Title
Essays in Development and Education Economics

Author
Ramirez Ritchie, Elizabeth A

Publication Date
2018

Peer reviewed|Thesis/dissertation

Essays in Development and Education Economics

by

Elizabeth A Ramirez Ritchie

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Elisabeth Sadoulet, Chair
Professor Alain de Janvry
Professor Michael Anderson
Professor Edward Miguel

Spring 2018

Essays in Development and Education Economics

Abstract

Essays in Development and Education Economics

by

Elizabeth A Ramirez Ritchie

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Elisabeth Sadoulet, Chair

This dissertation contains three empirical studies on the impact of three distinct government policies, ranging from the provision of agricultural insurance in Mexico to improved pollution controls on school buses in California. All three papers take advantage of administrative data from the respective programs and use selection on unobservables designs to obtain causal estimates for the impacts of these programs. Chapter 1 estimates the impact of an index-based agricultural insurance offered to small-holder farmers by the Mexican government on their income and consumption following a negative weather shock, as well as investment decisions for the subsequent growing season. Chapter 2 analyzes how the nutritional quality of the meals provided through the National School Lunch Program (NSLP) affects students' academic outcomes, as measured by standardized test scores. By studying the nutritional quality of these meals, as opposed to simply their availability, this analysis provides and important contribution to the body of literature on the educational benefits of the NSLP. In Chapter 3, I study the impact of the California Lower-Emission School Bus Program (LESBP), which seeks to lower the pollution emissions of school buses through replacements and retrofits, on students' school attendance and standardized test scores.

The first chapter of this dissertation, co-authored with Alain de Janvry and Elisabeth Sadoulet, examines the ex-post, shock-coping impact of weather index insurance from a pioneering, large-scale insurance program in Mexico to cover smallholder farmers as a social safety net. Exploiting insurance thresholds as a source of plausibly exogenous variation in insurance payments, we find evidence that these payments allow farmers to cultivate a larger land area in the growing season following a weather shock. Households in municipalities receiving payments also have larger per capita expenditures and income in the subsequent year. These results suggest that the insurance payments can make smallholder farmers more resilient to shocks, although some of the full impact may be offset by reductions in remittances from abroad that act as informal insurance.

In the second chapter, coauthored with Michael Anderson and Justin Gallagher, we provide evidence on a topic of intense policy interest: improving the nutritional content of public school meals. Debate on this topic is frequently motivated by the health of school children,

and, in particular, the rising childhood obesity rate. Medical and nutrition literature has long argued that a healthy diet can have a second important impact: improved cognitive function. We test whether offering healthier lunches affects student achievement as measured by test scores. We estimate difference-in-difference style regressions that take advantage of frequent changes in lunch-vendors California school districts and find that students served by healthy school-lunch vendor score higher on California state achievement tests. We do not find any evidence that healthier school lunches lead to a decrease in obesity rates. The test score gains, while modest in magnitude, come at very low cost, making this a cost-effective way to increase academic performance.

In the third and final chapter, I study the impact of a government program aimed at replacing old school-buses with inadequate pollution controls. Air pollution has been found to negatively impact children's health, which in turn can affect their academic achievement by causing absences from school, among other mechanisms. School buses are important sources of exposure, both because older models lack adequate pollution controls, and because they are more prone to self-pollution than other vehicles. As a result, the state of California established the LESBP to provide funding to replace and retrofit buses of model year prior to 1986. I find that bus replacements increase attendance in the average school district, with some evidence of larger effects in areas that are out of compliance for PM 10 and PM 2.5. I find no effect of the program on standardized test scores.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisors, Betty Sadoulet and Alain de Janvry, for their guidance, wisdom, and unending generosity to their students. I also thank Michael Anderson, Ted Miguel, Justin Gallagher, and Jeremy Magruder for the time that they have invested to improve my research. I am grateful to have shared this experience with classmates that have not only helped me grow as an economist, but also provided laughter and friendship. I dedicate this dissertation to my husband, Pedro, my parents, and my brothers who have been a source of love, support and encouragement without which this dissertation would not have been possible.

Chapter 1

Weather index insurance and shock coping: Evidence from Mexico's CADENA program[1]

1.1 Introduction

Weather is an important determinant of income for rural populations in developing countries. The short-term impacts of weather shocks can compound into long-term reductions in investments and growth, often resulting in poverty traps (Dercon and Christiaensen, 2011). In the absence of formal insurance markets, smallholder farmers typically resort to self-insuring through their choices of low-risk and low-profit investments (Rosenzweig and Binswanger, 1993; Barnett et al., 2008). They also turn to coping strategies, such as selling assets, that reduce their ability to generate income in the future (Rosenzweig and Wolpin, 1993). All in all, the risk management and shock coping mechanisms used by the poor to protect themselves against uninsured weather risks are typically insufficient and contribute to low production and poverty.

This relationship between weather shocks and poverty is of particular concern given the likelihood that extreme weather realizations, particularly unusually warm temperatures but also drought and floods in some regions, will become more common as a result of climate change (IPCC, 2013). Both extreme temperature and precipitation events have the potential to substantially depress rural incomes through their effects on agricultural production (Rosenzweig et al., 2002; Schlenker and Roberts, 2009; Lobell et al., 2011). Indeed, there is evidence that decline in yields attributable to increased climate variability has prompted outmigration from affected areas, suggesting that these impacts on agricultural production are economically significant (Cai et al., 2016; Feng et al., 2010). Agricultural insurance has

[1] An earlier version of the material in this chapter was made available as World Bank Policy Research Working Paper No. 7715. This version can be found online at https://openknowledge.worldbank.org/handle/10986/24632.

been proposed as a potential adaptation to mitigate the negative impacts of climate change on agricultural production (Lobell and Burke, eds, 2010), while recent scholarship has also acknowledged the challenges that these efforts have faced.

Traditional indemnity-based insurance, where individual damages have to be assessed by a certified loss adjuster, is considered too costly to apply to smallholder farmers in developing countries (Hazell, 1992). Index-based weather insurance has emerged as a tool with the potential of providing small-scale farmers with coverage against covariate weather shocks (Barnett and Mahul, 2007; Alderman and Haque, 2008). In this case, payments are triggered by indicators of weather events, such as rainfall or an average small area yield, falling below (or above) a verifiable threshold, without the need for individual assessment of losses. While greatly promising, demand for index insurance as currently defined and implemented has been low at market prices (Cole et al., 2013). See also the review papers by Miranda and Farrin (2012), Carter et al. (2017), and Jensen and Barrett (2016). The most important barriers to adoption are basis risk, whereby some risks are uninsured due to discrepancy between measured weather indicators and what happens in a farmer's field (McIntosh et al., 2015; Clarke, 2016), and high cost due in part to lack of data in calculating a fair price (Carter, 2012). Other research has found that low demand can be attributed in part to lack of trust and liquidity constraints (Cole et al., 2013), as well as the predominant format of individual policies for farmers who may be part of cooperatives (de Janvry et al., 2014).

Most of the evidence regarding the impact of weather index insurance has focused on its effect on ex-ante investment decisions. Many of these studies find that farmers offered index insurance take on riskier but more profitable investments consistent with risk management theory. Studies in India have found that these changes come in the form of shifting production towards riskier, but higher yielding crops or varieties (Mobarak and Rosenzweig, 2013; Cole et al., 2017). Others have found index insurance to increase fertilizer use on maize in Ghana (Karlan et al., 2014) and on tobacco in China (Cai, 2016). However, the impact of index insurance on the adoption of high-value crops is not uniformly positive. Giné and Yang (2009) find that bundling index insurance with a loan to purchase higher-yielding groundnut and maize varieties actually reduces adoption relative to the case where only the loan is provided. Additional theoretical work by Carter et al. (2016) has identified the conditions under which index insurance can have the largest impact on technology adoption. Janzen and Carter (2016) is one of the few papers that focuses on the ex-post, shock coping value of index insurance. They find that the provision of insurance reduces potentially costly strategies to cope with shocks by pastoralists in Kenya, such as lowering consumption (for the poorer households) and selling livestock assets (for the richer households).

We study the ex-post impact of payments provided through weather index insurance in the context of a large-scale government-funded insurance program. This program, which goes by the name of CADENA, was pioneered by the Mexican government as a social safety net for severe weather shocks. CADENA insures smallholder farmers and has achieved widespread coverage by having state and federal governments, rather than individual farmers, pay the insurance premiums. By 2013, CADENA insured approximately 12 million hectares of cropland (FAO, 2014). The expansive coverage and relatively long tenure of the CADENA program stand in contrast to the Randomized Control Trial settings in which the impact of index insurance has generally been studied.

Our paper also contributes to the more limited literature on the ex-post effect of weather index insurance on shock coping by smallholder farmers. While a more complete evaluation of the program would also incorporate its effect on risk management practices, the conditions of the program rollout do not support a research design that could identify these effects. Nevertheless, we believe that understanding the ex-post effects of insurance payments provides valuable insights for understanding index insurance. Moreover, we also contribute to the literature on the use of index insurance as part of a social safety net. A recent paper by Jensen et al. (2017) finds that investments resulting from purchases of livestock index insurance in Ethiopia generate benefits that persist over the medium term. They also find that the average cost of the index insurance is similar to that of cash transfers but the marginal cost is much lower, suggesting that it may be a more cost-effective method of expanding social protection schemes. To identify the effect of insurance payments, we exploit thresholds built into the insurance program in a regression discontinuity design. This design allows us to compare municipalities that received similar weather shocks such that differences in observed ex-post outcomes are attributable to insurance payments alone and do not rely on assumptions about how the program was rolled out over time. Consequently, the effects we estimate are net of any risk management effects induced by introduction of the insurance.

While certain data limitations affect the robustness of our estimates, this analysis provides evidence that insurance payments allow farmers to cultivate larger land areas in subsequent growing seasons, consistent with the presence of credit constraints that result in diminished investment following a weather shock, though we cannot completely rule out the hypothesis that receiving an insurance payment results in learning about the reliability of the insurance. The insurance payments also result in higher household expenditures per capita, indicating welfare gains, although some of the benefits may be offset by a reduction in private transfers from abroad.

The remainder of the paper is organized as follows. Section 1.2 provides background about the CADENA program and outlines the data used in the evaluation. Section 1.3 describes the empirical strategy, while sections 1.4 and 1.5 present and discuss the results, respectively. Section 1.6 provides falsification tests in support of the identification strategy, and section 1.7 concludes.

1.2 Background and Data

With a rural population of approximately 27 million and two-thirds of the country's poor living in rural localities (INEGI, 2010), reducing exposure to weather risks is an important component of poverty reduction efforts in Mexico, and climate change is only increasing its importance. Models linking crop yields to weather suggest that climate impacts have substantially reduced yield gains for wheat in Mexico over the period 1980-2008 (Lobell et al., 2011). Worldwide, the authors project that climate trends have reduced maize production, an important staple crop in Mexico, by approximately 3.8% over this same time period. To address the joint problems of poverty and uncertainty arising from climate change, the federal government began the CADENA weather index insurance program in 2003, administered through the Ministry of Agriculture (SAGARPA). The purpose of this program is to provide relief to smallholder farmers when crop failures occur and to do so in a way that makes government expenditures more predictable than standard relief programs. The federal government promotes the use of insurance by subsidizing up to 90% of premium payments paid by state governments,[2] while gradually reducing the percentage of funds it contributes to ex-post relief via the Direct Support scheme (*Apoyos Directos*).

CADENA began with drought index insurance covering small maize and sorghum farmers in one state of Mexico. It has expanded significantly since its inception in both geographic scope and breadth of coverage.[3] It now offers weather index insurance for a variety of perils (e.g., drought, flood, and hail), as well as area-based yield index insurance that provides payment when the average yield in an area, as determined by a random sample of plots, falls below a given threshold. CADENA also offers traditional and remote sensing index insurance for livestock (Arias et al., 2014). Between 2003 and 2011, the Mexican government, both state and federal, paid approximately USD 382 million in CADENA premiums. During this time period, the CADENA program made transfers of USD 353 million to insured farmers, of which index insurance accounted for about 20% (Arias et al., 2014). This analysis will focus on the drought index insurance because it has historically been CADENA's largest component, and it is the only type of insurance for which we have the weather station and thresholds associated with each policy. Drought insurance is also of particular interest given that 80% of the weather shocks resulting in severe agricultural losses in Mexico can be attributed to drought (Fuchs and Wolff, 2011). In what follows we simply refer to this insurance as index insurance.

Through the index insurance component of the program, CADENA currently insures farmers growing staple crops on less than 20 hectares of rainfed land, who are then automatically enrolled in the program at no cost to them (SAGARPA, 2014). Individual farmers are not directly insured. Instead, state governments buy insurance policies associated with a particular crop and municipality.[4] Each policy is assigned a corresponding weather station and provides coverage during three pre-determined phases of the production cycle that run from planting, to growing, and to harvesting. If precipitation as measured by the designated weather station falls below the threshold in any of the three phases, the insurer makes pay-

[2]Subsidies from the federal government depend on the marginality index of the insured municipality, as computed by CONAPO (Consejo Nacional de Población).

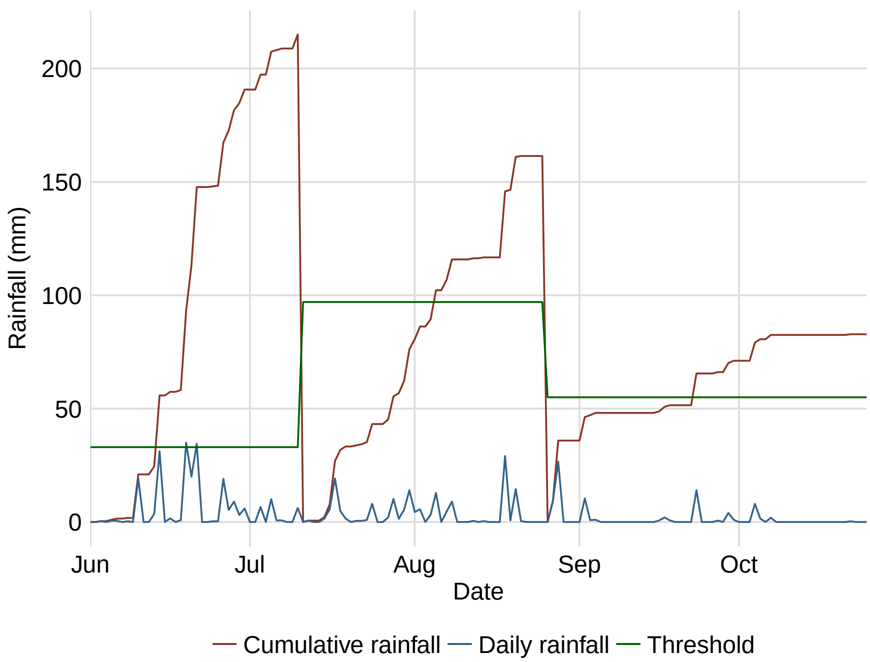
[3]CADENA's staggered rollout is a result of a number of factors. First of all, only municipalities sufficiently close to a weather station were eligible for index insurance. As the program expanded to traditional insurance, the coverage area also expanded. Secondly, climactic models to assess risk for individual municipalities were developed gradually. Lastly, states also had the choice of which municipalities to insure and these choices could have changed over time.

[4]Some policies are managed by the federal government for cases in which the state declines coverage, but the federal government deems the municipality a high priority area.

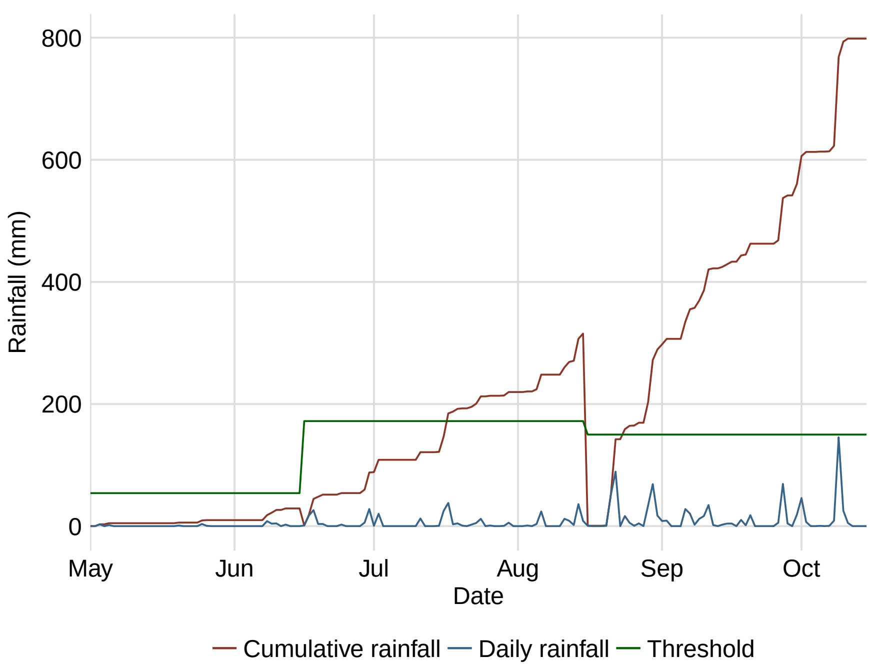
outs to the state, which in turn transfers these to eligible farmers in the insured area.[5] In figure 1.1, we provide an example of the policy thresholds and actual precipitation (cumulative and daily) for two municipalities. The municipality in panel a would not receive a payment from that policy because its cumulative precipitation has exceeded the insurance threshold in each phase of the growing cycle. In contrast, the municipality shown in panel b, would receive payment because its cumulative precipitation in the first phase fell below the insurance threshold for that phase. Because of restrictions regarding the maximum distance between the weather station and the insured area, a municipality may be insured by multiple policies each linked to a different station. Conversely, multiple municipalities can have policies linked to a single weather station.

[5]States have some discretion in using these funds, such as directing them towards investments that will benefit the affected communities or providing temporary employment to individuals in affected communities.

Figure 1.1: Example insurance thresholds and realized rainfall



(a) Municipality receiving no paymentt



(b) Municipality receiving payment

Note. Insurance thresholds obtained from policy data provided by SAGARPA and rainfall data comes from National Water Commission (CONAGUA).

The data for this evaluation come primarily from four sources. The policy data from SAGARPA include the area insured by crop type (i.e., maize, beans, sorghum, barley), the rainfall thresholds and corresponding stations, and record of all payments for each insured municipality for the period 2005 to 2013. The stations and thresholds associated with a given municipality can change from year to year. Weather data from the National Water Commission (CONAGUA) allow us to calculate the precipitation at each of the weather stations linked to an insurance policy, which in turn is compared to the policy thresholds and used to determine if that policy should have paid out. To determine the effect of insurance payments on yields and areas planted, we use agricultural production data from SAGARPA detailing the annual hectares sowed, hectares harvested, and total production in metric tons at the municipality-crop level. Lastly, to study the economic impacts of the insurance, we use national household income and expenditure surveys (ENIGH) which are carried out every other year with the latest one available in 2014. These household expenditure and income surveys are repeated cross-sections of households with a rotating sample of municipalities.[6] Households are surveyed between the months of August and November and the income, expenditure, and remittance variables refer to the previous three months. Lastly, the analysis sample when using ENIGH data is limited to households living in rural localities. The Mexican Statistical Institute (INEGI) defines rural localities as those with less than 2,500 inhabitants. Although the policy data is available at the level of the weather station, data on insurance payouts, and agricultural and economic outcomes are available only at the municipality level. Hence, our analysis will be carried out at the municipality level.

1.3 Empirical Strategy

This evaluation of the CADENA program focuses on the effect of insurance payouts on post-shock production decisions and other coping mechanisms. To identify this effect, we limit our sample to municipalities that were insured through drought index insurance policies between the years 2005 and 2013. Using weather data provided by CONAGUA, we match policies to their corresponding weather stations and calculate the total precipitation recorded at that station within each of the three phases of the production cycle designated in the policy. We then subtract the policy-specific precipitation threshold from the realized precipitation to obtain deviations from the threshold for each of the three phases, $X_{mcsti} = precip_{mcsti} - threshold_{mcsti}$, where m indexes the municipality, c the crop, s the weather station, t the year, and i the phase. A policy results in payment if precipitation at the corresponding weather station falls below the pre-determined threshold in any of the three phases.

In an ideal setting, we would observe insurance payments and all relevant outcomes for precisely the area covered by the policy. The running variable in this setting would be the

[6]The rotating sample means that in any given year some insured municipalities will not be included in the analysis sample because they do not have income or expenditure data, which substantially decreases the sample of municipalities for analyzing economic outcomes.

minimum deviation from the threshold at a given weather station over the three phases of the policy: $X_{mcst} = \min_{i\in\{1,2,3\}}\{X_{mcsti}\}$. However, we only observe payment and outcome variables at the municipality level, so we define the running variable at the municipality-crop-year level as the minimum deviation from the threshold over all the weather stations associated with a given municipality, $X_{mct} = \min_{s\in S(m)}\{X_{mcst}\}$. Given the data limitations, we have to account for the fact that there is substantial heterogeneity in the intensity of the treatment among municipalities receiving payment because some municipalities may have received payouts on 25% of their insured area and others may have received payment on 100%. In addition, there exists variation in terms of the percentage of agricultural land in the municipality that is insured via CADENA, since the policy is limited to land cultivated by producers with less than 20 hectares of land.

Thus, the treatment variable (T) is defined as the total number of hectares (across all weather stations linked to a given municipality) that received a payout divided by the number of hectares of land devoted to that crop in a given municipality:

$$T_{mct} = \frac{hapaid_{mct}}{\overline{ha}_{mc}} \tag{1.1}$$

To account for the fact that municipalities that received insurance payments may have different weather realizations than those that do not, we instrument treatment with an indicator for falling below the threshold and limit our sample to a narrow bandwidth from the normalized threshold of zero. This regression will give causal estimates so long as the precipitation measured at the weather station cannot be manipulated to obtain payouts, such that potential outcomes are smooth over the normalized threshold. We will provide evidence that this assumption holds in section 6.

We begin by estimating the first stage of our regression via the following equation:

$$T_{mct} = \alpha_0 + \beta_0 Z_{mct} + \gamma_0 X_{mct} + \pi_0 X_{mct} \cdot Z_{mct} + \delta_0^c + \nu_{mct} \tag{1.2}$$

where $Z_{mct} = \mathbf{1}\{X_{mct} < 0\}$ is an indicator for falling below the threshold amount of rainfall and serves as an instrument for treatment, while δ_0^c is a crop fixed effect. We then estimate the equation below, using the instrumented treatment variable $\widehat{T_{mct}}$ obtained from equation 1.[7]

$$y_{mct+1} = \alpha_1 + \beta_{2SLS}\widehat{T_{mct}} + \gamma X_{mct} + \pi X_{mct} \cdot \widehat{T_{mct}} + \delta^c + \varepsilon_{mct} \tag{1.3}$$

The coefficient β_{2SLS} measures the marginal effect of a change in the percentage area receiving payment in municipality m at time t on the outcome of interest y_{mct+1} at time $t+1$.

[7] For agricultural outcomes, we look at the effect on agricultural outcomes for a given crop, and the treatment variables is defined as the percentage of land devoted to that crop that received payment. For non-agricultural outcomes, we use the total percentage of land area receiving payment, irrespective of the crop.

Outcomes of interest include area planted, yield, expenditure, income, and remittances. The analysis is restricted to insurable crops (i.e., maize, beans, sorghum, and barley) unless we explicitly note otherwise. Household level regressions using ENIGH data are estimated with a similar set of regressions, where observations are no longer indexed by crop c, but are now indexed by h to indicate an individual household.

1.4 Results

1.4.1 Descriptive statistics

We begin by looking at some descriptive statistics for the analysis sample in table 1.1. The analysis sample for the agricultural outcomes is composed of 976 unique municipalities over 8 years with a total of 4,311 observations. As mentioned before, the way in which the ENIGH sample is constructed results in a smaller number of unique municipalities for our economic regressions (192),[8] but there are multiple observations (households) per municipality for a total sample size of 5,879. Given its roles as a social safety net, CADENA is designed to ensure relatively rare events. We confirm this in panel a, where we see that the mean of the running variable (deviation from the precipitation threshold) is 71 mm, suggesting that in an average year a municipality can expect to receive rainfall well in excess of the threshold. Additionally, only 11% of the observations in our sample have rainfall realizations that fall below the insurance threshold.[9] Turning to panels b and c, we see evidence that points to a strong first stage, which will be formally tested in a subsequent section. Specifically, the mean of the running variable is well above zero for municipalities that do not receive payment in a given year (panel b), and only four percent of these observations receive rainfall below the threshold. We contrast these results with municipalities that receive payment in that year (panel c), and we see that the mean of the running variable is negative (i.e., precipitation falls below the threshold) and 77% of these observations have rainfall realizations that should trigger insurance, as would be expected if these rainfall thresholds are generally enforced. Policy rules to deal with missing data may explain the imperfect compliance we observe. Specifically, the policy rules state that missing data is filled in with a secondary weather station, which we do not observe, so long as the number of days with missing information does not exceed 20% at a weather station. Despite the imperfect implementation, there is a strong first stage, which we will formally show by estimating equation 1.1.

[8]Given the large number of municipalities that get dropped from the ENIGH panel, we present summary statistics for the two different samples in appendix table A.1.1. While the municipalities in the ENIGH sample have much larger populations, they are very similar along other dimensions with only the education variables showing even marginally significant differences.

[9]If we repeat this calculation with rainfall data from before CADENA's rollout, we find that about 8.7% of observations fall below the most recent CADENA thresholds, a similarly low percentage.

Table 1.1: Summary statistics

Variable	Observations	Municipalities	Mean	Std. Dev.
Panel a: All				
Minimum deviation (mm)	4311	976	71.07	83.09
Min deviation < 0 (%)	4311	976	0.11	0.32
% ha receiving payment	4311	976	0.04	0.16
Hectares sowed (annual)	4311	976	4042.5	6995.6
Yield (metric tons/ha)	4311	976	2.10	1.74
Income p.c. (MXP)	5879	192	5983.1	7223.5
Expenditure p.c. (MXP)	5879	192	5879.7	7244.4
Remittances p.c. (MXP)	5879	192	253.4	1125.9
Panel b: Payment = 0				
Minimum deviation (mm)	3900	943	80.24	81.28
Min deviation < 0 (%)	3900	943	0.04	0.20
% ha receiving payment	3900	943	0	0
Panel c: Payment = 1				
Minimum deviation (mm)	411	267	-15.85	37.23
Min deviation < 0 (%)	411	267	0.77	0.42
% ha receiving payment	411	267	0.45	0.32

Note. Rainfall and agricultural variables in panel a come from the agricultural panel, which has municipal-level observations and consists of policy data merged with agricultural production data, as do the statistics in panels b and c. The economic variables in panel A come from the economic panel, which consists of household level observations from ENIGH merged with policy data. Economic variables are measured in nominal Mexican Pesos (MXP).

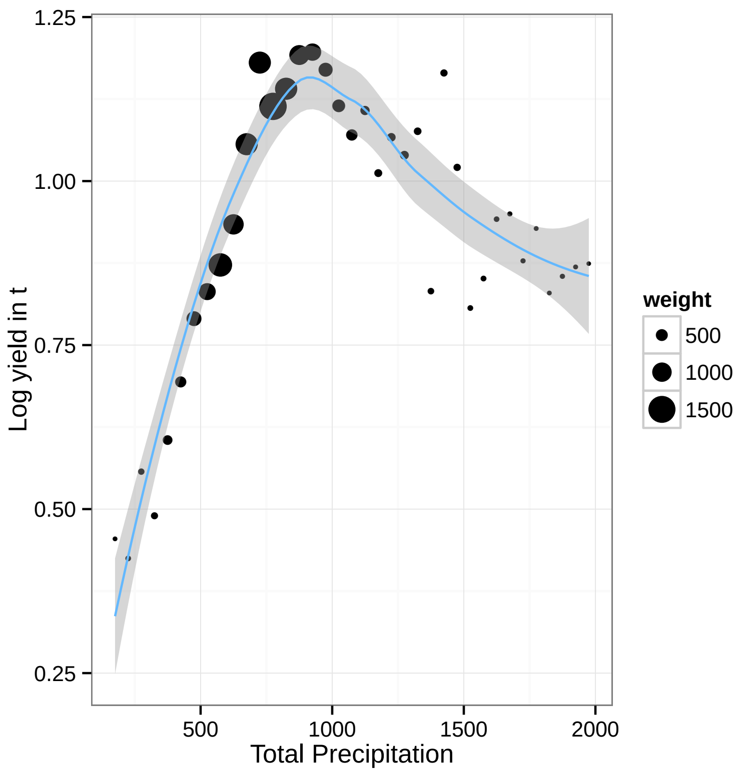
To understand the severity of shocks covered through CADENA's drought index insurance, we explore in more detail the relationship between rainfall and yields prior to the introduction of CADENA. In figure 1.2, we plot a local regression of log yield on total precipitation in the growing season.[10] The points on the graph represent the average yield in 50 mm bins, taking observations over all crops and municipalities in our sample period.[11] As can be seen, there is a positive relationship between precipitation and yields until about

[10] We define the growing season as the period covered by the insurance policy once CADENA comes into effect.

[11] A plot using residuals of a regression of log yield on crop fixed effects looks very similar.

1,000 mm of precipitation. To put this result in context, the average insurance threshold is about 230 mm of precipitation over the entire growing season, and the average yield at this level of precipitation is about 37% of the median yield.[12] We take this as evidence that farmers who depend on rainfed agriculture suffer losses when precipitation falls below the thresholds established in the insurance policies. This fact is useful to contextualize the effects we observe in subsequent analyses.

Figure 1.2: Pre-CADENA: log yield in t as a function of total precipitation



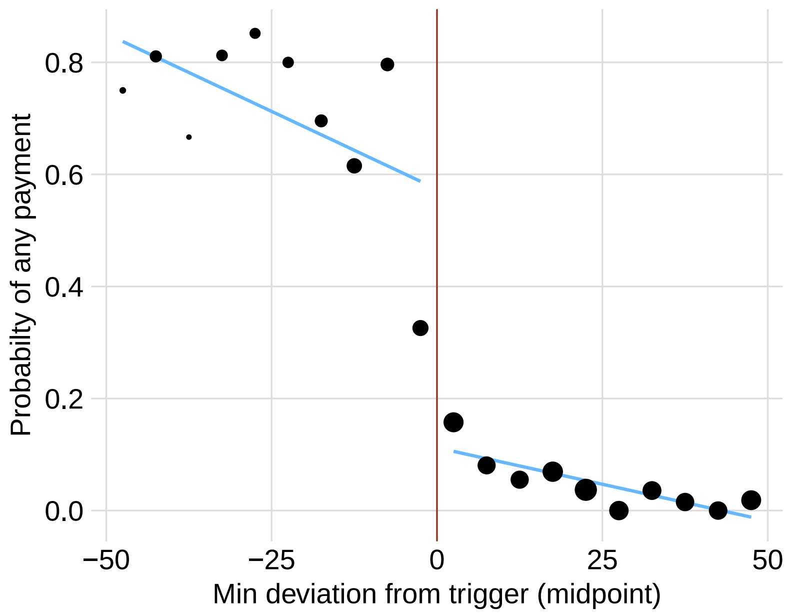
Note. This figure depicts the relationship between log yield and precipitation for insurable crops (maize, beans, barley, and sorghum) before CADENA is rolled out in each municipality beginning in 2001 (e.g., if a municipality is insured for the first time in 2008, we include observations up to 2007). The last year included in the sample for any municipality is 2012. Each circle depicts the average of log yield for all observations (municipality-crop-year) within 50 mm precipitation bins. The size of the circle represents the number of observations. The blue line depicts a loess smoother, while the gray shaded area represents the 5% confidence interval.

[12]The thresholds are defined separately for three different parts of the growing season. For illustrative purposes, we sum the three thresholds for each municipality and average over all municipalities to obtain 230 mm.

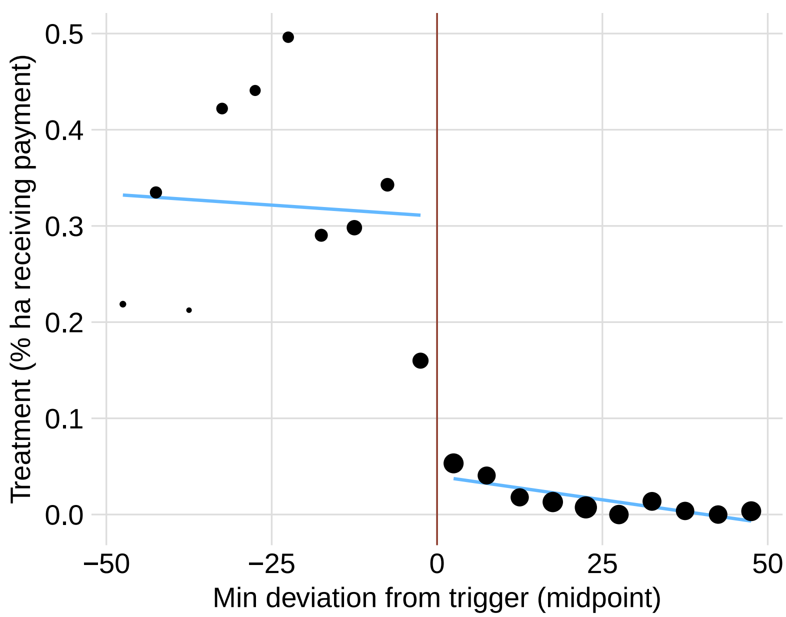
1.4.2 Agricultural outcomes

To ensure the validity of the research design, we must first verify that our treatment variable changes discontinuously across the insurance threshold. In panel a of figure 1.3, we group municipality-level observations into 5 mm bins based on the value of the running variable, and plot the probability that a municipality receives any insurance payment in each bin. We see that the probability of receiving any payment at the municipality level increases by approximately 50 percentage points at the insurance threshold. Meanwhile, panel 1.3b gives a graphical representation of the first stage, showing a 20 point increase in the average percentage of hectares that receive payment (at the municipality level) when rainfall falls below the insurance threshold. Given that the average treated municipality receives payment for about 45% of its agricultural land, this result is equivalent to the 50% increase in the probability of receiving any payment seen in panel a.

Figure 1.3: First Stage — Change in insurance payment at threshold



(a) Probability of any payment



(b) Percent of hectares receiving any payment

Note. The figures plot the local average of the respective outcome variable in 5 mm bins. The size of the circle indicates the number of observations in each bin. The lines represent linear regressions on the binned averages estimated separately on each side of the normalized threshold. The running variable is calculated by taking the minimum difference between realized rainfall and the threshold at the municipality level. The sample in these figures is limited to observations within the optimal bandwidth of 50 mm.

One of the stated aims of the CADENA program is to ensure that farmers who have suffered negative weather shocks have sufficient resources to purchase inputs for the next growing season. We cannot observe input purchase, but if the insurance is serving its intended purposes, these input purchases should be translating into improved agricultural outcomes in either the area planted or the yields achieved. Thus, we begin this analysis by estimating the effect of insurance payment on the change in log hectares sowed from t to $t+1$ and the log yield in $t+1$. Table 1.2 shows the results of estimating equations 1 and 2 with Δ log hectares planted and log yield of insured crops as the outcomes. Column 1 reports the results for Δ log hectares planted using the optimal bandwidth, which was calculated following the procedure in Calonico et al. (2014, 2015).[13] Using the 50 mm bandwidth, we find a reduced form effect of approximately 8% reported in panel b. Panel c reports the two-stage least squares estimate of 0.39, which is statistically significant at the 5% level. This estimate implies that for a given crop, increasing the percentage of hectares that received payment from 0 to 100% would increase the amount of land devoted to that crop by 39% relative to the previous year. Given that the average treatment municipality receives payment for 45% of its land, we would expect the average effect of receiving payment to be approximately 17%.

In columns 2 and 3, we show alternative specifications for robustness. Column 2 includes state fixed effects and a number of district level covariates,[14] which are not needed for identification but could provide greater precision. The point estimates do not change substantially for the area regression, but in this case, neither does the precision of the estimates. Column 3 shows results for a larger bandwidth of 70 mm, for which the coefficient estimate is similar in magnitude but no longer significant.[15] For the yield regressions, none of the results are statistically significant and the point estimates are less stable. A land response with no accompanying yield response is consistent with certain models of farmer behavior that we will discuss in section 5.

[13]The optimal bandwidth varied slightly for different outcomes, so we used 50 mm across all outcomes to maintain consistency.

[14]These covariates are state fixed effects, municipal land area cultivated by small (< 20 hectares) Procampo beneficiaries, as well as municipal population, average education, percentage of population that is indigenous, and female work force participation as reported in the 2010 census.

[15]The first stage is only significant for a smaller bandwidth of 30, but the reduced form results for land area remained quite similar.

Table 1.2: Agricultural outcomes (insured crops) at time $t+1$

	(1)	(2)	(3)	(4)	(5)	(6)
Panel a: First Stage						
	Treatment (% ha receiving payment)					
Below trigger	0.194***	0.193***	0.197***			
	(0.0578)	(0.0546)	(0.0673)			
Panel b: Reduced Form						
		Δ log ha planted			log yield	
Below trigger	0.0765**	0.0786**	0.0571	0.0733	-0.0416	0.107
	(0.0357)	(0.0334)	(0.0404)	(0.0596)	(0.0642)	(0.0755)
Panel c: 2SLS						
		Δ log ha planted			log yield	
$\hat{T}$	0.394**	0.407**	0.291	0.438	-0.215	0.545
	(0.194)	(0.197)	(0.201)	(0.344)	(0.329)	(0.466)
Bandwidth	50	50	70	50	50	70
Controls	N	Y	N	N	Y	N
N	1916	1916	2502	1916	1916	2502

Note. Standard errors are clustered at the state level. Asterisks indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Observations are at the municipality-crop-year-level. Agricultural production is for the corresponding insured crop (maize, beans, sorghum, and barley). All regressions include crop fixed effects. Panel a shows results from the OLS estimation of the first stage, which corresponds to estimating equation 1. Panel b shows results from the OLS estimation of the reduced form, which corresponds to estimating equation 2, and panel c displays 2SLS estimates. Columns 1 and 4 are estimated with a linear polynomial of the running variable on a sample of observations within the optimal bandwidth of 50 mm. Columns 2 and 5 include state fixed effects and municipality characteristics (e.g., population, % indigenous from the 2010 census). Columns 3 and 6 are estimated using a linear polynomial on a sample of observations within an alternative bandwidth of 70 mm.

In table 1.3, we analyze the impact of insurance payment on total area planted, including non-insured crops. Using a bandwidth of 50 mm, reported in column 1, we find a point estimate of 0.288. The estimates are robust and remain significant across the alternate specifications included in columns 2 and 3. These results suggest that farmers are not simply displacing uninsured crops to increase cultivation of insured crops, but actually expanding the total land area under cultivation, presumably by utilizing land that might otherwise have been left fallow. Nevertheless, the point estimate when considering all crops is about 73% as large as the estimate when considering only insured crops. In the average municipality, 76% of the land is devoted to insured crops, which suggests that the increase in total land area is coming almost exclusively from the expansion of insured crops.

Table 1.3: Total area planted at time $t + 1$

	(1)	(2)	(3)
Panel a: First Stage			
	Treatment (% ha receiving payment)		
Below trigger	0.245^{***}	0.243^{***}	0.235^{***}
	(0.0590)	(0.0559)	(0.0768)
Panel b: Reduced Form			
	Δ log ha planted (all)		
Below trigger	0.0704^{***}	0.0626^{**}	0.0662^{***}
	(0.0231)	(0.0225)	(0.0227)
Panel c: 2SLS			
	Δ log ha planted (all)		
$\hat{T}$	0.288^{**}	0.258^{**}	0.282^{*}
	(0.117)	(0.110)	(0.146)
Bandwidth	50	50	70
Controls	N	Y	N
N	1503	1503	1947

Note. Standard errors are clustered at the state level. Asterisks indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Observations are at the municipality-year-level. Hectares sowed are defined as the total hectares growing any rainfed agricultural crop as reported in SAGARPA production data. Panel a shows results from the OLS estimation of the first stage, which corresponds to estimating equation 1. Panel b shows results from the OLS estimation of the reduced form, which corresponds to estimating equation 2, and panel c displays 2SLS estimates. Column 1 is estimated with a linear polynomial on a sample of observations within the optimal bandwidth of 50 mm. Column 2 includes state fixed effects and municipality characteristics (e.g., population, % indigenous from the 2010 census). Column 3 is estimated using a linear polynomial on a sample of observations within an alternative bandwidth of 70 mm.

1.4.3 Economic outcomes

Next, we turn to economic outcomes as measured in the household income and expenditure surveys. Panel c of table 1.4 reports the 2SLS estimates of the effect of insurance payment on log expenditures per capita and log income per capita and remittances per capita.[16] In column 1, we see the point estimate for log expenditures per capita using the optimal bandwidth. The point estimate is large at 0.613, although it is only significant at the the 10% level. Adding covariates, somewhat reduces the magnitude of the point estimate, as seen in column 2, but it greatly improves precision. Column 4 reports a point estimate for log income per capita of 0.852, which is significant at the 5% level. Thus, going from a municipality that receives no payment to the average treatment municipality (i.e., 45% treated) we would expect increases in expenditure and income per capita of approximately 22-27% and 32-38%, respectively.[17] Columns 5 and 6 report the coefficients for remittances per capita received in the previous three months. The estimates suggest that the average treatment municipality receives between 242 and 278 pesos less remittances per capita relative to a municipality receiving no payment, although only the latter is statistically significant at the 10% level. The unconditional mean of remittances per capita is approximately 253 pesos. Thus, these estimates suggest a very large, albeit imprecisely estimated effect.

To better understand these effects, consider the fact that in a set of focus group interviews, the majority of farmers reported having between 4 and 8 hectares of land. On average, about half of the insured hectares receive payments of 1,500 pesos/ha, such that the average household could expect to receive between 3,000 and 6,000 pesos. Meanwhile, we observe that household income and expenditures per capita increase about 30 to 40% as a result of payment. The mean of both total household expenditures and total household income in the sample of rural localities is approximately 20,000 pesos[18], such that an increase of 30 to 40% implies 6,000 to 8,000 pesos in additional income. The timing of the survey is such that this increase in income should come primarily from increases in the value of the harvest or investments made in other income-producing activities using the insurance payments. A minority of the households in the sample may be receiving substantially delayed CADENA payments within this period of time. Thus, if we ignore the effect on remittances for a moment, these results suggest that every peso provided in insurance payment results in 1.0 to 2.7 additional pesos of income and/or expenditure for farmers, keeping in mind that the coefficients underlying this calculation are rather imprecise. Now, the point estimates suggest that remittances decrease by about 1100 pesos per household (mean household size of about 4). This result implies that the gross increase in other income sources for treated households must be approximately 7,100 to 9,100, which in turn changes the rate of return modestly to be in the range of 1.2 to 3.0 pesos of additional gross income per peso of insurance payment. This is a large multiplier effect on insurance payouts, comparable to the 1.5 to 2.6

[16] Remittances are in levels rather than logs due to the large number of zero values for this variable.

[17] Specifications with a bandwidth of 70 mm are qualitatively similar but are not included for the sake of brevity.

[18] This is income received in the past three months.

range observed for Procampo transfers to a similar smallholder farmer population in Mexico (Sadoulet et al., 2001).

Table 1.4: Economic outcomes at time $t+1$: expenditures and income

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First Stage						
	Treatment (% ha receiving payment)					
Below trigger	0.341^{**}	0.393^{***}				
	(0.121)	(0.120)				
Panel B: Reduced Form						
	log expenditures p.c.		log income p.c.		Remittances p.c.	
Below trigger	0.209	0.195^{**}	0.290^{*}	0.282^{***}	-208.4^{**}	-211.8
	(0.140)	(0.0839)	(0.165)	(0.0814)	(89.73)	(144.3)
Panel C: 2SLS						
	log expenditures p.c.		log income p.c.		Remittances p.c.	
$\hat{T}$	0.613^{*}	0.495^{***}	0.852^{**}	0.717^{***}	-611.9^{*}	-538.7
	(0.333)	(0.186)	(0.361)	(0.189)	(316.0)	(340.0)
Bandwidth	50	50	50	50	50	50
Controls	N	Y	N	Y	N	Y
N	2494	2391	2494	2391	2494	2391

Note. Standard errors are clustered at the state level. Asterisks indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Observations are at the household-municipality-level. The treatment variables vary only at the municipal level. Panel a shows results from the OLS estimation of the first stage, which corresponds to estimating equation 1. Panel b shows results from the OLS estimation of the reduced form, which corresponds to estimating equation 2, and panel c displays 2SLS estimates. Columns 1, 3 and 5 are estimated with a linear polynomial on a sample of observations within the optimal bandwidth of 50 mm. Columns 2, 4 and 6 include state fixed effects and municipality characteristics (e.g., population, % indigenous from the 2010 census).

The effect on remittances is in line with the research finding that formal transfers may substitute informal risk sharing mechanisms or other transfers. Albarran and Attanasio (2003) show that the cash transfers provided by the PROGRESA program in Mexico crowd out private transfers. Yang and Choi (2007) and Cox et al. (2004) study this phenomenon in the context of the Philippines and find that private transfers are highly sensitive to changes in income. These two studies point to the role of remittances as insurance. As such, it is reasonable to think that when formal insurance payments are made, the need for remittances to cope with shocks is diminished. An alternative way of interpreting this result is that remittances make up for the basis risk inherent in index insurance. Mobarak and Rosenzweig (2013) find that the presence of basis risk in weather insurance makes subcaste-based informal risk-sharing a complement to weather index insurance in India. Dercon et al. (2014) find similar complementarity between index insurance and informal risk sharing groups in Ethiopia. In the presence of basis risk, very risk averse individuals may have very low or no demand for index insurance (Clarke, 2016) because the disutility of the worst case scenario where they suffer a shock and receive no payouts outweighs the benefits of the insurance. However, if informal risk sharing networks can make transfers in this worst case scenario, then index insurance and informal insurance are actually complements.[19] It may be the case that migrants abroad continue to send transfers in response to shocks that do not result in insurance payments.

1.5 Mechanisms and Discussion

Two potential mechanisms can explain the observed behavior by farmers: learning about insurance policies and credit constraints. In the first mechanism, farmers either do not trust that payments will be made in case of a weather shock or there is some ambiguity about how the policy works. When they receive a payment following a weather shock the ambiguity and/or uncertainty is resolved, and given the promise of insurance coverage, they now find it optimal to expand the area under cultivation with insured crops, i.e., we observe what should have been an ex-ante risk management response. Such a mechanism would be consistent with recent findings that personal experience with infrequent disaster events is an important determinant of insurance demand, even when information about the probability of these events is available (Cai and Song, 2017; Gallagher, 2014a). The second alternative is that in the presence of credit constraints, farmers who received insurance payments will have more resources at the beginning of the next planting season relative to those that did not receive the payments and suffered a similarly bad harvest.

While we cannot definitively rule out either explanation, we show evidence that learning is unlikely to be the entire story. In table 1.5, we show the results of regressions with hectares sowed as the outcome that also include indicators for previous payments. Column 1 shows the result of a regression with three treatment lags. We see that the main effect remains significant and is similar in magnitude to previous estimates, while indicators for treatment in previous years are not significant. Column 2, includes an indicator for whether the municipality has ever received treatment in the past and an interaction of the main effect with this indicator. While the main effect is no longer significant, the point estimate remains large and the interaction is not significant, although it is large in magnitude. If learning where the only reason for the observed effects, we would expect the impact of an insurance payment to be zero if a municipality has received a payment in the past, and is thus likely to better understand or trust the insurance product. As a result, we conclude from these results that learning is unlikely to be the only channel through which insurance payments induce increased investment and income.

[19]Note that in the case of CADENA, farmers do not make premium payments, so this worst case scenario will never materialize. However, transfers from abroad can still serve to diminish the basis risk.

Table 1.5: Total area planted at time $t+1$

	(1)	(2)
	Δ log ha planted	Δ log ha planted
$\hat{T}_t$	0.441**	0.333
	(0.183)	(0.235)
$\hat{T}_{t-1}$	0.197	
	(0.167)	
$\hat{T}_{t-2}$	0.0150	
	(0.0267)	
$\hat{T}_{t-3}$	0.000270	
	(0.134)	
Treated before		0.0127
		(0.0273)
Treated before $\times \hat{T}_t$		0.222
		(0.316)
N	1916	1916

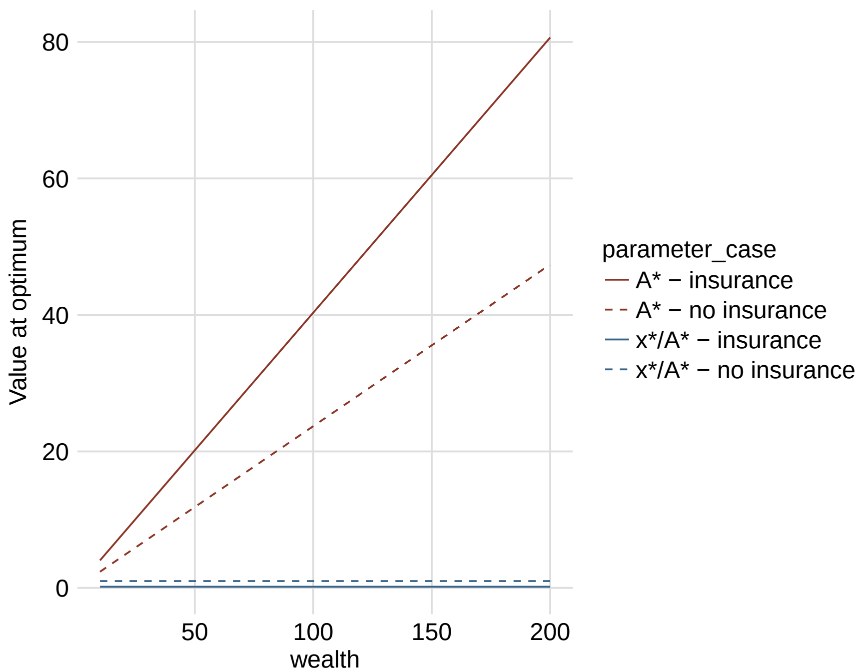
Note. Standard errors are clustered at the state level. Asterisks indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Observations are at the municipality-crop-year-level. Agricultural production is for the corresponding insured crop (maize, beans, sorghum, and barley). All regressions include crop fixed effects and are estimated with a linear polynomial of the running variable on a sample of observations within the optimal bandwidth of 50 mm.

The credit constraint mechanism is consistent with other papers studying the investment of cash transfers in Mexico. As previously noted, Sadoulet et al. (2001) find multipliers in the 1.5 to 2.6 range for Procampo transfers, suggesting that smallholder farmers faced credit constraints that were eased by the transfers resulting in productive investments. In the context of the Oportunidades programs, Gertler et al. (2012) find that beneficiaries invested part of the transfers and increase their agricultural income by about 10% after 18 months of benefits.

Lastly, we can also consider a model prediction of a farmer's behavior when receiving an income transfer, such as an insurance payment. Consider a risk averse farmer in a two-period model, who has a choice between spending his current wealth (w) on current consumption or inputs into production, land (A) and a second input such as fertilizer (x), which will determine his consumption in period two. If we further specify that the return to these inputs depends on a weather realization and we have a constant return to scale Cobb-Douglas production function, we can simulate this problem and obtain the optimal level of A and x as a function of income (see appendix A.1.2 for more details). In figure 1.4

we plot x^*/A^*, which in a constant return to scale production function is a good indicator of yield, and A^* as a function of wealth in the current period. These quantities are plotted for a scenario in which index insurance is present (solid line) and one where it is not (dotted line). In both cases, we see that x^*/A^* is constant with w, while A^* is increasing in w. Moreover, the increase in A^* in response to an increase in wealth is more pronounced in the scenario where index insurance is present. If we think of an insurance payment as increasing wealth at the time of planting, the predictions of this particular model are consistent with our findings of a null, although imprecisely estimated, result on yield, and a positive impact on land area under cultivation.

Figure 1.4: Optimum input values



1.6 Falsification Tests

The validity of the research design relies on the assumption that potential outcomes are smooth across the insurance threshold. Because potential outcomes are unobservable, we first test this assumption by ensuring that the density of the running variable is continuous across the threshold. If we were to observe bunching around the threshold, that would suggest that a municipality can manipulate the amount of rainfall recorded at the weather station to ensure that it receives a payout. Such a scenario would violate the assumptions on which the research design rests. Although this scenario is unlikely given that weather stations are run by a separate government entity and the information recorded is used for a variety of civilian and military purposes, we can test this assumption using the McCrary (2008) sorting test. Figure 1.5 is a visual representation of this test. The solid line is the estimated density of the running variable using local linear regression. The dashed lines represent confidence intervals for the estimated density. The overlapping confidence intervals imply that there is no discontinuous change in the density across the threshold. Formally, the p-value for the McCrary test is 0.49, so we fail to reject the null hypothesis that the density of the running variable is smooth across the threshold.

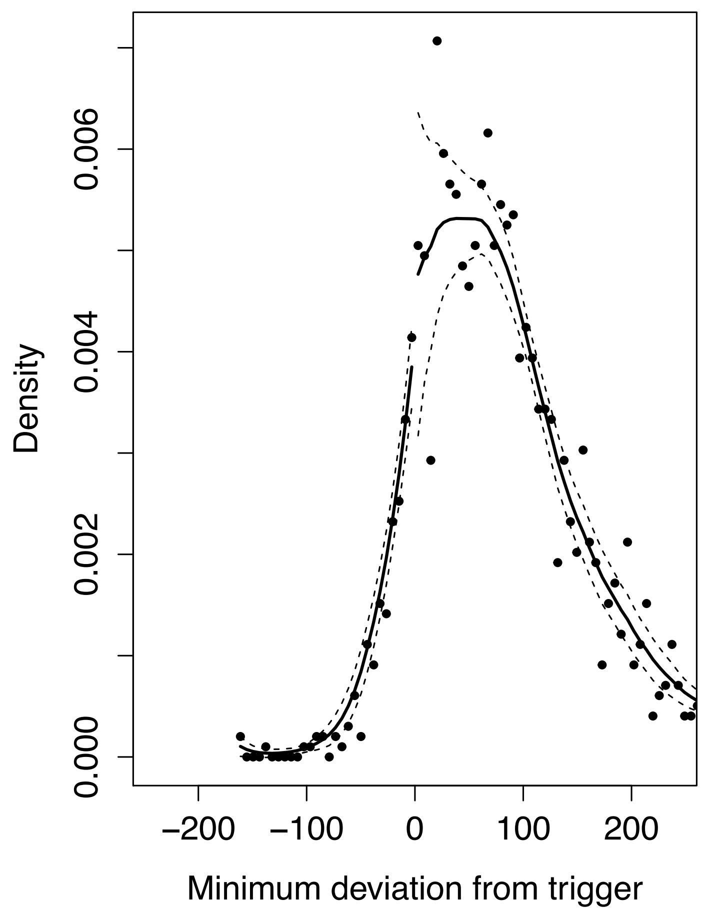
Another test of the identifying assumptions can be carried out using pre-CADENA data. Before CADENA is in place, we should not see any effect of crossing the insurance policy threshold on outcomes in the following season because no insurance payments would be made. Table 1.6 shows the results of this exercise. None of the different specifications show significant results for either outcome, and the point estimates are generally much smaller than what was obtained when estimating the same equation on the analysis sample. Overall, these tests suggest that our results are not driven by omitted variable bias.

Table 1.6: Placebo: Agricultural outcomes Pre-CADENA

	(1)	(2)	(3)	(4)	(5)	(6)
	Δ log ha planted				log yield	
Below trigger	-0.000422	0.00206	0.00999	-0.00132	-0.0123	0.0124
	(0.0494)	(0.0499)	(0.0299)	(0.0642)	(0.0635)	(0.0530)
Bandwidth	50	50	70	50	50	70
Controls	N	Y	N	Y	N	Y
N	3047	3047	4425	3047	3047	4425

Note. Standard errors are clustered at the state level. Asterisks indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Observations are at the municipality-crop-year-level and are restricted to pre-CADENA years for each municipality. Agricultural data are for the corresponding insurable crop (maize, beans, sorghum, and barley). All regressions include crop fixed effects. Columns 1 and 4 are estimated with a linear polynomial on a sample of observations within the optimal bandwidth of 50 mm. Columns 2 and 5 include state fixed effects and municipality characteristics (e.g., population, % indigenous from the 2010 census). Columns 3 and 6 are estimated using a linear polynomial on a sample of observations within an alternative bandwidth of 70 mm.

Figure 1.5: Density of running variable across threshold



Note. The points are binned values of the running variable. The black line represents a local linear smoother, estimated separately to the left and right of the threshold, while the dashed lines are the 5% confidence interval.

1.7 Conclusion

This paper contributes to our understanding of the use of index insurance as a social safety net for smallholder farmers, and more broadly the literature on policy responses to climate change, by exploring the effects of payments from weather index insurance on ex-post investment decisions and coping mechanisms. This analysis is valuable because it is carried out in the context of a large weather index insurance program with almost national coverage. In contrast to much of the existing evidence on index insurance, we are also able to observe effects over several years after the start of coverage. There exist some data limitations that negatively impact the precision and robustness of our results. Nevertheless, we believe that this analysis provides evidence that index insurance has the capacity of improving the welfare of rural households by providing them with resources to invest in the subsequent planting season, which in turn results in more land planted than would be the case without insurance payments. The payouts may also prevent households from resorting to costly coping mechanisms, such as reducing consumption, as evidenced by the result for household expenditures. Lastly, there appears to be some interaction between formal insurance payments and remittances that reduces the burden of private transfers for relief assumed by migrants. The potential benefits of index insurance are not only in coping with shocks but also in better managing risk. Because program implementation did not meet the conditions for a natural experiment, we were unable to measure these potential benefits. All in all, our results suggest that index insurance used as a social safety net may be a valuable policy tool to aid smallholder farmers who potentially face increased weather risk due to a changing climate.

Chapter 2

School Lunch Quality and Academic Performance[1]

2.1 Introduction

Improving the nutritional content of public school meals in the United States (US) is a topic of intense policy interest (Confessore 2014). A primary motivation underlying these nutritional improvements is to increase student health and reduce childhood obesity rates. A question of comparable import, however, is whether healthier meals affect student achievement. Recent research demonstrates that the provision of subsidized school meals can significantly increase school test scores (Figlio and Winicki 2005; Dotter 2014; Imberman and Kugler 2014b; Frisvold 2015), but to date little evidence exists on how the *quality* of school meals affects student achievement. This question is particularly important in light of recent arguments by policymakers that improved nutritional standards for school meals are ineffective or counterproductive (Green and Davis 2017).

To determine whether the quality of school meals affects student achievement, we exploit longitudinal variation in California school districts' meal vendors and estimate difference-in-differences type regressions. We combine two principal data sets from the California Department of Education, one covering breakfast and lunch vendors at the school level and the other containing school-by-grade-level standardized test results. Our five-year panel dataset includes all CA public elementary, middle, and high schools with non-missing state test score data (about 9,700 schools). For each California public school, we observe whether the district in which the school is located had an outside contract with a meal provider for the school year, and, if so, the name of the provider and the type of contract. The vast majority of schools provide meals using "in-house" staff, but a significant and growing fraction (approximately 12%) contract with outside vendors to provide meals. Crucially for our research design, there is substantial turnover in vendors at the school-district level during

[1]An earlier version of the material in this chapter was made available as NBER Working Paper No. 23218. This version can be found online at http://www.nber.org/papers/w23218

our sample period. Among schools in our panel that contract with an outside vendor, 62% switch between preparing meals in-house and contracting with a vendor.

A central obstacle in estimating the effects of healthy meal vendors on academic performance is accurate measurement of nutritional quality. We measure the nutritional quality of vendor school lunches using a modified version of the Healthy Eating Index (HEI). The HEI is a continuous score ranging from 0 to 100 that uses a well-established food-component analysis to determine how well food offerings (or diets) match the Dietary Guidelines for Americans (e.g., Guenther et al. 2013a). HEI is the measure of diet quality preferred by the United States Department of Agriculture (USDA) (USDA 2006) and has previously been used by researchers to evaluate menus at fast-food restaurants and child-care centers. We contracted with trained nutritionists at the Nutrition Policy Institute to calculate vendor HEI scores for this project.[2] Using their scores, we classify a vendor as healthy if its HEI score is above the median score among all vendors in our sample and as standard otherwise.

We find that contracting with a healthy meal vendor increases test scores by 0.03 to 0.04 standard deviations relative to in-school meal provision, after conditioning on school-by-grade and year fixed effects. This result is highly significant and robust to the inclusion or exclusion of our time-varying covariates, including demographic characteristics of the students, school district expenditures, student-teacher ratios, and changes in school leadership. The point estimates are also very similar whether they are estimated on the baseline sample of all CA schools or samples restricted to those schools that ever contract with an outside vendor. When estimating effects separately for economically disadvantaged and non-disadvantaged students, we find modest evidence that the effect of contracting with a healthy vendor is larger for economically disadvantaged students than for non-disadvantaged students. There is no evidence that contracting with a standard vendor affects test scores.

We conduct various tests to support the identifying assumption that the exact timing of vendor contracts is uncorrelated with other time-varying factors that may affect test scores. The frequent turnover in lunch vendor contracts makes it less likely that an unobservable factor could explain our test score results, as this factor would need to be highly correlated with with the timing of new contracts for healthy lunch vendors but not standard vendors. Event study specifications and "placebo" tests where the treatment activates one year prior to the actual treatment year provide evidence that test scores are not correlated with *future* changes in vendors (i.e., there are no differential trends preceding a new vendor contract). We also find that changes in observable characteristics of schools are uncorrelated with new vendor contracts. In particular, there is no evidence that changes in test scores predict when a school will contract with a vendor.

Introducing healthier school lunches does not appear to change the *number* of school lunches sold, which supports our interpretation that the change in test scores is due to the quality rather than the quantity of the food. At the same time, this result helps to alleviate concerns that offering healthier lunches could lead to lower consumption by economically disadvantaged students who qualify for free or reduced price school lunch. Similarly, we do

[2] http://npi.ucanr.edu/About_Us/

not find that healthier school lunches lead to a decrease in obesity rates. One explanation for the null effect on the percent of overweight students is that all school lunches – healthy vendor, standard vendor, and in-house – are subject to the same USDA calorie requirements.

Although our estimated test score effects are modest on an absolute scale, they are highly cost-effective for a human capital investment. We calculate a plausible upper bound on the cost of contracting with a healthy lunch provider, relative to in-house meal preparation, of approximately $85 (2013 $) per test-taker per school year. Using our preferred estimate of 0.031 standard deviations, this result implies that it costs (at most) $27 per year to raise a student's test score by 0.01 standard deviations. Despite assuming high costs, the cost effectiveness of contracting with healthy vendors matches the most cost-effective policies highlighted by Jacob and Rockoff (2011), and it compares very favorably when measured against interventions that achieve larger absolute effects, such as the Tennessee STAR class-size reduction experiment (Krueger 1999).

2.2 Background and Data

2.2.1 Related Literature

There is a large medical and nutrition literature examining the link between diet and cognitive development, and between diet and cognitive function (e.g., Bryan et al. 2004; Sorhaindo and Feinstein 2006; Gomez-Pinilla 2008; Nandi et al. 2015). Sorhaindo and Feinstein (2006) review existing research on the link between child nutrition and academic achievement and highlight how nutrition can affect learning through three channels: physical development (e.g., sight), cognition (e.g., concentration, memory), and behavior (e.g., hyperactivity). Gomez-Pinilla (2008) outlines some of the biological mechanisms regarding how both an increase in calories and an improvement in diet quality and nutrient composition can affect cognition. For example, "diets that are high in saturated fat are becoming notorious for reducing molecular substrates that support cognitive processing and increasing the risk of neurological dysfunction in both humans and animals" (Gomez-Pinilla 2008, p. 569). Most of the direct evidence on how nutrition affects academic achievement among school-age children comes from studies of children in developing countries (Alderman et al. 2007 and Glewwe and Miguel 2008 provide reviews).

A number of recent studies have estimated the effect of increased availability of either breakfast or lunch under the NSLP on student test scores in the US. Many of these studies find evidence that improved access to breakfast or lunch increased test scores (e.g., Figlio and Winicki 2005; Dotter 2014; Imberman and Kugler 2014b; Frisvold 2015), while others find no effect (e.g., Leos-Urbel et al. 2013; Schanzenbach and Zaki 2014). In all of these studies, the main hypothesized channel between the increased take-up of the school breakfast and lunch programs and test scores is an increase in calories consumed. The NSLP may also have broadly increased educational attainment by inducing children to attend school (Hinrichs 2010).

Our paper focuses on the nutritional quality of the calories provided. We are aware of just one other study that estimates the effect of food quality on academic test scores. Belot and James (2011) estimate the effect of introducing a new, healthier school lunch menu in 80 schools during the same academic year in one borough in London, as compared to schools in a neighboring borough. The authors estimate a positive effect on test scores for elementary school students, but find that the effect is larger for higher socioeconomic students who do not qualify for reduced price or free school lunch.

Relative to Belot and James (2011), we provide evidence from a much larger sample that includes all CA public schools (roughly 9,700 schools), of which 1,188 contract with an outside lunch provider. Our estimation approach uses within-grade and school variation in the introduction *and* removal of healthy and standard lunch providers that occurs in each of the five years of our panel. Thus, we can account for constant unobserved grade-by-school effects. Further, the staggered timing of the lunch contracts allows us to flexibly control for unobserved (calendar) time effects, and to conduct a series of robustness checks regarding the exogeneity of the timing of the contracts. Finally, like Belot and James (2011), we estimate the effect of healthier school lunches on the number of lunches served; however, unlike Belot and James (2011), we are also able to test whether healthier lunch provision changes obesity rates.

2.2.2 Data Sources

The data for this project come from the State of California Department of Education. We use information on school-level breakfast and lunch vendors, and school-by-grade-level standardized test results. We describe each type of information in detail below.

Vendor Data

The vendor meal contract information is provided by the California Department of Education for the school years 2008-2009 to 2012-2013.[3] All food vendor contracts with public (K-12) schools in California must be approved by the CA Department of Education. The CA Department of Education retains a list of the schools that contract with an outside meal provider for each school year, the name of the provider, and the type of contract. A total of 143 school districts covering 1,188 schools contracted with a total of 45 different vendors during our sample period. We merged the food vendor contract information with the list of all public schools (including charter schools) operating in CA during this time period to create our estimation panel. Overall, 12% of CA public schools contracted for at least one academic year with an outside company to provide school lunch.

Appendix Table A.2 lists the 45 vendors and the percent of students served by each vendor (conditional on being served by any vendor). For each vendor, we first calculate

[3]The data were received as part of an official information request. We thank Rochelle Crossen for her assistance in facilitating the request and in interpreting the data. Contract information for school years prior to 2008-2009 was not retained when the CA Department of Education switched computer database systems.

the number of (STAR) test-takers in districts that are being served by that vendor. This vendor-level total is then divided by the total number of test-takers being served by outside vendors. A single vendor serves just over 50% of the students. Altogether, the vendors with the ten largest student test-taker market shares serve 97.5% of CA students enrolled in schools that contract with outside school lunch providers.

Nearly all of the contracts (94%) are signed in the summer and cover the entire academic school year.[4] The CA Department of Education classifies all food provision contracts as one of four types: Vendor, Food Service Management Company (FSMC), Food Service Consulting Company (FSCC), and School Food Authority (SFA). A Vendor contract is when a school contracts with a private company to provide meals, but school employees (i.e., cafeteria staff) still handle and serve the food, including any additional prepping and cooking. In a FSMC contract, a private company prepares the meals and assists in staffing the school with cafeteria workers who serve the meals. In a FSCC contract, a private company provides "consulting services" on meal preparation and staffing, but does not provide any personnel for the jobs. SFA contracts usually denote that one public school district contracts with another district for meal provision. SFA contracts are unusual and account for just 1% of the contract-grade years. We do not distinguish between the four types of contracts in the main analysis of the paper and, unless otherwise specified, we refer to all such companies as "vendors."

Detailed vendor contract information is available for a subset of the contracts. Contract details include meals provided (either lunch or both breakfast and lunch), the dollar value of the contract, the number of other contract bidders (if any), the names of the companies which bid for the contract and were not selected, the dollar value of losing contracts, and the method by which the contract bids were solicited (i.e., sealed bid or negotiation). In the main analysis, we do not distinguish between vendors that provide both lunch and breakfast and those that provide only lunch, as this information is only available for a minority of the contracts.[5] We use the contract bid information to help construct counterfactual estimates for the cost to improve state test scores by contracting with healthy lunch providers.

The nutritional quality of the vendor school lunches is assessed using the Healthy Eating Index (HEI). The HEI is the US Department of Agriculture's (USDA) preferred measure of diet quality (USDA 2006), and the USDA uses it to "examine relationships between diet and health-related outcomes, and to assess the quality of food assistance packages, menus, and the US food supply" (USDA 2016). HEI has been used by researchers to assess both individual

[4]A small number of contracts cover less than the complete school year. These contracts correspond to the calendar year and thus cover only a fraction of the school year (August-December or January-June). Estimation results are insensitive to the inclusion of these contracts in our sample.

[5]The contract details are not available for all contracts for two reasons. First, school districts are only required to provide contract details to the state for the first year of a new contract. A contract can be renewed up to four times without having to issue a new contract. Second, school officials enter the contract information via a software program that electronically stores the data in the CA Department of Education database. In practice, many of the data fields are missing for most of the new contracts. This is because, until recently, the CA Department of Education didn't have the staff to review the contract price and bid data entered into the system.

diets (e.g., Volpe and Okrent 2012; Guenther et al. 2013b) and the diets of subpopulations (e.g., Hurley et al. 2009; Manios et al. 2009), as well as food offerings at fast food restaurants (e.g., Reedy et al. 2010) and child-care centers (e.g., Erinosho et al. 2013). The HEI scores range from 0 to 100, with higher scores representing healthier diets (or food offerings). Scores are calculated via a food component analysis done on a per calorie basis (Guenther et al. 2013a).

We contracted with nutritionists at the Nutrition Policy Institute to calculate vendor HEI scores using sample school lunch menus. Over the course of one year they collected detailed vendor data, refined the HEI score methodology, and computed the scores. Appendix A.2.1 provides details of the HEI score calculations, and a link to the Nutrition Policy Institute report includes examples of menus used as part of the analysis.[6] Menu information was not available for all of the vendors, and as a result some vendors were not assigned HEI scores. Appendix Table A.2 shows that this is mostly the case for vendors that contract infrequently with schools. Overall, HEI scores are calculated for 87.3% of student test-takers served under vendor contracts. The median vendor HEI score in our sample is 59.9. This median vendor score is similar to the average HEI score, 63.8, for the US population age two and older (USDA 2006, p. 21). We define a vendor as healthy if it has a vendor HEI score above the median vendor score. Healthy vendors are more likely to provide salad bars and sufficient amounts of fruits, vegetables, whole grains, dairy, and seafood or plant proteins. They serve fewer processed meats, fast-food items, fried potatoes, chocolate milk, sweets, and chips or other salty snacks, and their meals tend to contain less refined grains, sodium, and "empty calories."[7] Alternative classifications (e.g., coding any vendor with an HEI score above the mean vendor score as healthy) generate similar results. While no classification system is perfect, it is notable that misclassifying healthy vendors as standard, and vice versa, will only attenuate our estimates.

Academic Test Data and Covariates

To measure academic achievement, we use California's Standardized Testing and Reporting (STAR) test data. The STAR test is administered to all students in grades 2 through 11 each spring, toward the end of the academic year. The publicly available test scores are aggregated at the grade-by-school level. We use test score data from 1998 through 2013. Beginning with the 2013-14 school year, STAR testing was replaced with the California Assessment of Student Performance and Progress test.

The STAR test includes four core subject area tests (English/Language Arts, Mathematics, History/Social Sciences, and Science) and a set of end-of-course examinations (e.g.,

[6]In preparing their analysis, the nutritionists assumed that all vendors met the baseline USDA requirements, as they are obligated by law to do so. They also assumed that the average meal contains 650 calories, and they matched food items to foods available in the USDA food database. The classification of vendors as healthy or standard was not sensitive to any of these choices.

[7]Fast-food items include chicken nuggets, pizza or pizza pockets, hamburgers, fried chicken, nachos, hot dogs, and corn dogs. "Empty calories" are those that come from solid fats, alcohol, and added sugars.

Algebra II, Biology). We create a composite test score each year for each school-grade by calculating the average test score across all of the STAR subjects and the end-of-course tests taken by students in a particular grade in each school. We use the standard deviation of each test (which differs by grade and year of test) to standardize each subject and end-of-course test score before combining the scores into a single composite score.[8]

Average test scores are also available separately for students who qualify for reduced price and/or free school lunch under the NSLP. A student is eligible for a free school lunch if his family's income is less than 130% of the poverty level, and a reduced price lunch if her family's income is between 130% and 185% of the poverty level.[9] The CA Department of Education refers to these students, as well as students with parents who do not have high school diplomas, as "economically disadvantaged" (California Department of Education 2011, p. 48). Students eligible for the reduced price or free lunches are more likely to eat the lunch offered at the school, for two reasons: the price is lower for them than it is for ineligible students, and eligible students are less likely to have other lunch options. Furthermore, the nutritional quality of their home-provided meals may be lower than that of the average student. Thus, we hypothesize that the academic benefit of having healthier school lunches may be larger for these students.

Finally, district-level demographic and socioeconomic information is available from the California Department of Education, including enrollment by race, enrollment in English learner programs (i.e., English as a second language), and the number of enrolled students who are economically disadvantaged, as defined by eligibility for free or reduced price lunches. We use this information to control for time-varying differences within schools in our main econometric model.

2.3 Empirical Strategy

Our main empirical specification is a panel regression model.

$$y_{gst} = \beta_0 + \delta_H Healthy_{st} + \delta_S Standard_{st} + X_{st}\beta + \lambda_{gs} + \gamma_t + \epsilon_{gst} \tag{2.1}$$

The dependent variable y_{gst} is the mean STAR test score across all tests for grade g in school s in year t. The dependent variable is measured in STAR test standard deviation units.

Our independent variables of interest are indicators for whether a student test-taker is exposed to a standard or healthy outside lunch provider. Recall that a provider is classified as

[8] The qualitative results are robust to using only core test results, or in using just the English/Language Arts exam (which is the only exam taken by students in each grade). However, the point estimates are lower in specifications that only use test results from the English/Language Arts exam. This is consistent with other recent studies that separately measure the effect of access to school breakfast on test scores in different subjects (e.g., Dotter 2014; Imberman and Kugler 2014b).

[9] Note that eligibility does not imply participation. Dahl and Scholz (2011) estimate that 28% to 49% of eligible children participate in free or reduced-price school breakfasts, and 63% to 73% of eligible children participate in free or reduced-price school lunches. Even estimates for disadvantaged children thus represent "intent to treat" effects rather than full treatment effects.

healthy if its HEI score is above the median score among providers. The variable $Healthy_{st}$ equals one if school s contracts with a healthy outside lunch provider in year t and zero otherwise. The variable $Standard_{st}$ equals one if school s contracts with a standard outside lunch provider in year t and zero otherwise. The "omitted" category for our treatment indicators corresponds to the case in which the school does not contract with an outside lunch provider. In this case the school's employees (i.e., cafeteria workers) both prepare and serve the lunches.

The model includes school-by-grade (λ_{gs}) and year (γ_t) fixed effects. The school-by-grade fixed effects control for any characteristics in a given grade and school that are stable throughout the five-year estimation period (e.g., school catchment area characteristics, school infrastructure, STAR test differences by grade, or school staffing levels and leadership). Year fixed effects control for common state-wide factors such as state economic conditions and differences in the STAR test that vary by year throughout the panel. All specifications also include an indicator for contracting with a meal vendor of unknown HEI quality, $Unscored_{st}$. This category accounts for 13% of vendors on a test-taker weighted basis. In our regressions the coefficient on $Unscored_{st}$ lies, as we would expect, between the coefficients on $Healthy_{st}$ and $Standard_{st}$ and sometimes achieves statistical significance. However, we cannot interpret it since we do not know the fraction of unscored vendors that are healthy, so we treat it as an additional control.

Most specifications of the model also include X_{st}, a vector of district-level control variables that vary over time. These control variables include the racial composition of students in the district to which school s belongs, the proportion of students in English learner programs, and the proportion of economically disadvantaged students. Because the decision to contract with a lunch vendor (whether healthy or standard) almost always occurs at the district level, as opposed to the individual school level, it is sufficient to control for district-level covariates that may be correlated with this decision.[10]

Contract typically cover all schools in a district, so we estimate Equation (2.1) with standard errors clustered at the school-district level. Our preferred specification uses the number of test-takers for each grade-school-year observation as weights in the regression. Weighting by the number of test-takers allows us to recover the relationship between the type of school lunch served and academic performance as measured by the STAR test for the average student, rather than the average school.

The identifying assumption is that, after controlling for time-invariant school-by-grade factors, common state factors, and the vector of time-varying, school-level characteristics, a school's decision to contract with an outside vendor for school lunch provision is uncorrelated with other school-specific, time-varying factors that affect student test performance. If this is true, then we can interpret the estimate for δ_H (and δ_S) as the causal effect of contracting

[10]We also experimented with controlling for similar school-level covariates that we constructed directly from the STAR data. Controlling for these covariates at the school level has little impact on the coefficient estimates, but it results in many dropped observations because of frequent missing demographic information in the STAR data.

with a healthy (or standard) school lunch provider on student learning, as measured by performance on the STAR test.

2.4 Results

2.4.1 Vendor Choice and Test-Taker Characteristics

Appendix Table A.3 shows mean test-taker socioeconomic and racial characteristics for school districts in two different samples: the *All School* sample and the *Contract School* sample. The All School sample includes all school districts in the state of California. The Contract School sample is limited to the subset of districts that had a school lunch vendor contract for at least one year in our five-year panel. The means for each test-taker characteristic are calculated by first taking the five-year (2009-2013) district-level mean. In the All School sample, the average district mean is then calculated separately for districts that contract with a vendor (Column 1) and do not contract with a vendor (Column 2) during our panel (2009-2013). Column (3) calculates the difference in means and reports the standard error for this difference (in parentheses). The means are statistically different from each other at the 5% level for five of the six characteristics. For example, districts that contract with a vendor during our sample period tend to have fewer economically disadvantaged students and a higher proportion of Asian students.

Appendix Table A.3 also shows that, even among districts that contracted with a vendor at some time, those districts that contracted with a healthy vendor have different student characteristics (on average) than those districts that contracted with a standard vendor. These differences in test-taker characteristics in the two samples affect the generalizability of any association between test scores and vendor quality. Nevertheless, the differences in average characteristics between test-takers do not violate the identification assumptions of Equation (2.1).

Table 2.1 shows how *changes* in the test-taker characteristics correlate with the timing of a vendor contract. We cannot interpret an observed correlation between vendor adoption and test score changes as a causal effect if changes in test-taker characteristics at a school can predict when a school contracts with an outside vendor. Table 2.1 displays the coefficient estimates from 12 different regressions using a version of Equation (2.1). In each of the first five columns, we use a different test-taker characteristic as the dependent variable in place of test scores. In the last column, we use the fitted values from a regression of test scores on all five test-taker characteristics (and year and school-by-grade fixed effects) as the dependent variable. These fitted values summarize all of the test-taker characteristics, weighting each characteristic in relation to its correlation with test scores. All regressions in Table 2.1 include school-by-grade fixed effects and thus test whether within-school-by-grade changes in student characteristics correlate with the time at which a school adopted an outside lunch provider.

Table 2.1: Correlation: Test-Taker Covariates and the Timing of Vendor Contracts

Dependent variable:	(1) White	(2) Asian	(3) Hispanic	(4) Disadvantaged	(5) English Learner	(6) Predicted Test Score
Panel A. All School Sample						
Healthy Vendor	0.004	-0.006	0.004	0.000	-0.016	0.005
	(0.005)	(0.004)	(0.003)	(0.015)	(0.018)	(0.010)
Standard Vendor	-0.001	-0.001	-0.001	0.027	0.006	-0.024
	(0.006)	(0.002)	(0.005)	(0.023)	(0.008)	(0.019)
N	174,818	174,818	174,818	174,818	174,818	174,818
Panel B. Contract School Sample						
Healthy Vendor	0.001	-0.005	0.006	-0.024	-0.013	0.018
	(0.007)	(0.004)	(0.005)	(0.019)	(0.015)	(0.012)
Standard Vendor	-0.003	0.000	0.002	0.005	0.003	-0.006
	(0.007)	(0.004)	(0.005)	(0.022)	(0.007)	(0.018)
N	22,133	22,133	22,133	22,133	22,133	22,133

Note. Each column represents a separate weighted regression with weights equal to the number of test takers per observation. Observations are at the school-grade-year level. Standard errors clustered at the school district level appear in parentheses. All regressions include year and school-by-grade fixed effects. The contract school sample is the subset of schools that contract with a vendor at some point during our sample. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel A of Table 2.1 estimates models using the All School sample,[11] while Panel B uses the Contract School sample.[12] The point estimates are small in magnitude and precisely estimated. None of the estimated coefficients are statistically significant at conventional levels. The estimate in the last column of Panel A reveals that adoption of a healthy vendor correlates with a statistically insignificant 0.005 standard deviation increase in predicted test scores. The estimate for adoption of a standard vendor is also statistically insignificant. We interpret these results as initial evidence that changes in test-taker characteristics are uncorrelated with the timing of when a school contracts with a lunch provider. Sections 2.4.3 and 2.4.4 consider additional tests of the validity of our identifying assumption.

2.4.2 Vendor Choice and Test Scores

Table 2.2 shows estimation results for the effect of vendor quality on STAR scores. The first three columns estimate versions of Equation (2.1) on the All School sample, while the last three columns use the Contract School sample. Column (1) estimates the effect of contracting

[11] This includes all elementary, middle, and high schools in California that report STAR scores.

[12] This comprises all schools located in districts that had a school lunch vendor contract for at least one year in our five year panel.

with a standard or healthy lunch vendor on test scores and includes school and year fixed effects as controls. Column (2) adds school-by-grade fixed effects, while Column (3) adds the vector of student test-taker characteristics. The point estimate of the effect of having a healthy vendor on test scores, relative to no outside vendor, is 0.031 standard deviations and is highly significant at the 0.1% level in each of the three specifications. Moreover, we are able to reject the null hypothesis that the coefficients for the healthy and standard vendors are equal at the 10% level for each specification. The standard vendor coefficient is an order of magnitude smaller and not statistically different from zero in any specification.

Table 2.2: Vendor Choice and Standardized Test Scores

Dependent variable:	Standardized Test Score					
Sample:	All Schools			Contract Schools		
	(1)	(2)	(3)	(4)	(5)	(6)
Healthy Vendor	0.031***	0.031***	0.031***	0.037***	0.037***	0.033***
	(0.009)	(0.010)	(0.010)	(0.008)	(0.009)	(0.009)
Standard Vendor	-0.004	-0.005	-0.004	0.008	0.007	0.008
	(0.017)	(0.019)	(0.019)	(0.017)	(0.019)	(0.019)
School-by-grade FEs		X	X		X	X
Covariates included			X			X
R^2	0.714	0.930	0.930	0.700	0.922	0.922
N	174,818	174,818	174,818	22,133	22,133	22,133
Schools	9,719	9,719	9,719	1,188	1,188	1,188

Note. Each column represents a separate weighted regression with weights equal to the number of test takers per observation. All regressions are estimated on a common sample that excludes observations with missing covariates. Observations are at the school-grade-year level. Standard errors clustered at the school district level appear in parentheses. All regressions include year fixed effects and school fixed effects (unless school-by- grade fixed effects are specified). The contract school sample is the subset of schools that contract with a vendor at some point during our sample. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The estimates for a healthy vendor from the Contract School sample are also statistically significant at the 0.1% level and are similar in magnitude to those estimated with the All School sample (ranging from 0.033 to 0.037). The estimates for the standard vendor are again an order of magnitude smaller and not statistically significant.[13]

The fact that we observe very similar point estimates for the vendor coefficients in Columns (2) and (3) (and Columns (5) and (6)) is consistent with the conclusion from Table 2.1. If student characteristics were important in predicting when a school contracts with an outside vendor, then the coefficients in Table 2.1 would be statistically significant and the vendor estimates in Table 2.2 would likely be sensitive to the inclusion or exclusion of these controls.

[13]Our results are also qualitatively similar if we estimate Equation (2.1) without using student enrollment weights. These results can be found in Appendix Table A.4. The point estimate for those vendors with an unknown HEI score (not shown in table) are positive, range between 0.01 and 0.02 standard deviations, and are statistically significant in some specifications. This is not too surprising as we xpect these vendors to be a mix of standard and healthy vendors.

Table 2.3 investigates whether the effect of contracting with a lunch provider on STAR scores is different for economically disadvantaged and economically advantaged students. Recall that economically disadvantaged students are defined by the CA Department of Education as those students who qualify for reduced price and/or free school lunch under the NSLP based on family income. We expect that disadvantaged students would be more likely to eat a school lunch than their classmates who do not qualify for reduced price or free school lunch. Thus, we hypothesize that the effect on test scores of healthy school lunch vendors should be somewhat greater for disadvantaged students than for students who do not qualify for reduced price or free school lunch. Table 2.3 shows evidence consistent with this hypothesis.

Table 2.3: Vendor Choice and Standardized Test Scores by Socioeconomic Status

Dependent variable:	Standardized Test Score					
Sample:	All Schools			Contract Schools		
	(1)	(2)	(3)	(4)	(5)	(6)
Subgroup:	Disadvantaged	Advantaged	All	Disadvantaged	Advantaged	All
Healthy Vendor	0.035*	0.025	0.028*	0.046**	0.030*	0.035**
	(0.014)	(0.015)	(0.014)	(0.016)	(0.014)	(0.012)
Standard Vendor	0.004	-0.007	0.000	0.020	0.009	0.016
	(0.018)	(0.025)	(0.020)	(0.021)	(0.027)	(0.022)
R^2	0.897	0.906	0.928	0.880	0.902	0.919
N	103,432	103,432	103,432	13,259	13,259	13,259
Schools	7,607	7,607	7,607	940	940	940

Note. Each column represents a separate weighted regression with weights equal to the number of test takers in the subgroup indicated by the column name. All regressions are estimated on a common sample that excludes observations with missing test score data for any of the indicated subgroups. Observations are at the school-grade-year level. Standard errors clustered at the school district level appear in parentheses. All regressions include year and school-by-grade fixed effects and district-level demographic covariates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.3 again considers both the All School and Contract School samples, but limits the samples to those schools which report separate average STAR scores for both economically advantaged and economically disadvantaged students.[14] Column (1) of Table 2.3 estimates the effect of contracting with a lunch vendor on the average test score for economically disadvantaged students. Column (2) estimates the effect on the average test scores for economically advantaged students, while Column (3) estimates the effect for all students. The point estimates for contracting with a healthy vendor are about 40 to 50% larger for the disadvantaged students in both samples. In the All School sample the estimated coefficients are 0.035 and 0.025 respectively, while in the Contract School sample they are 0.046 and 0.030 respectively. There is again no evidence that a standard vendor has a statistically significant effect on test scores relative to having meals completely prepared by school staff.

[14]Due to privacy restrictions, the CA Department of Education releases the average test score (for a school-grade-year-subgroup) only if there are at least 10 students of the particular socioeconomic group who take the test. There is a 40% reduction in the size of the sample due to these sample restrictions.

2.4.3 Robustness Checks

Table 2.1 presents initial evidence that changes in test-taker characteristics are uncorrelated with the timing of when a school contracts with a lunch provider. In this section, we further test the validity of our identifying assumption that a school's decision to contract with an outside vendor for school lunch provision is uncorrelated with other school-specific, time-varying factors that affect student test performance.

Equation (2.2) is an event-study model that tests whether there is a correlation between test scores and contracting with a vendor in years before the vendor contract begins and in each year of the contract.

$$(2.2)\quad y_{gst} = \beta_0 + \sum_{\tau=-4}^{4} \delta_H^\tau Healthy_{st}^\tau + \sum_{\tau=-4}^{4} \delta_S^\tau Standard_{st}^\tau + X_{st}\beta + \lambda_{gs} + \gamma_t + \epsilon_{gst}$$

Equation (2.2) is identical to our main estimating equation, except that we replace the single indicator variables for whether a school contracted with a vendor ($Healthy_{st}$ and $Standard_{st}$) with a set of indicators ($Healthy_{st}^\tau$ and $Standard_{st}^\tau$). Indicators with $\tau < 0$ are indicators for the years before a contract with a healthy (or standard) vendor begins. Indicators with $\tau > 0$ are indicators for the consecutive years after a contract with a healthy (or standard) vendor begins, while the contract is still in effect.[15] The indicator variables for a year before a contract are normalized to zero when we estimate Equation (2.2). Thus, the estimated coefficients δ_H^τ and δ_S^τ are interpreted as the change in test scores for students in grade g, school s, and year t relative to the year before a contract.

Figure 2.1 plots the estimated healthy vendor event time coefficients and the 95% confidence intervals for the All School sample. The x-axis measures event time years (i.e., τ), and the y-axis measures test scores for all test takers.[16] In a healthy vendor contract year, there is an increase in test scores of 0.030 standard deviations relative to the year before a contract.[17] There is no evidence that increases in test scores precede contracting with a vendor, nor is there evidence of an upward pre-trend in test scores. Similarly, none of the estimated coefficients in the years after a contract begins are significantly different than the $\tau = 0$ coefficient, suggesting that there are not delayed impacts that take several years to materialize. Figure 2.2 plots analogous event time coefficients for standard vendors. None of the

[15] For example, $Healthy_{st}^{-3}$ equals unity if a school contracted with a healthy vendor three years later (and zero otherwise), and $Healthy_{st}^{3}$ equals unity if a school contracted with a healthy vendor three years earlier and has continued to contract with a healthy vendor each year since (and zero otherwise).

[16] The post-event coefficients in the figure compare the counterfactual of a contract that remains in effect indefinitely to one in which there is no contract; e.g., at $\tau = 2$ we plot $\delta_H^0 + \delta_H^2$. This aligns the figure with a standard event study figure in which the policy always remains in effect following the event.

[17] As a comparison, the estimate of the effect on test scores for the year of a healthy vendor contract from Equation (2.1) on the same sample is 0.031 (Column (3) of Table 2.2).

standard vendor coefficients, either before or after the contract's start date, are statistically significant.[18]

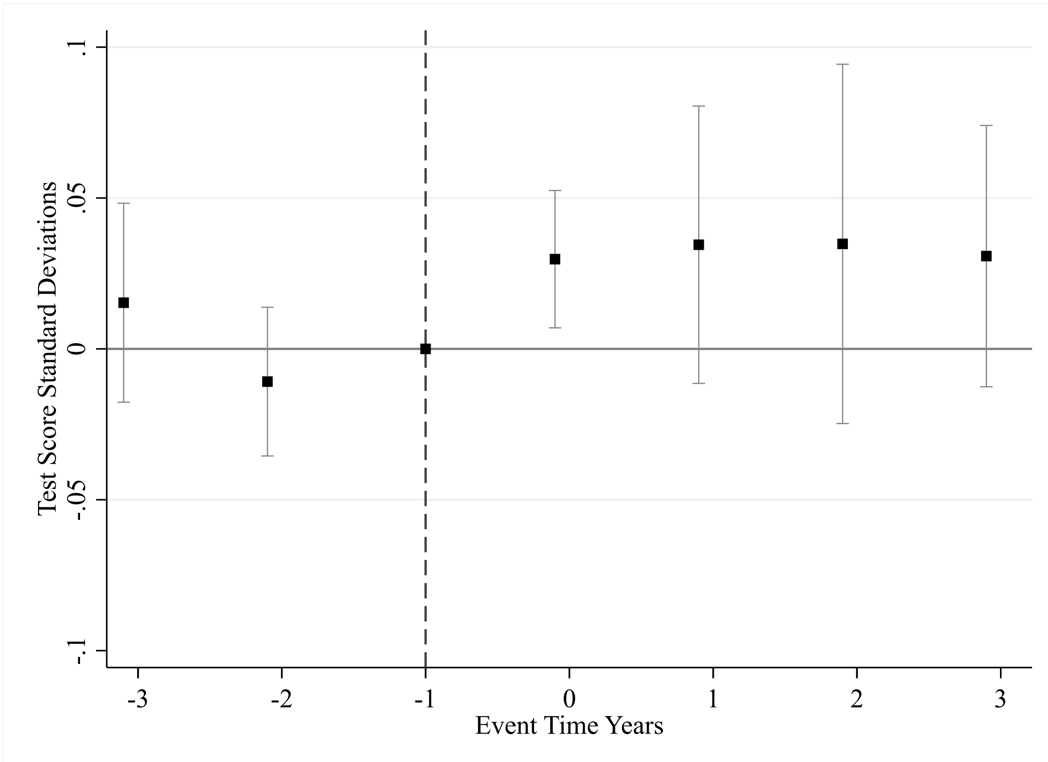
There are two caveats to the analysis in Figures 2.1 and 2.2. First, the event study coefficients toward the ends of our panel are imprecisely estimated because there are fewer observations available to identify these coefficients.[19] We address this concern by also estimating a model that pools the event time coefficients. Second, we do not know whether a school contracted with a vendor in the years before our five-year panel begins. This could attenuate our estimates if there are delayed impacts in the test score effect that appear several years after contracts begin, because the model would incorrectly assume zero treatment effect in the pre-vendor years.[20] We conclude that this is unlikely to be a concern, however, because our event study finds no evidence of delayed impacts.

[18]We also estimate a model, similar in spirit to Equation (2.2), that regresses an indicator for when a contract starts on lagged test scores. We run the model separately for healthy, standard, and unknown contracts, and include four lagged test score variables. This model tests whether changes in test scores predict vendor adoption, and it uses the full sample because test scores are available in years before 2008-2009 (the first year of our vendor data). In this model none of the lagged test score coefficients are statistically significant.

[19]For example, the indicator for four years before a vendor contract can equal one only if a school contracted with a vendor in the last year of our panel. By contrast, an indicator for one year after a vendor contract ends could equal one for four of the five years in our panel.

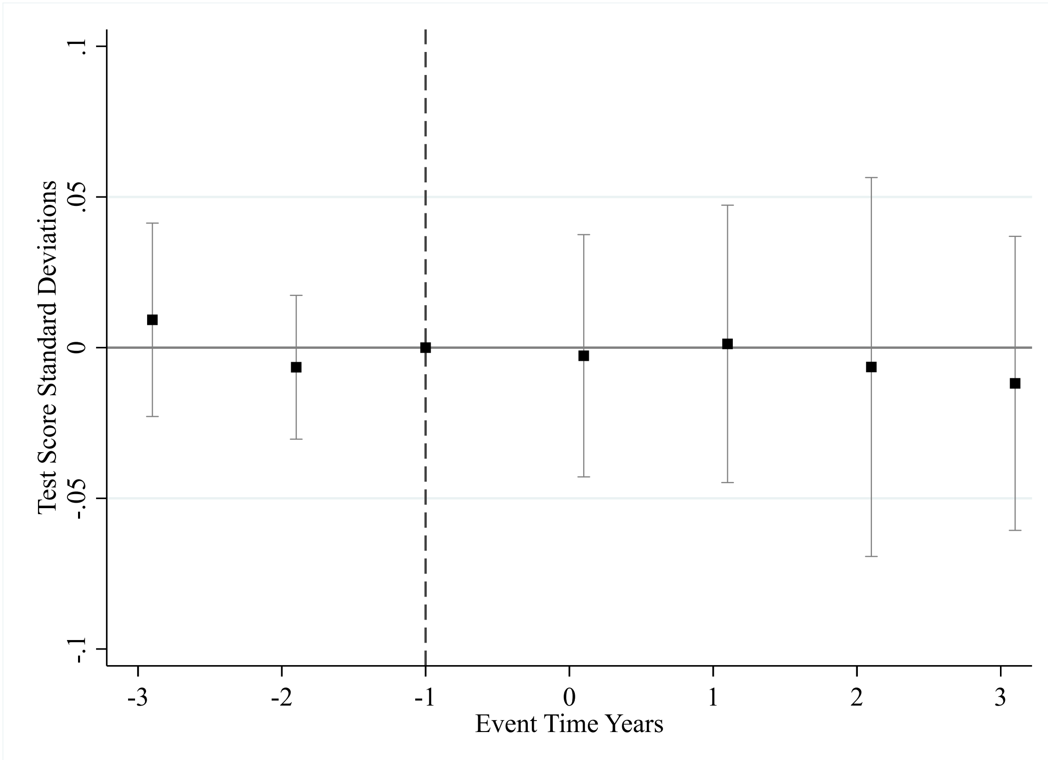
[20]For example, Gallagher (2014b) examines the effect on the take-up of flood insurance after a community is flooded, using a model similar to Equation (2.2). He shows that the estimate for flood insurance take-up in the year of a flood is about 20% lower if the model fails to control for the lagged effect of a flood that occurred before the panel period.

Figure 2.1: The Effect of Healthy Vendors on Test Scores



Note. This figure depicts point estimates for treatment leads and lags with their corresponding 95% confidence intervals. Negative event time years are for years prior to having a vendor. Positive event time years are for years after the vendor contract begins (assuming that it continues). The point estimates come from a weighted regression of Equation (2.2) using the number of test-takers per observation as weights and can be interpreted relative to having no vendor and relative to the year before a vendor. The event time period -3 (3) is pooled to include both period -3 and -4 (3 and 4) to improve precision. The regression includes the same control variables as Table 2.2, Column (3). Standard errors are clustered at the school district level.

Figure 2.2: The Effect of Standard Vendors on Test Scores



Notes: This figure depicts point estimates for treatment leads and lags with their corresponding 95% confidence intervals. Negative event time years are for years prior to having a vendor. Positive event time years are for years after the vendor contract begins (assuming that it continues). The point estimates come from a weighted regression of Equation (2.2) using the number of test-takers per observation as weights and can be interpreted relative to having no vendor and relative to the year before a vendor. The event time period -3 (3) is pooled to include both period -3 and -4 (3 and 4) to improve precision. The regression includes the same control variables as Table 2.2, Column (3). Standard errors are clustered at the school district level.

Table 2.4 presents the estimation results of six additional specifications that further test our identifying assumptions and the robustness of our main test score results. All six specifications can be interpreted relative to our baseline model in Table 2.2, Column (3). Column (1) presents results for a regression that estimates separate effects for vendors whose contracts run for the full academic year and those that end earlier. Students take STAR tests in April, and while most vendor contracts run though June, a small number end between November and February. If test scores increase solely because students have access to better food on test days, then we should expect no effect for contracts that end before the test date. The results in Column (1), however, reveal large and statistically significant effects for vendors whose contracts end several months before the test date. The coefficient for these contracts is larger than the coefficient for other healthy vendor contracts, but the difference is not statistically significant. These result suggest that healthy vendors may improve learning rather than simply improving performance on test days.

Columns (2) and (3) of Table 2.4 alter how we define a healthy vendor. Column (2) uses a second, slightly different, scoring method where the HEI score is supplemented by awarding additional points for healthy options that exceeded USDA requirements (e.g., salad bars) and subtracting points for unhealthy options (e.g., fast foods, certain processed foods, and high-sugar foods). The definition of a healthy vendor remains the same — one that receives a score above the median. The coefficient estimates for both the healthy and standard vendors are similar to those in the baseline model.

Column (3) uses the same HEI scores as the baseline model, but considers the scores as a continuous variable. Specifically, the model is adjusted to include an indicator variable for having a vendor with a known HEI score (rather than separate healthy and standard indicators) and an HEI score variable (divided by 100). The HEI score variable is zero if the school does not contract with a vendor that has a known HEI score, and we recenter the HEI score so that zero corresponds to the average HEI score for a standard vendor in our sample (56.0). With this recentering we may interpret the indicator for a vendor having a known HEI score as the average effect of a standard vendor. The estimated coefficient for the HEI score variable is positive and statistically significant. The difference in average HEI scores between healthy and standard vendors is 17.5, so the point estimate implies that a healthy vendor increases test scores by approximately 0.027 standard deviations relative to no vendor, or 0.022 standard deviations relative to a standard vendor. The highest and lowest rated vendors by HEI score are Revolution Foods (92.3) and Kid Chow (26.8). The model implies that the effects of contracting with each vendor on test scores are 0.050 and –0.032, respectively.

Column (4) reports a specification in which we aggregate the data to the district-by-year level, since the variation in vendor quality occurs almost exclusively at the district level. The estimated marginal effect of a healthy vendor on test scores is similar to that from our baseline specification.[21]

Column (5) of Table 2.4 considers a placebo test where we incorrectly consider the year before a vendor contract as the year of a contract (Currie et al. 2010). We define *healthy placebo* (*standard placebo*) as equal to one if the school contracts with a healthy (standard) vendor in the following year. The estimated coefficients for both vendor placebos are close to zero and not statistically different from zero after controlling for the actual vendor years. There is no evidence that test scores begin to rise in the year before a school contracts with a vendor.[22]

We also consider a second event-study style placebo test where we separately estimate the

[21]Note that the healthy and standard vendor variables are weighted averages of the school-by-grade level exposure to the vendors in each year and thus take on values between 0 and 1. The average value for the healthy vendor variable (conditional on having at least one school in the district that contracted with a healthy vendor for the year) is 0.95

[22]This specification also includes unknown vendor placebos. The estimated placebo coefficients are also close to zero and not statistically significant in a specification that does not condition on the actual vendor years.

Table 2.4: Vendor Choice and Standardized Test Scores: Robustness Checks

Healthy Vendor	0.031**	0.036***		0.037**	0.033**	0.031
	(0.010)	(0.009)		(0.013)	(0.012)	(0.028)
Standard Vendor	-0.004	-0.003		-0.010	-0.003	0.006
	(0.019)	(0.016)		(0.022)	(0.019)	(0.021)
Healthy Vendor with Contract Ending Before Test Date	0.063**					
	(0.018)					
Vendor with HEI Score			0.005			
			(0.014)			
HEI Score / 100			0.125*			
			(0.060)			
Healthy Vendor Placebo					0.009	
					(0.011)	
Standard Vendor Placebo					0.012	
					(0.010)	
R^2	0.930	0.930	0.930	0.985	0.930	0.931
N	174,818	174,818	174,818	4,057	174,818	12,117
Schools/Districts	9,719	9,719	9,719	908	9,719	649

Note. Each column represents a separate weighted regression with weights equal to the number of test takers per observation. In Column (1) the Healthy Vendor indicator equals zero when the Healthy Vendor with Contract Ending Before Test Date indicator equals unity. To construct the HEI Score regressor we subtract the mean standard vendor score (56) from HEI Score and rescale it by 1/100 (we also set the variable to zero for unscored vendors and schools with no vendors). Observations are at the school-grade-year level or, in Column (4), district-year level. Standard errors clustered at the school district level appear in parentheses. All regressions include year and school-by-grade (or, in Column (4), district) fixed effects and district-level demographic covariates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

correlation between test scores and panel years without a vendor. The econometric model is similar to Equation (2.2), except that we define $\tau > 0$ as years after *ending* a contract. In this specification, the $\tau = 0$ event study coefficient, $\delta_H^{\tau=0}$, is averaged across all years with a healthy vendor. If our model is correctly specified, we would only expect to measure a correlation between test scores and a vendor contract for healthy vendors when $\tau = 0$.[23] Appendix Figure A.1 shows that this is indeed the case. All of the estimates for the standard vendor coefficients are economically small and statistically insignificant. The same is true for the healthy vendor coefficients, except during years when a school contracts with a vendor ($\tau = 0$); in those years we estimate a statistically significant effect of 0.030.

Column (6) considers the sub-sample of schools from the Contract Sample that ever contracted with a standard vendor. As in the contract sample, this sample excludes schools that never had an outside vendor and further restricts the sample by excluding schools whose

[23] One potential exception is that we might expect the estimates for the healthy vendor to be positive when $\tau > 0$ if there is a "carry-over" effect on learning in years after the cancellation of a healthy vendor contract.

only outside vendors have been categorized as healthy (or unknown quality). Notably, the healthy vendor coefficient is estimated on a smaller sample of schools, only a fraction of which ever contract with both a standard and a healthy vendor. Nevertheless, the point estimate, while not statistically significant, is very similar to the estimate from the larger All School and Contract School samples. The similarity in the coefficient estimates across the three samples provides further evidence that our results are not driven by differential trends in test scores among the schools that contract with a healthy vendor.

2.4.4 Alternative Explanations

Table 2.5 tests several alternative explanations for the observed relationship between healthier school lunches and test scores. One concern that our robustness checks above do not necessarily test is that changes in vendor contracts may coincide with changes in school leadership or district-wide investments. To address this concern we test whether changes in school-district expenditures, student-to-teacher ratios, the percentage of students enrolled in charter schools, or school leadership can explain our results. To conduct these tests we merge additional data sources with our baseline panel. The new data sources, however, do not include information for all schools. Table 2.5 thus reports our vendor coefficient estimates when we restrict the estimation samples to observations with non-missing expenditure, student-teacher ratio, superintendent, or charter school enrollment data, and then shows how the vendor coefficient estimates change when we control for each factor.

Columns (1) and (2) consider a sample with non-missing information for school expenditures and student-teacher ratios. The estimated coefficients for healthy and standard lunch providers in Column (2) are virtually identical to those without the expenditure and student-teacher controls in Column (1). Neither the expenditure nor student-teacher coefficient is statistically significant, and in separate regressions we also find no statistically or economically significant relationship between either measure and healthy vendor adoption.[24]

Columns (3) and (4) consider a sample with non-missing information on changes in district superintendents.[25] Column (4) includes an indicator equal to unity in years when there is a new superintendent. The estimated coefficient for the new superintendent variable is economically small and statistically insignificant, and the vendor point estimates are unaffected by whether we control for having a new superintendent.

[24]The expenditure and student-teacher ratio data come from the CA Department of Education and are only available at the district level. We convert the expenditure data to thousands of dollars (2013 $) per average daily student attendance. Results are similar if we separately control for expenditures and student-teacher ratio.

[25]The CA Department of Education only retains information for current superintendents. We contacted each CA school district in our sample by email and phone between May and July 2017 to obtain historical information on whether there was a change of superintendent during our panel period. We successfully obtained information from 55% of the districts in our sample. Among these districts, 59% had at least one change of superintendent during our panel.

Table 2.5: Vendor Choice and Standardized Test Scores: Testing Alternative Explanations

Dependent Variable:	Standardized Test Score					
Sample:	School Expenditures		New Superintendent		Charter School Population	
	(1)	(2)	(3)	(4)	(5)	(6)
Healthy Vendor	0.028**	0.028**	0.031	0.031**	0.031***	0.031***
	(0.011)	(0.011)	(0.016)	(0.012)	(0.010)	(0.010)
Standard Vendor	-0.005	-0.004	-0.011	-0.011	-0.004	-0.004
	(0.019)	(0.018)	(0.014)	(0.027)	(0.019)	(0.019)
Expenditures		0.001				
		(0.001)				
Student-Teacher Ratio		-0.001				
		(0.001)				
New Superintendent				0.002		
				(0.003)		
Percent Charter School Enrolled						-0.036
						(0.062)
R^2	0.930	0.930	0.931	0.931	0.930	0.930
N	168,377	168,377	103,238	103,238	174,798	174,798
Schools	9,312	9,312	5,696	5,696	9,719	9,719

Note. Each column represents a separate weighted regression with weights equal to the number of test takers per observation. Columns (1) and (2) estimate our model on a sub-sample of the panel with non-missing information for school expenditures and student- teacher ratio. Columns (3) and (4) estimate our model on a sub-sample with non-missing information on whether there was a new school district superintendent during our panel period. Columns (5) and (6) estimate our model on a subsample of schools with non-missing information on charter school enrollment. Observations are at the school-grade-year level. Standard errors clustered at the school district level appear in parentheses. All regressions include year and school-by-grade fixed effects and district-level demographic covariates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Columns (5) and (6) estimate regressions on a sample of school districts with non-missing information on the percentage of students enrolled in charter schools. The model in Column (6) includes a variable for the fraction of students in the district enrolled in a charter school. School districts with a larger fraction of students enrolled in charter schools may have more flexibility to deviate from traditional school policies. Controlling for the fraction of charter school students, however, does not affect our vendor coefficient estimates.[26]

[26]We also estimate our model baseline after first splitting our sample into two groups based on whether the school is a charter school. The estimated healthy vendor coefficient is also the same to three decimal places for both samples. The source of the charter school and population data is the CA Department of Education. 306 districts include at least one charter school during our panel. The average yearly percent of students enrolled in a charter school for a district during this time period is 7%.

2.5 Discussion

2.5.1 Student Health

Congress passed the Healthy, Hunger-Free Kids Act (HHFKA) in 2010 with the aim of increasing the minimum nutritional standards that school lunches must meet. For example, the number of mandated servings of fruits and vegetables increased, while at the same time restrictions were placed on the number of servings of French fries (USDA 2012b). A major goal of the law is to improve the health of school-age children via a reduction in obesity (USDA 2013).

Previous research has shown that the source of a student's school lunch can affect obesity rates. Schanzenbach (2009) provides evidence that public school lunches have contributed to increases in childhood obesity rates. Students who are more likely to consume public school lunches, rather than other options such as bringing a brown-bag lunch, gain more weight. Currie et al. (2010) estimate that less than one academic year of exposure to fast-food restaurants near schools increases obesity rates for 9th grade students by about 5%. Currie et al. (2010) and Cutler et al. (2003) both emphasize that large increases in obesity rates could occur from as few as 100 excess calories per day.

We use the same source of physical fitness information as Currie et al. (2010) to test whether exposure to healthier school lunch options decreases student obesity. The Physical Fitness Test (PFT), also called FitnessGram®, is given to students in grades 5, 7, and 9 each spring in California. Six fitness areas compose the PFT, one of which is body composition. Schools have the option to complete the body composition portion using one of three measures: Body Mass Index (BMI), skin fold measurements, or bioelectric impedance analyzer. For each of these measurements, there is a defined "healthy zone" that varies by age and gender. The data are aggregated by school and grade level and indicate the percentage of students who have a body composition measurement in the healthy fitness zone. Following Currie et al. (2010), we define overweight as the percentage of students falling outside the healthy zone.

We do not find any evidence that contracting with a healthy lunch provider reduces the percentage of overweight students. We estimate Equation (2.1), except that we use the percentage of students who are outside the healthy fitness zone (whom we label as "overweight") as the dependent variable and restrict the sample to grades 5, 7, and 9.[27] Columns (1) and (2) of Table 2.6 show estimation results for all students and economically disadvantaged students, respectively. On average, 38.6% of the students (41.7% of disadvantaged students) in our sample are overweight. All four point estimates in Columns (1) and (2) are small and statistically insignificant. However, a lack of precision prohibits us from excluding an effect size as large as that found by Currie et al. (2010) (5.2%) for all but one of the coefficient estimates (All Students, Healthy Vendor). One explanation for why we find no effect on the percent of overweight students is that both standard and healthy lunch providers faced

[27]The Contract School sample for this regression includes 4,006 grade-year observations at 910 schools.

the same USDA calorie requirements, even prior to the HHFKA (USDA 2012a). Fast-food restaurants, in comparison, face no constraints on calories.

2.5.2 Number of Lunches Served

The first provisions of the HHFKA became binding beginning with the 2012-13 school year. One criticism of the law is that improving the health content of lunches may have the unintended consequence of reducing the number of students eating school lunches (Confessore 2014), possibly because of students' tastes. A decrease in the number of meals served to students eligible for reduced price or free lunches would be concerning because these students are considered most at risk for undernourishment and are a target population of the NSLP.

In order to estimate the impact of vendor quality on the number of lunches served in a district, we obtained NSLP data from the California Department of Education's Nutrition Services Division for the school years 2008-09 through 2012-13. The data report the average number of total NSLP lunches and the average number of free, reduced-price, and paid lunches served per operating day in each school district for each month. We use the monthly averages to calculate a single operating day average for the total number of lunches served and the number of free or reduced-price lunches served over the course of the academic year.[28]

Columns (3) and (4) of Table 2.6 report the impact of contracting with healthy and standard vendors on the number of daily lunches served per student. Column (3) estimates the effects on total lunches, while Column (4) estimates the effects on reduced price and/or free lunches. These regressions are run with observations at the level of district-year, rather than school-grade-year, because data on lunch purchases is only available aggregated at the district level. This constraint imposes minimal cost, however, because the treatment only varies at the district-year level. The dependent variable is the number of daily lunches sold per student (in Column (4), we only consider the number of reduced price and/or free lunches and the number of economically disadvantaged students). We do not find a significant effect of contracting with a healthy or standard vendor on the total number of lunches or the number of free and reduced price lunches. For example, the estimated coefficients for contracting with a healthy or a standard vendor in Column (3) are of a similar magnitude and imply a statistically insignificant reduction in the number of school lunches of approximately 10% (the sample mean of the dependent variable is 0.45).

The fact that disadvantaged students do not purchase more school lunches when the school contracts with a vendor supports the interpretation that the observed increase in test scores is due to the quality and not the quantity of school lunch meals consumed. At the same time, these findings help allay concerns that healthier lunches may actually lead to a reduction in the number of lunches served to students and are consistent with recent evidence in Johnson et al. (2016) that the Healthy Hunger-Free Kids Act increased nutritional quality without affecting student meal participation rates.

[28]We provide details of this calculation in Appendix A.2.2.

Table 2.6: Vendor Choice and BMI, Lunches Sold

Dependent variable:	Body Mass Index		Number Lunches Sold	
	(1)	(2)	(3)	(4)
Subgroup:	All	Disadvantaged	All	Disadvantaged
Healthy Vendor	-0.32	0.14	-0.049	-0.078
	(0.71)	(1.32)	(0.037)	(0.097)
Standard Vendor	-1.01	0.47	-0.035	-0.019
	(0.84)	(1.48)	(0.044)	(0.119)
Dependent variable mean	38.58	41.67	0.447	0.418
R^2	0.831	0.764	0.760	0.721
N	43,648	18,105	2,778	2,744
Schools/Districts	8,724	8,724	785	785

Note. Each column represents a separate weighted regression. The dependent variable in Columns (1) and (2) is the percentage of students who are overweight. Observations are at the school-grade-year level and the weights are equal to the number of physical fitness test takers. The dependent variable in Columns (3) and (4) is the number of school lunches sold. Observations are at the school district-year level and weights are the total student enrollment in the district. The first two regressions include school-by-grade fixed effects, while the latter two include district fixed effects. All regressions include year fixed effects and district-level covariates. Standard errors clustered at the school district level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The data on number of lunches served allow us to form a rough estimate of the effect of a school lunch on performance, or the "effect of the treatment on the treated" (TOT). If we view the treatment as exposure to a meal, rather than exposure to a vendor, the estimates in Tables 2.2 through 2.4 represent "intent-to-treat" (ITT) effects since a significant fraction of students do not eat school meals. Column (3) of Table 2.6 reveals that, on average, schools serve 0.45 lunches per test taker.[29] Since vendors do not affect lunches sold, this figure suggests that the TOT estimate is approximately $1/0.45 = 2.2$ times the ITT estimate. The meals sold data are at the district level. Using a district-level regression result (Column (5) of Table 2.4) we find an ITT effect of 0.037 that translates to a TOT effect of 0.082. This effect is statistically significant at the $p < 0.01$ level.

2.5.3 Policy Counterfactual

Public school administrators interested in improving the level of student learning and increasing test scores face decisions on how best to budget limited school resources. There are many potential changes in school policy that could improve learning. For example, school

[29] This approximation should be reasonably accurate as long as there are not many students that eat school breakfast but bring their own lunches. This group seems likely to be small because lunch participation rates are much higher than breakfast participation rates; Dahl and Scholz (2011) estimate that the number of lunches delivered was 2.24 times greater than the number of breakfasts delivered in the most recent year in their data (2002-2003).

administrators could hire more teachers to decrease average classroom size (Krueger 1999), lengthen the school day (Patall et al. 2010), increase teacher training (Angrist and Lavy 2001), give bonus pay to teachers based on student test scores (Fryer 2011), or increase student access to free or reduced price breakfast and lunch (Imberman and Kugler 2014b).[30]

Policies that direct resources toward teachers have been found to have a relatively large impact on student test scores in some settings. The Tennessee STAR experiment, which reduced average class size for primary school students by one-third and led to a 0.22 standard deviation test score increase, is a frequently cited benchmark. Nevertheless, these types of policies are often expensive and can be controversial (e.g., incentive pay). The Tennessee STAR experiment cost approximately $25 million (2013 $), with an implied cost of $3,009 (2013 $) per student placed in a smaller class.[31] Jacob and Rockoff (2011) highlight both the need and opportunity for cost-effective policies; lower-cost policies with modest effects on student test scores may generate a better return than costly policies with larger absolute effects.

We take advantage of contract-specific winner and loser bid information submitted to the CA Department of Education to calculate the cost of contracting with healthy lunch providers. The average price per lunch in healthy meal vendor contracts is $2.59. Two comparisons suggest that healthy vendor pricing is competitive with other alternatives. First, in seven contracts that appear in our data involving 18 bidders, both healthy and standard vendors bid on the same contract. A regression of log price per test taker on a healthy vendor indicator and contract fixed effects reveals that healthy vendors bid 19.9% less than other vendors in these cases, and we can reject the hypothesis that healthy vendors cost at least 6.6% more than other vendors at the 95% confidence level.[32] Second, the contract lunch price of $2.59 is close to the National School Lunch Program reimbursement rate of $2.93 per free lunch, suggesting that healthy school meal vendors do not cost dramatically more than in-house preparation.

To compute a plausible upper bound on the cost of increasing test scores via healthy school meal vendors, we assume that healthy school meal vendors cost 25% more than in-house preparation. This assumption implies that the average school makes a net profit of 30% on its school meal operations when reimbursed by the NSLP. A profit margin of this size or larger is unlikely because the NSLP specifically forbids the use of these revenues to fund non-food service operations (GPO n.d., 7 CFR, Section 210.14(a)).

The average healthy vendor meal contract is $426 (2013 $) per test-taker per school year. Over the 180-day academic year, a healthy school meal contract costs about $2.37 per test-taker day (this figure is different than the $2.59 meal cost because not every test taker eats a school meal, and some test takers eat both breakfast and lunch). We assume that

[30]This list highlights only a handful of policies and is not meant to be exhaustive.

[31]The original cost estimates reported by Krueger (1999) are adjusted to 2013 $ using the Consumer Price Index (CPI).

[32]This regression compares healthy vendors to all other vendors, including those with unknown quality. Limiting the comparison to known standard vendors, however, generates similar results, with a cost difference of −21.8% and a 95% confidence interval that reaches 12.5%.

the contract is 25% more expensive than in-house meal preparation, implying a difference of \$85 per test-taker per school year. To compare cost effectiveness we consider the dollar cost per 0.1 standard deviations of test score gains. This normalization does not imply that we can achieve a full 0.1 standard deviation increase with any given policy, including healthy vendor contracts (though the TOT estimate in Section 2.5.2 is close to 0.1). Rather, we recognize that policymakers have a menu of policies available for increasing test scores, and an optimizing policymaker will choose the combination of policies that yields a 0.1 standard deviation test score increase at the lowest possible cost.

Using a cost difference of \$85 per test-taker per school year and an estimated effect of 0.031 standard deviations (Column (3) of Table 2.2), we find that it would cost about \$274 per year to raise a student's test score by 0.1 standard deviations through switching from in-house preparation to a healthy lunch provider. In contrast, it cost \$1,368 per year to raise a student's test score by 0.1 standard deviations in the Tennessee STAR experiment. As additional context we consider recent estimates of the effects of changes in school spending. Lafortune et al. (2018) and Jackson et al. (2018) estimate that spending changes of \$556 and \$1,713 (2013 \$) per student-year respectively generate 0.1 standard deviation changes in test scores.[33] Thus, even an upper bound on the cost of raising test scores by 0.1 standard deviations through healthy meals is several times lower than a variety of benchmark estimates.

2.6 Conclusion

We exploit variation in the nutritional quality of school meals resulting from changes in meal providers to estimate the effect of nutritional quality on the academic performance of primary and secondary school students across the state of California. Using difference-in-differences type specifications, we find that switching to a healthy meal vendor is associated with a 0.031 standard deviation increase in test scores. While this effect is modest in magnitude, the relatively low cost of healthy vendors when compared to in-house meal preparation makes this a very cost-effective way to raise test scores.

We conduct a variety of robustness checks, including placebo tests and an event-study specification, to provide evidence that the timing of changes in meal providers is uncorrelated with omitted variables that could be driving changes in test scores. There is also no evidence that the introduction of healthier school lunches led to a change in the number of school lunches consumed. This supports our view that the observed relationship between healthier school lunches and test scores is due to the nutritional quality of the meals rather than the quantity of calories consumed. An analysis of the effects of healthy meal vendors on the

[33] Lafortune et al. (2018) find that a \$1,000 change in spending impacts test scores by 0.12 to 0.24 standard deviations. At the midpoint of 0.18 standard deviations, this result implies a cost of \$556 per 0.1 standard deviations. Jackson et al. (2018) find that a \$1,336 change in spending impacts test scores by 0.078 standard deviations, or \$1,713 per 0.1 standard deviations.

percentage of students who are overweight finds no effect, but it is possible that these effects could materialize on a longer time horizon.

Chapter 3

The Effect of Air Pollution on Children's Educational Outcomes: An Evaluation of California's Lower-Emission School Bus Program

3.1 Introduction

3.1.1 Background and Literature Review

An increasingly large body of literature has found a causal link between air pollution and various health outcomes (Schlenker and Walker, 2016; Anderson, 2015; Currie et al., 2009b; Moretti and Neidell, 2011; APA: Committee on Environmental Health, 2004). With this knowledge, many policies aimed at improving air quality have been introduced at both the federal and state level in the United States. Given that these policies necessarily entail costs, it is of great importance to quantify their impacts. A large proportion of the literature on the impact of air pollution has focused on naturally occurring variation (Moretti and Neidell, 2011; Currie et al., 2009b; Anderson, 2015) or changes induced by policies without the explicit intent of reducing pollution (Currie and Walker, 2011; Hanna and Oliva, 2015). Some notable exceptions include papers seeking to quantify the effect of the Clean Air Act of 1970 (Chay et al., 2003; Greenstone, 2004), which find small impacts. However, as Beatty and Shimshack (2011) note in their paper on school bus retrofits in Washington state, the effect of more localized policies may be different if they are better able to target pollution abatement to areas where exposure may be greatest.

This paper evaluates the impact of the California Lower-Emission School Bus Program (LESBP) on school attendance and academic test scores in California K-12 schools, and in doing so adds to the literature on the impact of pollution-reducing policies. The particular policy and setting are of interest for several reasons. Targeting school buses has the potential for large impacts because they contribute disproportionately to air pollution (Sabin et al., 2005; Hill et al., 2005; Thurston, 2000). The vast majority of school buses in the United States run on diesel fuel (Monahan, 2006). Diesel exhaust is a mixture of vapors and fine particulates, many of which are toxic or otherwise harmful to human health. In terms of the criteria pollutants determined by the U.S. Environmental Protection Agency (EPA), diesel exhaust contains high concentrations of NOx, PM 2.5 and PM 10. While the amount of these pollutants that is released into the air can be significantly reduced through technologies available on more recent bus models, many buses currently in circulation lack these pollution controls. According to the California Air Resources Board, an average 2007 model year bus emits 95% less toxic PM and 85% less NOx than a pre-1977 bus. Not only do these pollutants affect ambient air quality, but through a combination poor maintenance and flaws in design, the exhaust can enter the cabin of the bus, exposing children to levels of pollution that can be 5 to 10 times the level of background pollution (Hill et al., 2005; Wargo et al., 2002; Sabin et al., 2005). This level of exposure is of particular concern given that children have generally been found to be more vulnerable to the negative health effects of pollution (APA: Committee on Environmental Health, 2004; Schlenker and Walker, 2016; Thurston, 2000).

The LESBP addresses this problem by providing funds to replace all existing pre-1977 and a substantial portion of the 1977-1986 models currently in the state's fleet. The pre-1977 buses have no controls for nitrogen oxides (NOx) or PM emissions, while the 1977-1986 models have minimal controls (California Air Resources Board, 2008). Recent bus models can combine a number of technologies, such as diesel oxidation catalysts, soot traps, or crankcase filters. Diesel oxidation catalysts oxidize carbon monoxide and unreacted hydrocarbons to make carbon dioxide, which is not considered harmful to human health. Meanwhile, soot traps and crankcase filters, remove soot from the exhaust and prevent the exhaust from reaching the cabin, respectively. Another alternative is to use a cleaner burning fuel, such as natural gas instead of diesel. The LESBP program does not require particular technologies, but outlines a minimum standard in terms of NOx and PM emissions that replacement buses must meet. All in all, they estimate reductions of approximately 3,000 tons of NOx and 200 tons of PM as a result of the program.

The impact of a school bus retrofit program similar to the LESBP was studied in Beatty and Shimshack (2011). Using a differences-in-differences design, they find that the program reduces bronchitis, asthma, and pneumonia among children and adults with chronic conditions, which roughly translates into a cost to benefit ratio (in present value terms) between 7:1 and 16:1. A subsequent paper also studied the impact of bus retrofits and adoption of ultralow-sulfur diesel in the Puget Sound region and found reductions in a marker for inflammation in the lungs, increased lung function and reduced absenteeism among the children in the study (Adar et al., 2015).

While there are obvious similarities between my analysis and those previous studies, this paper still makes important contributions. Both Beatty and Shimshack (2011) and Adar et al. (2015) only analyze schools in the Puget Sound region of Washington State, while the analysis sample in this paper consists of all public schools in the state of California. The Puget Sound area has a higher average income, smaller minority population, and is less densely populated than most urban areas in the United States Beatty and Shimshack (2011). Thus, studying the effects of a bus replacement program in a population with greater income and ethnic heterogeneity may yield additional insights. Secondly, California has some of the highest levels of pollution in the United States. The five cities with the highest 24-hr PM2.5 and ozone concentrations in the country were all located in California (American Lung Association, 2014). Thus, effective policies aimed at reducing pollution are of great importance in the state of California. This paper is in some sense an extension of Beatty and Shimshack (2011) paper in that it looks at academic outcomes that may be affected through the channel of improved health. I also build on the findings of (Adar et al., 2015) by measuring not only school attendance but also academic test scores.

A substantial amount of research exists on the effects of pollution on health, particularly among vulnerable populations like children, but there is a smaller and more recent literature on how the health effects of pollution translate into effects on human capital and labor market decisions. Hanna and Oliva (2015) find that reductions in pollution resulting from a refinery closure in Mexico led to increases in hours worked per week among individuals living in close proximity to the refinery. Others have found decreases in worker productivity on high pollution days (Chang et al., 2016b,a). Thus, understanding the effect of pollution abatement on educational outcomes in school age children is a valuable addition to the literature about the effects of pollution on human capital.

Estimating the effect of the program on attendance is interesting for two reasons. First of all, school absences are thought to be sensitive to changes in children's health, particularly respiratory ailments that are often triggered by pollution (Currie et al., 2009a; Grossman and Kaestner, 1997). Thus, school attendance provides an indirect way of measuring health outcomes, and in particular may pick up effects not found using hospital data, since a child may be sick enough to miss school but not so sick as to require a hospital visit. Furthermore, attendance is an interesting outcome in and of itself given that school absences (i.e., low attendance) have been found to be negatively correlated with academic achievement (Gottfried, 2009). Among many reasons for this relationship is that chronically absent students receive fewer hours of instruction (Gottfried, 2009; Connell et al., 1994). If we find an effect on school attendance, then there is also reason to believe that academic performance could improve. Thus, our second outcome of interest is standardized test scores. Standardized test scores are frequently used in the education literature to measure the impact of interventions or as a measure of academic achievement (Rothstein, 2010; Roland G. Fryer, 2014; Imberman and Kugler, 2014a). I find some evidence that the program has a positive effect on attendance, although this effect is sensitive to the specification. However, I find no corresponding effect on test scores.

3.1.2 Description of Program

The LESBP received funding in the amount of $200 million from Proposition 1B passed in 2006. The first of the buses replaced through this program were delivered in spring of 2009 in time for the 2008-2009 school year. Funding was allocated to air districts based on the number of pre-1977 model buses in their fleet and their percentage share of the statewide 1977-1986 bus population (California Air Resources Board, 2008). Each of the 35 air districts was then tasked with disbursing the funds to the school districts within its boundaries.

While an application process was used to disburse funds, several aspects of the way the program was designed leads me to believe that the exact timing of replacements is plausibly exogenous, allowing me to exploit differences in timing to estimate the effect of the program. The priority of the program was to get all pre-1977 buses off the road by the start of the 2011-2012 school year. This goal was not universally achieved with some replacements occurring in late 2011. The legislation specified that air districts were required to reach out to school districts with pre-1977 model school buses to inform them of the program and encourage them to apply for the program, potentially reducing selection among the pool of applicants. Funds remaining after all pre-1977 buses replaced were to be used for the replacement or retrofit of 1977-1986 models with priority given to older models. As a "tie-breaker" program managers were allowed to prioritize districts with high levels of exposure as measured by total bus miles traveled per year but only after prioritizing older buses.

The rules of the program also stipulated that funds were to be used only for replacement of old school buses and not fleet expansion, with proof of disposal required. According to annual program reviews and program administrators, the new bus purchase funds were usually oversubscribed, such that the exact timing of replacements was generally determined by funding availability, logistical constraints, and the priority rules laid out in the legislation that funded the program. That is, several school districts may have applied to the program at approximately the same time, but their new buses would enter the fleet at different times. We will further explore the identifying assumption of exogenous timing in section 3 of the paper.

3.2 Background and Data

I use three sources of data to estimate the impact of LESBP on academic achievement. The first of these is administrative information provided by the Air Resources Board, which details replacements and retrofits.[1] The other two datasets come from the California Department of Education, and they provide information about attendance and standardized test scores.

3.2.1 Bus Replacement and Retrofit Data

The replacement and retrofit data are provided by the LESBP, which is overseen by the Air Resources Board, for the years 2008-2014, corresponding to the 2008-2009 through 2013-2014 school years. Given that the majority of the funds were allocated to replacements and the rules for prioritizing applications for replacements are clearly delineated, I will focus this analysis only on bus replacements. LESBP funds are apportioned to Air Pollution Control Districts (APCD) or Air Quality Management Districts (AQMD), henceforth air districts,

[1] I would like to thank Lisa Jennings at the California Air Resources Board for providing this information.

which then divide the funds among the school districts that comprise the air district. Thus, the data include both the air district and school district in which buses were replaced and the date on which the replacement took place. The data also include the number of school buses in each air district that were manufactured prior to 1977 and those dating from the 1977-1986 period, the criteria that determined funding at the air district level. Additional bus fleet information, including total fleet size, miles traveled and ridership at the school-district level was also obtained from the California Department of Education for the 2007-2008 through 2011-2012 school years.

3.2.2 Attendance and Test Score Data

In order to measure educational impacts, I use two different outcome measures.The first measure is attendance, which is reported by the California Department of Education at the school district level. Specifically, the Department of Education reports average daily attendance (ADA), which is calculated as the total days of student attendance divided by the total days of instruction.[2] To measure academic achievement, I use data from California's Standardized Testing and Reporting (STAR) for the school years 2006-2013. STAR tests are administered annually for grades 2 through 11. The core tests are Math (tested in grades 2 through 7), English Language Arts (tested in grades 2 through 11), Science (tested in grades 5, 8 and 10), and History/Social Sciences (tested in grades 8 and 11). Additional tests, denominated End of Course tests, are administered for various subjects, such as Geometry and Biology. The grade in which these subjects are tested depends on the grade in which the student finishes the course.

The outcome variable is a composite test score at the school-grade level, which averages all tests taken in a given school for a given grade and school year. Before averaging the scores, they are first standardized by the standard deviation of each test as reported by the California Department of Education. We standardize scores because these type of tests are often designed so that it is almost impossible to get a score very close to zero, such that estimating changes in raw scores may not be very informative. The data also include mean test scores for different subgroups at the school-grade level, which is used to estimate effects on particular subgroups of interest. In particular, I will study the effect of the program on test scores for disadvantaged students, who may have lower levels of health and thus may be more susceptible to pollution. Disadvantaged students are defined by the California Department of Education as students who are eligible for free or reduced lunches through the National School Lunch Program. The income threshold for eligibility for free or reduced lunches is 185% of the federal poverty line. Below I report summary statistics for academic outcomes, bus replacements, and demographic variables.

[2] An alternative way to measure attendance would be to look at reported absences. However, California only reports truancies, which is defined as the number of students who have three or more unexcused absences or tardies in a school year. Because health-related absences are frequently excused, I believe ADA provides a better measure for my purposes.

Table 3.1: Summary statistics

Variable	Mean	Std. Dev.	N
Enrollment	6945	24967	3378
Attendance Rate	0.920	0.135	3378
Stnd. STAR score: All	5.26	0.36	3385
Stnd. STAR score: Disadvantaged	5.04	0.28	3136
% Disadvantaged	0.542	0.257	3385
% Black	0.037	0.057	3385
% Asian	0.054	0.095	3385
% Hispanic	0.423	0.275	3385
% White	0.414	0.267	3385
% Parent ed: No high school	0.151	0.139	3385
% Parent ed: High school	0.205	0.098	3385
% Parent ed: College	0.167	0.100	3385
% Bus Riders	0.293	1.823	3358
Replaced buses per student	0.101	0.222	3385
Fleet Size	12.15	32.09	3356

Note. Summary statistics reported are at the school district level. Race, ethnicity, and parental education covariates come from STAR data. % Bus Riders is the average number of daily students transported divided by enrollment. % Replaced is the average percentage of eligible buses replaced over the four years in our sample.

3.3 Empirical Strategy

Our empirical design takes advantage of the differences in timing of replacements to estimate the effect on absences and test scores. The main specification used to estimate the effect on attendance is the following:

$$y_{dat} = \alpha + \beta\, No.\, Replaced_{dat} + \delta X_{dat} + \lambda_d + \gamma_t + \varepsilon_{dat} \tag{3.1}$$

The dependent variable is ADA divided by enrolled students in school district d, located in air district a, during year t. The independent variable is the number of buses that have been replaced in air district a by year t divided by the number of enrolled students (in thousands) within that air district. Thus, our independent variable varies only at the air district-year level. Given the durable nature of school buses, the number of replacements is calculated as the cumulative number of buses that have been replaced by year t.

The main specification includes year fixed effects γ_t and school district fixed effects α_d. The year fixed effects control for state-wide characteristics, such as budget cuts to education, that could affect both the test score outcomes and the timing of replacements. Meanwhile, the school district fixed effects control for time-invariant characteristics such as school infrastructure in a given district. Lastly, X_{dt} is a vector of controls at the school district and year level. These control variables are potentially time-varying characteristics of the test-takers, such as their race, parental education and economic status. Additional controls are included for the average percentage of students in a school district that ride buses to school, the percentage of replacements for which the air district provided matching funds to cover the school district's portion of the costs,[3] and school district budget expenditures in a given year.

Standard errors are clustered at the air-district level, since the variation in the percentage of eligibles buses replaced occurs at the air district level. While the data for bus replacements is available at the school-district level, aggregating to the air district level mitigates some endogeneity concerns, since air districts are automatically allocated funding to disburse, whereas school districts must apply for replacement funds. Moreover, if there exist spillovers between school districts, varying treatment at the air district level may be more appropriate. On the other hand, aggregating to the air-district level removes some potentially important variation, given that school bus fleets are managed at the school district level. For example, consider an air district that is comprised of two school districts with 50,000 students each. One school district does not replace any buses and the other replaces 10 buses. Then, both school districts will be assigned a value of the independent variable equal to 0.1 (.1 replacements per thousand students). I come back to this issue in a subsequent section.

[3] Program rules required school districts to contribute \$25,000 towards the cost of a bus replacement, but the air district could choose to cover this amount with their own funds, which is referred to as "District Match".

For the regressions estimating the program impact on test scores I use the following specification:

$$y_{gsat} = \alpha + \beta \, No. \, Replaced_{gsat} + \delta X_{gsat} + \lambda_{gs} + \gamma_t + \epsilon_{gsat} \tag{3.2}$$

This specification is almost identical to Equation 3.1, except that observations are at the school-grade level as opposed to the school district level. The dependent variable, y_{gsat}, is the mean standardized score across all STAR tests administered in grade g in school s for year t. Because observations are aggregated, and there is heterogeneity in the number of individuals in each unit of observation, I report regressions with and without regression weights. For the test score regressions, I will use the number of test-takers as regression weights, while for the attendance regressions, enrolled students per district will serve as weights.

The identifying assumption is that conditional on school district (or school-grade) and year fixed effects, as well as time-varying controls for school demographics, the exact timing of the school bus replacement is uncorrelated with school-specific trends in the outcome variable of interest. This assumption is plausible given that a school district's eligibility for the program depends on its population of pre-1986 and pre-1987 model year buses prior to the LESBP introduction, which would be taken care of by the school district (school-grade) fixed effects. In addition, program officials contend that the program was frequently oversubscribed, such that the exact timing of replacements was driven largely by logistical constraints.

One concern we might have is that districts with a larger number of pre-1977 school buses at baseline may be significantly different from other school districts, such that they may have been differentially targeted by other programs coinciding with bus replacements. Table 3.3 reports the difference in means between school districts with and without pre-1977 school buses across several baseline characteristics. I find that school districts with pre-1977 buses have a higher percentage of students that are disadvantaged, Hispanic or whose parents have less than a high school education. These districts also have lower percentages of students that are white and whose parents have a college education. Overall, school districts that have pre-1977 buses at baseline seem to be more disadvantaged economically and educationally than districts that do not. While this does not invalidate the research design, it implies that I must be careful in ruling out other programs that may have been targeted at disadvantaged schools. The first step towards addressing this issue is the inclusion of school district level annual expenditures in the regression, which should control for large budget increases that may come from other programs targeting struggling school districts. In section 4.3, I will also discuss how other programs that may have differentially targeted disadvantaged school districts are inconsistent with my results.

Table 3.2: School characteristics by pre-1977 bus ownership

	Number of pre-1977 Buses
% Disadvantaged	0.120**
	(0.0550)
% Black	0.00645
	(0.0139)
% Asian	0.00142
	(0.0557)
% Hispanic	0.259***
	(0.0547)
% White	-0.242***
	(0.0552)
% Parental ed: No HS	0.0900***
	(0.0289)
% Parental ed: HS	0.0374
	(0.0229)
% Parental ed: College	-0.0481**
	(0.0210)

Note. Table reports the coefficients from a regression of 2006 school district characteristics on an indicator for having at least one pre-1977 buses in the fleet.

3.4 Results

3.4.1 LESBP and Attendance

I begin by estimating equation 3.2 for the school years 2008-2009 through 2011-2012 in order to look at the effects of LESBP on attendance. Table 3.3 shows the results of this estimation. I report both weighted and unweighted regressions and regressions with and without covariates. Focusing on the specifications that include covariates, we see a substantial difference between the point estimates of the unweighted and weighted regressions. The main difference between these two specifications is that in the latter, observations from larger school districts, as measured by total enrollment, are given more weight. Thus, the results suggest that there is a difference in the effect of bus replacements between small and large districts. More specifically, it appears that the effect of bus replacements is larger in smaller districts.

Column (3) reports the results of the weighted regression, in which I find no effect of bus replacements. In the unweighted specification with covariates, column (4), I find a positive and statistically significant result. The point estimate of 0.0342 suggests that replacing one bus in an air district with 100,000 students would result in an increase in attendance of 0.000342, or approximately 0.034 % from a mean of 92%. The average air district has about 80 pre-1986 buses in its fleet, so replacing half of these would increase attendance be about 1.4 percentage points.

Table 3.3: The effect of bus replacements on attendance: All students

	(1) Weighted	(2) Unweighted	(3) Weighted	(4) Unweighted
No. replaced/student	0.00529	0.0265**	0.00122	0.0342***
	(0.0213)	(0.0108)	(0.0197)	(0.00977)
Covariates	N	N	Y	Y
N	3378	3378	3378	3378
R^2	0.898	0.833	0.902	0.846
School districts	869	869	869	869

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at air district level. Regressions include school district and year fixed effects. Number of students enrolled per school district are used as weights in weighted regressions.

Turning to the difference between the weighted and unweighted estimates, Solon et al. (2015) note that in the presence of unmodeled heterogeneity, weighted and unweighted OLS give us different weighted averages of the heterogeneous effects neither of which is necessary the population average treatment effect. They further note that when unweighted and weighted estimates are different, the correct course of action is to explore the source of the heterogeneity. I begin by excluding certain observations to understand if outliers are driving my results. In table 3.4, I report the results of regressions estimated without the Los Angeles School District, as well as regressions excluding districts that ever have more than 1 replacement per 1000 students greater than 1 or attendance less than 50%.

Los Angeles Unified District (LAUSD) is by far the largest district in our sample, approximately six times as large as the next largest school district, so it is weighted very heavily in the weighted regressions. As can be seen in columns (1) and (2), once I exclude LAUSD from the weighted regression, the point estimate increases by an order of magnitude. In column (4), we can see that excluding districts with number of buses replaced per student greater than one barely changes the point estimate in the unweighted regression although the results are no longer significant. Lastly, in columns (5) and (6), I report the results of regression excluding school districts that are outliers in terms of attendance. In column (6), we see that the unweighted estimates estimate on this sample are very similar in magnitude to the baseline results, and remain significant. I take this as evidence that the unweighted results we find in table 3.3 are not driven by outliers along the dimensions of number of bus replacements or attendance. However the large difference between the weighted and unweighted point estimates, with the weighted estimates being much smaller, is driven in part by a small treatment effect in the very large LAUSD.

Table 3.4: The effect of bus replacements on attendance: Excluding outliers

	Los Angeles		Outliers - replacements		Outliers - attendance	
	(1)	(2)	(3)	(4)	(5)	(6)
No. replaced/student	0.0129	0.0342***	-0.0240	0.0369	0.000498	0.0293***
	(0.0119)	(0.00976)	(0.0310)	(0.0242)	(0.0199)	(0.00698)
Covariates	Y	Y	Y	Y	Y	Y
Weights	Y	N	Y	N	Y	N
N	3374	3374	3196	3196	3266	3266
R^2	0.910	0.846	0.970	0.871	0.869	0.724
School districts	868	868	823	823	840	840

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at air district level. Regressions include school district and year fixed effects. Outliers in terms of replacements are defined as observations where the number of buses replaced per 1,000 students exceeds 1 at any point. This also roughly corresponds to the 99th percentile of bus replacements across the 4 years. Outliers in terms of attendance are defined as school districts with attendance rate less than 50% at any point in the sample. Number of students enrolled per school district are used as weights in weighted regressions.

Next, I look at heterogeneity in the treatment effect by interacting district characteristics with the number of buses replaced per student. I focus on the unweighted results with covariates, given that the unweighted estimates appear to be more stable. These results are reported in table 3.5. Column (1) reports the results of a regression where we interact treatment with the percentage of disadvantaged students in a district. While the point estimate is positive, the result is not significant. This result provides some evidence against the hypothesis that disadvantaged children may see larger impacts from the program because they may live in more polluted environments and be in generally worse health, making them more sensitive to additional factors that may negatively impact their health (APA: Committee on Environmental Health, 2004; Currie et al., 2009b). In column (2), I explore heterogeneity according to different levels of exposure as proxied by the total bus miles driven (in 100,000s). I find no indication that treatment varies by total miles driven.

Lastly, I test the hypothesis that there are heterogeneous treatment effects with respect to background levels of pollution. In particular, I would like to focus on background levels of the pollutants that are expected to see the largest decreases, NOx and particulate matter. To measure background exposure, I use non-attainment status for a particular pollutant as reported by the EPA. Counties out of attainment are assigned a value of 1, while counties only partially out of attainment are assigned a value of 0.5. None of the counties in California are out of attainment for NOx during the time period of interest, so I test heterogeneity with regards to PM10 and PM2.5. These results are reported in columns (3) and (4), respectively. In column (3), I find evidence that the treatment effect of bus replacements is larger for school districts located in counties that are out of attainment for PM 10, although the coefficient is significant only at the 10% level. The point estimate of 0.0585 suggests that going from partial non-attainment to attainment or non-attainment to partial attainment almost doubles the size of the treatment. The point estimate for PM 2.5 is positive but not significant.[4]

Table 3.5: Heterogenous effects of bus replacements on attendance

	(1) % Disadvantaged	(2) Tot. miles driven	(3) PM 10	(4) PM 2.5
No. replaced/student	0.0188	0.0354***	0.0326***	0.0346***
	(0.0166)	(0.0125)	(0.00959)	(0.00902)
% Disadvantaged	-0.00906			
	(0.0467)			
Replaced × % Disadv.	0.0303			
	(0.0352)			
Tot. miles		0.000842**		
		(0.000403)		
Replaced × Tot. miles		-0.000732		
		(0.00199)		
Replaced × PM 10			0.0585*	
			(0.0336)	
Replaced × PM 2.5				0.0288
				(0.0300)
Weights	N	N	N	N
Covariates	Y	Y	Y	Y
N	3378	3378	3378	3378
R^2	0.846	0.846	0.846	0.846
School districts	869	869	869	869

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at air district level. Regressions include school district and year fixed effects. Columns (3) and (4) report heterogeneous effects with regards to PM 10 and PM 2.5 non-attainment, respectively. Non-attainment status varies only at the county-level so these indicators are co-linear with school district fixed effects.

[4] Since attainment status varies only at the county level, the indicators for non-attainment are collinear with the school district fixed effects, so the effects of non-attainment alone cannot be reported.

3.4.2 LESBP and Test Scores

Table 3.6 shows the results of estimating equation (3.1) for all students in the sample. Column (1) shows the results of a regression using number of test-takers as regression weights, while column (2) shows the unweighted results. Columns (3) and (4) add test-taker characteristics as covariates to the weighted and unweighted regressions, respectively. While sample size decreases when I add covariates because some school-grade observations are missing information on test-taker characteristics, the changes in the point estimates suggest that it is important to control for test-taker and other time-varying characteristics. Thus, my preferred specification for test scores include covariates, similar to the attendance regressions. Overall, the point estimates are quite small and none of them are significant. Taking the upper bound of the 95 % confidence interval for the unweighted estimate, 0.030, suggests that replacing one school bus in an air district with 100,000 students[5] would result in a score increase of 0.00030 standard deviations, ruling out large positive effects. Even the weighted regression, which yields the largest estimate can reject effects larger than 0.00099 standard deviations per bus replacement. Table 3.7 reports the results of estimating Equation 3.1 on the subsample of students classified as disadvantaged in the STAR data. I find that the point estimates are of similar magnitude to those of estimated on the entire sample of students, and none of the point estimates across the four specifications are statistically significant, which is consistent with the finding in the attendance regressions.

Table 3.6: The effect of bus replacements on test scores: All students

	(1) Weighted	(2) Unweighted	(3) Weighted	(4) Unweighted
No. replaced/student	-0.00193	-0.0122	0.0225	-0.0206
	(0.0162)	(0.0176)	(0.0365)	(0.0257)
Covariates	N	N	Y	Y
N	138503	138503	86081	86081
R^2	0.935	0.914	0.940	0.931
School Districts	877	877	786	786

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at air district level. Regressions include school-grade and year fixed effects. Number of test-takers per observation are used as weights in weighted regressions.

[5] The mean number of students per air district is approximately 170,000

Table 3.7: The effect of bus replacements on test scores: Disadvantaged students

	(1) Weighted	(2) Unweighted	(3) Weighted	(4) Unweighted
No. replaced/student	-0.00372	-0.0144	0.0268	-0.0208
	(0.0190)	(0.0124)	(0.0353)	(0.0228)
Covariates	N	N	Y	Y
N	114901	120950	78487	78487
R^2	0.896	0.868	0.910	0.887
School Districts	750	825	736	736

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at air district level. Regressions include school-grade and year fixed effects. Number of test-takers per observation are used as weights in weighted regressions.

3.5 Identifying Assumptions and Robustness

To ensure that the results outlined above are not the product of some omitted variable bias, I perform two robustness checks. I first show that the timing of adoption is not correlated with changes in student characteristics. Table 3.8 shows the results of regressing student characteristics on our treatment variables and controls for the percentage of bus riders in a school district, percentage of replacements that received matching funds from the air district, and school district expenditures. Essentially, I am running our main specification with student characteristics as the dependent variable. If the identification assumption holds, one would expect the coefficient of *No. Replaced* to be small and not significant, indicating that adoption decisions are not correlated with changes in the student body.

Indeed, we see in Table 3.8 that only of the coefficients on the nine covariates tested is significant at the 5 % level, which suggests that the timing of replacements is not generally correlated with changes in school characteristics. While I directly control for these covariates in my main specification, finding significant results would cause us to be concerned that some unobservable characteristics are also correlated with the timing of replacements.

As another test of exogeneity for the timing of the replacements, I estimate the following specification:

$$y_{dat} = \alpha + \sum_{k=-3}^{0} \beta_k \, No.\ Replaced_{dat}^{k} + \delta X_{dat} + \lambda_d + \gamma_t + \varepsilon_{dat} \tag{3.3}$$

This specification adds three treatment "leads", such that $No.Replaced_{dat}^{k}$ for $k = -1$ is equal to $No.Replaced_{dat}$ in the subsequent year (t+1). Equation 3.2 is also modified to get an analogous specification for the regressions that use test scores as the dependent

Table 3.8: Correlation between test-taker covariates and timing of replacements

	(1) Adv	(2) Disadv	(3) Black	(4) Asian	(5) Hisp	(6) White	(7) No HS	(8) HS	(9) College
No. replaced/student	-0.0186	0.0137	0.00306**	-0.000344	0.00514	-0.00960	-0.000899	-0.00162	0.000577
	(0.0114)	(0.0138)	(0.00150)	(0.00181)	(0.00499)	(0.00921)	(0.00369)	(0.0119)	(0.00373)
% district match	0.00186	-0.000936	0.000489	-0.000319	0.00271	-0.00618	-0.00260	0.0000496	-0.00175
	(0.00640)	(0.00742)	(0.000922)	(0.00117)	(0.00330)	(0.00733)	(0.00403)	(0.00726)	(0.00402)
N	3378	3378	3378	3378	3378	3378	3378	3378	3378
R^2	0.964	0.961	0.981	0.993	0.992	0.962	0.955	0.852	0.930

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at air district level. Regressions include school district and year fixed effects.

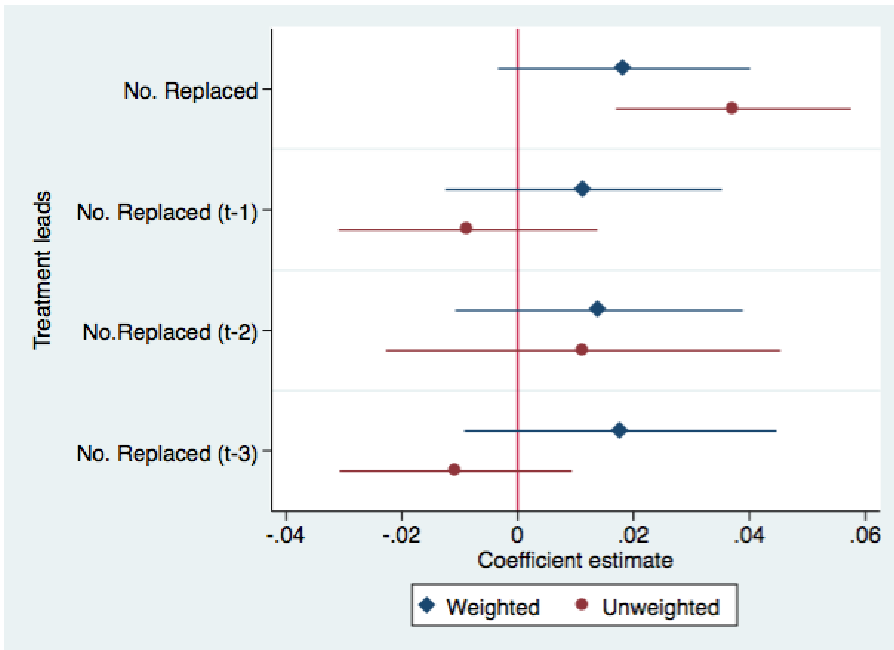
variable. If the timing of replacements is truly exogenous, I expect the coefficients β_k to be precisely estimated zeros for $k \in [-3, -1]$. Otherwise, the results in section could be driven by differential pre-trends. Further, the coefficient estimate for β_0, our true treatment effect, should not change very much relative to the estimation from the baseline specification.

Panel a of figure 2 plots the coefficients β_k for $k \in [-3, 0]$ for attendance as the outcome variable. For the unweighted regression, the coefficient on treatment remains very similar in magnitude and is still statistically significant. None of the treatment leads are significant, and they are about a third of the magnitude of the true effect, providing support for the assumption of exogeneity in the timing of replacements. For the weighted regression, neither the leads nor the treatment are statistically significant. Panel b of figure 2 shows the same coefficient plot for the test score regression. We see that the coefficients on the years prior to treatment are not significant in either the weighted or unweighted regression. The point estimates for β_0 become more negative relative to the main specification, but they are less precisely estimated. Overall, there seems to be little evidence of replacements occurring in response to changes in test scores.

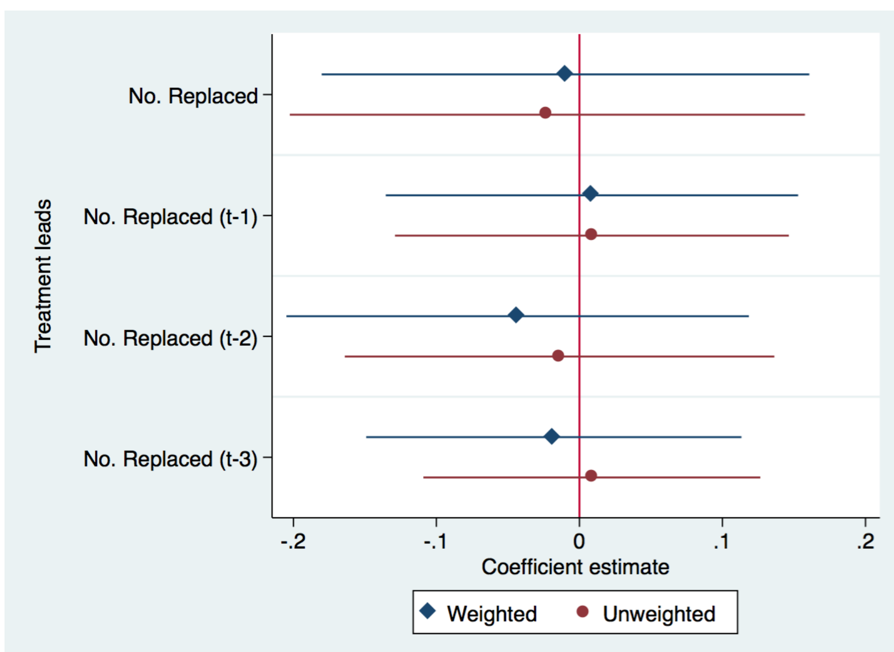
Administrative capacity and budget concerns prevented some school districts from getting school buses the same school year in which they applied for the replacement, adding additional random variation in the timing of bus replacements. Unfortunately, I do not have data regarding when the applications were filed. I only have information detailing when the bus order was placed after the application was approved, as well as the date the bus was delivered. In figure 3.2, I show the distribution in the number of days between the order and delivery date. While some school districts receive buses within 30 days of the order date, a not insubstantial number of purchases are delivered 6 months or more after the original order date, suggesting that there is variation in the delivery date that is plausibly unrelated to the date of application.[6]

[6]Note that it seems like some purchase orders were made retroactively as can be seen by the small number of districts where the delivery date is before the date of the purchase order.

Figure 3.1: Event study specification



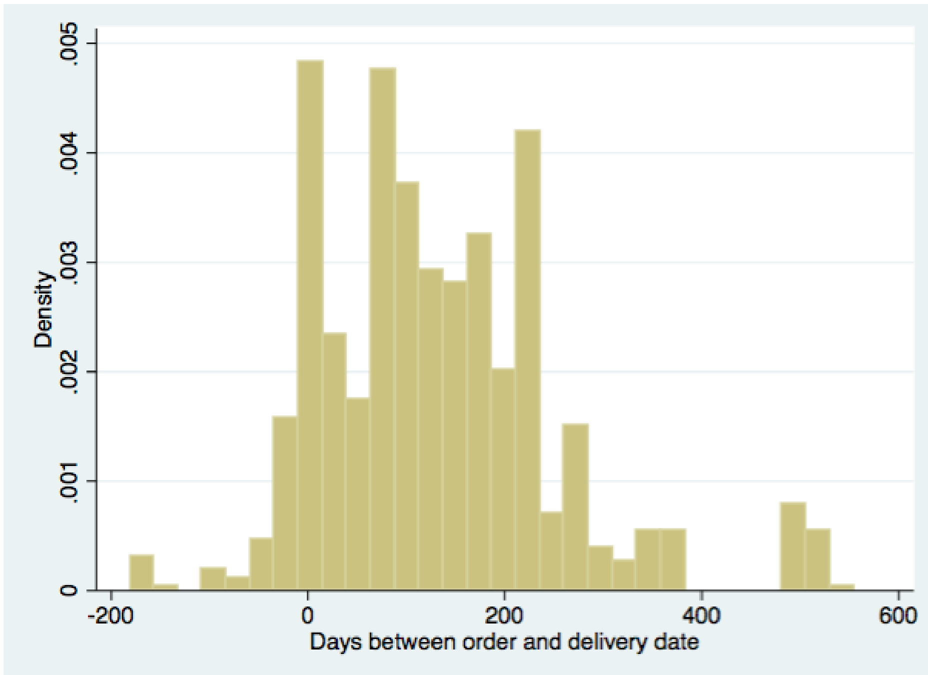
(a) Event study specification: Attendance



(b) Event study specification: Test scores

Note. Figure shows coefficients on treatment indicator and treatment leads (1-2) for test score and attendance regressions. Regressions include full set of covariates, year fixed effects, and district fixed effects in panel a or school-grade fixed effects in panel b. Standard errors are clustered at air district level.

Figure 3.2: Timing of bus delivery relative to order date



Note. This figure uses the distribution of the time between order and delivery for all the school buses replaced using bin-widths of 30 days. A small number of purchases report delivery dates that are before the order date, hence the negative numbers.

Lastly, I explore the idea that disadvantaged schools may have both been prioritized for replacements because of their older fleets and targeted by other programs. I tried to address this concern by controlling for total expenditures at the school district level, which should capture increased funds as a result of these programs. Through a search of California's bond accountability website I found other bonds approved for education expenditures during this period. The only program I found that may have directly targeted more disadvantaged school districts during this time are funds approved for overcrowding relief. These funds could be used to replace portable classrooms with permanent structures. However, it seems unlikely that the building of classrooms would reduce absences yet have no impact in test scores, as we find.

3.5.1 Robustness

All of the results presented up to this point have defined the treatment variable as the number of buses replaced per 1000 students at the air district level. This was done to mitigate endogeneity concerns and take into account the potential for spillovers within air districts, since many of these districts are drawn to coincide with air basins. However, I may also be using an important source of variation by aggregating to the air district level, so in this section of the paper, I estimate the results with treatment defined at the school district level. The results are reported in table 3.9. Focusing on column (4), which reports the unweighted results with covariates, we see that the point estimate is positive and significant, but at 0.00405 it is an order of magnitude smaller than the point estimate obtained when treatment was defined at the air district level. One possible explanation for this result is that the bus replacements significantly impact ambient air quality, which spills over beyond the school district where the bus is replaced. Such an effect would lead me to underestimate the true treatment effect (Miguel and Kremer, 2004). However, the spillover effects would have to be quite large to explain such a large difference in the results. Another potential explanation is that there is more endogeneity when defining the treatment at the school district level, which could lead to a downward bias. Results for test scores at the school district level, provided in the appendix, are very similar to those obtained when defining treatment at the air district level.

Table 3.9: The effect of bus replacements on attendance: School district analysis

	(1) Weighted	(2) Unweighted	(3) Weighted	(4) Unweighted
No. replaced/student	-0.0178	0.00259	-0.0194	0.00405**
	(0.0160)	(0.00158)	(0.0145)	(0.00149)
Covariates	N	N	Y	Y
N	3386	3386	3386	3386
R^2	0.904	0.832	0.908	0.845
No. districts	871	871	871	871

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at air district level. Regressions include school district and year fixed effects. Number of students enrolled per district are used as weights in weighted regressions.

3.6 Discussion and Conclusion

I find suggestive evidence that the bus replacements funded by LESBP have a positive effect on attendance, but no effect on STAR test scores. In the unweighted specification for attendance, bus replacements have a positive and statistically significant effect. This result is robust to excluding outliers across several dimensions, and generally robust to specifications allowing for treatment heterogeneity. I also find some evidence that the bus replacements have larger impacts for school districts located in non-attainment counties for PM10. The treatment effect reduces substantially when treatment is defined at the school district-level, potentially due to spillover effects, although endogeneity is also a concern. For STAR test scores, the point estimates for the cumulative number of buses replaced per student vary across specifications, but they all rule out large effects. The effect on disadvantaged students is similarly small and insignificant. However, the program does appear to have a positive impact on attendance rates.

While making direct comparisons to other studies is difficult, due to both differences in outcomes of interest and the nature of interventions, it is still instructive to understand how the estimates in this paper fit into the existing literature. I compare my estimates to two particularly relevant papers, one which studies the effects of a bus retrofit program on hospital discharges for respiratory illnesses (Beatty and Shimshack, 2011), and a second which looks at how naturally occurring variation in pollutions affects school absences (Currie et al., 2009a).

Beatty and Shimshack (2011) find that districts that retrofit their school buses see a 20 to 30 percent decrease in asthma and bronchitis, and pneumonia and pleurisy among children relative to control districts. These results translate approximately into a 2 % reduction in asthma and bronchitis cases (for both children and adults) for every 10 percent of the fleet that is replaced, with similarly sized effects for pneumonia and pleurisy. The estimate from my main specification translates into an increase in attendance of about 0.3 percentage points for every 10 percent of the eligible fleet that is replaced, from an average attendance rate of 92%. However, it is not clear how a 2% decline in asthma and bronchitis would translate into changes in attendance. For example, the attendance data does not specify what percentage of the absences are due to medical reasons, specifically respiratory diseases. Meanwhile, Currie et al. (2009a) find that the reduction in high CO days between 1986 and 2006 in particularly polluted areas of Texas reduced absences by 0.8 percentage points, an effect that based on the estimates in this paper would correspond to replacing about 27% of the eligible fleet.

A future line of research could try to more directly establish that the observed impacts on attendance are coming through a health channel by measuring the impact on respiratory diseases that other papers have observed (Beatty and Shimshack, 2011; Adar et al., 2015). Additional research could also help understand if bus replacements affect test scores in the long term, although we do not find concurrent impacts. Nevertheless, my estimates suggest that bus replacements have likely have benefits for the school children who are most exposed to their pollution, in addition to whatever impacts they may have on ambient air quality.

Bibliography

Adar, Sara D., Jennifer D'Souza, Lianne Sheppard, Joel D. Kaufman, Teal S. Hallstrand, Mark E. Davey, James R. Sullivan, Jordan Jahnke, Jane Koenig, Timothy V. Larson, and L. J. Sally Liu, "Adopting Clean Fuels and Technologies on School Buses. Pollution and Health Impacts in Children," *American Journal of Respiratory and Critical Care Medicine*, 2015, *191* (12), 1413–1421. PMID: 25867003.

Albarran, Pedro and Orazio P. Attanasio, "Limited Commitment and Crowding out of Private Transfers: Evidence from a Randomised Experiment," *The Economic Journal*, 2003, *113* (486), C77–C85.

Alderman, Harold and Trina Haque, "Insurance Against Covariate Shocks," *World Bank Policy Research Working Paper*, 2008.

Alderman, Harold, Jere R. Behrman, and John Hoddinott, "Economic and Nutritional Analyses Offer Substantial Synergies for Understanding Human Nutrition," *Journal of Nutrition*, 2007, *137* (3), 537–544.

American Lung Association, "State of the Air Report, 2014," 2014.

Anderson, Michael L., "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality," Working Paper 21578, National Bureau of Economic Research September 2015.

Angrist, Joshua and Victor Lavy, "Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools," *Journal of Labor Economics*, 2001, *19* (2), 343–369.

APA: Committee on Environmental Health, "Ambient Air Pollution: Health Hazards to Children," *Pediatrics*, 2004, *114* (6), 1699–1707.

Arias, Diego, Alejandro de la Fuente, Jesus Escamilla, and Charles Stutley, "Financial Innovations for Social and Climate Resilience: Mexico Case Study," *World Bank*, 2014.

Barnett, Barry J. and Olivier Mahul, "Weather index insurance for agriculture and rural areas in lower-income countries," *American Journal of Agricultural Economics*, 2007, *89* (5), 1241–1247.

Barnett, Barry J., Christopher B. Barrett, and Jerry R. Skees, "Poverty traps and index-based risk transfer products," *World Development*, 2008, *36* (10), 1766–1785.

Beatty, Timothy and Jay Shimshack, "School buses, diesel emissions, and respiratory health," *Journal of Health Economics*, 2011, *30*, 987–99.

Belot, Michele and Jonathan James, "Healthy school meals and educational outcomes," *Journal of Health Economics*, 2011, *30*, 489–504.

Bryan, Janet, Saskia Osendarp, Donna Hughes, Eva Calvaresi, Katrine Baghurst, and Jan-Willem van Klinken, "Nutrients for cognitive development in school-aged children," *Nutrition Reviews*, 2004, *62* (8), 295–306.

Cai, Jing, "The Impact of Insurance Provision on Household Production and Financial Decisions," *American Economic Journal: Economic Policy*, 2016, *8* (2), 44–88.

Cai, Jing and Changcheng Song, "Do disaster experience and knowledge affect insurance take-up decisions?," *Journal of Development Economics*, 2017, *124*, 83–94.

Cai, Ruohong, Shuaizhang Feng, Michael Oppenheimer, and Mariola Pytlikova, "Climate variability and international migration: The importance of the agricultural linkage," *Journal of Environmental Economics and Management*, 2016, *79*, 135–151.

California Air Resources Board, "Guidelines: Lower-Emission School Bus Program," 2008.

California Department of Education, "California Standardized Testing and Reporting, Post-Test Guide Technical Information for STAR District and Test Site Coordinators and Research Specialists," Technical Report 2011.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statistical Association*, 2015, *110* (512), 1753–1769.

Carter, Michael, "Designed for Development: Next Generation Approaches to Index Insurance for Smallholder Farmers," *ILO/MunichRe Microinsurance Compendium*, 2012, *2.*

Carter, Michael, Alain de Janvry, Elisabeth Sadoulet, and Alexandros Sarris, "Index Insurance for Developing Country Agriculture: A Reassessment," *Annual Review of Resource Economics*, 2017, *9*, 421–438.

Carter, Michael, Lan Cheng, and Alexandros Sarris, "Where and how index insurance can boost the adoption of improved agricultural technologies," *Journal of Development Economics*, 2016, *118*, 59–71.

Chang, Tom, Joshua Graff Zivin, Tal Gross, and Matthew Neidell, "The Effect of Pollution on Worker Productivity: Evidence from Call-Center Workers in China," Working Paper 22328, National Bureau of Economic Research June 2016.

Chang, Tom, Joshua Graff Zivin, Tal Gross, and Matthew Neidell, "Particulate Pollution and the Productivity of Pear Packers," *American Economic Journal: Economic Policy*, August 2016, *8* (3), 141–69.

Chay, Kenneth, Carlos Dobkin, and Michael Greenstone, "The Clean Air Act of 1970 and Adult Mortality," *Journal of Risk and Uncertainty*, 2003, *27* (3), 279–300.

Clarke, Daniel J., "A Theory of Rational Demand for Index Insurance," *American Economic Journal: Microeconomics*, 2016, *8* (1), 283–306.

Cole, Shawn, Xavier Giné, and James Vickery, "How does risk management influence production decisions? Evidence from a field experiment," *The Review of Financial Studies*, 2017, *30* (6), 1935–1970.

Cole, Shawn, Xavier Gine, Jeremy Tobacman, Petia Topalova, Robert Townsend, and James Vickery, "Barriers to Household Risk Management: Evidence from India," *American Economic Journal: Applied Economics*, 2013, *5* (1), 104–35.

Confessore, Nicholas, "How School Lunch Became the Latest Political Battleground," *The New York Times Magazine*, 2014.

Connell, James Patrick, Margaret Beale Spencer, and J. Lawrence Aber, "Educational Risk and Resilience in African-American Youth: Context, Self, Action, and Outcomes in School," *Child Development*, 1994, *65* (2), 493–506.

Cox, Donald, Bruce E. Hansen, and Emmanuel Jimenez, "How responsive are private transfers to income? Evidence from a laissez-faire economy," *Journal of Public Economics*, 2004, *88* (9-10), 2193–2219.

Currie, Janet and Reed Walker, "Traffic Congestion and Infant Health: Evidence from E-ZPass," *American Economic Journal: Applied Economics*, 2011, *3* (1), 65–90.

Currie, Janet, Erick A. Hanushek, Megan Kahn, Matthew Neidell, and Steven G. Rivkin, "Does Pollution Increase School Absences?," *The Review of Economics and Statistics*, 2009, *91* (4), 682–694.

Currie, Janet, Matthew Neidell, and Johannes Schmieder, "Air Pollution and Infant Health: Lessons from New Jersey," *Journal of Health Economics*, 2009, *28* (3), 688–703.

Currie, Janet, Stefano DellaVigna, Enrico Moretti, and Vikram Pathania, "The Effect of Fast Food Restaurants on Obesity and Weight Gain," *American Economic Journal: Economic Policy*, August 2010, *2*, 32–63.

Cutler, David M., Edward L. Glaeser, and Jesse M. Shapiro, "Why have Americans become more obese?," *Journal of Economic Perspectives*, 2003, *17*, 93–118.

Dahl, Molly W and John Karl Scholz, "The National School Lunch Program and School Breakfast Program: Evidence on Participation and Noncompliance," Technical Report, University of Wisconsin 2011.

de Janvry, Aalain, Vianney Dequiedt, and Elisabeth Sadoulet, "The demand for insurance against common shocks," *Journal of Development Economics*, 2014, *106*, 227–238.

Dercon, Stefan and Luc Christiaensen, "Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia," *Journal of Development Economics*, 2011, *96* (2), 159–173.

Dercon, Stefan, Ruth Vargas Hill, Daniel Clarke, Ingo Outes-Leon, and Alemayehu Seyoum Taffesse, "Offering rainfall insurance to informal insurance groups: Evidence from a field experiment in Ethiopia," *Journal of Development Economics*, 2014, *106*, 132–143.

Dotter, Dallas, "Breakfast at the Desk: The Impact of Universal Breakfast Programs on Academic Performance," Association for Public Policy Analysis and Management Online Paper Collection 2014.

Erinosho, Temitope O, Sarah C Ball, Phillip P Hanson, Amber E Vaughn, and Dianne Stanton Ward, "Assessing foods offered to children at child-care centers using the Healthy Eating Index-2005," *J Acad Nutr Diet*, Aug 2013, *113* (8), 1084–9.

FAO, "La gestion de riesgos climaticos catastroficos para el sector agropecuario en Mexico: caso del componente para la atencion a desastres naturales para el sector agropecuario," Technical Report 2014.

Feng, Shuaizhang, Alan B. Krueger, and Michael Oppenheimer, "Linkages among Climate Change, Crop Yields, and Mexico-US Cross Border Migration," *Proceedings of the National Academy of Sciences*, 2010, *107* (32), 14257–14262.

Figlio, David N. and Joshua Winicki, "Food for Thought: The Effects of School Accountability Plans on School Nutrition," *Journal of Public Economics*, 2005, *89*, 381–394.

Frisvold, David E., "Nutrition and Cognitive Achievement: An Evaluation of the School Breakfast Program," *Journal of Public Economics*, 2015, *124*, 91–104.

Fryer, Roland G., "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools," March 2011.

Fuchs, Alan and Hendrik Wolff, "Concept and unintended consequences of weather index insurance: the case of Mexico," *American Journal of Agricultural Economics*, 2011, *93* (2), 505–511.

Gallagher, Justin, "Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the United States," *American Economic Journal: Applied Economics*, 2014, *6* (3), 206–33.

Gallagher, Justin, "Learning About an Infrequent Event: Evidence from Flood Insurance Takeup in the US," *American Economic Journal: Applied Economics*, 2014, *6*, 206–233.

Gertler, Paul J., Sebastian W. Martinez, and Marta Rubio-Codina, "Investing Cash Transfers to Raise Long-Term Living Standards," *American Economic Journal: Applied Economics*, 2012, *4* (1), 164–92.

Giné, Xavier and Dean Yang, "Insurance, credit, and technology adoption: Field experimental evidence from Malawi," *Journal of Development Economics*, 2009, *89* (1), 1–11.

Glewwe, Paul and Edward A. Miguel, "The Impact of Child Health and Nutrition on Education in Less Developed Countries," *Handbook of Development Economics*, 2008, *4*, 3561–3606.

Gomez-Pinilla, Fernando, "Brain foods: the effects of nutrients on brain function," *Nature Reviews Neuroscience*, July 2008, *9*, 568–578.

Gottfried, Michael A., "Excused vs. Unexcused: How Student Absences in Elementary School Affect Academic Achievement," *Educational Evaluation and Policy Analysis*, 2009, *31* (4), 392–415.

GPO, US, "eCFR – Code of Federal Regulations."

Green, Erica L. and Julie Hirschfeld Davis, "Trump Takes Aim at School Lunch Guidelines and a Girls' Education Program," *The New York Times*, 2017.

Greenstone, Michael, "Did the Clean Air Act Cause the Remarkable Decline in Sulfur Dioxide Concentrations," *Journal of Environmental Economics and Management*, 2004, *47* (3), 585–611.

Grossman, Michael and Robert Kaestner, "The Effects of Education on Health," in Jere Behrman and Nancy Stacy, eds., *The Social Benefits of Education*, University of Michigan Press, 1997, pp. 69–125.

Guenther, Patricia M., Kellie O. Casavale, Jill Reedy, Sharon I. Kirkpatrick, Hazel A.B. Hiza, Kevin J. Kuczynski, Lisa L. Kahle, and Susan M. Krebs-Smith, "Update of the Healthy Eating Index: HEI-2010," *Journal of the Academy of Nutrition and Dietetics*, 2013, *113*, 569–580.

Guenther, Patricia M., Kellie O. Casavale, Sharon I. Kirkpatrick, Jill Reedy, Hazel A. B. Hiza, Kevin J. Kuczynski, Lisa L. Kahle, and Susan M. Krebs-Smith, "Diet Quality of Americans in 2001-02 and 2007-08 as Measured by the Healthy Eating Index-2010," Technical Report, US Department of Agriculture, Economic Research Service April 2013.

Hanna, Rema and Paulina Oliva, "The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City," *Journal of Public Economics*, 2015, *122* (0), 68 – 79.

Hazell, Peter B. R., "The appropriate role of agricultural insurance in developing countries," *Journal of International Development*, 1992, *4* (6), 567–581.

Hill, L Bruce, Neil J Zimmerman, James Gooch, and Clean Air Task Force, *A multi-city investigation of the effectiveness of retrofit emissions controls in reducing exposures to particulate matter in school buses*, Clean Air Task Force Boston, MA, 2005.

Hinrichs, Peter, "The Effects of the National School Lunch Program on Education and Health," *Journal of Policy Analysis and Management*, 2010, *29* (3), 479–505.

Hurley, Kristen M., Sarah E. Oberlander, Brian C. Merry, Margaret M. Wrobleski, Ann C. Klassen, and Maureen M. Black, "The Healthy Eating Index and Youth Healthy Eating Index Are Unique, Nonredundant Measures of Diet Quality among Low-Income, African American Adolescents," *Journal of Nutrition*, 2009, *139* (2), 359–364.

Imberman, Scott A. and Adriana D. Kugler, "The Effect of Providing Breakfast in Class on Student Performance," *Journal of Policy Analysis and Management*, 2014, *33* (3), 669–699.

Imberman, Scott A. and Adriana D. Kugler, "The Effect of Providing Breakfast on Student Performance," *Journal of Policy Analysis and Management*, 2014, *33*, 669–699.

INEGI, "Censo de Poblacion y Vivienda," Anexo Estadistico 2010.

IPCC, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2013.

Jackson, C. Kirabo, Cora Wigger, and Heyu Xiong, "Do School Spending Cuts Matter? Evidence from the Great Recession," Technical Report, National Bureau of Economic Research January 2018.

Jacob, Brian A. and Jonah E. Rockoff, "Organizing Schools to Improve Student Achievement: Start Times, Grade Configurations, and Teacher Assignments," *Brookings Institution*, September 2011.

Janzen, Sarah A. and Michael R. Carter, "After the Drought: The Impact of Microinsurance on Consumption Smoothing and Asset Protection," *NBER Working Paper 19702*, 2016.

Jensen, Nathaniel and Christopher Barrett, "Agricultural Index Insurance for Development," *Applied Economic Perspectives and Policy*, 2016, *39* (2), 199–219.

Jensen, Nathaniel D., Christopher B. Barrett, and Andrew G. Mude, "Cash transfers and index insurance: A comparative impact analysis from northern Kenya," *Journal of Development Economics*, 2017, *129*, 14–28.

Johnson, DB, M Podrabsky, A Rocha, and JJ Otten, "Effect of the healthy hunger-free kids act on the nutritional quality of meals selected by students and school lunch participation rates," *JAMA Pediatrics*, 2016, *170* (1), e153918–.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry, "Agricultural Decisions after Relaxing Credit and Risk Constraints," *Quarterly Journal of Economics*, 2014, *129* (2), 597–652.

Krueger, Alan B., "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 1999, *114* (2), 497–532.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach., "School Finance Reform and the Distribution of Student Achievement," *American Economic Journal: Applied Economics*, 2018, *Forthcoming.*

Leos-Urbel, Jacob, Amy Ellen Schwartz, Meryle Weinstein, and Sean Corcoran, "Not Just for Poor Kids: The Impact of Universal Free School Breakfast on Meal Participation and Student Outcomes," *Economics of Education Review*, 2013, *36*, 88–107.

Lobell, David B. and Marshall Burke, eds, *Climate Change and Food Security: Adapting Agriculture to a Warmer World*, Springer, 2010.

Lobell, David B., Wolfram Schlenker, and Justin Costa-Roberts, "Climate Trends and Global Crop Production Since 1980," *Science*, 2011, *333* (6042), 616–620.

Manios, Yannis, Georgia Kourlaba, Katerina Kondaki, Evangelia Grammatikaki, Manolis Birbilis, Evdokia Oikonomou, and Eleytheria Roma-Giannikou, "Diet Quality of Preschoolers in Greece Based on the Healthy Eating Index: The GENESIS Study," *American Dietetic Association*, 2009, *109*, 616–623.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

McIntosh, Craig, Felix Povel, and Elisabeth Sadoulet, "Utility, Risk, and Demand for Incomplete Insurance: Lab Experiments with Guatemalan Cooperatives," *CUDARE Working Paper*, 2015.

Miguel, Edward and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 2004, *72* (1), 159–217.

Miranda, Mario J. and Katie Farrin, "Index Insurance for Developing Countries," *Applied Economic Perspectives and Policy*, 2012, *34* (3), 391–427.

Mobarak, Ahmed Mushfiq and Mark R. Rosenzweig, "Informal Risk Sharing, Index Insurance, and Risk Taking in Developing Countries," *American Economic Review: Papers and Proceedings*, 2013, *103* (3), 375–80.

Monahan, P.A., *Pollution Report Card: Grading America's School Bus Fleets*, Union of Concerned Scientists, 2006.

Moretti, Enrico and Matthew Neidell, "Pollution, health, and avoidance behavior: evidence from the Ports of Los Angeles," *Journal of Human Resources*, 2011, *46* (1), 154–75.

Nandi, Arindam, Jere Behrman, Sonia Bhalotra, Anil Deolalikar, and Ramanan Laxminarayan, "The Human Capital and Productivity Benefits of Early Childhood Nutritional Interventions," *Disease Control Priorities*, 2015.

Patall, Erika A., Harris Cooper, and Ashley Batts Allen, "Extending the School Day or School Year: A Systemic Review of Research (1985-2009)," *Review of Educational Research*, September 2010, *80* (3).

Reedy, Jill, Susan M. Krebs-Smith, and Claire Bosire, "Evaluating the food envirionment, Application of the Healthy Eating Index-2005," *American Journal of Preventive Medicine*, 2010, *38* (5), 465–471.

Roland G. Fryer, Jr., "Injecting Charter School Best Practices into Traditional Public-Schools:Evidence From Field Experiments*," *Quarterly Journal of Economics (2014)*, 2014, *129* (3), 1355–1407.

Rosenzweig, Cynthia, Francesco N. Tubiello, Richard Goldberg, Evan Mills, and Janine Bloomfield, "Increased crop damage in the US from excess precipitation under climate change," *Global Environmental Change*, 2002, *12* (3), 197–202.

Rosenzweig, Mark R. and Hans P. Binswanger, "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments," *The Economic Journal*, 1993, *103* (416), 56–78.

Rosenzweig, Mark R. and Kenneth I. Wolpin, "Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India," *Journal of Political Economy*, 1993, *101* (2), 223–244.

Rothstein, Jesse, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *The Quarterly Journal of Economics*, 2010, *125* (1), 175–214.

Sabin, Lisa D, Eduardo Behrentz, Arthur M Winer, Seong Jeong, Dennis R Fitz, David V Pankratz, Steven D Colome, and Scott A Fruin, "Characterizing the range of children's air pollutant exposure during school bus commutes," *Journal of Exposure Science and Environmental Epidemiology*, 2005, *15* (5), 377–387.

Sadoulet, Elisabeth, Alain de Janvry, and Benjamin Davis, "Cash Transfer Programs with Income Multipliers: PROCAMPO in Mexico," *World Development*, 2001, *29* (6), 1043–1056.

SAGARPA, "Sistema de informacion para el seguimiento de la operacion de los seguros agropecuarios catastroficos," Technical Report, Mexico City: SAGARPA 2014.

Schanzenbach, Diane Whitmore, "Do School Lunches Contribute to Childhood Obesity?," *Journal of Human Resources*, 2009, *44* (3), 684–709.

Schanzenbach, Diane Whitmore and Mary Zaki, "Expanding the School Breakfast Program: Impacts on Children's Consumption, Nutrition and Health," National Bureau of Economic Research Working Paper 20308 July 2014.

Schlenker, Wolfram and Michael J. Roberts, "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change," *Proceedings of the National Academy of Sciences*, 2009, *106* (37), 15594–15598.

Schlenker, Wolfram and Reed Walker, "Airports, Air Pollution, and Contemporaneous Health," *Review of Economic Studies*, 2016, *83* (2), 768–809.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge, "What Are We Weighting For?," *Journal of Human Resources*, 2015, *50* (2), 301–316.

Sorhaindo, Annik and Leon Feinstein, "What is the relationship between child nutrition and school outcomes," Technical Report, Centre for Research on the Wider Benefits of Learning June 2006.

Thurston, George D, "Particulate matter and sulfate: Evaluation of current California air quality standards with respect to protection of children," *Prepared for CARB and OEHHA*, 2000.

USDA, "Strategic Plan for FY 2005-2010," June 2006.

USDA, "Comparison of Previous and Current Regulatory Requirements under Final Rule "Nutrition Standards in the National School Lunch and School Breakfast Programs"mparison of Previous and Current Regulatory Requirements under Final Rule "Nutrition Standards in the National School Lunch and School Breakfast Programs"," January 2012.

USDA, "Nutritional Standards in the National School Lunch and School Breakfast Programs," January 26 2012.

USDA, "USDA Proposes Standards to Provide Healthy Food Options in Schools," February 2013.

USDA, "Healthy Eating Index," 2016.

Volpe, Richard and Abigail Okrent, "Assessing the Healthfulness of Consumers' Grocery Purchases," Technical Report EIB-102, US Department of Agriculture, Center for Nutrition Policy and Promotion November 2012.

Wargo, J., D. Brown, M.R. Cullen, and Environment & Human Health Inc., *Children's Exposure to Diesel Exhaust on School Buses*, Environment & Human Health, Incorporated, 2002.

Yang, Dean and HwaJung Choi, "Are remittances insurance? Evidence from rainfall shocks in the Philippines," *The World Bank Economic Review*, 2007, *21* (2), 219–248.

Appendix A

A.1 Weather index insurance and shock coping: Evidence from Mexico's CADENA program — Appendix

A.1.1 Additional summary statistics

Table A.1: Municipality summary statistics

Variable	All municipalities	ENIGH municipalities	p-value
Population	50,115.3	102,765.5	0.0
Years of education	6.8	7.0	0.1
% Illiterate	13.3	12.3	0.1
% No running water	16.5	16.7	0.9
% Dirt floor	8.8	8.9	0.9
% Indigenous population	11.5	10.9	0.8
Female % of employed pop	25.7	25.8	0.9
Male % of employed pop	74.3	74.2	0.9

Note. The data used to created this table comes from the 2010 census. % Illiterate is defined as the number of illiterate individuals over the age of 15 divide by the total population over 15.

A.1.2 Model

Consider a farmer that holds some amount of wealth, w, at time $t = 0$. The farmer chooses to allocate his wealth w between consumption today c_0 and inputs into farm production that will result in output that he can consume in period $t = 1$. In terms of production, the farmer has two margins along which he can invest: expanding the amount of land sowed (A) at a cost r per unit of land and/or increasing the amount of fertilizer x at a cost of p per unit of fertilizer. Conditional on a given choice of inputs, the output at time $t = 1$ depends on which of the two possible states of nature, drought or no drought, is realized. We denote the production in a good year (no drought) as $f_g(x, A)$, and the production in a bad year (drought) to be $f_b(x, A)$. As a result of the area-based index insurance, the farmer also receives a payment of m per area sowed in the bad state of the world.[1] Thus, the farmer's maximization problem is:

$$\max_{A,x} u(c_0) + \delta[\alpha u(c_1^b) + (1-\alpha)u(c_1^g)]$$

subject to:

$$\begin{aligned} c_0 &= w - (Ar + px) \\ c_1^g &= f_g(x, A) \\ c_1^b &= f_b(x, A) + mA \end{aligned}$$

Furthermore, we place the following restrictions on the functional forms:

$$\begin{gathered} f_g(x, A) > f_b(x, A) \ \forall x \\ \frac{\partial f_g}{\partial x} > \frac{\partial f_b}{\partial x} \end{gathered}$$

The first order conditions for this maximization are:

$$\begin{aligned} -u'(c_0)r + \delta\left[\alpha u'(c_1^b)\left(\frac{\partial f_b}{\partial A} + m\right) + (1-\alpha)u'(c_1^g)\frac{\partial f_g}{\partial A}\right] &= 0 \\ -u'(c_0)p + \delta\left[\alpha u'(c_1^b)\frac{\partial f_b}{\partial x} + (1-\alpha)u'(c_1^g)\frac{\partial f_g}{\partial x}\right] &= 0 \end{aligned}$$

In order to generate figure 1.4, we chose the following functional form for the utility and production functions, and numerically solved for the optimal values x^* and A^* for different values of w:

$$\begin{aligned} u(c) &= ln(c) \\ f_g(x, A) &= x^{\frac{1}{2}} A^{\frac{1}{2}} \\ f_b(x, A) &= 0.8x^{\frac{1}{2}} A^{\frac{1}{2}} \end{aligned}$$

A.2 School Lunch Quality and Academic Performance — Appendix

A.2.1 Calculation of vendor HEI scores

Nutritionists at the Berkeley Nutrition Policy Institute conducted an analysis of menus for those vendors for whom this information was available. A copy of the com-

[1] This assumption ignores the issues of basis risk, but the main result holds so long as the probability of payment is higher during a drought than in a normal year because in this setting farmers are not responsible for paying insurance premiums.

plete report can be found here: `http://faculty.weatherhead.case.edu/jpg75/pdfs/Nutrition-Policy-Institute-July-2016.pdf` The menus were scored using the Healthy Eating Index (HEI). The process to calculate the HEI for each vendor was the following:

1. Nutrition information was gathered from vendors. This process included obtaining the full menu of offerings and nutritional information by contacting vendors. When this information was not available, sample menus from client school districts (with or without nutritional information) were used.

2. In order to calculate HEI, it was necessary to match foods listed in vendor menus to USDA food codes.

   a) For vendors with nutritional information available, vendor foods were matched to USDA foods using the What's in the Foods You Eat online search tool

      - Foods were first matched by names. Then, these matches were analyzed based on calories and fat content to determine how many USDA units corresponded to a vendor's portion.
      - Units were calculated so that calories, total fat, and saturated fat matched within 20% difference.
        - Entrees, meat/meat alternatives, and whole grain items were also matched by protein and fiber.
      - A coding system was created to denote the quality of the match.

   b) For vendors without nutritional information, the number of total calories and other nutrients had to be imputed to determine the number of USDA units corresponding to a vendor's portion. In these cases, a number of methods were tested, which included using the average calories for other vendors and USDA defaults (e.g., the necessary amounts to meet USDA guidelines)

   c) USDA HEI SAS macros were used to determine HEI scores (scoring system 1)

3. A supplemental scoring system was created to include additional food categories commonly found in school lunch menus. This method was reviewed by five nutrition experts.

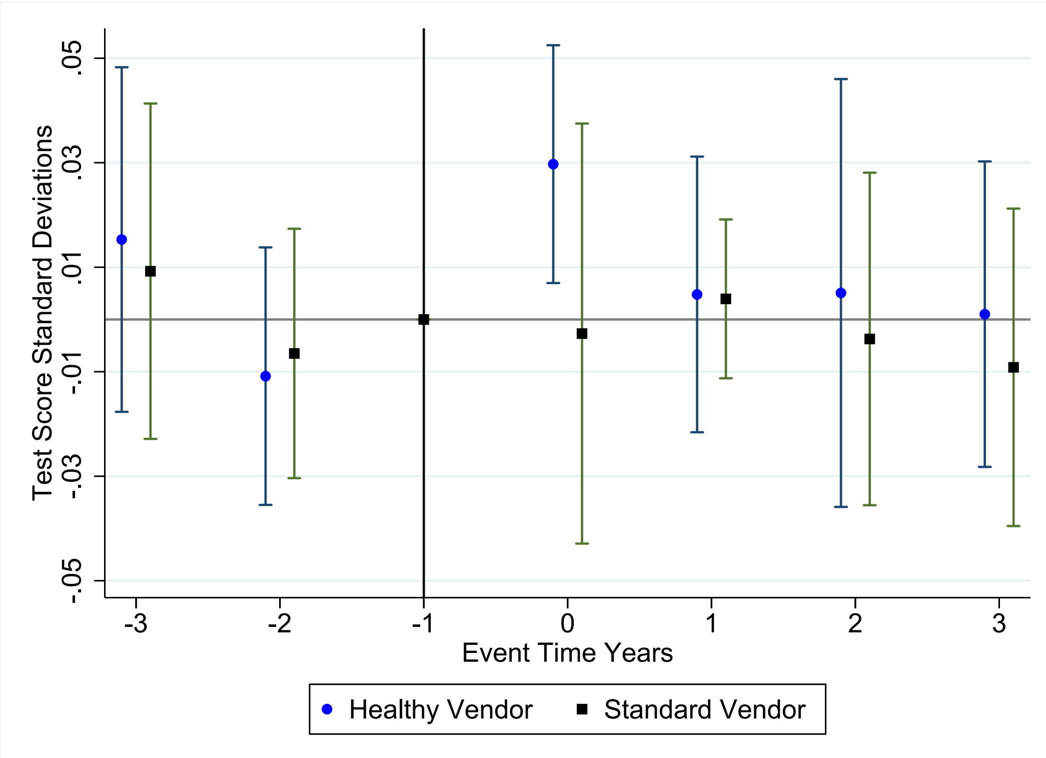
4. The HEI and supplemental scores were combined to calculate alternative total scores (scoring system 2).

A.2.2 Calculation of average school lunches served

National School Lunch Program (NSLP) data were obtained from the California Department of Education's Nutrition Services Division for the school years 2008-2009 through 2012-2013. The data report the average number of NSLP lunches served per operating day in each school district. Averages are calculated monthly, so in order to obtain an annual measure for the average number of lunches served per day, we multiply the monthly averages by the number of operating days in each month and sum the monthly totals. The months of June and July are excluded from the total because these months may correspond to summer lunch programs that are managed separately. The annual total is divided by the total number of operating days in the year, again excluding June and July, to calculate an annual average of lunches served per day. Lastly, we divide the number of lunches served per day by the total enrollment in the school district to eliminate changes in lunches served that are due only to changes in the number of enrolled students. Because we are interested in separately estimating the effect on economically disadvantaged students, we calculate averages for both total lunches served and free and reduced-price lunches. A student is eligible for a free school lunch if his family's income is less than 130% of the poverty level, and a reduced-price lunch if her family's income is between 130% and 185% of the poverty level. The CA Department of Education refers to these students as "economically disadvantaged."

A.2.3 Tables and Figures

Figure A.1: Placebo Test for the Effect of Contracting with a Vendor on Test Scores



Note. This figure depicts point estimates for treatment leads and lags with their corresponding 95% confidence intervals. Negative event time years are for years prior to having a vendor. Positive event time years are for years after a vendor contract ends. The point estimates come from a weighted regression of Equation (2.2), except that the coefficients for $\tau > 0$ are redefined to be years after the end of a contract. The event time period -3 (3) is pooled to include both period -3 and -4 (3 and 4) to improve precision. The regression includes the same control variables as Table 2.2, Column (3). Standard errors are clustered at the school district level.

Table A.2: Vendor Healthy Eating Index (HEI) Score by School Lunch Market Share

Vendor Name	(1) Percent of Students Served	(2) HEI Score	(3) Healthy	(4) SFA Contracts	(5) Only SFA Contracts
Sodexo	50.60	59.9	N	N	N
Compass	15.65	45.6	N	N	N
CSU Dominguez Hills	10.27	-	N	N	N
Preferred Meals	9.50	71.4	Y	N	N
Aramark	3.33	64.7	Y	N	N
Revolution Foods	2.57	92.3	Y	Y	N
Royal Dining	2.42	75.0	Y	Y	N
Choicelunch	1.21	37.1	N	N	N
The Lunchmaster	1.12	-	N	N	N
School Nutrition Plus	0.86	67.8	Y	N	N
Kid Chow	0.42	26.8	N	N	N
Morrison Management Specialists	0.32	-	N	N	N
Bellflower Unified School District	0.29	69.1	Y	Y	Y
CSU Chico	0.23	51.9	N	N	N
Unified Nutrimeals	0.16	-	N	N	N
Flour Creations	0.13	-	N	N	N
Feed You Well	0.13	-	N	N	N
La Luna On The Go	0.09	-	N	N	N
Fresno County EOC	0.09	-	N	Y	N
Preferred Choice	0.09	70.0	Y	N	N
Santa Clarita Food Services Agency	0.07	55.5	N	N	N
Oceanside Unified S.D. / Lighthouse Foods	0.07	-	N	Y	Y
Dulan's Catering	0.06	-	N	N	N
Arguello Catering	0.05	-	N	N	N
The Food Lady	0.03	-	N	N	N
Banyan Catering	0.03	-	N	N	N
Blue Lake Rancheria	0.03	72.2	Y	N	N
Food Management Associates	0.02	-	N	N	N
Good Day Cafe - San Lorenzo Unified S. D.	0.02	-	N	Y	Y
Brown Bag Naturals	0.02	-	N	N	N
Fieldbrook Family Market	0.02	39.7	N	N	N
Freshlunches	0.01	-	N	N	N
Happy Valley Conference Center	0.01	-	N	N	N
Aqua Terra Culinary	0.01	-	N	N	N
Trinidad Rancheria	0.01	-	N	N	N
Food 4 Thought	0.01	-	N	N	N
San Bernardino School District	0.01	-	N	Y	Y
Progressive Catering	0.01	-	N	N	N
Hesperia USD	0.01	-	N	Y	Y
Healthy Lunch And Lifestyle Project	0.01	-	N	N	N
James Aldrege Foundation	0.01	-	N	N	N
Taft City School District	0.00	-	N	Y	Y
Arcata School District	0.00	-	N	Y	Y
Celebrations Catering	0.00	-	N	N	N
Yosemite Unified School District	0.00	-	N	Y	Y

Note. The table lists the 45 vendors that contracted with schools during the 2008-2009 to 2013-2014 school years. The HEI scores for each vendor are based on lunch menus and calculated by nutritionists at the Nutrition Policy Institute. The percent of students served by each vendor is the number of test-takers (as reported in STAR data) in a school district served by a particular vendor divided by the total number of students served by any outside vendor.

Table A.3: Test-Taker Covariates for Schools that Contract with School Lunch Vendors

Sample:	All School Sample			Contract School Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Vendor	No Vendor	Difference	Healthy	Standard	Difference
Disadvantaged	0.461	0.528	-0.067***	0.459	0.420	0.039
			(0.012)			(0.028)
Asian	0.079	0.053	0.026***	0.041	0.100	-0.059***
			(0.004)			(0.010)
White	0.384	0.443	-0.059***	0.404	0.327	0.077**
			(0.012)			(0.025)
Hispanic	0.408	0.406	0.002	0.364	0.408	-0.044
			(0.012)			(0.026)
Black	0.053	0.040	0.012***	0.075	0.032	0.043***
			(0.003)			(0.008)
English Learner	0.217	0.196	0.021**	0.175	0.230	-0.054***
			(0.008)			(0.016)
Districts	573	4625		231	313	

Notes: Percentages are calculated by dividing the number of enrolled students in a given category by the total number of enrolled students in a district as reported by the California Department of Education. Standard errors reported in parentheses. The contract sample is the subset of districts that contract with any vendor at any point during our sample. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.4: The Effect of Vendor Choice on Standardized Test Scores (Unweighted)

Dependent variable:	Standardized Test Score					
Sample:	All Schools			Contract Schools		
	(1)	(2)	(3)	(4)	(5)	(6)
Healthy Vendor	0.031***	0.031***	0.030***	0.037***	0.028***	0.025**
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Standard Vendor	-0.011	-0.014	-0.013	0.008	-0.022	-0.021
	(0.020)	(0.022)	(0.022)	(0.017)	(0.023)	(0.022)
School-by-grade FEs		X	X		X	X
Covariates included			X			X
R^2	0.615	0.909	0.909	0.700	0.903	0.903
N	174,818	174,818	174,818	22,133	22,133	22,133
Schools	9,719	9,719	9,719	1,188	1,188	1,188

Notes: Each column represents a separate regression estimated on a common sample that excludes observations with missing covariates. Observations are at the school-grade-year level. Standard errors clustered at the school district level appear in parentheses. All regressions include school and year fixed effects. Regressions also include school fixed effects unless school-by-grade fixed effects are specified. The contract school sample is the subset of schools that contract with a vendor at some point during our sample. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

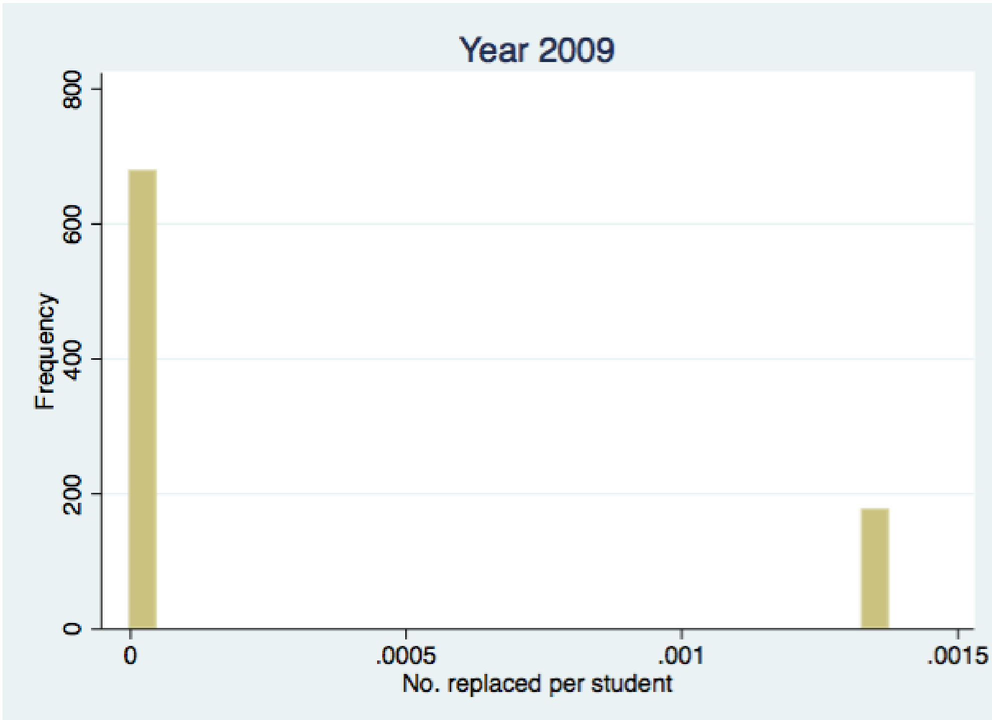
A.3 The Effect of Air Pollution on Children's Educational Outcomes: An Evaluation of California's Lower-Emission School Bus Program — Appendix

Table A.5: The effect of bus replacements on test scores: School district analysis

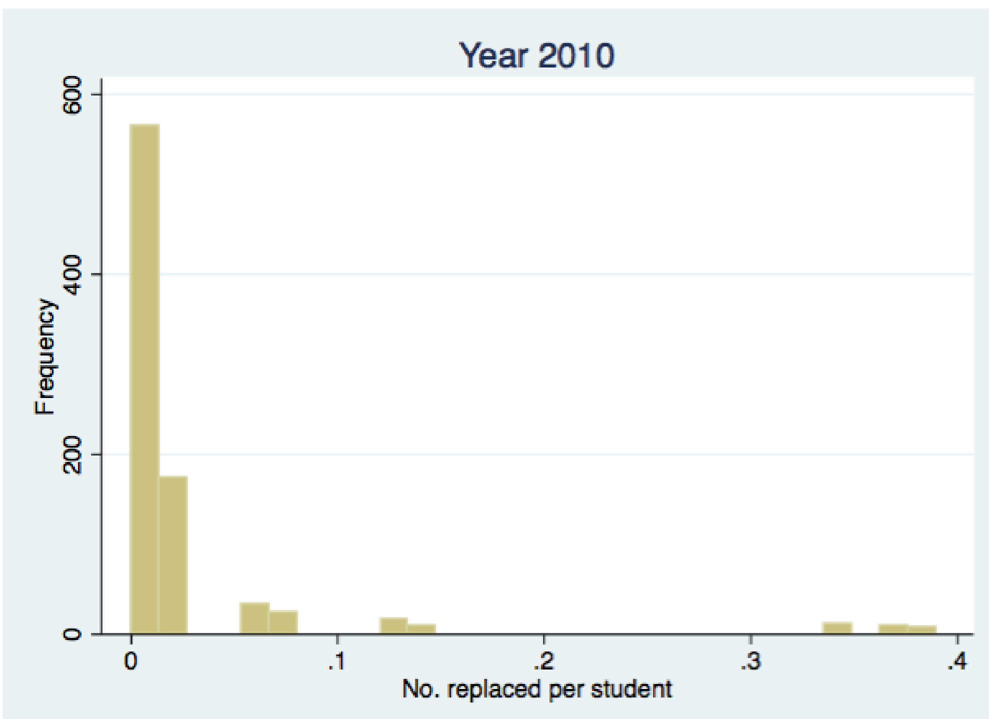
	(1) All	(2) All	(3) Disadvantaged	(4) Disadvantaged
No. replaced/student	0.00505	-0.00889	0.00720	-0.00124
	(0.0200)	(0.0210)	(0.0156)	(0.0194)
Weights	Y	N	Y	N
Covariates	Y	Y	Y	Y
N	3386	3386	3386	3386
R^2	0.940	0.931	0.910	0.887

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at air district level. Regressions include school district and year fixed effects. Number of students enrolled per grade-school observation are used as weights in weighted regressions.

Figure A.2: Variation in number of buses replaced per student by year (2009-2010)

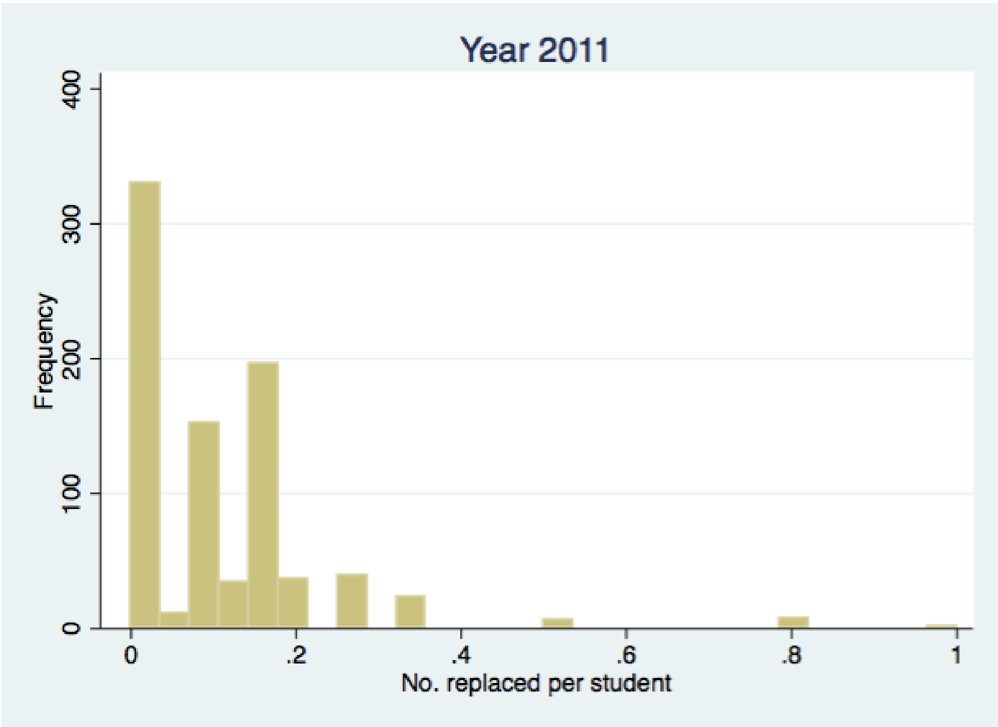


(a) Variation in treatment for 2009

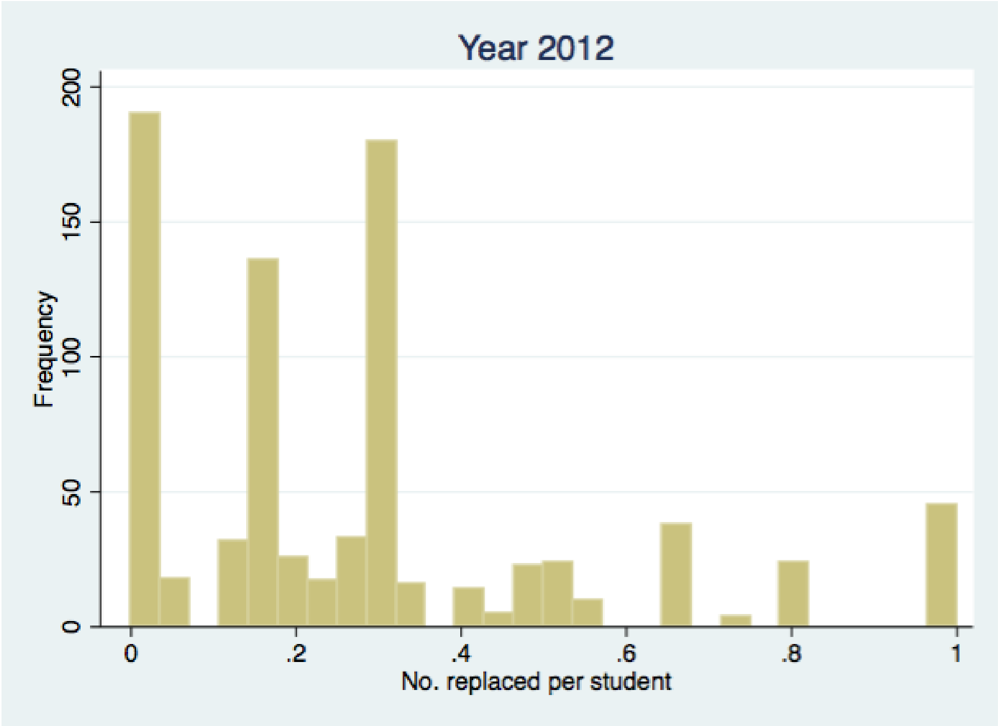


(b) Variation in treatment for 2010

Figure A.3: Variation in number of buses replaced per student by year (2011-2012)



(a) Variation in treatment for 2011



(b) Variation in treatment for 2012