of general equilibrium sorting, the benefits of class size reduction are likely to be significantly larger than previously realized, with an indirect sorting effect that is at least as great as the direct effect that has been the focus of much of the prior literature.

The indirect effects we estimate 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).

8.3 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, given that certain preconditions hold, each serving to increase its size.

First, the reform-related shock to public school quality needs to be large in size. Second, there needs to be a non-trivial share of households with children in private school pre-reform. Third, there needs to be a contrast between the characteristics of students in private versus public school, in order to generate changes in peer quality post-reform. And fourth, student peer characteristics need to be relevant in the production of student achievement. All these pre-conditions appear to hold in a Californian context.

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

To do justice to these features from a policy prediction perspective, a more fully articulated model of public and private school behavior and ‘consumer’ choice would be needed. To make estimating that worthwhile, one would need more disaggregated data pertaining to

\footnote{To the extent that more students are marginal, so the contrasts (under the third precondition) are less likely to be pronounced.}
the relevant economic agents than we have access to.

Never-the-less, in a Californian context, with more aggregated data, we have shown (as one would expect) that private school students differ from their public school counterparts. Further – and this serves to mitigate the size of the sorting effect we have estimated – there is clear evidence of adjustment on the part of private schools. That is, 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.$^61$

9 Conclusion

The merits of reducing class size have been the focus of perennial debates among policy-makers and economists. On the one hand, parents and teachers routinely and actively lobby for smaller classes, pressuring politicians to implement class size reduction initiatives. In the words of former President Clinton: “Reducing class size is one of the most important investments we can make in our children’s future. Recent research confirms what parents have always known – children learn better in small classes with good teachers, and kids who start out in smaller classes do better right through their high school graduation.” Economists, however, have often struggled to find significant benefits arising from class size reduction Hanushek (1999a) concludes: “The econometric evidence is clear. There is little reason to believe that smaller class sizes systematically yield higher student achievement.”

Yet econometric estimates of the direct, contemporaneous effects of class size reduction – the main focus in the literature – may not be 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, we have noted that general equilibrium

$^61$These results are available on request.
responses may also arise that can both dampen the effects of CSR, most notably through the need to hire new teachers, and also magnify the benefits, possibly through student sorting. Yet our understanding of these induced responses is limited, in no small part because of the difficulty in 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. Using the grade-specific timing of the reform, we documented a significant decrease in private school share in response to California’s CSR program, which in turn led to substantial 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.

Using similar sources of variation, we then estimated the direct effect of CSR and the indirect sorting effect on a common footing for the first time, while controlling for changes in observable teacher quality. We found the direct effect to be $0.11\sigma$ (in terms of mathematics scores), and the indirect effect to be even larger ($0.17\sigma$). Our estimation approach also allowed us to recover the persistence of the direct and indirect effects.

Our analysis indicates that estimates of the direct effect of class size on student achievement are by no means the final words on the efficacy of large-scale class size reduction policies. In light of the high costs of CSR, these general equilibrium effects are likely to be an important additional ingredient in determining the policy effectiveness of class size reduction in other contexts, under conditions we have outlined.
References


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, Grade 1 became eligible for the 1996-97 school year, grade 2 for the 1997-98 school year, and grade 3 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: 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 grade 1, the vertical line represents the 1996-97 school year, for grade 2, the 1997-98 school year, and for kindergarten and grade 3, 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. 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 4: Demographic Trends by Public Schools: Treated (Private School Within 3km) minus Untreated

Notes: Figure 4 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: 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 figure uses the treatment of grades (CSR versus non-CSR) and whether a school is close to a private school (within 3 km) as the two dimensions of differencing. The dashed vertical line represents the start of CSR implementation in a CSR grade: for grade 1, the vertical line represents the 1996-97 school year, for grade 2, the 1997-98 school year, and for kindergarten and grade 3, 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. 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 ‘treated’ districts (top three-quarter 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 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 year 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. The data generating each line have been weighted by housing transaction counts.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>·</td>
<td>·</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>1</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>2</td>
<td>·</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>3</td>
<td>·</td>
<td>·</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>4</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>5</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>6</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
</tbody>
</table>

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 first two rows and first column are in a lighter shading to reflect this.
## Table 2: Average Class Size by Grade and Year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten</td>
<td>24.2</td>
<td>21.0</td>
<td>19.9</td>
<td>19.6</td>
<td>19.5</td>
</tr>
<tr>
<td>Grade 1</td>
<td>19.2</td>
<td>19.2</td>
<td>19.2</td>
<td>19.2</td>
<td>19.2</td>
</tr>
<tr>
<td>Grade 2</td>
<td>19.4</td>
<td>19.2</td>
<td>19.1</td>
<td>19.0</td>
<td>19.0</td>
</tr>
<tr>
<td>Grade 3</td>
<td>22.4</td>
<td>20.1</td>
<td>19.6</td>
<td>19.4</td>
<td>19.3</td>
</tr>
<tr>
<td>Grade 4</td>
<td>29.1</td>
<td>28.9</td>
<td>28.9</td>
<td>28.7</td>
<td>28.5</td>
</tr>
<tr>
<td>Grade 5</td>
<td>29.4</td>
<td>29.3</td>
<td>29.2</td>
<td>29.3</td>
<td>29.0</td>
</tr>
</tbody>
</table>

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. 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)

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

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 students (K-12) within a district.
### Table 4: Descriptive Statistics

<table>
<thead>
<tr>
<th>A. School Data</th>
<th>Mean (1990-91 to 2012-13)</th>
<th>Pre-CSR (90-91 to 95-96)</th>
<th>CSR (96-97 to 98-99)</th>
<th>Post-CSR (99-00 to 12-13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary Student-Teacher ratio(^1)</td>
<td>22.3</td>
<td>25.5</td>
<td>23.9</td>
<td>20.7</td>
</tr>
<tr>
<td>Private School Share (%) (enrollment weighted)</td>
<td>9.0 (8.4)</td>
<td>9.9 (8.5)</td>
<td>9.9 (8.5)</td>
<td>8.4 (8.3)</td>
</tr>
<tr>
<td>CSR Intensity(^2)</td>
<td>84.7 (28.7)</td>
<td>85.4 (28.0)</td>
<td>84.5 (29.1)</td>
<td>84.5 (29.0)</td>
</tr>
<tr>
<td>% English Learner(^3)</td>
<td>27.7 (25.6)</td>
<td>24.8 (25.4)</td>
<td>26.2 (25.8)</td>
<td>29.2 (25.5)</td>
</tr>
<tr>
<td>% White</td>
<td>52.2 (29.1)</td>
<td>61.0 (27.6)</td>
<td>57.1 (28.5)</td>
<td>47.6 (28.8)</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>33.3 (27.6)</td>
<td>27.2 (25.2)</td>
<td>30.0 (26.4)</td>
<td>36.4 (28.3)</td>
</tr>
<tr>
<td>% Black</td>
<td>4.0 (7.5)</td>
<td>3.9 (7.8)</td>
<td>4.2 (7.9)</td>
<td>4.0 (7.3)</td>
</tr>
<tr>
<td>% Asian</td>
<td>4.5 (8.2)</td>
<td>4.1 (7.1)</td>
<td>4.4 (7.8)</td>
<td>4.7 (8.7)</td>
</tr>
<tr>
<td>Enrolment</td>
<td>582 (2246)</td>
<td>533 (2135)</td>
<td>572 (2249)</td>
<td>604 (2288)</td>
</tr>
<tr>
<td>% Free and Reduced Price Meals(^4)</td>
<td>40.5 (25.7)</td>
<td>37.0 (24.2)</td>
<td>43.1 (25.9)</td>
<td>43.1 (26.7)</td>
</tr>
<tr>
<td>Observations (District-Grade-Year)</td>
<td>253,056</td>
<td>63,983</td>
<td>32,761</td>
<td>156,312</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer Price (*10,000)</td>
<td>28.8 (25.8)</td>
<td>17.4 (12.1)</td>
<td>18.2 (14.2)</td>
<td>36.1 (29.3)</td>
</tr>
<tr>
<td>Lot Size (/1000)</td>
<td>42.0 (81.8)</td>
<td>41.4 (79.2)</td>
<td>44.3 (89.4)</td>
<td>41.5 (80.3)</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>2.9 (0.8)</td>
<td>2.9 (0.8)</td>
<td>2.9 (0.8)</td>
<td>2.9 (0.8)</td>
</tr>
<tr>
<td>Square Feet (/1000)</td>
<td>1.7 (0.4)</td>
<td>1.7 (0.4)</td>
<td>1.7 (0.4)</td>
<td>1.7 (0.4)</td>
</tr>
<tr>
<td>Observations (District-Year)</td>
<td>11,321</td>
<td>2,508</td>
<td>2,009</td>
<td>6,804</td>
</tr>
</tbody>
</table>

\(^1\) 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Untreated Grades (Grades 4-12)</th>
<th>Treated Grades (Grades K-3)</th>
<th>Difference (Untreated-treated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>8.88 (8.80)</td>
<td>11.76 (7.62)</td>
<td>-2.87 (0.25)</td>
</tr>
<tr>
<td>After</td>
<td>8.16 (8.71)</td>
<td>9.63 (7.31)</td>
<td>-1.47 (0.28)</td>
</tr>
<tr>
<td>Change</td>
<td>0.73 (0.15)</td>
<td>2.13 (0.25)</td>
<td>-1.41 (0.17)</td>
</tr>
<tr>
<td>Observations</td>
<td>165,950</td>
<td>87,106</td>
<td>253,056</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment*Post</td>
<td>-1.41***</td>
<td>-1.35***</td>
<td>-1.45***</td>
<td>-1.78***</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.18)</td>
<td>(0.28)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Treatment</td>
<td>2.87***</td>
<td>2.82***</td>
<td>3.95***</td>
<td>5.29***</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.52)</td>
<td>(0.50)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>Post</td>
<td>-0.73***</td>
<td>0.26*</td>
<td>0.15</td>
<td>0.44**</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.16)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>Year/Grade FE</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Demographic Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>253,056</td>
<td>253,056</td>
<td>215,139</td>
<td>215,139</td>
</tr>
</tbody>
</table>

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 (%)

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

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: Regression-Discontinuity Estimates by Grade Span

<table>
<thead>
<tr>
<th>Outcome Variable: Private School Share in Grade Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten (Placebo) (1)</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>Average Effect</td>
</tr>
<tr>
<td>(0.22)</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

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, a separate local linear regression allowing for a different functional form on either side of the cutoff (see Equation 5.4) is run for each grade. We then average these grade-level estimates to find the average effect over the grade span. The bandwidth used is three. Standard 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 9: Triple-Differences Estimates of Compositional Changes

**Outcome Variable: Student Demographics (%)**

<table>
<thead>
<tr>
<th>Treatment<em>Post</em>I{Buffer &lt; 1.5 km}</th>
<th>Percent White (1)</th>
<th>Percent Hispanic (2)</th>
<th>Percent Black (3)</th>
<th>Percent Asian (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.16***</td>
<td>-1.69***</td>
<td>-0.65***</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>(0.73)</td>
<td>(0.48)</td>
<td>(0.19)</td>
<td>(0.24)</td>
<td></td>
</tr>
<tr>
<td>Treatment<em>Post</em>I{Buffer &lt; 3 km}</td>
<td>3.13***</td>
<td>-1.64***</td>
<td>-0.70***</td>
<td>0.06</td>
</tr>
<tr>
<td>(0.74)</td>
<td>(0.49)</td>
<td>(0.19)</td>
<td>(0.26)</td>
<td></td>
</tr>
<tr>
<td>Treatment<em>Post</em>I{Buffer &lt; 5 km}</td>
<td>3.15***</td>
<td>-1.62***</td>
<td>-0.71***</td>
<td>0.05</td>
</tr>
<tr>
<td>(0.74)</td>
<td>(0.49)</td>
<td>(0.19)</td>
<td>(0.26)</td>
<td></td>
</tr>
</tbody>
</table>

| % 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 (5.5). I{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>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSRd*Post</td>
<td>9.33***</td>
<td>11.33***</td>
<td>13.46***</td>
<td>12.04***</td>
<td>7.72**</td>
<td>6.40*</td>
</tr>
<tr>
<td></td>
<td>(3.14)</td>
<td>(3.14)</td>
<td>(3.66)</td>
<td>(4.02)</td>
<td>(3.18)</td>
<td>(3.66)</td>
</tr>
<tr>
<td>CSRd</td>
<td>18.73***</td>
<td>14.66***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(3.44)</td>
<td>(3.61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>7.04***</td>
<td>5.90**</td>
<td>3.92</td>
<td>6.50**</td>
<td>34.86***</td>
<td>35.98***</td>
</tr>
<tr>
<td></td>
<td>(2.65)</td>
<td>(2.52)</td>
<td>(3.12)</td>
<td>(3.26)</td>
<td>(4.25)</td>
<td>(4.36)</td>
</tr>
<tr>
<td>House Characteristics</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Teacher Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Demographic Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>14,754</td>
<td>14,754</td>
<td>14,754</td>
<td>12,265</td>
<td>13,181</td>
<td>11,054</td>
</tr>
</tbody>
</table>

Notes: Observations are at the district-year level. Columns (1)-(3) and (5) cover the 1990-91 through 2011-12 school years, while Columns (4) and (6) cover the 1994-95 through 2011-12 school years. House characteristics are square feet, lot size, and number of bedrooms. Teacher controls include experience and education levels. Demographic controls include student race, gender, free and reduced price meal eligibility, English second language, enrollment, and enrollment squared. All regressions include year fixed effects and are weighted by housing transaction counts. Standard errors are clustered at the district level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.
### Table 11: Structural Estimates

<table>
<thead>
<tr>
<th></th>
<th>With $\psi = \frac{2}{3}$</th>
<th>With $\psi = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$\gamma_R$</td>
<td>2.10***</td>
<td>2.22***</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>$\gamma_S$</td>
<td>2.27</td>
<td>3.26**</td>
</tr>
<tr>
<td></td>
<td>(1.69)</td>
<td>(1.54)</td>
</tr>
<tr>
<td>$\delta_R$</td>
<td>0.64**</td>
<td>0.57**</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>$\delta_S$</td>
<td>0.49*</td>
<td>0.45**</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>County FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>147,636</td>
<td>147,636</td>
</tr>
</tbody>
</table>

Notes: Observations are at the school-grade-year level, and cover the 1997-98 through 2003-04 school years. All parameter estimates include controls for teacher quality. Standard errors for $\gamma_R$ and $\gamma_S$ are computed using the delta method and are clustered at the school level. Standard errors for $\delta_R$ and $\delta_S$ are bootstrapped. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.
Table 12: Estimates of Teacher Quality

Outcome Variable: Mathematics Test Scores

<table>
<thead>
<tr>
<th></th>
<th>CSR</th>
<th>non-CSR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>(Q_{CSR/non,01-02})</td>
<td>1.123***</td>
<td>-0.089***</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>(Q_{CSR/non,00-01})</td>
<td>0.929***</td>
<td>-0.361***</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>(Q_{CSR/non,99-00})</td>
<td>0.520***</td>
<td>-0.643***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>(Q_{CSR/non,98-99})</td>
<td>0.041***</td>
<td>-0.678***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>(Q_{CSR/non,99-00,K5})</td>
<td>0.997***</td>
<td>-1.132***</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>(Q_{CSR/non,99-00,K6})</td>
<td>0.632***</td>
<td>-0.673***</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.053)</td>
</tr>
</tbody>
</table>

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. 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.
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.
Notes: This figure shows the estimated effect of CSR on private school share for each grade using the RD design described in Subsection 5.1. 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: Difference in Grade 6 and Grade 5 Math Scores by School Type

Notes: The vertical line indicates the year the first treated CSR cohort enters grade 6.
### Table A.1: Data Availability and Sources

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>California Department of Education Data</td>
<td><a href="http://www.cde.ca.gov/ds/sd/fileenr.asp">www.cde.ca.gov/ds/sd/fileenr.asp</a></td>
</tr>
<tr>
<td>DataQuick House Price Data</td>
<td><a href="http://www.cde.ca.gov/ds/sh/cw/filesafdc.asp">www.cde.ca.gov/ds/sh/cw/filesafdc.asp</a></td>
</tr>
</tbody>
</table>

#### Data Type: California Department of Education Data

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public School Enrollment Data (includes race)</td>
<td><a href="http://www.cde.ca.gov/ds/sd/fileenr.asp">www.cde.ca.gov/ds/sd/fileenr.asp</a></td>
</tr>
<tr>
<td>Private School Enrollment Data</td>
<td><a href="http://www.cde.ca.gov/ds/si/ps/index.asp">www.cde.ca.gov/ds/si/ps/index.asp</a></td>
</tr>
<tr>
<td>Public School ESL Data</td>
<td>wwww.cde.ca.gov/ds/sd/fileelsch.asp</td>
</tr>
<tr>
<td>Public School Free or Reduced-Price Meal Data</td>
<td>wwww.cde.ca.gov/ds/sh/cw/filesafdc.asp</td>
</tr>
<tr>
<td>CSR Implementation Data (grades K-3 only)</td>
<td>wwww.cde.ca.gov/ds/si/ps/index.asp</td>
</tr>
<tr>
<td>Standardized Testing and Reporting Data (grades 2-11 only)</td>
<td>star.cde.ca.gov</td>
</tr>
<tr>
<td>Teacher Assignment and Demographic Data</td>
<td>wwww.cde.ca.gov/ds/sd/df/filesassign.asp</td>
</tr>
<tr>
<td>Teacher Demographic and Experience Data</td>
<td>wwww.cde.ca.gov/ds/sd/df/filecertstaff.asp</td>
</tr>
</tbody>
</table>

#### Data Type: DataQuick House Price Data

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing Price Data (District-level)</td>
<td>Private use data</td>
</tr>
<tr>
<td>Housing Price Data (School-level)</td>
<td>Private use data</td>
</tr>
</tbody>
</table>

---

**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.

---

*a* Only non-zero grade-level observations are included in this observation count.

*b* Private school enrolment data for 1990-91 through 1998-99 inclusive are not available on the CDE website. They were provided upon request by the CDE.

*c* 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.
Table A.2: Mathematics Test Score Summary Statistics

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

Notes: Test scores are for the Stanford 9 test and report the mean percentile ranking of students relative to a nationally representative reference group.
### Table A.3: Triple-Differences on Number of Private Schools

**Outcome Variable:** Private Schools per 1000 School-Aged Children

<table>
<thead>
<tr>
<th></th>
<th>D-in-D (CSR Schools)</th>
<th>D-in-D (non-CSR Schools)</th>
<th>Triple-D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>β</td>
<td>-0.078*** (0.015)</td>
<td>-0.013** (0.006)</td>
<td>0.065*** (0.013)</td>
</tr>
<tr>
<td>State and Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>663</td>
<td>663</td>
<td>1,326</td>
</tr>
</tbody>
</table>

Notes: Observations are at the state-by-biennial year level and cover 1989-90 through 2013-14 school years. 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-CS) and state (CA vs rest-of-country) as the two layers of differencing and restricts to private school primarily serves CSR grades in column (1) and private school primarily does not primarily serve CSR grades in column (2). The triple-D adds whether the private school primarily serves CSR grades as an additional layer of differencing. Standard errors are clustered at the state level. ***,*** and *** denote significance at the 10%, 5% and 1% levels, respectively.
**Table A.4:** School Statistics By Grade Span Configuration

<table>
<thead>
<tr>
<th>Grade Span</th>
<th>Number of Schools</th>
<th>% Implementing CSR in First Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-5</td>
<td>2183</td>
<td>95.9</td>
</tr>
<tr>
<td>K-6</td>
<td>1954</td>
<td>92.5</td>
</tr>
<tr>
<td>K-8</td>
<td>455</td>
<td>90.1</td>
</tr>
<tr>
<td>K-12</td>
<td>49</td>
<td>48.3</td>
</tr>
<tr>
<td>Other</td>
<td>289</td>
<td>90.3</td>
</tr>
</tbody>
</table>

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

**Table A.5:** Percent of Inexperienced Teachers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>27.3</td>
<td>26.7</td>
<td>21.6</td>
<td>18.8</td>
<td>17.0</td>
<td>15.2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>26.8</td>
<td>26.3</td>
<td>22.0</td>
<td>17.9</td>
<td>16.4</td>
<td>14.0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>26.9</td>
<td>33.1</td>
<td>32.9</td>
<td>30.1</td>
<td>27.6</td>
<td>24.5</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>24.0</td>
<td>27.8</td>
<td>28.8</td>
<td>28.4</td>
<td>25.7</td>
<td>22.5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>23.5</td>
<td>26.7</td>
<td>27.4</td>
<td>26.5</td>
<td>27.0</td>
<td>23.2</td>
</tr>
</tbody>
</table>

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.
B School Level Implementation Evidence on Public School Composition and House Prices

This Appendix Section provides evidence on CSR-induced changes to public school student composition and house prices by exploiting variation induced by schools choosing whether or not to implement CSR. This additional evidence is brought to bear on the effects of CSR on both public school composition and house prices from Section 5 and is placed here for brevity. The results in this Appendix Section complement the main text by corroborating the evidence in Section 5.

**Public School Composition:** The main text focuses on proximity to a private school for 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.\(^{62}\) The weighted estimating equation is:

\[
demo_{sgt} = \beta_0 + \beta_1 post_{gt} + \beta_2 treat_{gt} + \beta_3 CSR_s + \beta_4 (post_{gt} \cdot treat_{gt}) + \beta_5 (post_{gt} \cdot CSR_s) + \beta_6 (treat_{gt} \cdot CSR_s) + \beta_7 (post_{gt} \cdot treat_{gt} \cdot CSR_s) + \eta_s + \theta_t + \delta_g + \epsilon_{dgt}, \tag{B.1}
\]

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 \(\eta_s\), \(\theta_t\) and \(\delta_g\) are school, time and grade fixed effects, respectively. The triple-differences coefficient of interest is \(\beta_7\). 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 B.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

\(^{62}\)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. Thus, 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.
To that end, we estimate triple-differences specifications in Panel A of Table B.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 that are over-represented in private schools relative to public schools (white and Asian students) being drawn into the public system.

As a validity check, we compute difference-in-differences estimates by year, using the treatment of grades (CSR versus non-CSR) and whether the school had implemented CSR or not as the two dimensions of differencing. Figures B.2(a) and B.2(b) plot these estimates for the fraction of white and Hispanic students, respectively. In both cases, the point estimates are indistinguishable from zero in the pre-reform years.

**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 data does not cover the entire state, but rather is focused on urban population centers. A school level analysis, however, may be illuminating since aggregating to the district level obscures important variation. Figure B.3 highlights the CSR implementation variation in four dense school districts for which we have data.\(^{63}\) Visually, we see that non-CSR implementing schools tend to be clustered within certain regions of the four districts.

To conduct our analysis at the school level, we use a school-level CSR implementation variable (\(\text{CSR}_s\)) to conduct a difference-in-differences regression. In that vein, we estimate the following weighted regression:

\[
\text{price}_{st} = \beta_0 + \beta_1 \text{post}_t + \beta_2 \text{CSR}_s + \beta_3 (\text{post}_t \ast \text{CSR}_s) + \phi X_{st} + \eta_s + \theta_t + \epsilon_{st},
\]

where \(\text{price}_{st}\) is the average house price in the attendance zone for school \(s\) at time \(t\), \(\text{CSR}_s\) 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 \(\eta_s\) and \(\theta_t\) are school

---

\(^{63}\)These four large districts (Los Angeles, Orange, Sacramento and Riverside) account for 52 percent of our total observations.
and time fixed effects, respectively. As before, in a difference-in-differences context, the coefficient of interest is $\beta_3$, 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 absence of CSR.

**School Level House Price Results:** Figure B.4(a) displays trends among treated and control groups for the school level variation. Visually, there does 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 counterparts.

This evidence maps directly into the econometric specifications given by equation (B.2), which Table B.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 to capture the full effect of the reform. The point estimate of $33,600 implies that schools implementing CSR saw around a 17 percent increase in house prices.\(^\text{64}\) While comparing with the district-based estimate is not straightforward since nearly every district implemented CSR to at least some extent, the 17 percent increase we find here is consistent with the district-based finding of 10 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 effect 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 teacher characteristics and 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 6(b) and B.4(b), respectively. In both cases, 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.

\(^{64}\)The point estimate is $33,600, which represents a 17 percent increase in house prices from the 1995 pre-CSR (weighted) mean of $195,000.
Notes: Figure B.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 B.2: Effect of CSR on Public School Demographics for Years Before, During, and After Implementation

(a) Percent White (School Implementation Design)

(b) Percent Hispanic (School Implementation Design)

Notes: The 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 schools (CSR versus non-CSR schools) as the two dimensions of differencing. The dashed vertical line represents the start of CSR implementation in a CSR grade: for grade 1, the vertical line represents the 1996-97 school year, for grade 2, the 1997-98 school year, and for kindergarten and grade 3, 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. Grade, year and school fixed effects are included. Standard errors are clustered at the district level.
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 county region includes Santa Ana, Orange Unified, Garden Grove, and Newport school districts. Treated schools had CSR implemented in at least grade 3 or kindergarten in 1998, while control schools had no kindergarten or grade 3 CSR implementation in 1998. These two counties and regions cover 53.2% of all school used in the school-level analysis.
Figure B.4: 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 B.4(a) shows average house prices in ‘treated’ districts (top three-quarter of CSR implementing districts in 1996-97) and ‘control’ districts (bottom quartile of CSR implementing districts in 1996-97). Figure B.4(b) reports the effect of district CSR treatment intensity on house prices 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 year 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. The data generating each line have been weighted by housing transaction counts.
Table B.1: Triple-Differences Estimates of Compositional Changes (School Level Implementation Design)

<table>
<thead>
<tr>
<th>Outcome Variable: Student Demographics (%)</th>
<th>Percent White</th>
<th>Percent Hispanic</th>
<th>Percent Black</th>
<th>Percent Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
</tbody>
</table>

**A. School Implementation Design**

<table>
<thead>
<tr>
<th>Treatment<em>Post</em>CSR</th>
<th>0.90</th>
<th>-2.33**</th>
<th>-0.39</th>
<th>1.62***</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.43)</td>
<td>(1.11)</td>
<td>(0.48)</td>
<td>(0.57)</td>
<td></td>
</tr>
</tbody>
</table>

| % 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 (B.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 B.2: Difference-in-Differences Estimates of Impact on House Prices (School CSR Implementation Design)

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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CSR_s \times \text{Post})</td>
<td>2.59</td>
<td>3.01*</td>
<td>3.36**</td>
<td>2.04</td>
<td>1.47</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(1.82)</td>
<td>(1.72)</td>
<td>(1.70)</td>
<td>(2.19)</td>
<td>(2.49)</td>
<td>(2.68)</td>
</tr>
<tr>
<td>House Characteristics</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District/School FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Teacher Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Demographic Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>42,142</td>
<td>42,142</td>
<td>42,142</td>
<td>34,106</td>
<td>40,385</td>
<td>33,402</td>
</tr>
</tbody>
</table>

Notes: Observations are at the school-year level. Columns (1)-(3) and (5) cover the 1990-91 through 2011-12 school years, while Columns (4) and (6) cover the 1994-95 through 2011-12 school years. House characteristics are square feet, lot size and number of bedrooms. Teacher controls include experience and education levels. Demographic controls include student race, gender, free and reduced price meal eligibility, English second language, enrollment and enrollment squared. All regressions include year fixed effects and are weighted by housing transaction counts. Standard errors are clustered at the district level. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.
C California State Testing – a Quick Primer

Statewide testing in California started in 1961 for mathematics, reading and writing in grade 5, 8 and 10. In 1972, the California Assessment Program was created, which tested reading in grade 2 and 3 and mathematics, reading and writing in grades 6 and 12, lasting 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 grade 4, 5, 8 and 10.

In 1994, under public pressure from civil rights groups that the CLAS was inaccurate and intruded into students' privacy (this was due to numerous race-based questions on the test), Governor Pete Wilson vetoed a Senate bill to extend CLAS, though the Governor stated 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). Therefore, there were no statewide tests for the 1994-95 and 1995-96 school years, though 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 Governor Pete Wilson – 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.

D Housing Data Linkage: a Description

This appendix describes how the house transaction data obtained from DataQuick was mapped into school districts and school attendance zones. First, the DataQuick data 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 from http://gis.ats.ucla.edu/.

With the house transactions now identified with a specific latitude and longitude, we 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 districts 2000 Census shapefiles. For school attendance zone coverage, we used the 2009 kindergarten school attendance zone shapefile from The College of William...

---

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

Finally, 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 and the California private school directory 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.

## E Estimating Teacher Quality

Let $Q_{CSR,t}^l$ and $Q_{non,t}^l$ 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. 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.

We incorporate variation in teacher inexperience into our estimation strategy. We estimate the teacher quality parameters $Q_{CSR,t}^l$ and $Q_{non,t}^l$ for each lag $l$ according to the following two-step procedure. First, we regress test scores in 1997-98 + $l$ ($y_{s,g,1997-98+l}$) on the

---

66Since virtually all California schools cover grades K-3, it makes no difference which K-3 grade attendance zone shapefile is used.
68http://www.cde.ca.gov/ds/si/ps/
69Inexperience in grades 5 and 6 is close to but slightly lower than for second through fourth grades.
70It 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.
share of teacher inexperience in 1997-98 ($X_{s,g,1997-98}$), including grade fixed effects ($\phi_g$):

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

Second, we use the resulting estimate $\hat{\gamma}_l$ to compute teacher quality relative to the 1997-98 baseline:\footnote{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.5 shows that the share of teacher inexperience is essentially identical across third and fourth grades in 1997-98.}

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

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 $Q_{CSR,2001-02}^0 = \hat{\gamma}_0 \times (X_{3,2001-02} - X_{3,1997-98})$, $Q_{non,2001-02}^0 = \hat{\gamma}_0 \times (X_{4,2001-02} - X_{4,1997-98})$ and $Q_{CSR,1997-98}^4 = \hat{\gamma}_4 \times (X_{4,1997-98} - X_{4,1997-98}) = 0$. The necessary parameters to compute $\hat{\gamma}_S$ are estimated analogously.\footnote{We estimate the parameters $Q_{CSR,2000-01,K6}^2$, $Q_{non,2000-01,K6}^2$, $Q_{CSR,2000-01,K6}^3$ and $Q_{non,2000-01,K6}^3$. Due to a lack of test score data in 1996-97, the parameters $Q_{CSR,1996-97,K5/K6}$ and $Q_{non,1996-97,K5/K6}$ cannot be estimated. However, as with $Q_{CSR,1997-98}^4$, we can assume that they are negligible since teacher quality across grades is likely to be similar in 1996-97 and 1997-98.}
F Structural Equations

F.1 Without Teacher Effects

\( \gamma_R \): To identify \( \gamma_R \), we subtract \( \Delta y_{4,01-02} \) from \( \Delta y_{3,01-02} \):

\[
\Delta y_{3,01-02} - \Delta y_{4,01-02} = \gamma + \delta_R^2 \gamma R \Delta R_{0,98-99} + \delta_R^2 \gamma R \Delta R_{1,99-00} + \delta_R \gamma R \Delta R_{2,00-01} + \gamma R \Delta R_{3,01-02} \\
+ \delta_S^2 \Delta X_{0,97-98}^S + \delta_S^2 \Delta X_{1,98-99}^S + \delta_S \Delta X_{2,99-00}^S + \gamma S \Delta X_{3,00-01}^S + \Delta \epsilon_{3,01-02} \\
- (\gamma + \delta_R^3 \gamma R \Delta R_{1,98-99} + \delta_R^2 \gamma R \Delta R_{2,99-00} + \delta_R \gamma R \Delta R_{3,00-01} \\
+ \delta_S^3 \Delta X_{1,98-99}^S + \delta_S^2 \Delta X_{2,99-00}^S + \delta_S \Delta X_{3,00-01}^S + \gamma S \Delta X_{4,2001-02}^S + \Delta \epsilon_{4,2001-02}) \\
= \gamma_R \Delta R_{2001-02}, \hspace{1cm} (F.1)
\]

where the final equality comes the fact that CSR affected all grades equally once it was implemented, so that \( \Delta R_{g,t} = \Delta R_{g',t} \) and \( \Delta X_{g,t}^S = \Delta X_{g',t}^S \forall g, g' \).

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:

\[
y_{3,01-02} - y_{4,01-02} - (y_{3,97-98} - y_{4,97-98}) = \gamma_R \Delta R_{01-02} \hspace{1cm} (F.2)
\]

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

\[
\Delta y_{6,01-02,K6} - \Delta y_{6,01-02,K5} = \gamma + \delta_R^2 \gamma R \Delta R_{1,96-97,K6} + \delta_R^3 \gamma R \Delta R_{2,97-98,K6} + \delta_R \gamma R \Delta R_{3,98-99,K6} \\
+ \delta_S^3 \gamma S \Delta X_{1,96-97,K6}^S + \delta_S^3 \gamma S \Delta X_{2,97-98,K6}^S + \delta_S^2 \gamma S \Delta X_{3,98-99,K6}^S \\
+ \delta_S^3 \gamma S \Delta X_{4,99-00,K6}^S + \delta_S \gamma S \Delta X_{5,00-01,K6}^S + \gamma S \Delta X_{6,01-02,K6}^S + \Delta \epsilon_{6,01-02,K6} \\
- (\gamma + \delta_R^3 \gamma R \Delta R_{1,96-97,K5} + \delta_R^2 \gamma R \Delta R_{2,97-98,K5} + \delta_R \gamma R \Delta R_{3,98-99,K5} \\
+ \delta_S^3 \gamma S \Delta X_{1,96-97,K5}^S + \delta_S^3 \gamma S \Delta X_{2,97-98,K5}^S + \delta_S^2 \gamma S \Delta X_{3,98-99,K5}^S \\
+ \delta_S^3 \gamma S \Delta X_{4,99-00,K5}^S + \delta_S \gamma S \Delta X_{5,00-01,K5}^S + \gamma S \Delta X_{6,01-02,K5}^S + (1 - \psi) \gamma S \Delta X_{6,01-02,K5}^S + \Delta \epsilon_{6,01-02,K5}) \\
= \psi \gamma S \Delta X_{6,01-02}^S, \hspace{1cm} (F.3)
\]

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 grade 6) so that \( \Delta R_{g,t,K6} = \Delta R_{g,t,K5} \) and \( \Delta X_{g,t,K6} = \Delta X_{g,t,K5} \forall g \).

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 grade 5 test scores in K-5 ($y_{5,01-02,K5}$) and K-6 schools ($y_{5,01-02,K6}$) as are first level of differencing. Then, we use the pre-reform test scores for both grades 5 and 6, $y_{5,97-98}$ and $y_{6,97-98}$, in K-5 and K-6 schools as counterfactuals for the observed test scores in grades 5 and 6 in the 2001-02 school year. Therefore, we have:

$$
\psi\gamma_S \Delta X_{5,01-02} = [y_{6,01-02,K6} - y_{5,01-02,K6} - (y_{6,97-98,K6} - y_{5,97-98,K6})]
- [y_{6,01-02,K5} - y_{5,01-02,K5} - (y_{6,97-98,K5} - y_{5,97-98,K5})]
$$

(F.4)

$(\delta_R, \delta_S)$: Identification of $\delta_R$ and $\delta_S$, uses Equations 6.10 and 6.11 from the structural section which yielded:

$$
\Delta y_{4,00-01} - \Delta y_{3,00-01} = \gamma + \delta^3_{R} \gamma R \Delta R_{1,97-98} + \delta^2_{R} \gamma R \Delta R_{2,98-99} + \delta R \gamma R \Delta R_{3,99-00}
+ \delta^3_{S} \gamma S \Delta X_{1,97-98}^S + \delta^2_{S} \gamma S \Delta X_{2,98-99}^S + \delta R \gamma S \Delta X_{3,99-00}^S + \gamma S \Delta X_{4,00-01}^S + \Delta \epsilon_{4,01-02}
- (\gamma + \delta^2_{R} \gamma R \Delta R_{1,98-99} + \delta R \gamma R \Delta R_{2,99-00} + \gamma R \Delta R_{3,00-01}
+ \delta^3_{S} \gamma S \Delta X_{1,98-99}^S + \delta R \gamma S \Delta X_{2,99-00}^S + \gamma S \Delta X_{3,00-01}^S + \Delta \epsilon_{3,01-02})
= \delta^3 R \gamma R \Delta R_{1,97-98} - \gamma R \Delta R_{3,00-01} + \delta^3 S \gamma S \Delta X_{1,97-98}^S
$$

(F.5)

$$
\Delta y_{5,00-01} - \Delta y_{4,00-01} = \gamma + \delta^3_{R} \gamma R \Delta R_{1,96-97} + \delta^2_{R} \gamma R \Delta R_{2,97-98} + \delta^2_{R} \gamma R \Delta R_{3,98-99} + \delta^4_{R} \gamma S \Delta X_{1,96-97}^S
+ \delta^3_{S} \gamma S \Delta X_{2,97-98}^S + \delta^2_{S} \gamma S \Delta X_{3,98-99}^S + \delta R \gamma S \Delta X_{4,00-01}^S + \gamma S \Delta X_{5,00-01}^S + \Delta \epsilon_{5,00-01}
- (\gamma + \delta^3_{R} \gamma R \Delta R_{1,97-98} + \delta^2_{R} \gamma R \Delta R_{2,98-99} + \delta R \gamma R \Delta R_{3,99-00}
+ \delta^3_{S} \gamma S \Delta X_{1,97-98}^S + \delta^2_{S} \gamma S \Delta X_{2,98-99}^S + \delta R \gamma S \Delta X_{3,99-00}^S + \gamma S \Delta X_{4,00-01}^S + \Delta \epsilon_{4,00-01})
= \delta^3 R \Delta R_{1,96-97} - \delta R \Delta R_{3,99-00} + \delta^3 S \gamma S \Delta X_{1,96-97}^S
$$

(F.6)

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 $\Delta X_{1,96-97}^S = \Delta X_{1,97-98}^S$. Suppressing the grade and year notation on the $\Delta R_{gt}$

---

73 Here, we are overidentified since we could use 1997-98, 1998-99 and 1999-00 as counterfactuals since those cohorts in grades 5 and 6 were not subject to CSR in those years. In practice, we use all three and take an average of the estimates, although estimates are quantitatively similar regardless which counterfactual year we use.

74 $\Delta y_{3,99-00} - \Delta y_{4,99-00}$ yields the same structural equation as $\Delta y_{1,00-01} - \Delta y_{4,00-01}$ 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 both equations and take an average of the estimates, although estimates are quantitatively similar regardless which structural equation is used.
and $\Delta X_{gt}$ variables yields the following two equations with two unknowns ($\delta_R$, $\delta_S$):

\begin{align*}
  y_{4,00-01} - y_{3,00-01} - (y_{4,97-98} - y_{3,97-98}) &= \gamma_R \Delta R(\delta_R^2 - 1) + \delta_S^3 \gamma_S \Delta X^S \\
  y_{5,00-01} - y_{4,00-01} - (y_{5,97-98} - y_{4,97-98}) &= \delta_R \Delta R(\delta_R^2 - 1) + \delta_S^3 \gamma_S \Delta X^S
\end{align*}

(F.7) (F.8)

F.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 grade 3 test scores in 2001-02 as:

\begin{align*}
  \Delta y_{3,01-02} &= \gamma + \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_{CSR,98-99} + Q^2_{CSR,99-00} + Q^1_{CSR,00-01} + Q^0_{CSR,01-02}) + \Delta \epsilon_{3,01-02}
\end{align*}

(F.9)

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

\begin{align*}
  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}).
\end{align*}

(F.10)

$\gamma_S$: Similarly, the differences between observed and counterfactual test scores that yield
\( \gamma_s \) can be expressed in terms of the parameters in the following way:

\[
[y_{6,01-02,K6} - y_{5,01-02,K6} - (y_{6,97-98,K6} - y_{5,97-98,K6})]
- [y_{6,01-02,K5} - y_{5,01-02,K5} - (y_{6,97-98,K5} - y_{5,97-98,K5})] = \psi \gamma_s \Delta X_{6,01-02}^S
+ \gamma_Q(Q_{CSR,96-97,K6} + Q_{CSR,97-98,K6} + Q_{CSR,98-99,K6} + Q_{non,99-00,K6} + Q_{non,00-01,K6} + Q_{non,01-02,K6})
- \gamma_Q(Q_{CSR,97-98,K6} + Q_{CSR,98-99,K6} + Q_{CSR,99-00,K6} + Q_{non,00-01,K6} + Q_{non,01-02,K6})
- [\gamma_Q(Q_{CSR,96-97,K5} + Q_{CSR,97-98,K5} + Q_{CSR,98-99,K5} + Q_{non,99-00,K5} + Q_{non,00-01,K5} + Q_{non,01-02,K5})
- \gamma(Q_{CSR,97-98,K5} + Q_{CSR,98-99,K5} + Q_{CSR,99-00,K5} + Q_{non,00-01,K5} + Q_{non,01-02,K5})]
= \psi \gamma_s \Delta X_{6,01-02}^S + \gamma_Q(Q_{CSR,96-97,K6} + Q_{non,99-00,K6} - Q_{CSR,99-00,K6})
- [\gamma_Q(Q_{CSR,96-97,K5} + Q_{non,99-00,K5} - Q_{CSR,99-00,K5})].
\] (F.11)

**\((\delta_R, \delta_S)\):** Finally, to solve for \(\delta_R\) and \(\delta_S\), we incorporate teacher effects into the final two regressions:

\[
y_{4,00-01} - y_{3,00-01} - (y_{4,97-98} - y_{3,97-98}) = \gamma_R \Delta R_{00-01} (\delta_R^3 - 1) + \delta_S^3 \gamma_S \Delta X_{02-03}^S
+ \gamma_Q(Q_{non,96-97} + Q_{CSR,97-98} + Q_{CSR,98-99} + Q_{CSR,99-00} + Q_{non,00-01})
- \gamma_Q(Q_{non,97-98} + Q_{CSR,98-99} + Q_{CSR,99-00} + Q_{CSR,00-01})
= \gamma_R \Delta R_{00-01} (\delta_R^3 - 1) + \delta_S^3 \gamma_S \Delta X_{02-03}^S
+ \gamma_Q(Q_{non,96-97} + Q_{CSR,97-98} - Q_{non,97-98} + Q_{non,00-01} - Q_{CSR,00-01}).
\] (F.12)

\[
y_{5,00-01} - y_{4,00-01} - (y_{5,97-98} - y_{4,97-98}) = \delta_R \gamma_R \Delta R_{00-01} (\delta_R^3 - 1) + \delta_S^4 \gamma_S \Delta X_{02-03}^S
+ \gamma_Q(Q_{CSR,96-97} + Q_{CSR,97-98} + Q_{CSR,98-99} + Q_{CSR,99-00} + Q_{non,00-01})
- \gamma_Q(Q_{non,96-97} + Q_{CSR,97-98} + Q_{CSR,98-99} + Q_{CSR,99-00} + Q_{non,00-01})
= \delta_R \gamma_R \Delta R_{00-01} (\delta_R^3 - 1) + \delta_S^4 \gamma_S \Delta X_{02-03}^S
+ \gamma_Q(Q_{CSR,96-97} - Q_{non,96-97} + Q_{non,99-00} - Q_{CSR,99-00}).
\] (F.13)