Unintended Consequences of Education and Housing Reform Incentives

by

David P. Sims

B.S. Economics
Brigham Young University, 1999

Submitted to the Department of Economics in
Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the
Massachusetts Institute of Technology

September 2004

Signature of Author..
Department of Economics
August 10, 2004

Certified by..
Joshua Angrist
Professor of Economics
Thesis Supervisor

Certified by..
Daron Acemoglu
Charles P. Kindleberger Professor of Applied Economics
Thesis Supervisor

Accepted by..
Peter Temin
Elisha Gray II Professor of Economics
Chairperson, Departmental Committee on Graduate Studies

Unintended Consequences of Education and Housing Reform Incentives

by

David P. Sims

Submitted to the Department of Economics in
Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Economics.

ABSTRACT

This thesis measures some unintended consequences of government education and housing policies. Chapter 1 estimates the net educational effect, measured by student test scores, of the California Class Size Reduction Program on second and third graders. This program inadvertently created incentives for schools to combine students in multiple grade classrooms as well reduce class size. Using the non-linear relationship between enrollment and combination classes I estimate that students placed in combination classes by the program suffered a large, significant drop in test scores. I also find little evidence of positive achievement effects due to smaller class size suggesting that the program's net effect may have been negative.

Chapter 2 seeks to identify the effects of rent control on cities in the Boston area using the variation provided by a 1995 Massachusetts ballot initiative banning rent control. My findings support the intuition economist derive from simple economic models of price ceilings. Though rent controls achieve their stated aim of lowering rents, they also decrease the willingness of owners to rent apartments, lead to housing unit deterioration, and result in inefficiently long tenancy durations. I also find suggestive evidence that the deterioration in rent controlled housing quality may lower the rent in nearby non-controlled units.

Chapter 3 examines an unintended strategic response of school districts to accountability testing. Using Wisconsin data I show that some school districts advance the starting date of their school year to allow their students more time to prepare for state accountability tests. I find that this leads to small test score gains in math, but may lead to higher absence rates and reduced reading scores among low achieving children.

Thesis Supervisor: Joshua Angrist
Title: Professor of Economics

Thesis Supervisor: Daron Acemoglu
Title: Charles P. Kindleberger Professor of Applied Economics

Acknowledgements

I am grateful to Joshua Angrist, Daron Acemoglu and David Autor for comments and advice on the papers in this dissertation and on many other aspects of academic work. Their time and dedication was an immense asset in all aspects of graduate school. I appreciate the insight and advice of Melissa Boyle, Patrick Buckley, Alexis Leon, Cindy Perry, and Michael Steinberger whose optimism and positive attitude were more helpful than they realize. I also thank numerous other participants in field lunches and seminars. Most of all, I thank my wife Susan, for her patience, understanding and support.

Contents

Introduction

This thesis measures some unintended consequences that arise when government policy makers fail to account for the incentives provided by their legislative implementation and institutional structure. In the case of rent control, the possible perverse incentives of this government policy are well understood by economists, but the magnitude of the unintended consequences has not been adequately measured. In the case of two important educational reforms, class size reduction and accountability testing, I explain some perverse incentives that were unrecognized by policy makers as well as measuring their impact.

In 1996 the State of California implemented an extensive and costly program to reduce class sizes for K-3 students by an average of ten pupils. The program's non-linear reward scheme provided incentives for schools to both reduce class size by creating new classes and to smooth class size across grades by creating combination classes. Chapter 1 uses the rules created by the class size reduction policy to generate instruments for both class size and the percentage of students taught in combination classes. I find that the use of combination classes has a negative and significant effect on the test scores of second and third grade students in California. Furthermore, this negative effect counteracts any positive effect of class size reduction for students placed in combination classes due to program incentives. While smaller classes may be beneficial, the negative effects of combination classes are large enough that the net effect of the California Class Size Reduction Program may have been negative.

Chapter 2 uses the sudden end of rent control in Massachusetts in 1995 to estimate the effects of rent control. I examine Boston MSA data from the American

Housing Survey years 1985-1998 to determine how rent control affected the supply, price and quality of rental housing. My results suggest rent control had little effect on the construction of new housing but did encourage owners to shift units away from rental status. Rent control also led to a small deterioration in the quality of rental units, operating primarily through smaller items of physical damage. Larger maintenance items such as plumbing or heating systems were unaffected. I also examine specifications that allow rent control to affect rent levels both directly through controlled status and indirectly through spillover effects from nearby rent controlled units. These estimates imply that rent control may have small negative effects on the price of the non-controlled rental housing stock

The adoption of state accountability testing in the 1990s coincided with the movement of school start dates from September into August. Using data from Wisconsin and Texas, Chapter 3 connects these phenomena, showing that some low scoring districts advanced their school start dates to allow their students more time to prepare for exams. I use a 2001 Wisconsin state law that restricted districts to start dates after September 1st to identify the effects of this extra time on student achievement. Extra classroom days led to small increases in Math scores for 4th graders, but did not increase students' average reading or language scores. Extra classroom time may also have increased third grade reading scores for students in the upper portion of the ability distribution while reducing achievement for those of lower ability. This could be due to an increased absence rate caused by early school start dates.

Chapter 1 - How Flexible is Educational Production? Combination Classes and Class Size Reduction in California

1.1 Introduction

The California Class Size Reduction Program, adopted in 1996, was one of the largest state education reforms of the decade. Though a number of states adopted measures to reduce the size of elementary school classes, the California program was the most ambitious in scope, affecting millions of kindergarten through third grade students at a cost of several billion dollars. Instead of providing a scale of rewards based on reductions in average class size, the program rules provided an "all or nothing" payment for schools that met a threshold requirement of fewer than twenty students per class. Though policy makers intended to provide schools with incentives to hire more teachers and create more classes, the non-linear reward structure they created also provided schools with incentives to smooth class size across grades by creating combination classes, thereby reducing maximum class size without increasing the number of teachers or lowering average class size.

The effect of incentive schemes in public education has been of recent interest to economists. Acemoglu, Kremer and Mian (2003) argue that high-powered incentives create distortions in educational production. They claim that the government provides most elementary education because it can avoid the high powered incentives offered by firms. [1] Recent research by Jacob (2002) and Jacob and Levitt (2003), shows that

[1] There is a large literature on the perverse effects of non-linear incentives in firms. For example, the multitasking literature beginning with Holmstrom and Milgrom (1991) details the costly nature of providing high powered incentives to an increasingly flexible workforce. Chevalier and Ellison (1997) demonstrate that mutual funds alter their holdings at the end of the year in response to returns earlier in the

perverse effects can come from adding high-powered incentive programs in public education. They find that the adoption of high stakes testing in Chicago led to teachers teaching to the test and helping their students cheat.

This paper also contributes to the literature on combination classes. A combination class is an otherwise normal, self-contained classroom in which multiple grades are generally taught by the same instructor. Though there is little work by economists on the subject, there is a large education literature. Despite the large volume of empirical work, the conclusions are far from uniform. Some studies such as Russell, Rowe, and Hill (1998) find negative consequences of combination classes, while others such as Pavan (1992) find that combination classes enhance academic achievement.[2]

Likewise, theoretical arguments over combination classes remain unresolved. The advocates of combination classes insist that they foster cooperation and critical thinking. Opponents argue that such classes breed confusion and resentment among students who have difficulty working together. Teachers often claim that combination classes are more difficult to instruct.

The debate over the desirability of combination classes can be seen as a debate about academic tracking. Mixing students across grade level leads to a wider range of student interest and ability levels within a classroom due to students' differing levels of experience in an academic setting and prior exposure to certain material. Since diversity

---

year, since managers do not have incentives to maximize fund value but to meet targets. Oyer (1998) shows that managers and salespeople respond to quota incentives by varying sales and prices over the fiscal year, particularly in the final quarter.

[2] A review of this education literature can be found in Veenman (1995), and Guttierez and Slavin (1992).

is a primary aim of non-tracking initiatives, a program that expands combination classes can be thought of as an experiment in a variety of extreme non-tracking.[3]

This paper develops a simple model that relates two forms of classroom organization, class size and student homogeneity, to illustrate the effect that creating combination classes may have in the context of class size reduction. I then estimate the impact of class size and the percentage of students in combination classes on student achievement using instruments derived from the non-linear relationship between enrollment and classroom organization.

I find that the use of combination classes reduces student test scores. The negative effects are larger for third graders than second graders. Estimates of class size effects are small and statistically insignificant. Using generous outside estimates for the effect of class size, I conclude that the 4-5% of students placed in combination classes by the program are worse off than in the absence of such a program. If class size effects are small then the program had a net negative effect on student achievement.

The remainder of the paper is organized as follows: Section two describes the institutional background and theoretical framework, section three discusses the data and identification strategy, section four presents the results and interpretation, and section five concludes.

[3] The literature on tracking has long considered it a disastrous policy for poor students (eg. Slavin 1990). However, recent empirical findings cast doubt upon this conventional wisdom. For example, Figlio and Page (2002) show that corrections for endogeneity and selection lead to positive estimates of the impact of tracking on the test scores of poor students.

1.2 Background

1.2.1 The California Class Size Reduction

The California Class Size Reduction Program arose from an unexpected political alliance in the summer of 1996. At the time, the state had a budget surplus and there was widespread interest in using the money to improve primary education. A large portion of the legislature favored a program to reduce class size, then about thirty students per class. Governor Wilson, on the other hand, supported a program offering school vouchers to students. Other initiatives were also discussed.

As the 1996-97 school year approached with the prospect of no major reform, the governor gave his support to the class size reduction advocates. The resulting law took effect less than a month before the school year began. Schools scrambled to adopt the program but many did not have time to fully implement it in the first year.

The Class Size Reduction Program provided incentives for schools to voluntarily reduce their class sizes in the early grades. The state committed to pay each school district $650 dollars for every student in a participating program grade. A school grade was considered a participant if it was in a participating district and had all of its students in that grade in a class of twenty students or fewer. This payment was sizeable relative to California's 1995-96 per pupil expenditure of $6,068. The payment amount steadily increased in subsequent years and stood at $906 in 2002-03. Anticipating a lack of classroom space the state also arranged to subsidize the procurement of temporary classrooms with payments of $25,000. After the first year this subsidy rose to $40,000.

Schools were required to reduce class sizes in a particular order. To participate in the program, schools were required to reduce the class size of first graders. Only when first grade class sizes were below twenty could a school receive money for reducing the size of their second grade classes. A school that had reduced class sizes for first and second graders could receive program money for reducing the size of either kindergarten or third grade classes. After the first year, the program was amended to allow reduction of both kindergarten and third grade classes, though schools still had to reduce first and second grade classes.

The large awards offered by the state led to high program participation rates. Table 1 shows the level of participation and overall participation percentage in the first few years of the program. In the first year, two-thirds of first graders were in classes that qualified for a subsidy, though few kindergarteners and third graders were. However, in subsequent years participation became nearly universal in first and second grade and reached considerable levels in the other grades. By year three, all grade levels exceeded eighty percent participation and by the fourth year all grade levels had a participation rate of over ninety percent.

California's Class Size Reduction Program was extraordinarily expensive. In its first year, including payments for classroom space, the program committed the state to over $1.3 billion in payments. This number increased as per student award and participation levels rose so that by 2001 the subsidy payments constituted six percent of the state education budget.

Previous research on the California Class Size Reduction includes a state commissioned evaluation by a consortium of five companies. Their report found modest gains in student achievement associated with the reduction in class size. Since the program had been offered to all school districts, there was not a clear control group for the study. The consortium's primary solution was to use a difference-in-difference estimator based on the difference in fifth versus third grade test scores in adopting and non-adopting schools. This strategy assumes that the program did not affect fifth grade students. This seems unlikely since fifth grade students may have seen their class size increase and teacher characteristics change as teachers with seniority were transferred to lower grades.

The study by Rivkin and Jepsen (2003) uses variation in the timing of program adoption to identify the effects of smaller class size on test scores. They find large and significant effects, especially for students in poorer districts. They also investigate a potentially perverse effect mentioned in the consortium report. The Class Size Reduction Program forced many districts to hire new teachers with little experience and incomplete credentials. Hanushek, Rivkin and Kain (1998) suggest that inexperienced teachers reduce student achievement. Rivkin and Jepsen argue that the influx of inexperienced teachers in California reduced student test scores, especially in heavily African-American schools. They suggest that the CSR program had net positive effects but increased educational inequality. Nevertheless, the influx of inexperienced teachers represents a short run adjustment rather than a lasting problem.

A feature of the Class Size Reduction program which has not drawn attention is the incentives it provided to use combination classes. Students from eligible grades in combination classes qualified for program money as long as the size of the combination class was below twenty students. This applied even when some of the students in the class were not from eligible grades. For example, a class of eighteen third graders and two fourth graders received a payment for the eighteen third graders. In practice schools were far more likely to combine classes within eligible grades as this reduced the inefficiency of putting students that did not qualify for the subsidy in a smaller class.

The remainder of the paper demonstrates that these incentives led to reduced educational achievement and smaller class size reductions.

1.2.2. Theoretical Background

This section outlines an education production function that relates classroom structure to student outcomes. Following Lazear (2001), classroom instruction time is considered a public good and the amount of classroom time available to the teacher is assumed to be scarce and fixed. Any time that a teacher must spend working with an individual student, whether on material the rest of the class understands, individual questions, or discipline, is time not producing classroom instruction. This is a type of congestion effect. If students in combination classes are more likely to require individual teacher time than students in a single-grade class, then combination classes will have less time for learning and lower achievement. In addition to its simplicity, this model agrees with anecdotal evidence provided by teachers that combination classes are harder to control and short on time.

Consider the following setup: With probability q a student does not require individual attention from the teacher in a one unit period of classroom time. If each student's behavior is independently determined, the probability that a classroom of n such students has no interruptions and that learning takes place is q^n. In this hypothetical classroom the learning of one time period has a value of w. The per student value of classroom educational production is:

$$wq^n \qquad (1)$$

This simple formulation highlights two features of classroom production. First, since $q<1$, the marginal effect of increasing class size is negative. Second, the marginal effect of increasing q is positive. Decreasing class size has a positive effect on per student educational production while increasing the need for teacher time spent with individual students reduces learning. Now envision two hypothetical classrooms, both the same size, one of which has only students from grade g, while the other combines a percentage α students from g with students from another grade j. This combination class has lower per student educational production if:

$$q_g^{\,n} > q_g^{\,\alpha n} q_j^{\,(1-\alpha)n} \qquad (2)$$

This inequality depends on the relative magnitudes of q_g and q_j. One assumption that conforms to common ideas about childhood behavior is that disruptiveness is a function of age. If the need for teacher time is strictly a function of age then the older grade will have a larger q than the younger one. This means that the inequality in (2) will be true if $j<g$ but false in the opposite case. In this situation, the older students lose by being in a combination class, while the younger students benefit.

Another plausible assumption is that the need for teacher attention may be higher when a student is in a classroom where the teacher covers multiple curricula than in a classroom with a common curriculum. In this case, combination classes are disruptive. The inequality in (2) becomes:

$$q_g^{\,n} > q_{gc}^{\,\alpha n} q_{jc}^{\,(1-\alpha)n} \qquad (3)$$

where $q_{gc} < q_g$ is the probability that a student in grade g does not require teacher attention.

This inequality depends both on the relative magnitudes of student disruptiveness and on the previous assumption about student age. If age is not a factor in determining q, then the inequality in (3) will be true and the combination class will have lower educational output. If age is a factor, students in grade g will have lower per student output when combined with students from lower grades. However, if they are combined with students from a higher grade, the effect is ambiguous and depends on the ordering of q_g, q_j, q_{gc}, and q_{jc}. The model predicts that students combined with those in lower grades will always be worse off, while students combined with those in higher grades may fare better or worse depending on how disruptive the various groups are.

The model can also show the class size reduction level required to offset the effects on grade g students of their class becoming a combination class. Assuming that $q_g > q_j$ this quantity can be found by solving a slight variation on the inequality in (2), namely:

$$q_g^{n^*} = q_g^{\alpha n} q_j^{(1-\alpha)n} \qquad (2')$$

where n* represents the original class size that would make the students in grade g indifferent to moving into a smaller combination class. After some algebra this becomes:

$$n^* = (1-\alpha)n\,[(\log q_j)/(\log q_g)] + \alpha n \qquad (2'')$$

1.3 Data and Identification

1.3.1 Data

This paper draws upon two data sources. Data from the Standardized Testing and Reporting (STAR) program were provided by the assessment division of the California Department of Education. The STAR program began with the administration of standardized tests to students in grades two and above in the spring of 1998.[4] In 1998-2000, the test years used in this paper, the elementary STAR included the Stanford 9 norm-referenced test.

I use scores for second and third graders from both the mathematics and language sections of the test to measure educational achievement. These scores are available on a school by grade level basis rather than a classroom by classroom basis. I use the National Percentile Rank (NPR) of a hypothetical mean student in a particular grade for a specific school in math or language as dependent variables.

The rest of the data came from the Educational Demographics Office of the California Department of Education. These included detailed reports from schools and teachers about their classes, and contained information on a variety of teacher characteristics such as experience, education level, class sizes, and demographics. The data also provided demographic information including the number and ethnicity of students in each grade, the number of English learners, and the number of students receiving free or subsidized meals. I aggregate this data to the school-grade level where necessary and match it to test scores.

The dataset used in my analysis consists of observations on second graders from the 1998-2000 test years and third graders from the 1999-2000 test years.[5] I eliminate observations for which the necessary demographic and testing information is unavailable

[4] Immediately before 1998 there is no reliable statewide testing data for the early elementary grades. This makes it impossible to estimate preprogram test scores, discussed later.

[5] Corresponding to the 1997-98 through 1999-2000 school years and 1998-99 to 1999-2000 school years respectively. Third graders from test year 1998 were omitted because of their lower participation rate, and the inability to classify all of them as participants on non-participants in the program.

and observations for which average class size cannot accurately be figured.[6] The bulk of the analysis also excludes approximately 1,500 observations of grades for which the school did not participate in Class Size Reduction that year.

The dataset's size and detail is greater than that of the data used in most previous studies of the effects of combination classes on student achievement. However, an important limitation in the data is the inability to measure outcomes on the classroom level. Largely because of this, I am not able to look at the detailed workings of combination classes, but rather look at the percentage of students in a school and grade that are in combination classes.

Another limitation is the lack of outcome data for pre-program years. Pre-program data would provide a valuable check on the identification strategy and allow estimation of "value-added" models. Finally, for confidentiality reasons, test scores are unavailable for any school and grade where ten or fewer students were tested. Thus, extremely small schools are excluded from the sample. Fortunately, the vast majority of the schools in California are larger than this cutoff.

Descriptive statistics are found in Table 2, reported separately by grade. Program participants scored close to the national average on standardized tests, though they scored slightly above average in math but below average in reading. The two grades also had fairly similar characteristics. However, a higher percentage of second grade students (15.05%) than third grade students (12.23%) were in combination classes. Also, second graders in program schools appear to have slightly less experienced teachers and slightly higher poverty and English learner percentages than third graders.

---

[6] Neither of these seems to be a systematic error. I also eliminate the few schools that had more than 240 students in a grade. Because my approach relies on non-linear variations in enrollment, it requires a sufficient density of observations to be effective.

1.3.2 Graphical Analysis

The California Class Size Reduction Program provided schools with an incentive to create combination classes. Figures 1 and 2 illustrate two effects of these incentives. Figure 1 plots the average class size for a school and grade against the number of students in that grade. The figure also plots a predicted class size function similar to that of Angrist and Lavy (1999). To obtain this function, I begin by estimating the smallest number of equal sized classes under twenty students that would accommodate the enrollment of each school grade. The predicted number of classes is:

$$CLN_{sgt} = (\text{int}[(STUDENTS_{sgt}\text{-}1)/20]\,) \qquad (4)$$

where s indexes school, g indexes grade and t indexes time. Int(.) represents the integer function, meaning that int(n) in the largest integer less than or equal to n. Using this variable, the predicted class size can be defined as:

$$PCS_{sgt} = [(ENROLLMENT_{sgt}/\ CLN_{sgt}] \qquad (5)$$

where $ENROLLMENT_{sgt}$ is the total number of students in that school and grade observation.

This predictive function indicates what the average class size would be if students were actually divided into the predicted number of classes. Figure 1 shows that actual class size is generally greater than the predicted class size. The two come closest to matching at twenty student intervals.

Figure 2 graphs the percentage of students in a grade in combination classes against the total school enrollment in that grade. The pattern is striking. The proportion of students in a combination class decreases markedly whenever the size of that grade approaches a multiple of twenty. It is easy two see how these two patterns are related.

Instead of lowering class size the predicted amount, administrators kept class size under the twenty student maximum without opening as many new classes by putting the excess students into combination classes. The closer a grade is to having natural multiples of twenty students, the less administrators can employ this type of shifting.

To illustrate the process of choosing class sizes and combination class levels, consider a school that had thirty students in a single class, in each of two first and second grade classes prior to the program. Without combination classes this school would have to implement the program for first and second graders by hiring two new teachers and providing four classes with fifteen students each. However, the additional money that would be paid to the school for implementing the program might be as little as $39,000.[7] Obviously, it would be impossible to hire two additional teachers with this amount. To adopt, this school must shift money from other areas to pay for the extra cost. However, if the school is allowed to count combination classes toward its goal it can hire only one new teacher and have three classes of twenty students. One of these classes is a combination class with equal numbers of students from each grade. This way, the school covers more of its costs from the program bonus.

A negative effect of combination classes on achievement may explain a puzzling pattern in the Class Size Reduction adopters' test scores. Figures 3 and 4 present this pattern in two different formats. Figure 3 plots math scores against the number of students enrolled in a grade. For convenience, the plot shows only schools with between 16 and 124 students in a grade, but the same pattern continues at higher enrollment levels. The vertical lines show multiples of twenty students that would require the

[7] This is figured at the initial payment rate of $650. Even at the 2002 rate of $906, it is hard to imagine paying the salaries and benefits of two new teachers for $54,360.

creation of an additional class if combination classes were not available. If the only effect of the program were a reduction in class size, and if the reduction in class size from forming a new class led to higher test scores, we would observe test scores rising after crossing a twenty student threshold.

In contrast, test scores appear to rise as they approach an enrollment threshold and drop off immediately afterward. Figure 4 isolates this pattern by plotting test scores against the distance from an enrollment threshold. This figure plots test scores against the distance in students from an enrollment threshold. Two facts emerge: First, test scores rise as enrollment levels approach a threshold. Second, test scores fall discontinuously as the threshold is crossed, and continue to fall thereafter. This suggests that something other than class size is driving the variation in test scores with enrollment.

One possible explanation for this pattern is the use of combination classes. Figure 2 shows that the percentage of combination students in a grade follows a pattern opposite to test scores. Combination class percentage falls as it approaches the threshold and rises discontinuously afterward.

An alternative explanation involves the Rivkin and Jepsen findings on teacher experience. Schools forming an extra class when crossing a threshold have to hire a new teacher. The new teachers may have had less experience. However, there is little graphical evidence that this explains the test score pattern. Figure 5 plots average teacher experience against grade level enrollment. The drop in teacher experience appears to be a smooth function of enrollment. Teacher experience rises as often as it falls when a threshold is crossed. Intuitively this may happen because some experienced teachers from other grades are willing to switch grades to teach smaller classes. Figure 6 shows a similar pattern for another measure of teacher experience, the percentage of novice teachers in a grade. Although the Class Size Reduction Program may have led to an overall experience decrease, teacher experience does not vary in a non-linear pattern with enrollment in participating grades.

The Class Size Reduction Program did not reduce class sizes by as much as grade level enrollments would predict. It also led to the use of combination classes and a decrease in teacher experience. The use of combination classes may have had a negative effect on test scores that would help explain the unexpected pattern observed in the data.

1.3.3 Identification

I use the data described above to capture the effects of combination classes and class size on achievement by exploiting the non-linear relationship between these variables and enrollment. The idea of using a program induced discontinuity as a source of identification is not new. Campbell (1969) discusses the use of regression-discontinuity designs in empirical research. More recently Hoxby (2001), Angrist & Lavy (1999), and Guryan(2001) make use of regression discontinuities to form instruments for instrumental variables estimation in education related investigations.

The causal relationship of interest is:

$$TESTSCORE_{sgt} = X_{sgt}'\alpha + \phi CLASSIZE_{sgt} + \delta COMBINATIONPCT_{sgt} + \gamma_g + \tau_t + \eta_{sgt} \quad (6)$$

where s indexes school, g indexes grade and t indexes year. X is a vector of demographic controls including grade level enrollment, percentage black and hispanic students, percentage of English learners, and percentage of students that qualify for free or subsidized meals. Also, γ is a grade effect, τ a time effect and η_{sgt} is the error term.

OLS estimates of equation (6) are unlikely to have a causal interpretation because the demographic variables included in the regression are unlikely to completely control for all the factors that relate classroom organization to test scores. For example, parental incomes and levels of involvement are likely to be negatively correlated with combination classes and class size and positively correlated with test scores. Omitting these factors from the regression biases the OLS estimates toward zero.

The presence of two variables with potential causal interpretations in the regression is also a concern. Consider estimation of a two stage least squares model which uses a non-linear enrollment function such as Predicted Class Size to instrument for combination class percentage but allows class size to be exogenous. The first stage relationship is:

$$COMBINATIONPCT_{sgt} = X_{sgt}'\pi_1 + \pi_2 Z_{sgt} + \pi_3 E_{sgt} + \pi_4 E^2_{sgt} + \varepsilon_{sgt} \qquad (7)$$

where X now includes demographic controls, year effects and class size , E is enrollment in grade g, and Z is the instrument, in this case predicted class size. This leads to the second stage:

$$TESTSCORE_{sgt} = X_{sgt}'\varphi + \rho COMBINATIONPCT^{*}_{sgt} + \mu E_{sgt} + \theta E^2_{sgt} + \omega_{sgt} \qquad (8)$$

where COMBINATIONPCT* is the predicted combination percentage produced by the first stage.

This specification assumes the instrument Z does not affect test scores except through its effect on combination class percentage. The assumption is likely violated in this case, since Figure 1 shows an apparent correlation between predicted class size and actual class size. An administrator who crosses an enrollment threshold does not have to put the extra students in a combination class, since she may opt instead to create a new class and reduce class size.

A potential solution to this "two causes" problem exploits the non-linear relationship between combination class percentage, class size, and enrollment to construct instruments for both class organization variables. Each of these instruments is a different non-linear function of enrollment. I have already introduced two potential instruments, Predicted Class Size and Predicted Number of Classes, in the graphical analysis. In principle these two variables can be used to instrument for both endogenous regressors. In practice, both variables are different functions of the same underlying enrollment variable, and this strategy is unlikely to provide precise estimates of both coefficients.

Another approach is to generate an instrument that is correlated with combination classes but uncorrelated with class size, conditional on enrollment. The predicted class size of a lower grade is a candidate. If a school reduced class size at the second or third grade level, the grade immediately below the observed grade must also have participated. To form a combination class there must have been students at two grade levels available to combine. Since schools were far less likely to combine students with non-participating grades, the ability of a school to form combination classes depended on the predicted class size of the immediately lower grade. The Lower Grade Class Size Predictor is:

$$PCS_{s(g-1)t} = [(STUDENTS_{s(g-1)t} / CLN_{s(g-1)t}] \,. \qquad (9)$$

I also construct a Combination Classes Predictor (CSP) that by design is purged of correlation with class size. To do this I calculate the number of classes of fewer than twenty students required for the students in a grade and the grade below. Then I subtract this number from the predicted number of classes of fewer than twenty students the school would require to avoid mixing these two grades. Intuitively the predictor counts additional classes a school would need to form to participate in both grades and avoid combination classes. The formula is:

$$CSP_{sgt} = [\ (CLN_{sgt} + CLN_{s(g-1)t}) - (CLN_{s(g+(g-1))t}] \qquad (10)$$

I present results that use a variety of methods to deal with the potential confounding effects of class size, including instrumenting for both potentially endogenous variables, and using the Combination Classes Predictor as an instrument. In all cases the estimates of the effects of combination students on achievement are similar. The effects of class size are generally small and always statistically insignificant.

1.4 Results

1.4.1 OLS

Table 3 presents ordinary least squares estimates of equation (6) for second graders. The dependent variable is the national percentile rank of the hypothetical average student in mathematics. Panel A provides results for the sample of participating schools discussed in the data section. Column (1) presents a specification comparable to many previous studies of class size. The class size coefficient is small and insignificant. However, some measures of teacher experience are correlated with student achievement. Both the percentage of first year teachers and the percentage of teachers with credential waivers have negative significant coefficients.[8]

[8] To get a teaching credential in California, a candidate must take 30 credit hours beyond a bachelors degree in a recognized education program. This is often referred to as the "fifth year". Teachers with a bachelors degree who pass other certification requirements such as the competency test can get an emergency credential which allows them to teach for a few years under the understanding they will use the

Columns (2) – (5) add the percentage of students in combination classes to the specification. Estimate of this coefficient are consistently in the neighborhood of -.073, implying that a five percentage point change in students in combination classes would leads to a drop of about one-third of a percentile in math scores. This seems like a modest effect. This result is robust to the level of control and addition of smooth enrollment controls. Teacher inexperience continues to play a negative role, with estimates of a similar magnitude. The class size coefficient falls toward zero once higher order demographic controls are added to the model.[9] The small coefficient on class size differs from recent research documenting large class size effects[10]

These results do not conclusively demonstrate that the class size changes caused by the Class Size Reduction Program had no effect on test scores. There is no pre-treatment versus post-treatment element in any of these estimates. Since the sample is composed of schools that have all implemented the reduction program, the variation in class size is smaller than in most populations. Most schools in the sample lie within a 2.5 student range of class size. This may account for the failure to find large class size effects in this paper. This result is similar to Hoxby (2001), which finds no class size effects when examining natural population variation.

If the lack of variation in class size is partially responsible for the small class size coefficient estimates, then OLS estimation using a sample with more variation should

---

time to complete the other requirements. A credential waiver, is more radical and releases the teacher from even more requirements of the credentialing process. Because a large part of credentialing is gaining classroom understanding, experience, and performing student teaching, these variables can still be thought of as a type of experience measure.

[9] Throughout the following tables the estimated models contain the full set of demographic controls up to third order terms, except where specified otherwise, as well as year effects and grade effects when relevant. These coefficients are not reported because they are not a primary object of interest in this paper and because the follow a pattern that previous research predicts. Namely, test scores drift upward over time, and schools with a high percentage of disadvantaged students perform perform poorly.

[10] For example Angrist and Lavy (1999), Krueger (1999), Finn and Achilles (1990).

yield larger coefficients. Panel B of Table 3 confirms this. It re-estimates the specifications used in panel A on a sample which adds the 445 second grade classes that did not participate in the Class Size Reduction Program to the previous data. The coefficient on class size is larger and consistently negative. However, the class size coefficient is still insignificant in all but one of the specifications. Panel B also shows slightly attenuated teacher experience and combination class effects when compared to the CSR sample estimates.

Table 4 demonstrates that the percentage of students in combination classes has larger effects on second grade language scores than on math scores. The coefficient estimates of combination class effects are -.094, about thirty percent greater than the math coefficients. Increased class size effects are also larger, about half the magnitude of the combination class effect, but still statistically insignificant. Teacher experience remains important, as do credential waivers.

The results in Table 5 show that OLS regressions using third grade test scores produce similar patterns. Columns (1)-(3) show that the effect of combination class percentage on math scores is about the same as for second graders. Class size effects are now positive but are still small and imprecisely estimated. In contrast with the second grade results the percentage of first year and credential waiver teachers have no significant relationship with test scores. Columns (4)-(6) show similar findings for class size and combination class percentage when the dependent variable is language scores.

1.4.2 Instrumental Variables

First Stage Estimation

The OLS results provide a reference point but are unlikely to have a clear causal interpretation. Any omitted variable, such as parental education or involvement that is positively correlated with test scores and negatively correlated with percentage of students in combination classes, will bias the OLS estimates toward zero.

The first five columns of Table 6 present estimates of the first stage relationship described in equation (7) for various instruments. Column (1) shows a significant positive correlation between the Combination Class Predictor and the percentage of students in a combination class, conditional on enrollment and demographic controls.[11] Columns (2) and (3) show that Predicted Class Size and Predicted Number of Classes are also correlated with combination class percentage. This confirms the graphical evidence of Figure 2.

Columns (4) and (5) present regression results from specifications that contain two non-linear functions of enrollment. All these functions, including the Lower Grade Class Size Predictor have significant coefficients. However the use of two instruments does not substantially improve the fit of the prediction. Also, the presence of Predicted Class Size in the regression attenuates the coefficient on the Combination Class Predictor by half. This relationship makes sense as both are functions of the same underlying enrollment variable.

The last five columns of Table 6 repeat these regressions, with average class size as the dependent variable. Columns (6) and (9) demonstrate that the Combination Classes

[11] Though the tables only report results of quadratic enrollment controls, regressions using quartic enrollment controls yield essentially the same results, with the higher enrollment terms having insignificant coefficients.

Predictor is not a significant predictor of class size, even in specifications which include Predicted Class Size as a regressor. The results also indicate that class size is correlated with the other non-linear enrollment functions, including Lower Grade Predicted Class Size.

Additionally, the table shows that the non-linear enrollment functions are better predictors of combination class percentage than of class size. The coefficients for these three instruments are smaller (in an absolute value sense) and less precisely estimated than their counterparts in the first half of the table.

The non-linear enrollment functions are also poor predictors of teacher experience. Columns (1) – (4) of Table 7 present estimates of the relationship between the non-linear enrollment instruments and average years of teacher experience. In all specifications the quadratic enrollment controls are significant predictors of teacher experience, while the non-linear enrollment functions have coefficients that are not statistically distinguishable from zero. Columns (5) – (8) show the same pattern using a different teacher experience measure, the percentage of novice teachers in a grade and school. Despite a positive and significant OLS relationship with test scores, teacher experience does not vary in a non-linear fashion with enrollment. While teacher experience may have predictive power for test scores on average, it cannot explain the pattern of scores shown in Figure 4.

2SLS

Table 8 presents the results of a two stage least squares estimation of equation (8) for second graders. The first six columns treat class size as exogenous. The final three impose a zero coefficient on class size. The first three columns present results for language achievement. Instrumenting the percentage of combination classes with the Combination Class Predictor yields a significant coefficient estimate of -.195. This implies that a five percentage point increase in students in combination classes leads to a one percentile fall in test scores. This estimate is about 2.5 times the magnitude of the OLS estimates, suggesting that the OLS estimates suffer from omitted variables bias. Columns (2) and (3) report results from estimation using the predicted class size instruments. The coefficient estimates are slightly smaller than those in column (1) but are highly significant.

Columns (4) – (6) show an effect of similar magnitude on math scores. Instrumenting with the Combination Class Predictor, Column (4) gives coefficient estimates of -.180 for the effect of combination classes on math scores. This implies a five percentage point increase in combination class students again results in a one percentile drop in average test scores. Estimation using the predicted class size instruments provides similar results.

In all of these specifications, the class size estimates range from one-third to one-half the magnitude of the combination class estimates. The class size estimates are also very imprecise. In addition to the insignificant class size estimates the table reveals that the coefficients on the smooth enrollment controls are not significantly different from zero in any specification.

As a specification check, the final three columns of Table 8 repeat the estimation of columns (4) – (6), imposing a coefficient of zero on class size. These regressions yield almost identical results. Estimates of the effect of combination classes on second grade math scores yield coefficients of -.18 to -.20 whether class size is treated as exogenous, zero, or an instrument uncorrelated with class size is used.

Table 9 presents similar two stage least squares regressions results for third graders. The estimated coefficient for percentage of students in a combination class is consistent across different instruments and larger than the second grade estimates. Estimates for math and language scores are about -.36 and highly significant. This implies that a five percentage point increase in combination class students corresponds to a one and a half percentile drop in average test scores for the entire grade. The larger third grade estimates may be due to third graders greater propensity to be placed in combination classes with lower graded students. Estimates of average class size effects are positive but extremely imprecise.

As a further specification check, Columns (7) – (9) estimate the effect of classroom organization on the math scores of third graders with the class size coefficient constrained to equal the Rivkin and Jepsen estimates.[12] The resulting coefficient estimates of the effect of combination class percentage on math scores are very similar to to the estimates in columns (4) – (6) which treat class size as exogenous.

Table 10 presents the results of two stage least squares estimation with both class size and combination class percentage treated as endogenous regressors. The combination of instruments used in each specification is shown at the bottom of the column. The second grade results for combination class percentage presented in Panel A are similar to my previous two stage least squares estimates, -.18 for math and -.19 for

[12] See Rivkin and Jepsen (2002) Table 4. p36. The coefficient is adjusted to reflect the different scale of the test score measure used by Rivkin and Jepsen.

language. The pattern holds for third graders with estimated coefficients about -.36 for math and -.38 for language.

At first the class size coefficient estimates in this table might seem implausibly large. These coefficients are much larger than the estimates shown in previous tables. However, all the instruments rely on the variation in the same underlying enrollment variable. Because of their colinearity they are unlikely to provide precise estimates for both coefficients. The instruments have a stronger relationship with combination classes than with class size. Thus, the large class size coefficients come with very large standard errors. Such large standard errors make it impossible to rule out any of the earlier class size estimates or a zero effect.

Two stage least squares estimates provide consistent evidence that combination class students explain the perverse effect seen in Figure 4. Furthermore, the effect of combination classes on test scores is larger than OLS estimates suggest. The coefficient estimates are robust across different approaches to the potential confounding effects of class size. Class size effects on the other hand are small or zero and very imprecisely estimated.[13]

1.4.3 Comparison and Interpretation

Why do the results show unambiguous negative effects of combination classes while previous research has generated mixed results? Two factors seem important. The first is the variety of organizational structures that might be considered combination classes. Combination classes in California might be different in some important respects

[13] This is not surprising. Class size had very small OLS coefficients and a weaker relationship with the instruments than combination classes. This does not necessarily mean there was no effect on test scores from the class size reduction, but rather there is no discernable effect in this sample.

from those studied in other contexts. California exercises an unusual amount of centralized control over school curriculum. In addition, since 1998, the state has required grade-specific standardized tests be administered to all students in grade two and above. These tests reinforce the need to teach distinct skills to students at different grade levels. This structure and testing mean that a combination class teacher in California is less likely to rely on thematic or common curriculum elements than teachers in other settings. In effect, the teacher must teach two separate classes within one classroom. The education literature suggests that a prime source of benefit in combination classes is the ability of students to work together in accomplishing mutual tasks, an advantage that is lost if the students are involved in different tasks.

Though it might seem that this rigid structure limits the applicability of this study to other combination class contexts, education policy trends indicate otherwise. These emerging trends involve a shift toward centralized standards and curriculum and greater grade-specific testing. This makes the California model of the combination class a good approximation for what many states might choose in the future.

A second consideration distinguishing this study from previous work is study design. Many studies rely on small samples of classrooms and are limited by the lack of important data on school characteristics. Additionally, there is no clear source of exogenous variation in the use of combination classes. In contrast, this study uses a clearly defined source of exogenous variation and a relatively large sample of schools.

Did the use of combination classes make the California Class Size Reduction Program at net loss in academic achievement terms? To answer this question, I provide some estimates of the net effect of the program. These estimates are illustrative, requiring assumptions about the effect the program had on test scores through changing class size in the absence of combination classes.

In order to estimate a net effect, I first translate my coefficient estimates into "effect size" (i.e. standard deviation) units. Let β be the estimate of the effect of combination students on test scores, X_{pre} and X_{post} measures of combination percentage before and after the program, and σ the standard deviation of student math scores, then the effect size of the California Class Size Reduction, working through combination classes is:

$$[(X_{post} - X_{pre})\beta] / \sigma \qquad (11)$$

Often, researchers have a choice of σ when calculating effect sizes. Effect sizes calculated using the standard deviation among student test scores will always be smaller than effect sizes that use the standard deviation between groups of students, because the latter has a smaller variance. Because within-student test score information is not available, I calculate effect sizes using the between-grade variation. Because they are presented in standard deviations, effect sizes can be compared across outcome measures.

Table 11 presents estimates of the effect sizes of the increase in combination class percentage due to the Class Size Reduction Program. The first row shows the effect size of a five percentage point increase in combination class students. These results seem modest, representing only 4-10% of a standard deviation decrease in test scores. Unlike other policies that affect all the students in a grade, combination classes may only affect the test scores of combination class students. This is reflected in the second row of the table, which presents effect size estimates scaled by the proportion of combination students in that grade. These adjusted figures indicate that achievement losses for students placed in combination classes are between .24 and .36 of a standard deviation for second graders and .58 to .66 of a standard deviation for third graders.

To figure the net effect of the California Class Size Reduction Program I combine its effects on test scores through three channels, combination classes, class size and teacher experience. I measure changes in these variables from 1995-96 to 1999-2000. During this time average class size dropped by about 10 students for affected grades, average teacher experience decreased by a year and the population of novice teachers grew by seven percentage points. Using the effect sizes implied by my estimates of class size and teacher experience I find that the net effects of the program for second graders are slightly negative with magnitudes of 2-4 percent of a standard deviation. The net effects on third graders are also negative, but more substantial, 10-13 percent of a standard deviation.

I next figure net effects using the Rivkin and Jepsen estimates of the class size effects of the program[14]. Their estimates imply a math score effect size of .199 standard deviations for a ten student decrease in class size among third graders. They do not estimate models using language scores, but find a .1167 standard deviation effect for reading scores.[15] These larger class size estimates imply a positive program net achievement effect. Third graders experienced a .10σ increase in math scores and a .03σ increase in language scores. Second graders experienced a similar positive effect, assuming they faced the same class size effects.

Finally, I calculate the net effect of a hypothetical policy that implements all the Class Size Reduction Rules but does not allow combination classes. In this scenario,

[14] See Rivkin and Jepsen (2002) Table 4. p36.
[15] I use reading scores and language scores interchangeably in this comparison. In fact student scores on the two tests are similar but not identical.

there is no perverse combination class effect, class size equals predicted class size, on average more than a student lower than under the real program, and teachers are slightly less experienced.

This hypothetical program has a larger positive net effect than the actual program.. Third grade students experience an increase in math scores of .2σ, almost double the net effect of a program that allows combination classes. The effect on language scores of third graders quadruples, to .12σ. This calculation assumes that all schools that participated in the real program would participate in the hypothetical one. In practice, some schools might not join the program if they were unable to use combination classes. This would diminish the number of students that received the increasd benefits.

Though the program may well have had net positive effects, the effect on the students put into combination classes by the program was almost certainly negative. Angrist and Lavy (1999) and Krueger (1999) consider classroom settings outside California and find larger effect sizes than Rivkin and Jepsen (approximately .3σ). Even if class size effects this large occurred in California, an increase in second grade combination students of five percentage points would roughly offset a contemporaneous class size decrease of eight students. It would take a much larger class size decrease to offset the negative impact of a five percentage point increase in combination students at the third grade level. .

1.5 Conclusion

The California Class Size Reduction Program spent billions of dollars to reduce class sizes for early elementary school children. However, the program used a non-linear incentive scheme that rewarded schools for meeting a target threshold. These incentives led schools to shift students into combination classes as well as add classes to meet program requirements. This study offers strong evidence against the use of combination classes. Combination classes have an unambiguously negative effect on student achievement and the effect is greater for third graders than second graders. Students placed in combination classes by the program were almost certainly worse off in achievement terms.

The sign of the overall net effect of the Class Size Reduction Program depends crucially on the actual class size effect of the program. My estimates of class size effects were small and not significantly different from zero. These small class size effects lead to the conclusion that the program had negative net effects in the first few years. However, other studies have found larger class size effects that would imply a positive net effect. Further research might examine cross state variation in class size reduction policies to better estimate the magnitude of the class size effect.

References

Acemoglu, Daron, Michael Kremer, and Atif Mian, Incentives in Markets, Firms and Governments. *NBER Working Paper # 9802.* June 2003.

Angrist, Joshua D. and Victor Lavy, Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement,. *Quarterly Journal of Economics* 114. (1999), 533-575.

California Department of Education. *Standardized Testing and Reporting Data.* Years 1998-2000.

______, *Class Size Reduction: Final Participation and Funding Data.* Years 1997-2003

California Department of Education – Educational Demographics Office. *Demographic Data Files.* Years 1995-2001.

Campbell, Donald T., Reforms as Experiements. *American Psychologist* 24. (1969) 409-429.

Card, David and Alan Krueger, Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States,. *Journal of Political Economy* 100. (February 1992) 1-40.

Chevalier, Judith and Glenn Ellison, Risk Taking by Mutual Funds as a Response to Incentives. Journal of Political Economy 105. (December 1997), 1167-1200

CSR Research Consortium, *Class Size Reduction in California: Early Evaluation Findings, 1996-98.* (1999).

______. *Class Size Reduction in California: The 1998-99 Evaluation Findings.* (2000).

Figlio, David N. and Marianne E. Page, School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality? *Journal of Urban Economics 51.* (May 2002), 497-514

Finn, Jeremy D. and Charles M. Achilles, Answers and Questions About Class Size: A Statewide Experiment, *American Educational Research Journal* 27. (Fall 1990), 557-77.

Guryan, Jonathan, Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts. *NBER Working Paper #8269.* (May 2001).

Gutierrez, R., Slavin, R.E. (1992). Achievement Effects Of The Non-Graded Elementary School: A Best Evidence Synthesis. *Review of Educational Research* 62. (1992), 333-376.

Hanushek, Eric A., .The Economics of Schooling: Production and Efficiency in Public Schools,. *Journal of Economic Literature* 24. (September 1986) 1141-1177.

______., John F. Kain and Steven G. Rivkin. Teachers, Schools and Academic Achievement. *NBER Working Paper #6691.* (August 1998).

Holmstrom, Bengt, and Paul Milgrom, Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. Journal of Law, Economics and Organization 7. (1991), 24-52.

Hoxby, Caroline M, The Effects of Class Size on Student Achievement: New Evidence from Natural Population Variation. *Quarterly Journal of Economics* 115. (November 2000), 1239-85.

Jacob, Brian, and Steven Levitt, Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *Quarterly Journal of Economics 118.* (August 2003), 843-77.

Jacob, Brian, Accountability, Incentives, and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools. *NBER Working Paper # 8968* (2002).

Krueger, Alan B., .Experimental Estimates of Education Production Functions,. *Quarterly Journal of Economics* 114. (May 1999), 497-532.

Lazear, Edward P., Educational Production. *Quarterly Journal of Economics* 116 (August 2001), 777-803

Oyer, Paul, Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality. Quarterly Journal of Economics 113. (February 1998), 149-85.

Pavan, B. N., The benefits of non-graded schools. *Educational Leadership* 50. (1992) 22-25.

Rivkin, Steven G. and Christopher Jepsen, What is the Tradeoff Between Smaller Classes and Teacher Quality? *NBER Working Paper #9205.* (September 2002).

Russell, V.J., Rowe, K.J., Hill, P.W. Effects of Multi-Grade Classes on Student Progress in *Literacy and Numeracy: Quantitative Evidence and Perceptions of Teachers and School Leaders*. Melbourne, Australia: University of Melbourne (1998).

Slavin. R. E., Achievement Effect of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. *Review of Education Research.* (Fall 1990), 471-499.

Slavin, R. E.. Synthesis of research on grouping in elementary and secondary schools *Educational Leadership* 46.(1988), 67-77.

U.S. Department of Education, National Center for Education Statistics. *Common Core of Data (CCD)*, online version. http://nces.ed.gov/ccd/

Veenman, S. Cognitive and Noncognitive Effects of Multi-grade and Multi-age Classes: A Best Evidence Synthesis. Review of Educational Research 65. (1995), 319-381.

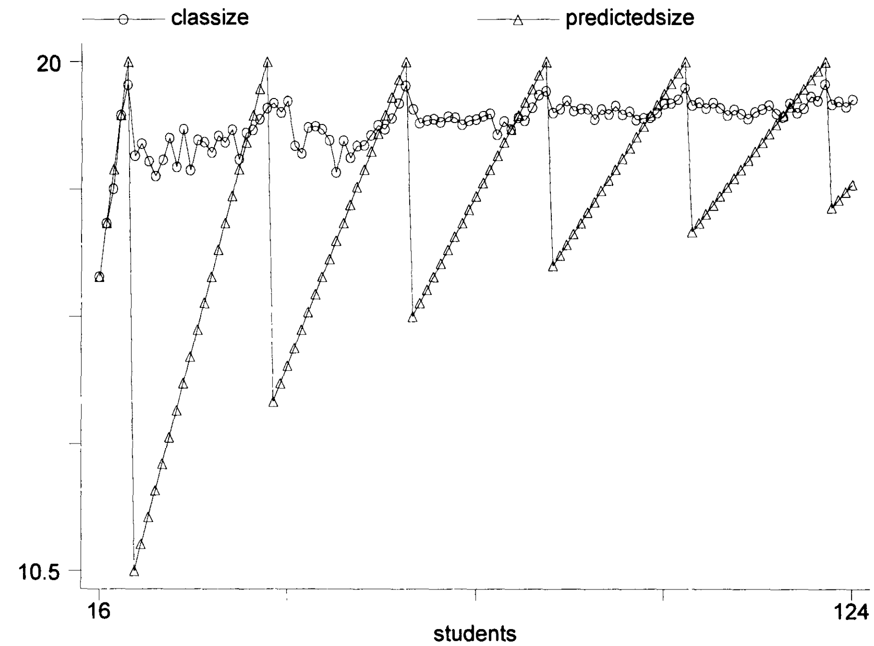


Figure 1: A Plot of the predicted class size function versus actual class size.

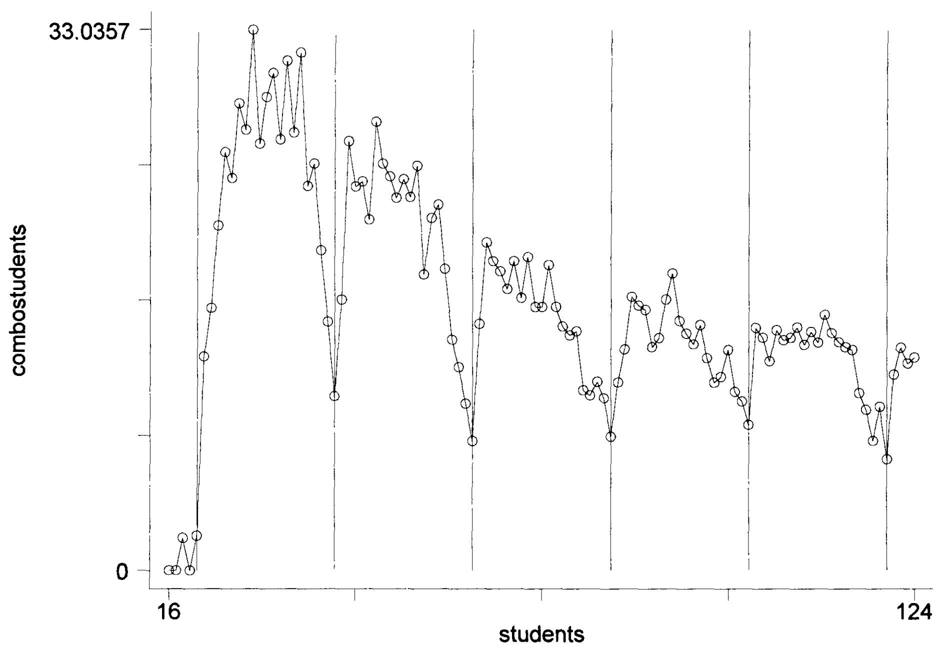


Figure 2: A plot of the percentage of students in combination classes versus total grade enrollment. Vertical lines are at 20 student intervals.

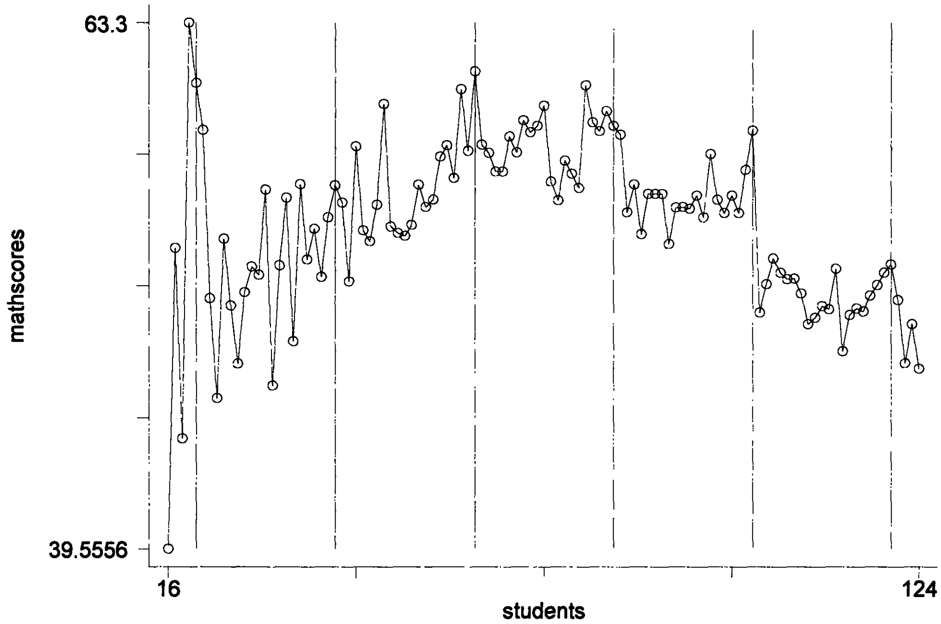


Figure 3: A plot of math test scores by grade level enrollment. Vertical lines are at 20 student intervals.

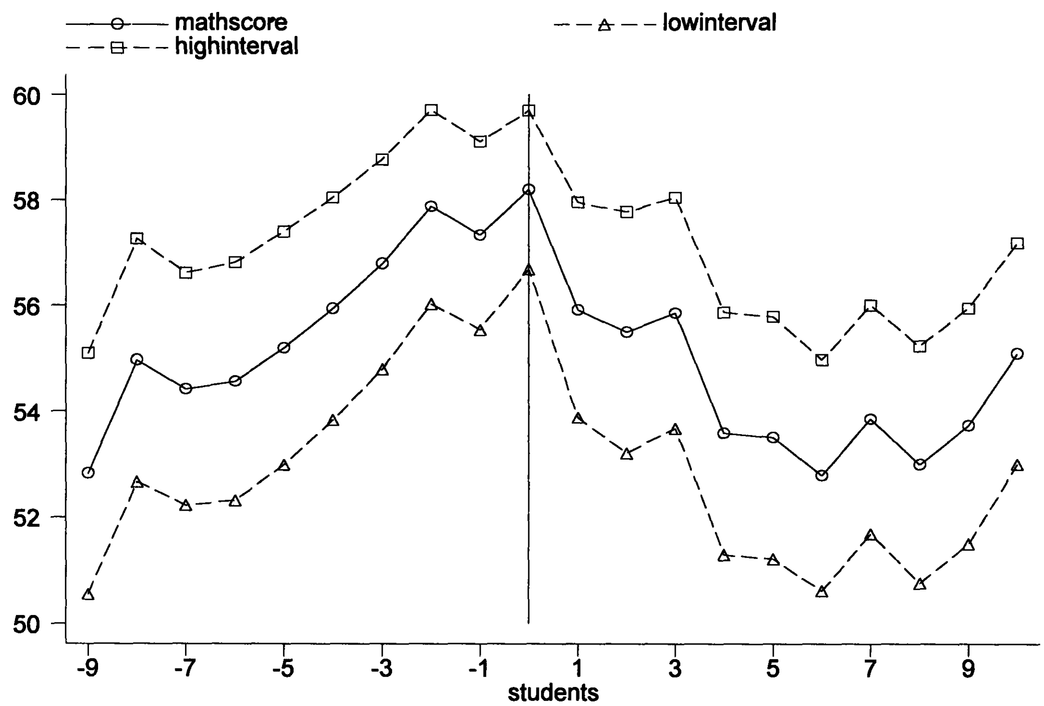


Figure 4: Plot of math test scores versus school grade enrollment measured as distance from a multiple of twenty students. Dashed lines give a 95% confidence interval.

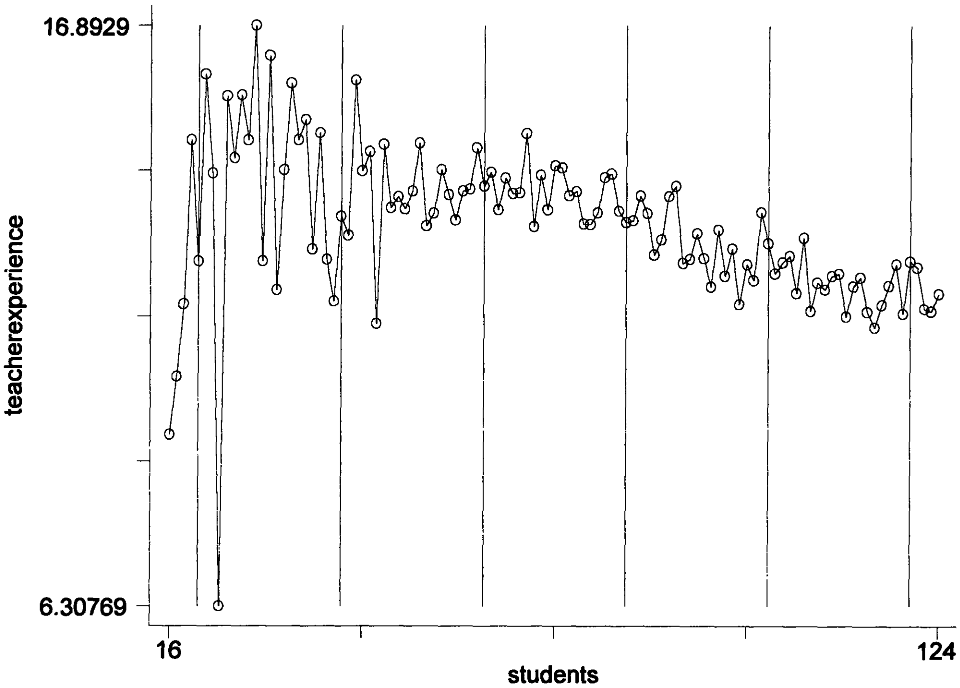


Figure 5: plot of years of average years of teacher experience by grade level enrollment. Vertical lines are at 20 student intervals.

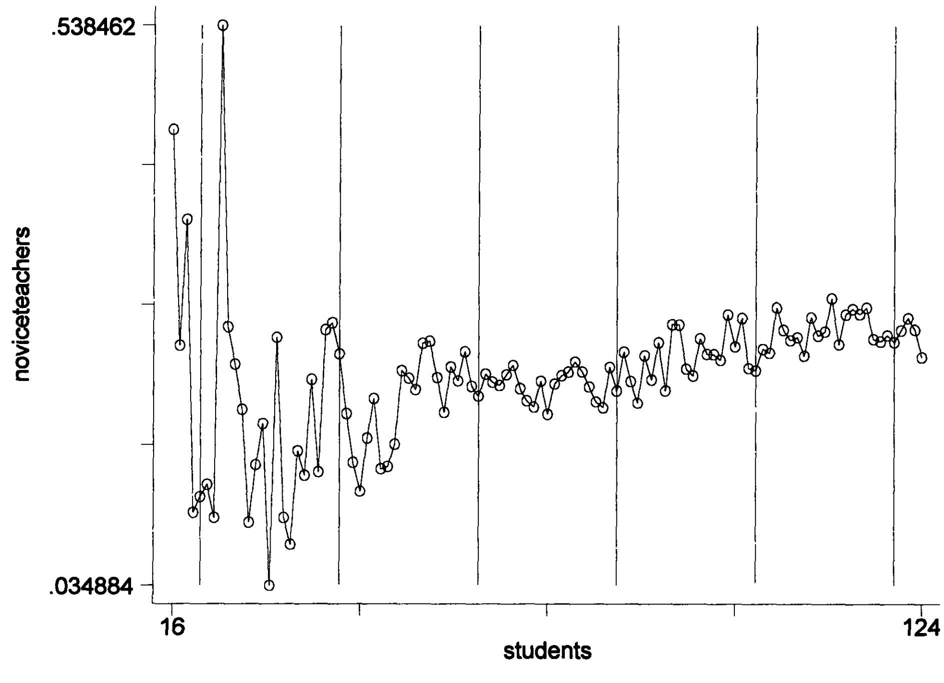


Figure 6 : A plot of the percentage of novice teachers by grade level enrollment. Vertical lines are at 20 student intervals.

Table 1 - Participation in the California Class Size Reduction Program

Year:	1997	1998	1999	2000	2001
Kindergarten:					
Students	64,779	321,209	393,036	421,943	439,439
Percentage	14%	69%	86%	92%	96%
First Grade:					
Students	428,242	484,518	483,714	477,150	480,307
Percentage	88%	99%	99%	98%	99%
Second Grade					
Students	262,074	468,103	475,477	472,842	475,702
Percentage	57%	96%	98%	97%	97%
Third Grade					
Students	79,062	309,828	410,089	444,136	458,040
Percentage	18%	67%	84%	91%	91%

Participants are those students in a school that applies for program funding for the student's grade and has all class sizes in that grade at twenty or fewer pupils. The year listed corresponds to the spring portion of the school year. Thus 1997 is the first program year.
Source: California Department of Education

Table 2 - Descriptive Statistics

	Grade 2	Grade 3
Math NPR	51.80	55.30
	(19.56)	(18.52)
Language Arts NPR	47.94	49.63
	(20.80)	(18.63)
Percentage of students in the grade in combination classes	15.05	12.23
	(15.66)	(13.76)
Number of combination classes in the school	2.68	2.55
	(2.90)	(2.90)
Average Class Size	18.03	18.15
	(1.22)	(1.42)
Average Teacher Experience	12.58	12.90
	(6.38)	(6.37)
Percentage novice teachers	25.15	23.66
	(25.66)	(25.22)
Percentage first year teachers	7.42	6.42
	(14.81)	(13.61)
Percentage second year teachers	9.56	8.90
	(16.51)	(15.94)
Percentage third year teachers	8.17	8.35
	(15.12)	(14.69)
Percentage teachers with emergency credential	9.32	9.27
	(17.52)	(17.39)
Percentage teachers with credential waiver	0.57	0.49
	(4.81)	(4.26)
n=	9974	6079

Table continues on next page. Standard Errors are in parentheses below means. Unit of observation is the school grade year. NPR is the National Percentile Rank of the hypothetical average student.

Table 2 - Descriptive Statistics - continued

	Grade 2	Grade 3
Percentage of free/reduced Meal students in school	52.41	52.01
	(29.71)	(29.70)
Percentage African-American students in grade	9.16	9.65
	(13.50)	(14.34)
Percentage Hispanic students in grade	39.83	38.69
	(28.04)	(27.74)
Percentage English Learner students in grade	28.77	27.15
	(23.92)	(23.44)
Grade enrollment	98.87	97.84
	(39.51)	(39.35)
School enrollment	621.70	617.00
	(229.79)	(236.03)
n=	9974	6079

Standard errors are in parentheses below means. Throughout the paper and remaining tables percentage of students with subsidized meals, percentage of minority students and percentage english learners are used as the demographic controls.

Table 3 - OLS Estimates of the Effect of School Characteristics on Second Grade Math Scores

	(1)	(2)	(3)	(4)	(5)
A. Class Size Reduction Schools Only (n=9974)					
Percentage of students		-0.073***	-0.075***	-0.070***	-0.069***
in combination classes		(0.010)	(0.010)	(0.010)	(0.010)
Average Class Size	0.028	-0.022	-0.013	0.001	0.002
	(0.115)	(0.115)	(0.115)	(0.114)	(0.114)
Average Teacher	0.054	0.066**	0.063*	0.079**	0.078**
Experience	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Percentage of first year	-0.038***	-0.036***	-0.036***	-0.038***	-0.038***
Teachers	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Percentage of second	-0.011	-0.009	-0.009	-0.009	-0.009
year teachers	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Percentage of teachers with	0.021*	0.018	0.018	0.008	0.008
emergency credentials	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
Percentage of teachers with	-0.117***	-0.108***	-0.109***	-0.104***	-0.104***
credential waivers	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Enrollment in grade			-0.003	-0.006	0.019
			(0.005)	(0.005)	(0.054)
Enrollment in grade					-0.024
squared*100					(0.045)
B. Class Size Reduction Schools and Non-Participants (n=10419)					
Percentage of students		-0.070***	-0.073***	-0.068***	-0.068***
in combination classes		(0.010)	(0.010)	(0.010)	(0.010)
Average Class Size	-0.120	-0.142*	-0.135	-0.111	-0.112
	(0.084)	(0.084)	(0.084)	(0.084)	(0.084)
Average Teacher Experience	0.050	0.062*	0.058*	0.074**	0.074**
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Percentage of first year	-0.036***	-0.034***	-0.034***	-0.036***	-0.036***
Teachers	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Percentage of teachers with	0.015	0.011	0.013	0.004	0.004
emergency credentials	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Percentage of teachers with	-0.114***	-0.105***	-0.107***	-0.102***	-0.102***
credential waivers	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Demographic controls	Y	Y	Y	Y	Y
higher order controls				Y	Y

Estimates are of Equation (6) in the text. *** Indicates 1% significance level, ** 5% and * 10%. Reported standard errors are adjusted to correct for clustering at the school level. All regressions are weighted by the number of test takers.

Table 4 - OLS Estimates of the Effect of School Characteristics on Second Grade Language Scores

	(1)	(2)	(3)	(4)	(5)
Percentage of students in combination classes		-0.098***	-0.099***	-0.094***	-0.094***
		(0.008)	(0.008)	(0.008)	(0.008)
Average Class Size	-0.002	-0.069	-0.058	-0.050	-0.050
	(0.104)	(0.104)	(0.104)	(0.103)	(0.104)
Average Teacher Experience	0.073**	0.090***	0.086***	0.099***	0.099***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Percentage of first year Teachers	-0.037***	-0.035***	-0.035***	-0.036***	-0.036***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Percentage of second year teachers	-0.010	-0.007	-0.007	-0.007	-0.007
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Percentage of teachers with emergency credentials	0.022**	0.017*	0.018*	0.008	0.008
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Percentage of teachers with credential waivers	-0.142***	-0.130***	-0.132***	-0.125***	-0.125***
	(0.026)	(0.027)	(0.027)	(0.026)	(0.026)
Enrollment in grade			-0.004	-0.007*	0.022
			(0.004)	(0.004)	(0.048)
Enrollment in grade squared*100					-0.027
					(0.041)
n	9974	9974	9974	9974	9974
level demographic controls	Y	Y	Y	Y	Y
higher order controls				Y	Y

The table mirrors the estimates of Table 3 Panel A using language scores as the dependant variable. *** Indicates 1% significance level, ** 5% and * 10%. Reported standard errors are adjusted to correct for clustering at the school level. All regressions are weighted by the number of test takers. Level controls refer to controls for percent minority, percent subsidized lunch and percent English learners as well as year effects. Higher order controls add quadratic and cubic terms as well as interactions for the demographic control variables.

Table 5 - OLS Estimates of the Effect of School Characteristics on Third Grade Test scores

Dependent Variable:	math scores			language scores		
	(1)	(2)	(3)	(4)	(5)	(6)
Percentage of students in combination classes	-0.077*** (0.013)	-0.073*** (0.013)	-0.072*** (0.013)	-0.085*** (0.011)	-0.079*** (0.011)	-0.079*** (0.011)
Average Class Size	0.016 (0.121)	0.023 (0.119)	0.024 (0.119)	0.022 (0.101)	0.026 (0.099)	0.039 (0.099)
Average Teacher Experience	0.073** (0.036)	0.083** (0.036)	0.081** (0.036)	0.121*** (0.031)	0.130*** (0.031)	0.126*** (0.030)
Percentage of first year Teachers	0.000 (0.014)	-0.002 (0.014)	-0.002 (0.014)	0.014 (0.012)	0.012 (0.012)	0.013 (0.011)
Percentage of second year teachers	-0.005 (0.011)	-0.010 (0.011)	-0.011 (0.011)	0.004 (0.010)	-0.002 (0.009)	-0.003 (0.009)
Percentage of teachers with emergency credentials	0.016 (0.013)	0.014 (0.013)	0.014 (0.013)	0.016 (0.010)	0.011 (0.010)	0.010 (0.010)
Percentage of teachers with credential waivers	-0.033 (0.036)	-0.029 (0.036)	-0.029 (0.037)	-0.071** (0.030)	-0.065** (0.031)	-0.064** (0.031)
Enrollment in grade	-0.002 (0.005)	-0.004 (0.005)	0.084 (0.055)	-0.005 0.004	-0.007* 0.004	0.064 0.048
Enrollment in grade squared*100			-0.083* (0.047)			-0.080** 0.041
n=	6079	6079	6079	6079	6079	6079
level controls	Y	Y	Y	Y	Y	Y
higher order controls		Y	Y		Y	Y

The table estimates equation (6) in the text for third graders. *** Indicates 1% significance level, ** 5% and * 10%. Reported standard errors are adjusted to correct for clustering at the school level. All regressions are weighted by the number of test takers. Level controls refer to controls for percent minority, percent subsidized lunch and percent English learners as well as year effects. Higher order controls add quadratic and cubic terms as well as interactions for the demographic control variables.

Table 6 - First Stage Estimates for pooled second and third graders

Dependent Variable:	Percentage of Students in Combination Classes					Average Class Size				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Instruments:										
Combination Classes Predictor	1.720***			0.734***		0.001			0.034	
	(0.174)			(0.174)		(0.016)			(0.018)	
Predicted Class Size		-1.794***		-1.599***	-1.632***		0.061***		0.070***	0.057***
		(0.102)		(0.110)	(0.100)		(0.009)		(0.010)	(0.009)
Predicted Number of Class Groups			5.515***					-0.158***		
			(0.393)					(0.033)		
Predicted Class size of lower grade					-1.172***					0.019**
					(0.111)					(0.008)
Enrollment Controls:										
Own Grade Enrollment	-0.115***	-0.026	-0.368***	-0.048**	-0.070***	0.013***	0.010***	0.020***	0.010***	0.012***
	-0.023	(0.017)	(0.026)	(0.024)	(0.025)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Own Grade Enrollment Squared*100	0.018**	-0.003	0.015**	0.002	0.006	-0.003***	-0.003***	-0.004***	-0.002**	-0.002**
	(0.009)	(0.007)	(0.006)	(0.009)	(0.010)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Lower Grade Enrollment	0.028			0.015	0.080***	0.001			-0.001	-0.003
	(0.021)			(0.021)	(0.023)	(0.002)			(0.002)	(0.002)
Lower Grade Enrollment Squared*100	-0.005			-0.002	-0.019**	-0.001			-0.001	-0.001
	(0.008)			(0.008)	(0.008)	(0.001)			(0.001)	(0.001)
Root MSE	13.797	13.677	13.715	13.705	13.64	1.2105	1.2076	1.2085	1.2085	1.2083
N	15,976	16,053	16,053	15,976	15,976	15,976	16,053	16,053	15,976	15,976

The table estimates equation (7) in the text for both second and third graders *** Indicates 1% significance level, ** 5% and * 10%. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions though results are not reported.

Table 7 - Pooled Second and Third Grade Predictions of Teacher Experience Using Non-linear Enrollment Instruments

Dependent Variable:	Average Years Teacher Experience				Percent Novice teachers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Instruments:								
Combination Class Predictor	0.034 (0.099)				-0.0071 (0.0041)			
Predicted Class Size		0.003 (0.044)		0.002 (0.0432)		-0.0009 (0.001)		-0.0001 (0.0017)
Predicted Number of Class Groups			0.038 (0.155)				0.0025 (0.0065)	
Predicted Class size Of Lower Grade				0.012 (0.0398)				0.0000 (0.0015)
Enrollment controls:								
Own Grade Enrollment	-0.023** (0.011)	-0.036*** (0.008)	-0.037*** (0.011)	-0.0359*** (0.0078)	0.0008** (0.0004)	0.0003*** (0.0001)	0.0008* (0.0004)	0.0012*** (0.0003)
Own Grade Enrol. Squared*100	0.002 (0.004)	0.008*** (0.003)	0.008*** (0.003)	0.0075** (0.0029)	-0.0001 (0.0002)	-0.0002** (0.0001)	-0.0002** (0.0001)	-0.0002* (0.0001)
Lower Grade Enrollment	-0.011 (0.009)			0.002 (0.0030)	0.0001 (0.0004)			-0.0004*** (0.0001)
N=	15,976	16,053	16,053	15,976	15,976	16,053	16,053	15,976

The table shows the relationship between teacher experience and my instruments for pooled second and third graders *** Indicates 1% significance level and ** 5%. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions though results are not reported.

Table 8 - 2SLS Estimates of the Determinants of Second Grade Test Scores

Dependent Variable:	Language Scores			Math Scores					
Class Size Control:	Class Size Exogenous						Class Size Effect Zero		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Classroom Variables:									
Percentage of students in combination classes	-0.195*** (0.060)	-0.163*** (0.053)	-0.177*** (0.045)	-0.180*** (0.070)	-0.199*** (0.059)	-0.167*** (0.049)	-0.179*** (0.070)	-0.197*** (0.058)	-0.166*** (0.048)
Average Class Size	-0.095 (0.110)	-0.083 (0.110)	-0.086 (0.109)	-0.051 (0.122)	-0.062 (0.124)	-0.044 (0.121)			
Enrollment Controls:									
Enrollment in grade	-0.034 (0.027)	-0.019 (0.017)	-0.032 (0.027)	-0.021 (0.029)	-0.022 (0.019)	-0.019 (0.029)	-0.020 (0.029)	-0.022 (0.019)	-0.020 (0.029)
Enrollment in grade squared*100	0.005 (0.011)	0.003 (0.007)	0.005 (0.011)	-0.001 (0.012)	0.003 (0.008)	-0.009 (0.012)	-0.001 (0.012)	0.004 (0.008)	-0.001 (0.012)
Enrollment in lower Grade	0.016 (0.021)		0.016 (0.021)	0.004 (0.024)		0.003 (0.024)	0.005 (0.024)		0.003 (0.024)
Enrollment in lower grade squared*100	-0.003 (0.008)		-0.003 (0.008)	0.003 (0.009)		0.003 (0.009)	0.004 (0.009)		0.003 (0.009)
N	9950	9974	9950	9950	9974	9950	9950	9974	9950
Instruments	CSP	PCS	PCS PCS-1	CSP	PCS	PCS PCS-1	CSP	PCS	PCS PCS-1

This table presents the results of 2SLS regressions for the second graders Combination class percentage is the endogenous regressor. CSP stands for the Combination Class Predictor, PCS for Predicted Class Size and PCS-1 for the Predicted Class Size of the lower grade. All these instruments are defined in the text. Columns (1) –(6) treat class size as an exogenous regressor while columns (7) –(9) impose a zero coefficient. *** Indicates 1% significance level and ** 5%. All regressions are weighted by the number of test takers. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions though results are not reported.

Table 9 - 2SLS Estimates of the Determinants of Third Grade Test Scores

Dependent Variable:	Language Scores			Math Scores					
Class Size Control:	Class Size Exogenous						Class Size Effect = -.37		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Classroom Variables:									
Percentage of students in combination classes	-0.368***	-0.301***	-0.328***	-0.360***	-0.302***	-0.323***	-0.362***	-0.322***	-0.331***
	(0.102)	(0.066)	(0.060)	(0.121)	(0.076)	(0.069)	(0.125)	(0.077)	(0.074)
Average Class Size	0.102	0.098	0.093	0.078	0.084	0.069	-0.370	-0.370	-0.370
	(0.120)	(0.112)	(0.115)	(0.141)	(0.133)	(0.135)			
Enrollment Controls:									
Enrollment in grade	-0.078**	-0.072***	-0.075**	-0.054	-0.045**	-0.051	-0.025	-0.040	-0.015
	(0.031)	(0.018)	(0.030)	(0.036)	(0.021)	(0.035)	(0.078)	(0.021)	(0.069)
Enrollment in grade squared*100	0.026**	0.021***	0.026**	0.017	0.011	0.017	0.011	0.010	0.011
	(0.012)	(0.007)	(0.012)	(0.014)	(0.008)	(0.014)	(0.010)	(0.008)	(0.010)
Enrollment in lower grade	-0.008		-0.007	-0.001		-0.001	-0.001		-0.001
	(0.028)		(0.028)	(0.033)		(0.033)	(0.033)		(0.033)
Enrollment in lower grade squared*100	-0.002		-0.007	-0.002		-0.002	-0.003		-0.003
	(0.011)		(0.011)	(0.014)		(0.013)	(0.014)		(0.013)
N	6026	6079	6026	6026	6079	6026	6026	6079	6026
Instruments	CSP	PCS	PCS PCS-1	CSP	PCS	PCS PCS-1	CSP	PCS	PCS PCS-1

This table presents the results of 2SLS regressions for the third graders. Combination class percentage is the endogenous regressor. CSP stands for the Combination Class Predictor, PCS for Predicted Class Size and PCS-1 for the Predicted Class Size of the lower grade. All these instruments are defined in the text. All regressions treat class size as exogenous. *** Indicates 1% significance level and ** 5%. All regressions are weighted by the number of test takers. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions.

Table 10 - 2SLS Estimates for Test Scores with Double Instrumenting

	math		language	
	(1)	(2)	(3)	(4)
		A. Second Graders		
Percentage of students in combination classes	-0.174*	-0.180***	-0.202**	-0.193***
	(0.096)	(0.068)	(0.087)	(0.059)
Class Size	-0.383	-0.158	-1.085	-1.223
	(2.829)	(1.902)	(2.571)	(1.683)
N	9950	9950	9950	9950
Instruments	PCS PCS-1 NCL	PCS CLN CSP	PCS PCS-1 NCL	PCS CLN CSP
		B. Third Graders		
Percentage of students in combination classes	-0.368**	-0.355***	-0.395**	-0.369***
	(0.193)	(0.139)	(0.177)	(0.118)
Class Size	-1.263	-1.315	-1.844	-1.638
	(5.186)	(3.017)	(4.709)	(2.546)
N	6026	6026	6026	6026
Instruments	PCS PCS-1	PCS CLN CSP	PCS PCS-1	PCS CLN CSP

This table presents the results of 2SLS regressions where both class size and combination class percentage are treated as endogenous regressors CSP stands for the Combination Class Predictor, CLN for the predicted number of class groups, PCS for Predicted Class Size and PCS-1 for the Predicted Class Size of the lower grade. All these instruments are defined in the text. *** Indicates 1% significance level and ** 5%. All regressions are weighted by the number of test takers. Reported standard errors are adjusted to correct for clustering at the school level. The full set of demographic controls up to cubic terms is included in all regressions though results are not reported.

Table 11 - Effect Sizes of a five percentage point increase in combination class percentage

Grade Level:	second grade		third grade	
Dependent Variable:	math	language	math	language
Effect on all students	.035-.051σ	.039-.049σ	.081-.093σ	.082-.097σ
Effect on only combination class students	.24-.36σ	.25-.35σ	.58-.63σ	.59-.66σ

Effect sizes are calculated by dividing the change in test scores implied by a five percentage point increase in combination class students by the standard deviation of test scores. In this case the between class standard deviation is used since the within student standard deviation is unavailable. The second row is obtained through scaling the first row by the reciprocal of the proportion of students in combination classes.

Chapter 2 - Out of Control? What Can we Learn From the End of Massachusetts Rent Control?

2.1 Introduction

In a 1995 article Richard Arnott called for "revisionism on rent control," contending that most economists hold traditional views about the effects of rent control that are founded upon unrealistic models and scant empirical evidence. Indeed, few government policies have united economists in opposition as effectively as rent control. A well known 1990 survey of economists found that over ninety percent believed that rent control decreased the quantity and quality of rental housing in an area (Alston, et al. 1992). In contrast to this virtual unanimity of opinion, the work of Olson (1988) and Kutty (1996) predicts that the effect of rent control on housing supply and rental quality is theoretically ambiguous and must be empirically determined.

Most recent empirical evidence on the effects of rent control in the U.S. comes from studies of New York City.[1] While New York City has had the most domestic experience with rent control, measured in volume of rent controlled apartments, its rent controls consist of complex overlapping sets of regulation regimes enacted at different points in time. This makes it atypical of American cities that have experienced rent control and makes measurement of the effects of any one set of rent controls difficult.

However, Rent control policies have also played an important role in hundreds of American cities from San Francisco to Boston. Many of these policies rose out of social unrest and tenant activism of late 1960's and early 1970's and remained the focus of animosity between activists and landlords. In 1990, a group of homeless advocates led

[1] E.g. Moon and Stotsky (1993), Gyourko and Linneman (1989, 1990), Glaeser and Luttmer (1997)

the seizure of an apartment building in Cambridge that had been left vacant in violation of the city control ordinance. Local landlords complained to the city council that with rents set below $150, the units were not worth the trouble of renting and threatened to raise a private army to prevent further seizures by force. Angered by the loss of property rights, these owners sought a political climate in which they could end rent control.

The opportunity came in November 1994, when landlords succeeded in placing an initiative on the Massachusetts ballot to ban rent control statewide. Though this initiative, known as Question 9, passed statewide, it was overwhelmingly defeated in the three Massachusetts cities with rent control, Boston, Brookline, and Cambridge. This externally imposed end to rent control provides an opportunity to study the effects of rent control on housing unit supply and quality in Eastern Massachusetts.

This paper examines the effects of rent control in Massachusetts on the willingness of owners to rent housing units, on the rent and cost levels of renter occupied apartments, the maintenance of those apartments, and length of tenancy. My results suggest rent control has little effect on new construction, but does induce owners to remove their units from the rental market. Additionally, rent control leads to large rent decreases and small but significant decreases in the maintenance of rental units. I also examine whether the effect of rent control on rent operates partly through spillover effects on non-controlled housing. I conclude that rent control may lower the rent of non-controlled units, possibly through spillover effects of decreased unit quality. These estimates allow calculation of Massachusetts' approximate welfare losses in due to rent control.

These estimates indicate little need for revisionism in the traditional economic assessment of rent control. Rather, they confirm the simple intuition that economists derive from very basic microeconomic models; rent control artificially lowers price, decreases supply and decreases unit quality. Although inefficiencies are inherent in any price control, rent control is an opportunity to study a price control that is large in magnitude and has large effects on the behavior of many tenants and landlords.

These findings also provide important evidence to policy makers. Almost a decade after the abolition of Massachusetts rent control the debate over its legacy continues. The Mayor of Boston, Thomas Mennino, often lobbies for a return to rent control and blames the repeal of rent control laws for the skyrocketing Boston rental rates of the late 1990's. Opponents argue that this price increase was due to an economic boom and that renewed rent control will exacerbate the fundamental problem of unit undersupply. Both sides previously lacked the clear evidence on rent control effects that this study provides.

This study of rent control is also relevant to the policy process of other cities. From San Francisco to Washington D.C. there is a need to assess the potential impact of further weakening or completely repealing existing controls. This study provides a clearer empirical understanding of what repealing rent control might mean for these cities.

The remainder of the paper is organized as follows: Section 2 provides additional institutional detail about rent control in Massachusetts and its repeal, Section 3 discusses the data, Section 4 presents the results and Section 5 concludes.

2.2 Rent Control in Massachusetts

2.2.1. The Rent Control Laws

Only 3 cities in Massachusetts maintained rent control ordinances from 1985 to 1994: Boston, Cambridge and Brookline. Though the laws in each city varied in the details, all rent control policies shared four common elements. The first was a centralized board or commission which was empowered to set maximum rents. Though a rent control board would occasionally approve general rent increases for all controlled apartments in the city, most increases were approved on an individual basis, requiring the landlord to provide proof of an operating cost increase such as a rise in property taxes or utility rates.

The second common element was a set of removal regulations, designed to keep the supply of rental apartments from shrinking. The most important removal laws were passed in the early 1980's to reduce condominium conversions. In Cambridge, for example, condominium conversions required the express approval of the rent control board. In Boston, the owners of buildings destined for condominium conversion were also required to give tenants up to three years advance notice before conversion, help the tenants find new housing, and pay a severance fee. These regulations made it very difficult to remove a controlled unit by converting it into a condominium.

A third set of regulations concerned tenant protection. These forbade the eviction of tenants without approval of the board. A short list of grounds for eviction was codified in the laws, and landlords faced the burden of proving violations. Another regulation set monetary punishments for landlords who failed to maintain the provision of essential services such as heat and running water to their units. Most controversially, Cambridge forbade landlords from leaving controlled units vacant for more than three months.[2]

The final common element was a system for removing units from further control. Each city exempted newly constructed housing and housing units that were completely remodeled, provided a certain amount of money was spent updating the units.[3] Boston and Cambridge also exempted owner-occupied 2 and 3 family homes from rent control.

Boston and Brookline also had forms of vacancy decontrol during this period. Boston's decontrols were adopted in 1984, just before the period covered by this study. This regulation allowed a rental unit to leave active control once it was completely vacated by its present tenants. It then passed into a state of passive control, still subject to the removal and eviction protections. Landlords were free to raise rent on a yearly basis, but tenants in such apartments could appeal to the rent control board to override unfair rent increases. In Brookline, a simpler rent decontrol system, adopted in 1991, exempted most vacated apartments from all rent controls.

2.2.2 The repeal

Throughout the early 1990's the fight to end rent control was led by the Small Property Owners Association (SPOA). Massachusetts cities with rent control also had a majority of tenant residents as well as government officials with a favorable attitude toward rent control. Thus SPOA had little success in repealing rent control laws through local action. Finally, in 1994 they changed tactics and proposed a statewide ban of rent control in the form of a ballot initiative, Question 9. The ensuing political campaign was

[2] There is little evidence that the Cambridge ordinance banning vacancies was ever actively enforced.
[3] In Boston the required investment ranged from $15,000 to over $35,000 per unit depending on the age of the unit, and the year of renovation.

bitterly contested on both sides, with a number of tenant activist groups organizing opposition to the proposed law.

In the November election the voters approved the Question by a narrow margin (51% favored it). After failing to obtain judicial intervention, tenant advocates sought help from the state legislature. They argued the end of rent control on January 1, 1995 would be marked by wholesale evictions of the poor and elderly, and a massive rise in homelessness. Finally, the governor brokered a compromise whereby tenants that met certain age or poverty guidelines could retain their controlled unit status for a 1-2 year transition period. Thus, most units were immediately decontrolled, but a few remained in control until January 1997.[4]

2.3 Data & Estimation

2.3.1 Data Description

The primary data for this study come from the American Housing Survey – Metropolitan Sample (AHS-MS) for the Boston MSA for the years 1985, 1989, 1993, and 1998. The AHS-MS interviews a sample of households drawn from census long forms in several cities across the United States on a rotating basis. The unit of observation is the housing unit, not the inhabitant. This survey is particularly useful because it asks residents a wide variety of questions about the rent, maintenance, and physical characteristics of their units and provides a wealth of information on the inhabitants. There are, however, disadvantages to using this survey data.

[4] Fewer than 2,000 people applied for these extension waivers, supporting the case of rent control opponents that most rent controlled tenants were not poor or elderly.

One primary limitation is the level of geographical identification available to the researcher. Due to confidentiality concerns, the smallest identifiable geographical unit in the Public Use files of the AHS-MS is the zone. Though vaguely defined as, "a roughly homogenous region of greater than 100,000 population" by the Census Bureau, in practice a zone in the Boston MSA corresponds to a small group of towns or cities. An exception is the city of Boston where a zone corresponds to a group of neighborhoods. Because it is impossible to make more precise geographical distinctions than the zone, I treat entire zones as controlled or non-controlled even though two zones are only partially subject to rent control. These are zone 112 which encompasses Cambridge and Somerville and zone 110 which includes Brookline and Newton. In both cases the geographical proximity and similarity of the cities involved supports the idea that they comprise a rental market. Nevertheless, the effect of being in a rent controlled zone, explored by this paper, is not precisely the same as the effect of being in a rent controlled city.

Furthermore some of the Boston MSA zones in the survey changed geographical boundaries as the composition and extent of the MSA changed over time. A list of the 1990 census tracts within each zone allows me to construct zone boundaries for constant geography zones in the 1985-98 time period. I limit my sample to these constant boundary zones which include almost all of the interior suburbs surrounding Boston. In practice this excludes many outlying regions of the Boston MSA from the sample. However, the excluded areas comprise the portion of the MSA least comparable with the rent controlled areas. Figure 1 provides a map of the zones in the Boston MSA used in this study as well as a list of the towns or cities that comprise each of the 21 Zones.

Figure 2 shows the zone boundaries and lists the neighborhoods that make up the zones in the city of Boston.

In addition to geographical limitations there are other potential difficulties with the AHS-MS. In 1998 the survey switched from personal interviews to computer assisted interviewing, which may have affected the comparability of responses over time.[5] Also, the 1998 questionnaire rephrased some questions from previous survey years. Most notably, certain items about the external condition of the unit that were previously answered by the surveyor were now asked to the tenant. I have excluded these items from my analysis.

I examine two different samples of Boston AHS data. The first sample includes all housing units in each of the four survey years that were properly interviewed, no matter the tenure of their residents. The second is a subsample that includes only the units in which the resident is a tenant. This renters sample also excludes public housing units and housing units where the tenant pays a non-monetary rent. Tables 1-1a give descriptive statistics for both samples. The study also uses data on building permits issued by the localities in Eastern Massachusetts provided by the Department of Housing and Urban Development.

2.3.2 Empirical Strategy

This study seeks to identify the effect of rent control on controlled zones. Specifically, I examine the effects on rent control on several housing unit characteristics

[5] Because I compare zones that were decontrolled with zones that remained control free, survey changes will only bias my results if they have a differential effect on these two types of zones.

including unit supply, level of rent and housing costs, unit quality, and length of renter tenure.

I attempt to identify the these effects of rent control by comparing outcomes in zones that were decontrolled with zones that did not change status. Consider an example where unit rent (P) is the outcome of interest. Next define two types of local markets or zones, those that were at one time controlled (C_j=1)and those that were never controlled (C_j=0). Further define years as pretreatment years if they are before 1994 (T=0), and define 1998 as the treatment year (T=1). Then the treatment indicator (D_{ijt}) is the product of C_j and T. This indicator variable picks out units in zones decontrolled by the 1994 law, when seen in 1998. The units in all other zones serve as a control group (D_{ijt}=0). I then assume we can write a unit's expected rent as:

$$E(P_{ijt} \mid C_j, T, D) = \alpha C_j + \beta T + \delta D_{ijt} \qquad (1)$$

where i indexes unit, j zone and t time. Adding an error term produces the familiar difference in differences estimator.

This approach allows units to differ only by location and date of observation. I introduce more flexibility into the analysis by adding a series of housing unit characteristics X_{ijt} as control variables in the regression. Furthermore, I include zone level effects ϕ_j for each zone rather than assuming zone effects are equal within once-controlled and never-controlled groups. Finally the dichotomy between pre-treatment and post-treatment periods is broken by allowing each period a unique effect γ_t. These changes produce the following estimating equation

$$P_{ijt} = \phi_j + \gamma_t + \delta D_{ijt} + X_{ijt}\theta + \varepsilon_{ijt} \tag{2}$$

If the included controls are sufficient to account for unit specific factors that change over time then δ can be interpreted as the causal effect of decontrolling a zone on the rent of units in the zone.

Rent control might affect rent levels in a zone by exclusively suppressing the rent of controlled units. Alternatively, the presence of controlled units may influence the rent of non-controlled rental housing. To test for the existence of such spillover effects I estimate:

$$P_{ijt} = \phi_j + \gamma_t + \eta RC_{ijt} + \lambda PCT_{ijt} + X_{ijt}\pi + \nu_{ijt} \tag{3}$$

where RC_{ijt} is a dummy variable equal to one if the unit is controlled and equal to zero otherwise and PCT_{ijt} gives the percentage of units in zone j that are controlled at time t.

Because rent controlled status is not randomly assigned, OLS estimates of η may be biased even if the spillover term λ is truly zero. I use instrumental variables estimation to account for this problem. Instrumental variables estimation also corrects for the mechanical connection between η and λ that arises since the PCT variable is a zone level average of the RC variable. If there are no spillover effects (λ=0) the policy instrument is sufficient for identification. However, in the presence of spillover effects multiple instruments are necessary to identify both direct and spillover effects. I construct these instruments using the legal requirements for rent controlled units. For example, because owner-occupied 2 family houses are excluded from control, the interaction of dummy variables for 2 family houses and owner-occupied premises with a pre-treatment year indicator and a controlled zones indicator provides a potential instrument. Similar instruments are constructed from clauses in the rent control laws exempting new construction and single family homes.

2.4 Results

2.4.1 Supply

Opponents of rent control in Massachusetts often cite housing permit data as evidence that rent control led to an undersupply of rental housing. Their argument is summarized by Figure 3, which shows the number of building permits granted in Boston, Brookline and Cambridge for multi-family housing in the years 1990-2000. It appears that the number of permits granted rose markedly after the end of rent control. This graph, however, does not show the entire story.

Figure 4, shows the same data but for the entire 1980-2000 period. With these added years to provide context, the low permit years of the early 1990s appear to be the anomaly. Of course, the permit graphs are not definitive statements about supply, since they do not count actual construction activity, do not account for demolition and abandonment, and fail to consider the condominium market which builds permitted units that are not rented.

In order to provide a more detailed analysis, I split the supply issue into two components. Rent control might affect the extensive supply of housing units, that is the number of housing units available in total, or it might affect the intensive margin of rental housing units, that is the percentage of housing units that are rented.

Table 2 provides estimates of the effect of rent decontrol on extensive supply, intensive supply and the number of condominium units. All regressions flexibly control for the number of bedrooms, number of rooms, age and other characteristics of the unit. The first two columns address the effect of rent decontrol on the total supply of housing units in a zone. The extensive supply measure is constructed using the probability weights assigned to the units and census estimates to reconstruct the number of housing units in a zone in each sample year. I then divide the result by the total number of units in all 21 zones for that year. Thus, the dependent variable represents the percentage of the MSA housing units in that year in that zone. A zone that grows at the same rate as the MSA will retain the same supply percentage over time.

Table 2 indicates that being in a decontrolled zone leads to an increase of about .2 percent in the relative housing supply of that zone. This small effect, however, is not statistically different than zero. When Boston and the other decontrolled zones are treated separately Boston has a positive growth in housing supply due to rent decontrol that is not shared by the zones of Cambridge and Brookline. However, the standard error of this estimate remains large.

In contrast the results on intensive supply seem clearer. Columns (3) – (6) provide evidence that rent control decreases the number of rental units and condominiums. Reported standard errors are corrected for clustering at the zone year level. Columns (3) – (4) report estimation results of a linear probability model with the dependant variable equal to one if the unit a rental and zero otherwise. The coefficients in column (3) indicate that rent decontrol is associated with a 6 percent increase in the probability of a unit being a rental. In regressions that consider condominium status in a similar fashion, rent decontrol is associated with an 8 percent increase in the probability of a unit being a condo.

These results seem counterintuitive. They indicate that removal laws had the intended effect of reducing condominium conversion. But what were the non-rented units' alternate uses if they were not converted into condominiums? There are a couple of possible solutions to this quandary. The first is that the results may apply to housing units on two separate margins. Some condominium units were not rented due to the rent control laws, while some rental units failed to convert to condominium status due to the laws. The prevalence of 2-3 family housing in the Boston area suggests a second possibility: because such units were exempt from control if the owner maintained a residence there, some of those units might not have been rented to avoid controlled status for the rest of the building.

Table 2a provides additional evidence that rent control caused units to be removed from rental markets. It uses the limited longitudinal information available for the 1998 AHS-MS units showing whether the unit was owned or rented, its size, value, and level of rent in 1990. The first three rows of the table consider all units in the 1998 sample that were rented in 1990. The dependent variable is a dummy equal to one if the unit is owner occupied in 1998 and zero otherwise. These linear probability model regressions control for rent, value and unit size as indicated in the table. In general, there is no significant difference between decontrolled zones and other zones in the number of renter units that converted to owner occupancy. Columns (4) – (6) however, use the owner occupied units in 1990 as the sample and consider whether they converted to renter occupancy. The results indicate that units in zones that ended rent controls were 7 percent more likely to convert to renter status and that this difference is significant.

In summary, there is weak evidence that rent control affected the extensive supply of housing units in Boston , but much stronger evidence that rent control led owners to shift units away from renting. The 6-7% change in rental probability between controlled and uncontrolled zones may seem small, but when applied to all three cities it implies that rent control kept thousands of units off the market.

2.4.2 Rent, Maintenance, and Tenure

If rent control operates by capping rents below market rates, it seems sensible that controls would have negative rent effects. This section measures the magnitude of these effects. Additionally it presents findings about the effects of rent control on unit maintenance and the length of renter tenure. Table 3 presents results from the estimation of equation (2) using a variety of outcome variables. Column (1) measures the effect of being in a decontrolled zone on rent levels. It indicates that the end of rent control is associated with an $84 jump in rent. Additionally, the year effect coefficients show that rents increased at an increasing rate over this time period. The reported zone fixed effects are relative to the base zone 114, which includes a number of areas just outside Boston.

Column (2) uses a more comprehensive measure of housing costs as the dependent variable. It adds is the cost of utilities to the rent. The results imply that ending rent control causes housing costs to jump by $61. The smaller coefficient on cost makes sense as rent control provides strong incentives to landlords not to include any utilities in the rent, a practice they might forgo when controls are lifted. Both the cost and rent coefficients are significantly different from zero.

Ideally, to estimate the effect of rent control on unit maintenance, I would use a full history of the maintenance expenditures by the landlord over the life of the building. Unfortunately, this data is not available. The rent control literature traditionally addresses this problem by looking at the physical condition of rental units not at maintenance spending.

I adopt this practice in columns (3) – (5), which demonstrate how the end of rent control affected various measures of unit condition. Column (3) is a linear probability model where the dependent variable takes on a value of 1 if the unit experienced a major maintenance problem in the last season. These problems include plumbing and heating failures as well as pipe leaks, wiring shorts or other electrical problems. The results indicate that decontrol reduces the probability of a unit experiencing such problems, but the coefficient is not significantly different from zero. The dependent variable in column (4) is the number of these major problems experienced by units. Again, the estimated effect is negative but insignificant.

In contrast, column (5) considers the effect of decontrol on what might be thought of as less functional maintenance items such as broken paint or plaster, holes in the walls or floors, and loose railings. The dependent variable is a dummy indicating whether one of these problems is present. The estimates demonstrate that ending rent control leads to a significant reduction in these maintenance problems. A unit was almost 6% less likely to experience such problems once decontrolled. Though rent control does not seem to lead to catastrophic maintenance failures, it appears to reduce the maintenance performed on rental units. As landlords can be fined for allowing water and heat failures, but not for cracked paint, this result is not surprising.

The final column of table 3 considers the effect of rent control on the length of time a renter stays in a unit. To the extent that artificially low rents reduce the mobility of the population they impose inefficiency. People who would otherwise move away decide to stay in a controlled unit to keep the advantage of an artificially low rent. Column (8) shows that decontrol is associated with a decrease of renter stays of 1.74 years. This is sizeable when compared to the mean renter stay of 6 years in the sample.

Although results of Table 3 are evocative, they could be an artifact of sample composition. As discussed previously, the end of rent control caused units to shift from owner to renter status. Thus the zone level changes in 1998 might not be due to changes of unit quality or rent caused by rent control, but due to higher quality units entering the renter sample. To deal with this possible criticism, Table 4 presents the same regressions using a composition constant sample. This sample is obtained by omitting all the 1998 units that were owner occupied in 1990. This new sample is composition constant because all of its units were rented during the rent control regime as well as after control ended.

The results in Table 4 indicate that compositional issues are not a significant worry in interpreting the original evidence. The results are very similar with the exception of the renter stay estimate, which is significantly larger (in absolute value) in the composition constant sample. This makes sense. Only including apartments rented in 1990 excludes newer renters in converted buildings from the sample.

Table 5 presents a series of specification checks on the Table 3 results. The first four columns divide the decontrolled zones into Boston and other categories. The rent and cost effects are somewhat smaller in the Boston zones while the maintenance and renter stay results are almost the same across all zones. The final four columns add interaction terms to mimic hypothetical policy changes conducted in 1993 and 1989 and compare them to the real end of rent control in 1998. In all cases the real policy change provides significant results while the imaginary experiments result in small and insignificant coefficients.

The results in this section indicate that there are rent, maintenance, and renter stay effects associated with being in a decontrolled zone. However, the magnitude of the rent effects is potentially troubling. If the end of rent control only raises the rents in the 12.5% of units under active control in the average controlled zone then my estimates imply a $640 average rent reduction due to rent control for each controlled unit. It is possible, however, that rent control also affects units that are not actively controlled. In the case of Boston, this may operate through the passive rent appeal system discussed earlier. Alternatively, the reduced care given to rent controlled units may make the zones with rent control less desirable for those living in non-controlled housing. This spillover effect due to sub-optimal maintenance may decrease all rents in an area.

2.4.3 Spillover Effects of Rent Control

Estimates of Equation (3), which permits rent control to affect both controlled and non-controlled units, are presented in Table 6. These regressions contain the same unit characteristic controls as earlier regressions as well as zone and time fixed effects, though these are not reported in the table. Columns (1) and (2) show estimates of the effect of being a rent controlled unit on rent with the potential spillover coefficient, λ constrained to equal zero.

The OLS Estimates in panel A indicate that rent control is associated with a \$170-\$180 monthly decrease in rent levels.[6] However, because rent controlled status is not randomly assigned, this coefficient may not have a clear causal interpretation. Panel B addresses this through Two stage Least Squares Estimation of Equation (3). In this model the instruments are interactions between certain unit characteristics and whether the unit is in a zone that had rent control, as well as the policy instrument. The characteristics include whether the unit is a single family house, whether the unit is an owner occupied 2-3 family house, and whether the unit was built before the rent control regulations were adopted. Each of these characteristics is used to define eligible units for rent control, and hence is correlated with the rent control status of the unit. The key identifying assumption is that the main effect of these characteristics on rent is the same in the rent controlled zones as in the non-controlled zones.

The IV estimates are much larger than the corresponding OLS coefficients. They indicate that rent control decreases the rent of a controlled unit by about \$450 a month, and decreases housing costs by about \$340 dollars a month. This represents a large proportion of the 750 dollar average rent for decontrolled zones in 1998.

Columns (3) and (4) relax the constraint on λ, allowing the percentage of rent controlled units in a zone to affect the rent of the non-controlled units. The OLS Results suggest that the effects of control on controlled units are sizeable, decreasing rent about

[6] All rent and cost data is in 1998 dollars.

$200 a month. The spillover effects are smaller but also statistically significant. The coefficients imply that having 10-12% rent controlled units in your zone will decrease your rent by $23-$28 a month. As previously mentioned, a possible explanation for this result might be chronic maintenance problems caused by rent control.

The Instrumental Variables estimates of this relationship are also suggestive. They indicate that rent control lowers the rent of controlled units by about $320, while the estimated spillover coefficient is about the same as under OLS. However, these estimates are insufficiently precise to conclude that either coefficient is significant at anything smaller than a ten percent level. The coefficient on the potential spillover effect for housing costs is actually positive but very imprecisely estimated. These results are far from definitive, but provide some evidence of a moderate spillover effect from rent control.

2.4.4 Welfare

These estimates allow a rough calculation of the excess burden of Massachusetts rent control regulations. For this calculation, I conservatively assume that the price elasticity of demand for housing is -.5. I also account for the reduction in housing quality of rent controlled units. Since the chronic damage problems found in rent controlled housing would lead to an approximately 20% rent reduction if that unit was in a non-controlled zone, I assume that rent controlled units have about 8% of their rental loss compensated for by a decline in quality. This reflects the rent controlled units approximately 40 percent greater chance of having these problems.[7]

[7] This assumes the entire 5% increase in maintenance problems works through the 12.5% rent controlled apartments.

Under these assumptions, the deadweight burden of actively controlling approximately 50,000 units in these three Boston cities amounted to about 3.17 million dollars a month, or about 63 dollars per controlled apartment. This excess burden from underprovision of rental housing is in addition to the excess burden from misallocation of rental units discussed by Glaeser and Luttmer (1997).

Proper policy analysis compares the costs of such a program with the benefits it generates. The potential benefits of rent control arise through the transfer of surplus, but are only realized if social preferences are such that more weight is placed on the recipients of this surplus than its original owners. A rough calculation, using the above assumptions indicates that the rent control policies in these three cities transferred roughly $15.1 million a month from landlords to tenants.

However, there is little evidence that rent control programs effectively transfer this surplus to tenants society might wish to help, such as the poor or minority households.[8] Table 7 provides some initial evidence that the recipients of rent control benefits in Boston might not have been those that society traditionally seeks to target in housing assistance programs. Only 26% of rent controlled apartments were occupied by renters in the bottom quartile of the household income distribution, while 30% of units were occupied by tenants in the top half of this distribution. Minorities, especially Hispanics, were similarly underrepresented in rent controlled housing, comprising only 4% or the rent controlled population. This suggests that much of the transferred surplus may have been received by wealthier households.

[8] Olsen (1998) reviews several such papers.

2.5 Conclusion

The sudden end of rent control in Massachusetts in 1995 provides a natural experiment to study the effects of rent control. My results indicate that the intuition presented in simple microeconomic models is correct. Rent control decreases the supply of rental units, as well as rent and unit maintenance. It also lengthens renter stays. In addition, some evidence suggests that rent control produces small spillover effects that decrease the rent of uncontrolled units in controlled areas.

Further work could describe in greater detail the distribution of the surplus shifted by the program. If much of the benefit accrues to white upper income households, rent control may prove to be an ineffective transfer program as well as an inefficient one. Researchers might also focus on investigation of spillover effects from rent control policies and whether they are an artifact of the lower maintenance levels in rent controlled areas.

References

Alston, R.M., Kearl, J.R. and M.B. Vaughan. 1992. Is There a Consensus Among Economists in the 1990's? *American Economic Review 82*, 203-209.

Arnott, R., 1995. Time for revisionism on rent control. *Journal of Economic Perspectives 9*, 99–120.

Ault, R.W., Jackson, J.D., Saba, R.P., 1994. The effect of long-term rent control on tenant mobility. *Journal of Urban Economics 35*, 140–158.

Early, D. W., 2000. Rent Control, Rental Housing Supply, and the Distribution of Tenant Benefits. *Journal of Urban Economics 48*, 185-204.

Early, D. W., Phelps, J.T. 1999. Rent Regulations' Pricing Effect in the Uncontrolled Sector: An Empirical Investigation. Journal of Housing Research 10, 267-285

Glaeser E.L., Luttmer, E.F.P. 1997. The Misallocation of Housing Under Rent Control. NBER Working Paper #6220.

Gyourko, J., Linneman, P., 1989. Equity and efficiency aspects of rent control: An empirical study of New York City. *Journal of Urban Economics 26*, 54–74.

Gyourko, J., Linneman, P., 1990. Rent controls and rental housing quality: A note on the effects of New York City's old controls. *Journal of Urban Economics 27*, 399–409.

Kutty, M. K., 1996. The Impact of Rent Control on Housing Maintenance. Journal of Housing Studies 11, 69-88.

Moon, C., Stotsky, J.G., 1993. The effect of rent control on housing quality change: A longitudinal analysis *Journal of Political Economy 101*, 1114–1148.

Olsen, E.O., 1972. An econometric analysis of rent control. *Journal of Political Economy 80*, 1081–1100.

Olsen, E.O., 1988a. What do economists know about the effect of rent control on housing maintenance? *Journal of Real Estate Finance and Economics 1*, 295–307.

Olsen, E.O., 1988b. Economics of Rent Control. *Regional Science and Urban Economics 28*, 673–678.

U.S. Census Bureau, U.S. Department of Housing and Urban Development. A*merican Housing Survey, Metropolitan Sample-Boston and documentation,* 1985, 1989, 1993, 1998.

U.S. Department of Housing and Urban Development. *Building Permits Database Boston MSA* 1980-2000.

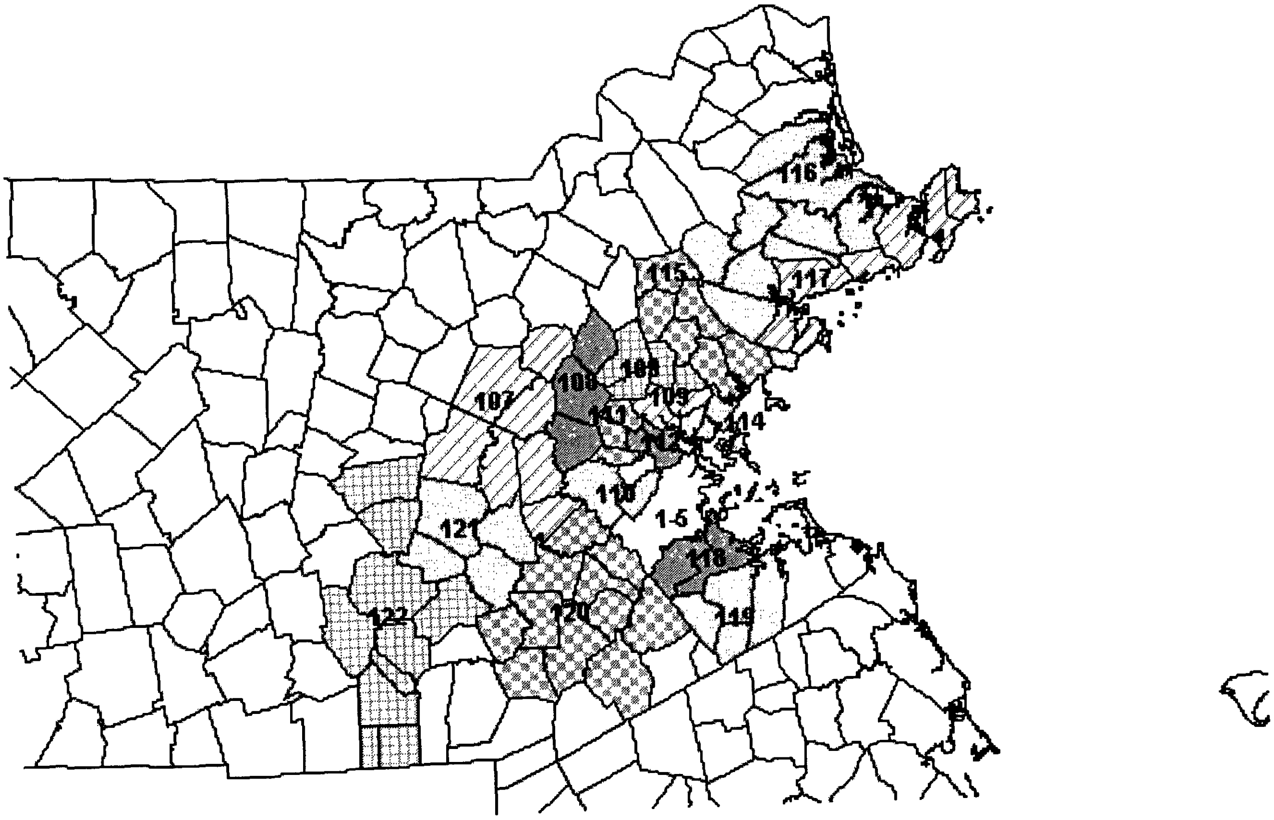


Figure 1: AHS-MS Geography Constant Zones

Zones 1-5
Boston

Zone 107
Bedford
Concord
Lincoln
Sudbury
Wayland
Wellesley
Weston

Zone 108
Burlington
Lexington
Waltham

Zone 109
Melrose
Stoneham
Winchester
Woburn

Zone 110
Brookline
Newton

Zone 111
Arlington
Belmont
Watertown

Zone 112
Cambridge
Somerville

Zone 113
Malden
Medford

Zone 114
Chelsea
Everett
Revere
Winthrop

Zone 115
Nahant
N. Reading
Lynn
Lynnfield
Reading
Saugus
Wakefield
Wilmington

Zone 116
Swampscott
Peabody
Danvers
Middleton
Topsfield
Hamilton
Wenham
Essex
Ipswich
Rowley

Zone 117
Marblehead
Salem
Beverly
Manchester
Rockport
Gloucester

Zone 118
Milton
Quincy

Zone 119
Braintree
Holbrook
Randolph
Weymouth

Zone 120
Dedham
Needham
Dover
Medfield
Millis
Norfolk
Walpole
Westwood
Norwood
Sharon
Canton

Zone 121
Ashland
Framingham
Natick
Sherborn

Zone 122
Blackstone
Holliston
Hopedale
Hopkinton
Marlborough
Mendon
Milford
Millville
Southborough
Upton

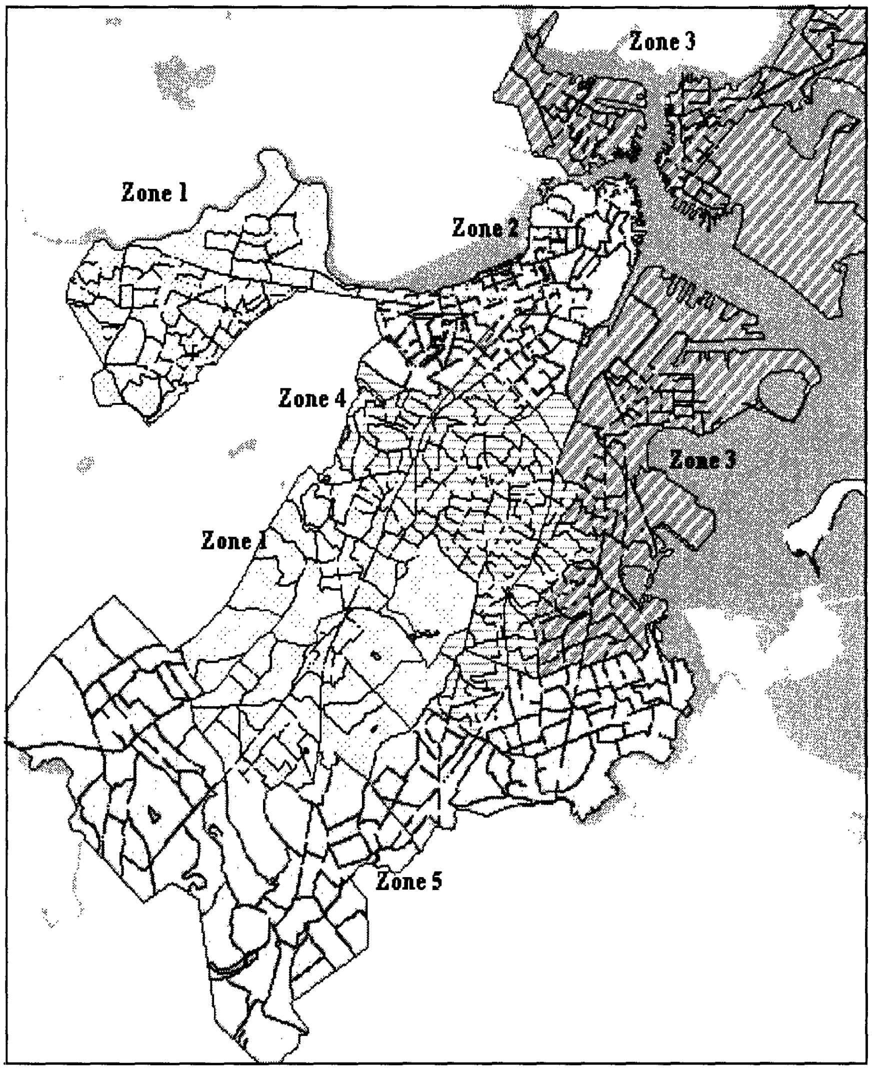


Figure 2 – Boston city zones AHS-MS
Source: author's calculations based on Census Bureau data

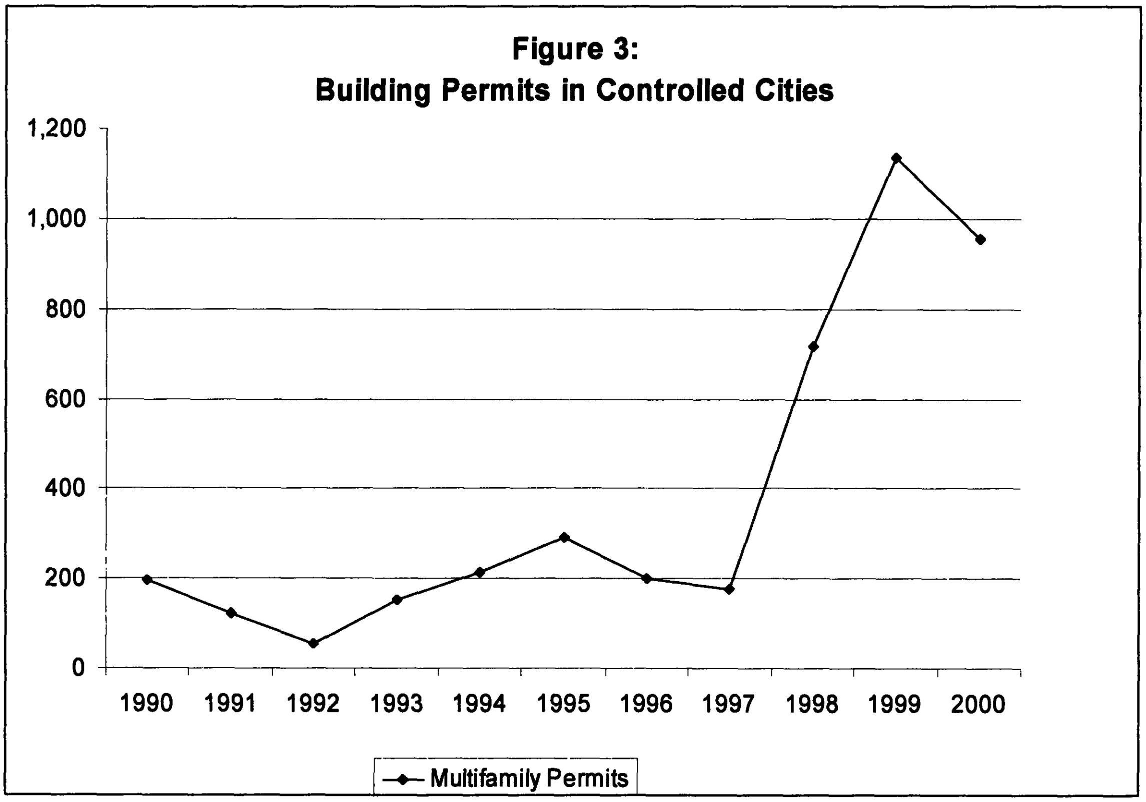
Zone 1
Allston
Brighton
Roslindale
Jamaica Plain

Zone 2
Back Bay
Fenway
Beacon Hill
Downtown
North End
South End
Chinatown

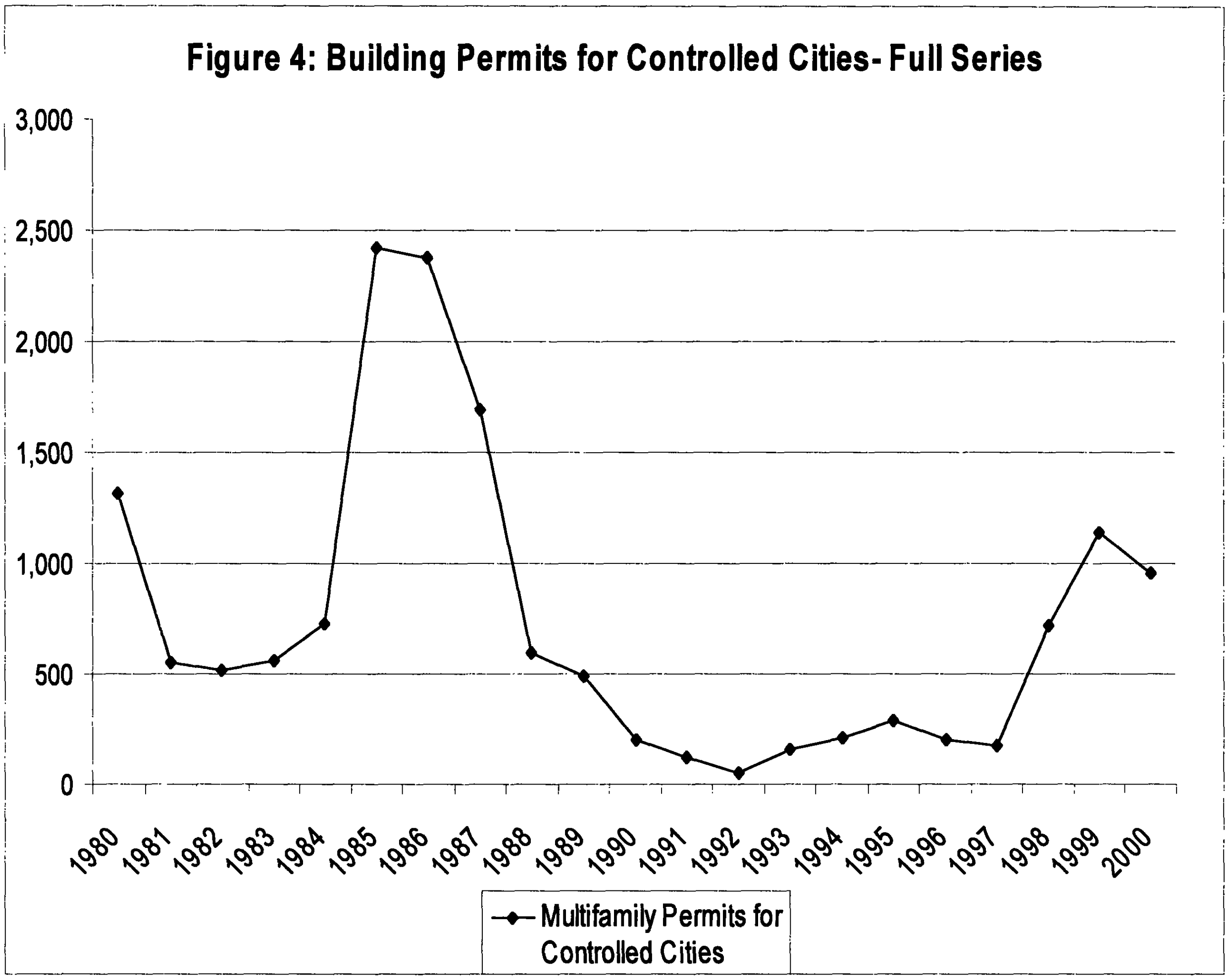
Zone 3
Charlestown
East Boston
South Boston
Roxbury (part)
North Dorchester
Harbor Islands

Zone 4
Mission Hill
Roxbury (part)
Dorchester (part)

Zone 5
Neponsett
Mattapan
West Roxbury
Hyde Park



The figure shows the number of units approved under building permits in Boston, Brookline and Cambridge from 1990-2000. Source: Department of Housing and Urban Development building permits database.



The figure shows the number of units approved under building permits in Boston, Brookline and Cambridge from 1980-2000. Source: Department of Housing and Urban Development building permits database.

Table 1 - Descriptive Statistics - Whole Sample

	mean	std. dev.
Unit Characteristics		
Condominium	0.075	0.263
Rent (1998 dollars)	486.559	293.283
Monthly Housing Cost	755.705	521.244
Length of stay	11.859	13.304
Percentage controlled	0.033	0.076
Number of bedrooms	2.46	1.189
Number of other rooms	3.078	1.258
Year built	1959	21.564
Single Family detached	0.419	0.493
2-4 Family	0.298	0.458
5-12 unit building	0.102	0.303
13-25 unit building	0.043	0.203
26-50 unit building	0.033	0.179
51+ unit building	0.067	0.25
Unit is gas heated	0.399	0.49
Unit is oil heated	0.47	0.499
Unit is electrically heated	0.121	0.326
Other source of heat	0.01	0.101
Central air in unit	0.125	0.331
Off street parking included	0.451	0.498
Owner/manager present	0.14	0.347
Public housing	0.049	0.215
Rent/service adjustment	0.036	0.187
Gas included in rent	0.118	0.323
Electricity in rent	0.097	0.297
Oil/coal/kerodsene in rent	0.143	0.35
Sanitation in rent	0.776	0.417
n=	10,512	

Table 1 - Descriptive Statistics - Whole Sample - continued

	mean	std. dev.
Maintenance Measures		
Maintenance Problem	0.354	0.478
Number of Problems	0.478	0.809
Physical Damage	0.084	0.278
Interior Leak	0.102	0.303
Exterior Leak	0.182	0.386
Plumbing Problem	0.03	0.17
Heating Problem	0.043	0.256
House self-rating	8.328	1.758
Wall holes	0.011	0.102
Flr. holes	0.054	0.227
Large peeling paint/plaster	0.047	0.212
Occupant Characteristics		
Number of occupants	2.462	1.416
Number of adults	1.913	0.918
reference person Age	48.925	17.727
reference person Male	0.566	0.496
HHIncome	43012	37602
reference person Black	0.073	0.26
reference person White	0.88	0.325
Owned in 1990		
Rented in 1990		
n=	10,512	

Source: Author's calculations based on AHS-MS Boston 1985, 1989, 1993, 1998 surveys. Dataset excludes units that were not properly interviewed

Table 1a -Descriptive Statistics - Renter Sample

	all renters		Pre 1995				1998			
			Controlled Zones		Non-controlled zones		Was controlled zones		Non Controlled zones	
	mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.
Unit Characteristics										
Condominium	0.053	0.224	0.046	0.209	0.051	0.219	0.087	0.282	0.046	0.211
Rent (1998 dollars)	486.559	293.283	398.181	239.274	413.658	211.488	757.868	369.674	677.276	310.314
Monthly Housing Cost	617.99	318.159	547.534	282.558	582.33	261.995	795.95	401.258	738.69	370.087
Length of stay	5.987	8.594	6.839	9.711	5.495	8.034	5.639	7.857	5.568	7.581
Percentage controlled	0.046	0.084	0.133	0.093	0	0	0	0	0	0
Number of bedrooms	1.829	0.991	1.822	1.004	1.788	0.899	1.961	1.3	1.85	0.879
Number of other rooms	2.424	0.862	2.394	0.791	2.437	0.781	2.478	1.248	2.415	0.836
Unit Type										
Single Family detached	0.08	0.271	0.014	0.119	0.07	0.255	0.232	0.422	0.139	0.347
2-4 Family	0.431	0.495	0.433	0.496	0.487	0.5	0.281	0.45	0.401	0.49
5-12 unit building	0.189	0.392	0.211	0.408	0.187	0.39	0.158	0.365	0.167	0.373
13-25 unit building	0.08	0.271	0.11	0.313	0.058	0.234	0.081	0.274	0.063	0.244
26-50 unit building	0.056	0.23	0.065	0.246	0.046	0.21	0.058	0.235	0.06	0.238
51+ unit building	0.121	0.326	0.131	0.338	0.117	0.321	0.127	0.334	0.102	0.303
Unit is gas heated	0.395	0.489	0.362	0.481	0.409	0.492	0.416	0.493	0.418	0.494
Unit is oil heated	0.413	0.492	0.457	0.498	0.433	0.496	0.333	0.472	0.319	0.466
Unit is electrically heated	0.181	0.385	0.159	0.366	0.153	0.36	0.246	0.431	0.262	0.44
Other source of heat	0.011	0.103	0.022	0.148	0.005	0.074	0.005	0.073	0.002	0.039
Central air in unit	0.1	0.3	0.07	0.256	0.097	0.296	0.142	0.349	0.146	0.353
Off street parking included	0.542	0.498	0.322	0.467	0.728	0.445	0.37	0.483	0.711	0.454
Owner/manager present	0.287	0.453	0.325	0.468	0.304	0.46	0.195	0.396	0.226	0.419
Public housing	0.103	0.304	0.122	0.328	0.083	0.276	0.112	0.315	0.107	0.309
Rent/service adjustment	0.063	0.244	0.057	0.231	0.07	0.255	0.064	0.244	0.062	0.241
Gas included in rent	0.222	0.415	0.234	0.424	0.219	0.414	0.257	0.437	0.167	0.373
Electricity in rent	0.196	0.397	0.202	0.402	0.181	0.385	0.219	0.414	0.203	0.402
Oil/coal/kerosene in rent	0.27	0.444	0.301	0.459	0.229	0.42	0.303	0.46	0.277	0.448
Sanitation in rent	0.86	0.347	0.954	0.209	0.917	0.276	0.646	0.479	0.655	0.476
n=	3590		1290		1406		397		462	

Source: Author's calculations based on AHS-MS Boston 1985, 1989, 1993, 1998 surveys

Table 1a -Descriptive Statistics - Renter Sample - continued

	all renters		Pre 1995				1998			
			Controlled Zones		Non-controlled zones		Was controlled zones		Non Controlled zones	
	mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.
Maintenance Characteristics										
Maintenance Problem	0.350	0.477	0.360	0.480	0.338	0.473	0.366	0.482	0.346	0.476
Number of Problems	0.526	0.934	0.574	1.016	0.476	0.860	0.532	0.884	0.545	0.955
Physical Damage	0.114	0.318	0.129	0.335	0.099	0.298	0.110	0.313	0.124	0.330
Interior Leak	0.126	0.332	0.138	0.345	0.125	0.331	0.113	0.317	0.108	0.311
Exterior Leak	0.118	0.323	0.101	0.301	0.124	0.330	0.126	0.332	0.140	0.347
Plumbing Problem	0.041	0.198	0.043	0.202	0.034	0.180	0.058	0.233	0.041	0.199
Heating Problem	0.062	0.314	0.085	0.341	0.046	0.336	0.060	0.238	0.048	0.214
House self-rating	7.874	1.937	7.862	1.958	8.072	1.937	7.414	1.857	7.735	1.873
Wall holes	0.016	0.126	0.017	0.131	0.012	0.109	0.016	0.125	0.025	0.156
Flr. holes	0.076	0.265	0.096	0.294	0.059	0.235	0.071	0.257	0.082	0.275
Large peeling paint/plaster	0.066	0.249	0.078	0.268	0.062	0.240	0.042	0.202	0.073	0.260
Tenant Characteristics										
Number of occupants	2.127	1.332	2.158	1.446	2.117	1.254	2.119	1.300	2.084	1.276
Number of adults	1.645	0.803	1.679	0.878	1.627	0.738	1.658	0.837	1.604	0.756
reference person Age	44.384	18.747	44.153	18.689	45.520	19.253	40.573	17.227	45.063	18.296
reference person Male	0.468	0.499	0.476	0.500	0.496	0.500	0.393	0.489	0.437	0.496
HHIncome	26362	22544	23187	19410	25034	19603	29979	27246	34886	29531
reference person Black	0.112	0.316	0.199	0.400	0.034	0.182	0.179	0.383	0.057	0.233
reference person White	0.818	0.386	0.715	0.452	0.935	0.246	0.681	0.466	0.864	0.343
Owned in 1990							0.131	0.337	0.167	0.374
Rented in 1990							0.869	0.337	0.833	0.374
n=	3590		1290		1406		397		462	

Source: Author's calculations based on AHS-MS Boston 1985, 1989, 1993, 1998 surveys. Dataset excludes units occupied by owner, as well as public housing units and units where the tenant pays non-cash rent.

Table 2- Effects of rent decontrol on housing supply

	Extensive Supply		Intensive Supply		Condominiums	
Dependent variable	Pct. units in the MSA		Prob a given unit is rented		Prob Unit w/ condo status	
	(1)	(2)	(3)	(4)	(5)	(6)
Decontrol Effects						
Ever controled zone*1998	0.0018		0.0608		0.0812	
	(0.0038)		(0.0237)		(0.0090)	
Boston *1998		0.0080		0.0829		0.0867
		(0.0045)		(0.0299)		(0.0105)
Zone Effects						
Other controlled zone*98		-0.0005		0.0170		0.0700
		(0.0035)		(0.0312)		(0.0084)
Allston - Brighton - Roslindale	-0.0117	-0.0134	0.0469	0.0418	0.0169	0.0156
Jamaica Plain	(0.0075)	(0.0066)	(0.0243)	(0.0264)	(0.0085)	(0.0082)
Back Bay - Fenway - Downtown	-0.0220	-0.0287	0.0067	0.0014	0.0399	0.0385
North End - South End	(0.0111)	(0.0102)	(0.0260)	(0.0277)	(0.0126)	(0.0121)
Charlestown - E. Boston -	-0.0058	-0.0092	0.0260	0.0198	-0.0246	-0.0262
S. - Boston - N. Dorchester	(0.0067)	(0.0061)	(0.0267)	(0.0239)	(0.0077)	(0.0082)
Mission Hill - Roxbury (part)	-0.0333	-0.0372	0.1219	0.1169	-0.0446	-0.0458
Dorchester	(0.0075)	(0.0069)	(0.0420)	(0.0388)	(0.0114)	(0.0119)
Mattapan - W. Roxbury	-0.0219	-0.0208	0.0174	0.0119	-0.0421	-0.0435
Hyde Park	(0.0106)	(0.0093)	(0.0264)	(0.0258)	(0.0144)	(0.0144)
Brookline - Newton	-0.0088	-0.0077	0.0109	0.0225	0.0519	0.0548
	(0.0074)	(0.0065)	(0.0291)	(0.0236)	(0.0097)	(0.0099)
Cambridge - Somerville	0.0051	0.0036	0.0234	0.0337	-0.0037	-0.0011
	(0.0046)	(0.0041)	(0.0180)	(0.0189)	(0.0082)	(0.0081)
Year Effects						
1989	-0.0203	-0.0246	0.0437	0.0428	0.0212	0.0212
	(0.0175)	(0.0155)	(0.1395)	(0.1397)	(0.0160)	(0.0162)
1993	-0.2790	-0.2672	0.1500	0.1494	0.2729	0.2728
	(0.2654)	(0.2320)	(0.1223)	(0.1225)	(0.0938)	(0.0938)
1998	-0.1649	-0.1321	0.0489	0.0485	0.1781	0.1776
	(0.1660)	(0.1459)	(0.0828)	(0.0831)	(0.1024)	(0.1024)
n=	84	84	10480	10480	10512	10512
Unit of observation	Zone	Zone	Unit	Unit	Unit	Unit

All regressions include quadratics and cubics for number of bedrooms and other rooms, categorical age and number of building units, age of building interactions with year, and controls for type of heat and presence of central air as well as zone and year fixed effects. Standard errors are corrected for clustering on the zone year level. Col (3)-(6) are linear probability models with the dependent variable a dummy for rental tenure (3) & (4) or condo status (5) & (6).

Table 2a - Transition of unit tenure 1990-1998

	Pr(transition renter---owner)			Pr(transition owner---renter)		
	(1)	(2)	(3)	(4)	(5)	(6)
Zone decontroled in 1995	-0.036	-0.024	-0.024	0.119	0.082	0.071
	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)	(0.018)
control for number of rooms	no	yes	yes	no	yes	yes
control for categorical 1990 rent	no	yes	yes	no	no	no
control for categorical 1990 value	no	no	no	no	yes	yes
interaction of rooms/rent or value	no	no	yes	no	no	yes
n=	1648	1648	1648	1875	1875	1875

Regressions are linear probability models without outcome dummies =1 if the unit changed tenure status from 1990 to 1998. The sample in columns (1)-(3) is all units from the 1998 survey that were rented in 1990. The sample in (4) – (6) is all 1998 units that were owner occupied in 1990. Categorical controls reflect value intervals provided in the data.

Table 3- Effects of rent decontrol on Zone housing characteristics

dependent variable	rent	cost	severe	# probs	chronic	length
	(1)	(2)	(3)	(4)	(5)	(6)
Decontrol Effects						
Ever controled zone*1998	83.78	64.25	-0.02	-0.11	-0.05	-1.84
	(20.63)	(24.73)	(0.05)	(0.08)	(0.02)	(0.59)
Zone Effects - Controlled Zones						
Allston - Brighton - Roslindale	100.50	109.27	0.08	0.07	0.02	0.61
Jamaica Plain	(26.86)	(32.95)	(0.04)	(0.12)	(0.03)	(0.78)
Back Bay - Fenway - Downtown	200.14	199.51	0.04	0.02	0.00	1.11
North End - South End	(25.37)	(31.36)	(0.05)	(0.11)	(0.03)	(0.87)
Charlestown - E. Boston -	-75.25	-77.30	-0.04	-0.19	-0.06	3.23
S. - Boston - N. Dorchester	(32.35)	(33.94)	(0.05)	(0.11)	(0.03)	(0.85)
Mission Hill - Roxbury (part)	-111.84	-71.60	0.00	-0.04	0.04	1.12
Dorchester	(25.69)	(33.71)	(0.06)	(0.14)	(0.05)	(1.39)
Mattapan - W. Roxbury	-109.26	-91.23	0.10	0.13	0.05	2.88
Hyde Park	(25.81)	(33.54)	(0.05)	(0.10)	(0.03)	(0.66)
Brookline - Newton	197.93	207.45	0.09	0.02	0.00	-0.36
	(35.53)	(39.05)	(0.04)	(0.12)	(0.02)	(0.70)
Cambridge - Somerville	8.02	15.93	0.14	0.25	0.08	1.24
	(23.94)	(29.51)	(0.05)	(0.10)	(0.02)	(0.76)
Year Effects						
1989	-18.48	156.15	-0.05	-0.36	-0.07	1.77
	(87.09)	(71.50)	(0.14)	(0.38)	(0.03)	(1.74)
1993	22.10	123.80	-0.20	-0.29	-0.20	2.27
	(90.94)	(123.03)	(0.19)	(0.22)	(0.09)	(1.55)
1998	119.73	150.37	-0.18	0.05	0.01	4.67
	(61.67)	(64.38)	(0.24)	(0.34)	(0.06)	(0.84)
n=	3541	3445	3145	3047	3543	3542
number of clusters	84	84	84	84	84	84

All regressions include quadratics and cubics for number of bedrooms and other rooms, categorical age and number of building units, age of building interactions with year, and controls for type of heat and presence of central air as well as zone and year fixed effects. Column (1) also controls for whether various utilities are included in the rent. Standard errors are corrected for clustering on the zone year level. All regressions weighted by inverse probability that the housing unit is in the sample. Severe problems include pipe and plumbing failures, heating failures and electrical problems. Chronic problems include holes in wall or floor, chipped or peeling paint, plaster damage, loose railings, etc. Rents and housing costs are in 1998 dollars.

Table 4 - Effects of rent decontrol on housing - composition constant sample

dependent variable	rent	cost	severe	# probs	chronic	length
	(1)	(2)	(3)	(4)	(5)	(6)
Decontrol Effects						
Ever controled zone*1998	82.37	56.20	0.00	-0.08	-0.06	-2.51
	(22.97)	(25.93)	(0.05)	(0.08)	(0.02)	(0.60)
Zone Effects - Controlled Zones						
Allston - Brighton - Roslindale	76.94	84.41	0.07	0.04	0.03	-0.33
Jamaica Plain	(23.87)	(26.90)	(0.04)	(0.10)	(0.03)	(0.80)
Back Bay - Fenway - Downtown	150.38	151.93	0.01	-0.01	0.00	0.44
North End - South End	(22.10)	(22.49)	(0.05)	(0.10)	(0.03)	(0.82)
Charlestown - E. Boston -	-46.91	-45.32	-0.06	-0.23	-0.06	2.88
S. - Boston - N. Dorchester	(26.87)	(27.05)	(0.05)	(0.11)	(0.03)	(0.61)
Mission Hill - Roxbury (part)	-76.34	-60.00	0.00	-0.04	0.04	0.63
Dorchester	(31.29)	(32.82)	(0.07)	(0.14)	(0.06)	(1.22)
Mattapan - W. Roxbury	-48.08	-38.61	0.10	0.11	0.05	2.61
Hyde Park	(22.64)	(26.58)	(0.05)	(0.10)	(0.03)	(0.61)
Brookline - Newton	65.57	86.80	0.02	-0.04	-0.03	-2.14
	(36.34)	(49.58)	(0.07)	(0.17)	(0.04)	(0.65)
Cambridge - Somerville	124.37	129.26	-0.01	-0.11	0.01	-1.60
	(25.55)	(27.33)	(0.04)	(0.09)	(0.03)	(0.57)
Year Effects						
1989	214.18	221.62	-0.05	-0.29	-0.08	1.01
	(35.75)	(38.36)	(0.13)	(0.19)	(0.03)	(0.69)
1993	359.33	652.06	-0.20	-0.52	-0.06	1.78
	(33.87)	(142.42)	(0.19)	(0.43)	(0.11)	(2.00)
1998	549.34	559.77	0.08	-0.09	-0.03	4.89
	(78.02)	(77.19)	(0.16)	(0.24)	(0.05)	(0.94)
n=	3476	3388	3098	3000	3490	3490
number of clusters	84	84	84	84	84	84

All regressions include quadratics and cubics for number of bedrooms and other rooms, categorical age and number of building units, age of building interactions with year, and controls for type of heat and presence of central air as well as zone and year fixed effects. Column (1) also controls for whether various utilities are included in the rent. Standard errors are corrected for clustering on the zone year level. All regressions weighted by inverse probability that the housing unit is in the sample. Severe problems include pipe and plumbing failures, heating failures and electrical problems. Chronic problems include holes in wall or floor, chipped or peeling paint, plaster damage, loose railings, etc. Rents and housing costs are in 1998 dollars.

Table 5 - Effects of rent decontrol - some specification checks

dependent variable	rent	cost	chronic	length	rent	cost	chronic	length
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Decontrol Effects								
Boston*1998	75.03	44.70	-0.06	-1.73				
	(25.51)	(29.41)	(0.03)	(0.67)				
Non-Boston Controlled* 1998	104.96	93.40	-0.06	-1.78				
	(31.41)	(34.66)	(0.02)	(0.53)				
Comparison of real and hypothetical experiments								
Ever controled zone*1998					97.25	70.18	-0.09	-1.46
					(27.30)	(31.28)	(0.03)	(0.77)
Ever controled zone*1993					7.89	2.48	-0.06	0.61
					(23.00)	(23.52)	(0.04)	(0.94)
Ever controled zone*1989					24.64	20.14	-0.03	0.29
					(19.89)	(21.57)	(0.03)	(0.73)
n=	3541	3445	3543	3543	3541	3445	3543	3543
number of clusters	84	84	84	84	84	84	84	84

All regressions include quadratics and cubics for number of bedrooms and other rooms, categorical age and number of building units, age of building interactions with year, and controls for type of heat and presence of central air as well as zone and year fixed effects. Columns (1) and (5) also control for whether various utilities are included in the rent. Standard errors are corrected for clustering on the zone year level. All regressions weighted by inverse probability that the housing unit is in the sample. Chronic problems include holes in wall or floor, chipped or peeling paint, plaster damage, loose railings, etc. Rents and housing costs are in 1998 dollars.

Table 6 - Own unit and Spillover Effects of Rent Control

	no spillover effects		with spillover effects	
dependent variable	rent	cost	rent	cost
	(1)	(2)	(3)	(4)
	A. Ordinary least squares			
Unit is Controlled	-172.26	-182.47	-204.6	-205.96
	-28.53	-26.86	-25.01	-26.3
Percent Controlled in Zone	--	--	-230.53	-165.94
			-94.07	-100.23
	B. Instrumental Variables			
Unit is Controlled	-457.56	-341.19	-327.37	-351.38
	-137.1	-148.89	-185.6	-204.25
Percent Controlled in Zone	--	--	-238.91	19.04
			-191.69	-213.28
	C. First-Stage			
dependent variable			controlled	% controlled
Decontrolled Zone * 1998			-0.11	-0.10
			(0.04)	(0.02)
Built before construction exemption * Controlled Zone			0.06	0.01
			(0.03)	(0.01)
Exemptable by owner occupied provision * Controlled Zone			-0.03	0.007
			(0.02)	(0.004)
Exemptable as single family dwelling * Controlled Zone			-0.21	-0.05
			(0.04)	(0.02)
F-value (4, 83 degrees of freedom)			14.17	11.45
n=	3544	3488	3544	3488
number of clusters	84	84	84	84

All regressions include quadratics and cubics for number of bedrooms and other rooms, categorical age and number of building units, age of building interactions with year, and controls for type of heat and presence of central air as well as zone and year fixed effects. Columns (1) and (3) also controls for whether various utilities are included in the rent. Standard errors are corrected for clustering on the zone year level. All regressions weighted by inverse probability that the housing unit is in the sample. Rents and housing costs are in 1998 dollars. The F tests have as null hypotheses that the coeffcients on all instruments jointly equal zero.

Table 7 - Distribution of rent controlled units by income and race

A. Income				
Income Quartile	1	2	3	4
Percent of controlled units	0.26	0.41	0.19	0.11

B. Race	white	black	hispanic
Percent of controlled units	0.81	0.08	0.04

Numbers reflect the percentage of rent controlled tenants that fit in each characteristic category.

Chapter 3 - Strategic Responses to School Accountability Measures: It's all in the Timing

3.1 Introduction

The development of assessment-based accountability programs is one of the key features of education reform in the last decade. By the year 2002, over 30 states offered punishments or rewards to schools based on their students' scores on standardized exams. The provisions of the 2001 federal No Child Left behind Act include extensive testing requirements, making it likely that accountability policies will continue to play an important role in public education.

Opponents of this assessment policy charge that it leads to a narrowing of the curriculum, or teaching to the test, that decreases overall learning and is unfair to minorities. Advocates of high stakes accountability tests argue that teaching to the test content is appropriate if tests are properly constructed to measure achievement. They claim that a yardstick for student achievement provides teachers and administrators with incentives to help students learn. Empirical studies generally find that the introduction of an accountability program raises test scores.[1]

Such analyses often overlook the possibility that low scoring schools may act strategically, responding in ways not envisioned by policy makers, in order to raise student test scores. The importance of identifying such strategic responses is highlighted by the finding of Jacob and Levitt (2003) and Jacob (2002) that 4-5% of teachers in Chicago helped their students cheat on "high stakes" examinations. They conclude that teacher abetted cheating occurs most often in poorly performing schools.

[1] e.g. Grissmer (2000), Ladd(1999), Richards and Sheu(1992).

This paper identifies a previously unexamined strategic response to accountability programs: changes in the school-year calendar. When faced with low test scores some administrators move the starting date of their school year forward, so students have more class time before state-mandated testing occurs. I find that this change in school calendars raises 4^{th} grade math scores. However, the effect is small relative to the total score gains in low-achieving schools, suggesting that most improvement in scores comes from other school actions and mean reversion. Furthermore early school-start dates may increase 3^{rd} grade reading scores for top students, but appear to decrease the scores of lower-achieving students. This decline in scores may be linked to a drop in school attendance for lower achieving students caused by the calendar change.

Thus, my results are roughly consistent with Pischke (2003) who examines a substantial variation in the length of the German school year and finds only small negative effects on future wages. On the other hand they appear to contrast with the results of Card and Krueger (1992) and Grogger (1996) who find no link between the length of the school year and student achievement.[2] It should be noted, however, that this comparison is imperfect because the above studies examine changes in the length of the school year while I consider changes in the number of school days prior to a standardized test, with overall school-year length held constant.

The remainder of the paper is organized as follows: Section 2 briefly explains the institutional background and data, Section 3 presents the analytical methods and results, and section 4 concludes.

[2] Other studies that find no mean effect include Eide and Showalter (1998) and Rizzuto and Wachtel (1980) on U.S. data and Lee and Barro (2001) on cross country data.

3.2 Background and Data

Beginning in the late 1990s, parents, farmers and tourism officials in states such as South Carolina, Texas and Wisconsin complained that school districts were moving the beginning of classroom instruction from traditional dates near Labor Day into the middle of August. Some schools with otherwise traditional calendars adopted start dates as early as August 2nd.[3] The resulting district calendars left the length of the school year unchanged. They merely shifted the timing of the school year from a September – June schedule to an August – May term. Opponents of this shift claimed that the change in the school year was costly for schools, families, and local industry[4]

Investigating the evolution of school start dates is difficult since most states do not keep long term records of calendar information. However, the Texas Education Agency collected survey data on the start dates of its 50 largest districts in 1990 and 1999. The 47 respondents serve more than half the students in Texas and illustrate a clear shift in school start dates. Figure 1a shows that in 1990, the majority of the surveyed districts started classroom instruction approximately one week, or five school days, before Labor Day. According to figure 1b, by 1999 most of the schools had moved their start dates up by at least two weeks. The modal school start date by the end of the decade was three weeks before Labor Day. Some schools began class as much as five weeks before the traditional start date.

Though Texas implemented an assessment based accountability system in the early 1990s, the two year survey (n= 94) represents all of the data the Texas Education Agency collected on school starting dates. Without the ability to follow school start dates

---

[3] Despite recent moves toward state control, most states cede authority to make many calendar decisions such as when to start school to local districts.
[4] An analysis of the costs of early school start dates is presented by Strayhorn (2000).

over consecutive years it is difficult to determine if there is a causal relationship between accountability testing and the timing of the school year in Texas. Of the states that adopted accountability programs and noted shifts in the timing of the school year, only Wisconsin kept extensive records of the change in school start dates over time. Thus the remainder of the paper will examine the link between assessment and the timing of the school year in Wisconsin.

There are other advantages to focusing on Wisconsin's experience with changing school start dates. Beginning with the 2002 school year Wisconsin school districts were prevented by law from adopting start dates prior to September 1st.[5] This law, a result of aggressive advocacy by parent and industry groups, provides a potential shift in school start dates that is uncorrelated with district characteristics or history . Since Wisconsin collected data on the starting date of each of its 426 districts from the mid 1990s to present it is possible to examine Wisconsin school start dates before and the legal mandate. Figure 2a shows that prior to the legal restriction many Wisconsin districts started about 10 days (2 weeks) before Labor Day, though there was a significant range of start dates. After the law's passage, most districts (361) began classroom instruction for 2002 the day after Labor Day as reflected in figure 2b.

In order to determine the relationship between start dates and accountability testing, I combine data on school start dates with data on district resources, demographics

[5] A few districts later received a waiver of the requirement. Most waivers were granted to schools that had externally imposed calendaring requirements such as the International Baccalaureate program (IB). However, the number of waivers was small.

and test scores from 1997-2003 provided by the Wisconsin Department of Public Instruction.[6] Where necessary, data is aggregated to the district level.

To determine the effect of having additional classroom days to prepare for accountability tests, I construct two measures of school time for each district-year observation. The first measure is the number of days before Labor Day that classroom instruction begins in a school district. This measure, shown in figures 2a and 2b, captures the extent to which schools advance their starting date. The second measure is the number of school instruction days between the beginning of the school year and the state mandated testing window. This second measure also incorporates variation induced by differences in the state testing period as well as variation due to differing school start dates. The most important change in the state testing window occurred in the 2002 school year, when the Wisconsin Department of Public Instruction moved its accountability assessments from the spring of the school year to the fall. Thus the 2001-2002 tests were given in March 2002, but the 2002-2003 tests were administered in November 2002. Figures 3a and 3b show that this timing change reduced the preparation time of most districts from over 110 days to about 45 days.

In addition to school time measures, Table 1 provides descriptive statistics for a variety of school resources. Table 1 demonstrates that Wisconsin provides a generous level of support to its schools. The average school has a pupil teacher ratio of about 12.4 (this figure counts school level administrative personnel as teachers), with average per pupil instructional spending of $8,744 (2002 dollars) and average teacher salaries of

[6] The WKCE accountability test results are available from the 1997-98 school year on. I match these with calendar data from the 1998-99 school year onward. Hereafter I will refer to each school year by its fall date. Hence 1998 will be the 1998-99 school year.

$40,645. Additionally, the minority population of the average school is only about 6.6 percent. In each case Wisconsin districts compare favorably with national averages.

I use scores from two different tests to measure student achievement. The Wisconsin Knowledge and Concepts Examination (WKCE) is Wisconsin's main accountability assessment, testing 4th, 8th and 10th graders in mathematics, reading, language arts, science, and social sciences. It was administered in the spring of each year prior to 2002, when it was moved to the fall. Unfortunately, these tests were only normed to a national sample of students for the 1996-2001 school years. For the 2002 school year, a key year for this study, the tests added a couple of items that were not tested on a norming population. Consequently the test data for this year are available only in mean scaled score format.

I also use scores on the Wisconsin Reading Comprehension Test (WRCT), which is administered to all 3rd graders in the spring of each year. Though this test does not change timing in 2002, it suffers from other potential drawbacks. Most importantly, it groups students into achievement categories rather than giving them flat numerical scores. This test is also less prominent in the Wisconsin accountability system and could possibly be seen by schools as a lower priority than the WKCE tests.

Summary statistics for all test score measures are provided in Table 1a. In general, Wisconsin students perform well relative to the national average. The average Wisconsin student scores near the 70th percentile of the national distribution in math and the 65th percentile in language arts. Almost 96 percent of the third graders achieve a passing mark on the WRCT.

3.3 Methods and Results

3.3.1. Accountability tests and school start dates.

How might the shift of school start dates into August relate to the rise of state accountability programs? If districts believe that classroom time before an assessment test is an important factor in improving scores, they may strategically adjust their start date to gain an instructional time advantage over other schools. This strategic action, however, is costly. Evidence suggests early start dates raise utility and transportation costs and run the risk of angering parents. Because Wisconsin's accountability program punishes schools with repeatedly low scores, districts with low scores have the most to gain from raising test scores through strategic actions. It is plausible that these poorer scoring schools will be most willing to bear the costs of advancing their start dates.

Since schools and districts are ranked relative to each other, districts that notice their rivals have advanced start dates may follow suit in subsequent years. Depending upon magnitude of the cost to advance start dates, this could create a "race to the top" scenario where the de facto school calendar changes as each district refuses to allow others a strategic advantage.

To investigate the connection between test scores and start dates, I examine a statistical model of the district decision about when to start school. If the school start choice is a function of district demographics, the year, and past district test scores the relationship can be estimated as:

$$E_{it} = \theta_1 Y_{it-1} + \theta_2 Y_{it-2} + X'_{it}\beta + \tau_t + \varepsilon_{it} \tag{1}$$

where i indexes districts and t time. E_{it} is the number of days before Labor Day the district starts school that year, the achievement variables Y_{it-1} and Y_{it-2} are lagged test scores from the WKCE, and X is a vector of school resource controls.[7] Because school starting dates in the 2002 school year are determined by state law rather than district choice, that year is dropped from the sample for these regressions.

If schools are adjusting their schedules in response to poor test results, OLS estimation of equation (1) should produce a negative θ_2 coefficient. Prior to 2002, test results from the WKCE were not available to schools until early June, after the start date for the following year was set. Thus, schools responding to low test scores face a year delay in adjusting start dates. This is reflected in the use of two lags in the estimation.

Table 2 presents the results of regressions of school start dates on lagged 4^{th} and 8^{th} grade math scores. Columns (1) and (2) constrain θ_1, the effect of the immediately preceding year's test scores, to be zero and measure the effect of 4^{th} and 8^{th} grade test scores separately. In both cases the coefficients indicate that low prior math scores translate into earlier future start dates. A decrease in average math scores of 5 percentile points for 4^{th} graders advances the school start date .38 days. A similar score decline for 8^{th} graders corresponds to starting school .28 days earlier.

Columns (3) and (4) drop the constraint on θ_1, allowing the prior year's (t-1) test scores to effect start dates. Column (5) includes jointly 4^{th} and 8^{th} grade achievement measures and columns (6) and (7) add variables that control for district resources. Since school start dates for the following year are set before the scores from time t-1 are known, I expect, a priori, that θ_1 will have a zero coefficient. However, in all

[7] Scores in this section of the paper are measured as the national percentile ranking of the hypothetical average student. Regressions using mean scaled test scores change the scale of the results but the significance and relative magnitudes (in terms of test score standard deviations) are similar.

specifications the estimated θ_1 coefficient is positive and significantly different from zero. On the other hand the θ_2 coefficient is negative, statistically significant, and about double the magnitude (in absolute value) of θ_1. This suggests that, controlling for school resources, a test score loss of about five percentile points is associated with school start dates moving up one-third to two fifths of a day in the second school year following the test.

The positive θ_1 coefficient seems troubling. How is it that the results of a test graded after calendaring decisions are made can be correlated with those decisions? The simplest explanation is that test scores are mean reverting. In this scenario a district with poor test scores in one year is more likely to improve test scores the following year than a district with average test scores. These districts are also the most likely to later decide to advance their school calendars, producing a positive association between early start dates and prior test scores.

Another possible explanation is that schools which seek to improve test scores by moving up start dates are also likely to implement other unobserved programs to improve test scores. These may include increased supervision of teachers or a change in time allocation to some subjects. A school with poor scores may immediately implement changes that improve scores in the following year. These will be the same schools that move up their start dates in subsequent years, producing a positive correlation.

Table 3 presents the results of similar regressions that use lagged language scores of 4^{th} and 8^{th} graders as achievement measures. Although lagged 4th grade language scores are not significant predictors of future start dates, the observed coefficients have the same signs as their math counterparts in Table 2. Additionally, 8^{th} grade language scores have a statistically significant relationship with future school start dates. Once again, twice-lagged scores have a negative effect, implying that a 5 percentile drop in test scores leads to a future advance in the school calendar of half a day. The one-year-lagged test score is again about half this magnitude and positively related to future start dates.

Tables 2 and 3 also demonstrate that most traditional measures of school resources have little relation to schools' early start dates. For example, pupil-teacher ratio, district expenditures, and percentage of minority students all have negative coefficients but none are statistically significant. The resource measures that do appear to be related to school start dates are teacher salary and experience. The estimation results imply that districts with more experienced but lower paid teachers are more likely to start the school year earlier. One possible explanation is that schools in rural areas have lower staff salaries, less teacher movement, and are more likely to adopt early start dates.

Thus, low test scores can induce districts to change their calendar, though they might make other, unobserved changes to increase test scores. This suggests that district administrators believe early start dates increase student test scores.

3.3.2. The Effect of Early School Start Dates on Test Scores

The fact that schools strategically move start dates in response to poor test scores does not indicate how effective this strategy is in improving scores. To what extent do earlier start dates improve student test scores? In this section I attempt to answer this question.

The causal relationship of interest is:

$$Y_{it} = \varphi E_{it} + \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + X'_{it}\gamma + \tau_t + \mu_{it} \qquad (2)$$

where variable definitions and indexing remain as in Equation (1).

Table 4 reports estimation results of equation (2) using 4th grade math mean scaled scores from the WKCE as achievement measures. As mean scaled scores are difficult to interpret outside the context of the test I translate the results into effect sizes (standard deviation units). In all specifications an early start date has a significant, positive effect on math scores. The effect diminishes in magnitude when lagged test scores are included as regressors yet is robust to the addition of higher order resource controls. The results suggest that starting school one calendar week earlier would raise average math scores by 0.02 – 0.06 standard deviations.

The results also indicate a significant positive correlation of 0.26 to 0.53 between past math results and present scores. Curiously the correlation is stronger for the two period lag than for the immediately preceding test results. The coefficients on most of the resource controls exhibit the expected signs, though not all are statistically significant predictors of test scores. One exception is teacher experience, which is negatively correlated with math scores. This might be due to high colinearity between the teacher experience and teacher salary variables.

These OLS estimates are unlikely to have a causal interpretation. The main obstacle to identification is the existence of unobserved factors that influence both test scores and school start dates. For example, schools with greater parental involvement in education choices are likely to have higher test scores and to avoid early start dates. Omitting this factor from the regression would bias the OLS estimates of φ towards zero. Intuitively, the problem can be thought of as the lack of a proper control group to compare with schools that are changing their starting dates.

The Wisconsin Law that mandated districts begin the 2002 school year after September 1, provides potential variation in start dates that is not driven by unobserved school or district characteristics. As seen in figures 2a and 2b, the law changed the starting dates of virtually all school districts in Wisconsin. Furthermore, different schools had to move their starting dates by differing amounts depending on where they had previously been starting. The core of my identification strategy is to compare the change in test scores of schools that had extreme changes in their start dates due to the law with schools that experienced mild changes.

To implement this I take first differences across my data. For example, the difference operator, Δ is defined for the test score variable Y as:

$$\Delta Y_{it} = Y_t - Y_{t-1} \qquad (3)$$

Taking first differences across equation (2), produces the estimating equation:

$$\Delta Y_{it} = \lambda \Delta E_{it} + \omega \Delta Y_{it-1} + \Delta X'_{it}\alpha + (\tau_t - \tau_{t-1}) + \rho_{it} \qquad (4)$$

where the variables are defined as in equation (1) and λ is the parameter of interest. Note I have eliminated the second differenced lag to preserve sample size. When estimated using data from the years bracketing the law change (2001-02 and 2002-03), this differenced regression compares schools according to the degree their start dates were moved by the law.[8] Additionally, it removes the influence of any fixed but unobservable differences among districts.

The 2002 change in WKCE accountability testing from the spring to the late fall complicates the analysis. This second change in student test preparation time is contemporaneous with the identifying law change. To account for this I assume that all days of classroom time prior to the testing window have an equal effect on test scores and estimate:

$$\Delta Y_{it} = \lambda \Delta P_{it} + \omega \Delta Y_{it-1} + \Delta X'_{it}\alpha + (\tau_t - \tau_{t-1}) + \rho_{it} \qquad (5)$$

where P_{it} measures the number of school days between the start of the year and the accountability testing window.

Table 5 presents estimates of the effect of changes in the number of school days preceding a test on test scores. The Regressions in odd-numbered columns use all years in the sample while even-numbered columns present similar regressions using only the school years bracketing the law change. Column (1) indicates that increasing class time prior to testing by a week would lead to a statistically significant 0.07 standard deviation increase in math scores. When the sample is restricted to the years bracketing the law change the estimated coefficient increases in magnitude but the standard errors increase due to loss of sample size.

[8] In other words t=2002 and t-1=2001.

However, this effect does not appear to translate to language or reading scores. Columns (4) – (6) use these other test measures as dependent variables. In all cases estimates of the effect of school time are positive but imprecise. Even if the coefficients were taken as a non-zero causal effect they would translate into a very small effect size of 0.02 – 0.03 standard deviations score increase for a one week school time increase. Table (5) also suggests that WKCE test scores might exhibit mean reversion. Prior test score gains are significant predictors of test score decreases in the current period

3.3.3. Strategic Timing Decisions and Test Scores

Taken at face value, these results indicate that there are small but significant math score gains associated with the decision to add a week of test preparation time. These gains do not extend to reading or language scores.

Are these effects truly small relative to other potential educational reforms? By way of comparison, the class size literature finds an effect size of about 0.29 – 0.33 standard deviations from an 8 student reduction in class size. My results indicate that generating a comparable math score gain through advancing school start dates would require starting school 4–5 weeks early. Thus the effect of early start dates, though significant, is small.

Table 6 shows the average score decline in the law change year of 2002. My results imply that changes in school start dates due to the law change account for .18 standard deviation or 19% of the math score decrease. The remainder is attributable to the movement in testing window and district characteristics.

3.3.4. Alternative Test Measures and Attendance

My results in the previous section depend on the assumption that days added to test preparation by manipulating the school start date have the same effect as days lost when the state shifts the testing window. They also require that tests given in the spring do not differ from tests administered in the fall. Additionally, it is possible that changing school start dates effect some portion of the conditional reading and language score distribution despite a measured mean effect of zero.

I explore these issues by examining the effect of school starting dates on a different test score measure. The 3rd grade WRCT was administered in the spring throughout my study period thus avoiding the issue of shifting testing windows. Table 7 examines the effects of early school start dates on the percentage of students who pass the WRCT. Columns (1) – (2) present OLS estimations for comparison with the first differenced estimates of columns (3) – (4). Though the OLS estimates are smaller in magnitude than the differenced estimates, the implied negative effect of starting school early on reading scores is puzzling. The differenced estimates imply that starting school five days early actually reduces the percentage of students that pass the reading test by 0.6 percent. With an average state failure rate just over four percent, this represents a very large negative effect associated with starting school early.

WRCT students are assigned a failing grade or one of four passing competency levels. Each column in Table 8 reports differenced regression results where the dependent variable is the differenced percentage of students in the district that attained a particular competency level. The results indicate that the percentage of students in each of the three lowest achievement levels fell in schools that adopted earlier start dates. Meanwhile, the percentage of students in the top category rose significantly, but by a smaller amount.

One interpretation of these results is that 1 percent of students move upward from the proficient to advanced categories for each 5 days earlier school starts, but 0.35 to 0.7 percent of students exit each of the three lower competency levels. Many of these students end up failing. This peculiar result, where increased school time has a positive effect on students at some ability levels and a negative effect on those of lower ability levels is similar to the findings of Showalter and Eide (1998) that lengthening the school year has positive effects only on those students in the top half of the ability distribution.

A change in school attendance might explain this pattern of results. If lower ability students are more likely to miss school days when the school year starts in August, then the movement of the school year forward might improve test scores among higher ability students by giving them more preparation time while simultaneously reducing the scores of lower ability students by causing them to miss school instruction.

Table 9 examines the effect of early school start dates on attendance rates. The results imply that advancing the school start date by 5 days leads to a .1 percentage point decline in average attendance over the year. Assuming the entire effect of an early school day on attendance works through student absences on the early day itself, a day of school before Labor Day will have an attendance rate 4.3 percentage points lower than a normal school day.[9] The additional students who miss instruction on early school days could account for the increased failure rates on the WRCT.

[9] This is calculated using a 180 day school year.

3.4 Conclusion

Many school districts moved their school start dates earlier into August during the 1990s. My results indicate that the adoption of statewide accountability exams explains part of this trend. Some school districts faced with low test scores acted strategically to raise their relative scores by starting the school year earlier. Early school start dates increase the math scores of 4^{th} graders, but have little effect on their average reading or language scores. The math effect is fairly small in an absolute sense and relative to other educational reforms. The 2002 law that moved school start dates back to September 1, only accounts for about one-fifth of the test score decline in that year. Early school start dates may also increase the reading scores of third graders at the top of the skill distribution and reduce the scores of those closer to the bottom. A possible explanation for this pattern is that early school start dates increase absences among lower skilled children.

These findings contrast with earlier studies that fail to find a significant role for school time in student achievement. This may due to the differing measure of school time used in this paper and the fact that raising the student achievement measures I use are seen by schools as important goals.

My results suggest that accountability tests to assess the relative performance of schools may not be on a level playing field if some schools are allowed to start preparation at an earlier time than others. They also indicate that if accountability tests are truly indicators of student human capital, administrating these tests at the end of the year might increase students' human capital and allow districts to more effectively use classroom time.

References

Angrist, Joshua D. and Victor Lavy (1999). Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement,. *Quarterly Journal of Economics* 114, 533-575.

Card, David and Alan Krueger, (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States,. *Journal of Political Economy* 100, 1-40.

Ehrenberg, R.G. and Brewer, D.J., (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review* 13, 1–17

Eide, Eric and Mark H. Showalter (1998) .The Effect of School Quality on Student Performance: A Quantile Regression Approach., *Economics Letters* 58, 345-350.

Grissmer, D.W. et. al. (2000). *Improving Student Achievement: What NAEP Test Scores Tell Us*. MR-924-EDU. Santa Monica: RAND Corporation.

Grogger, Jeff (1996) .Does School Quality Explain the Recent Black/White Wage Trend? *Journal of Labor Economics* 14, 231-253.

Hanushek, Eric., (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature* 24, 1141–1177

Hanushek, E., (1996). School resources and students performance., in Gary Burtless (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success.* Brookings Institution, Washington, D.C., 43–73

Heckman, James, Anne Lane-Farrar, and Petra Todd (1986) .Does Measured School Quality Really Matter? An Examination of the Earnings Quality Relationship,.in Gary Burtless (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success.* Washington, DC: Brookings Institution Press, 192-289.

Jacob, Brian, and Steven Levitt, (2003). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *Quarterly Journal of Economics 118*. 843-77.

Jacob, Brian, (2002). Accountability, Incentives, and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools. *NBER Working Paper # 8968.*

Ladd, H. F. (1999). The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes. *Economics of Education Review* 18: 1-16.

Lee, Jong-Wha and Robert Barro, (2001) .School Quality in a Cross-Section of Countries,. *Economica* 68, 465-488.

Pischke, Jorn-Steffen, (2003). The Impact of Length of the School Year on Student Performance and Earnings:Evidence from the German Short School Years. *NBER Working Paper #9964.*

Richards, Craig E. and Sheu, Tian Ming, (1992). The South Carolina School Incentive Reward Program: A Policy Analysis. *Economics of Education Review 11.* 71-86.

Rizzuto, Ronald and Paul Wachtel (1980) .Further Evidence on the Returns to School Quality,. *Journal of Human Resources* 15, 240-254.

Strayhorn, Carole Keeton.(2002) An Economic Analysis of the Changing School Start Date in Texas. Texas Comptroller of Public Accounts Special Report. Available at http://www.window.state.tx.us/specialrpt/ssd/

Wisconsin Department of Public Instruction, Library and Statistical Information Center. District Demographic Files , years 1997-2003.

Wisconsin Department of Public Instruction, Office of Educational Accountability. Wisconsin Knowledge and Concepts Examination Results – District Level Files, years 1997-2003.

Wisconsin Department of Public Instruction, Office of Educational Accountability. Wisconsin Reading Comprehension Test Results – District Level Files, years 1997-2003.

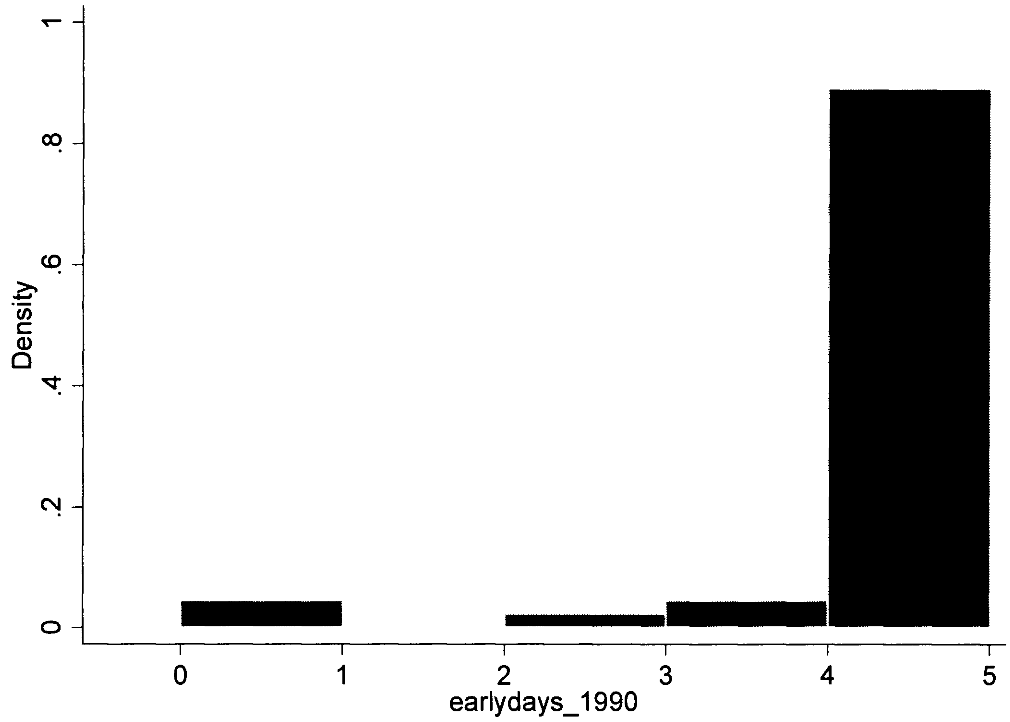


Figure 1a. – 1990 Texas School Start Dates – Days Before Labor Day

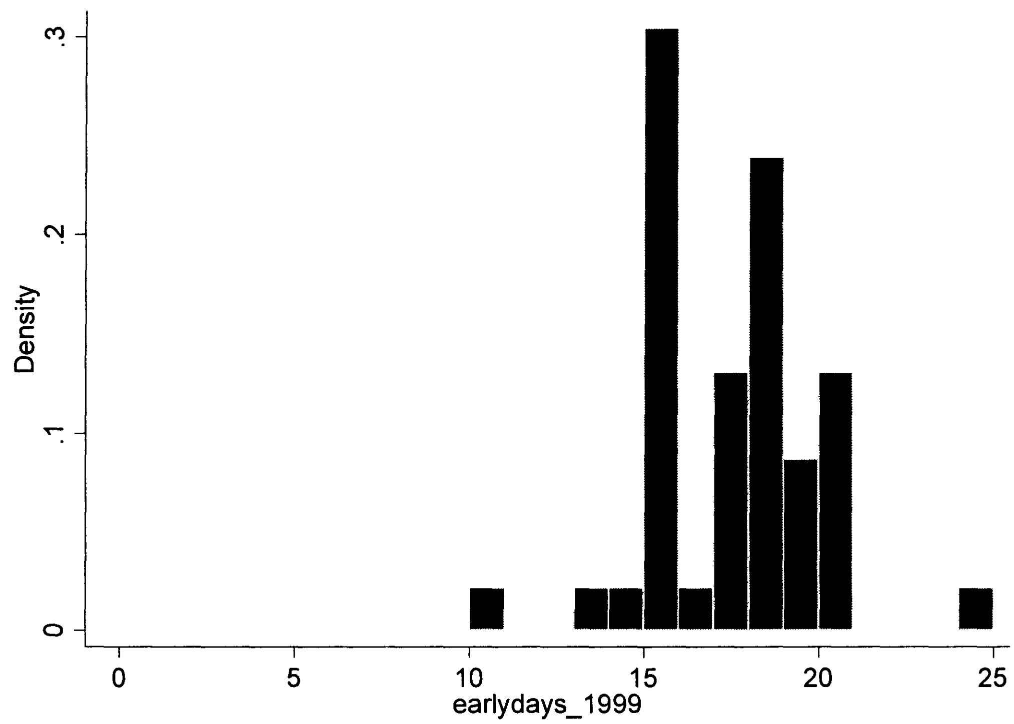


Figure 1b. – 1999 Texas School Start Dates – Days Before Labor Day
(Source: Author's calculations on data from Strayhorn (2002))

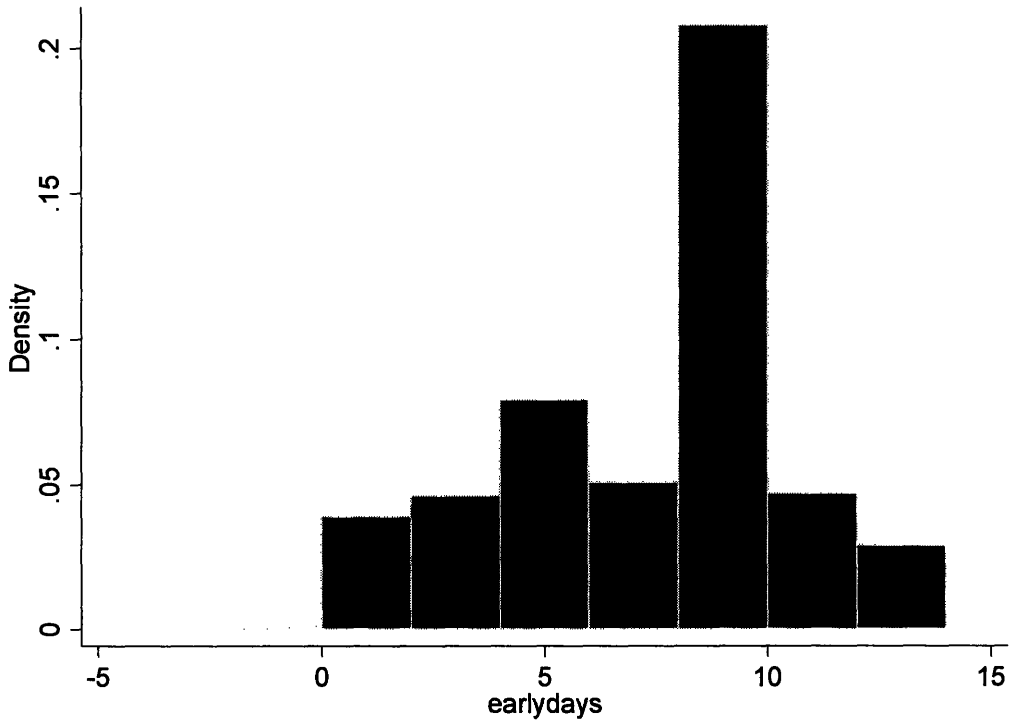


Figure 2a: School Start Dates Relative to Labor Day in Wisconsin 1998-2001.

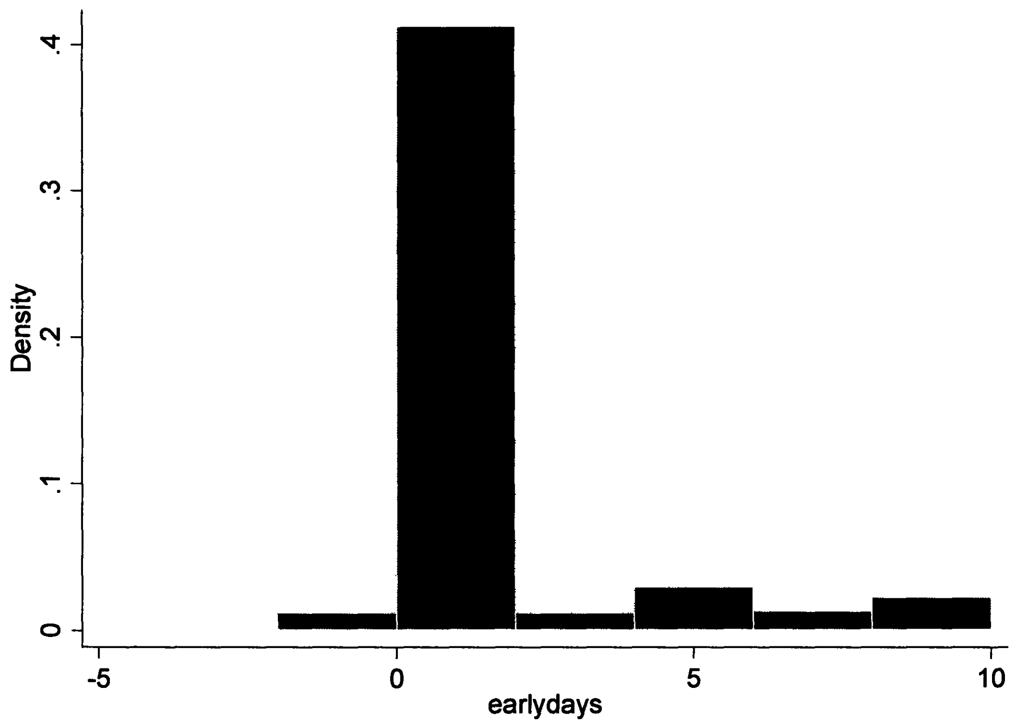


Figure 2b: School Start Dates Relative to Labor Day in Wisconsin 2002.
(Source: Author's calculations on data from Wisconsin DPI)

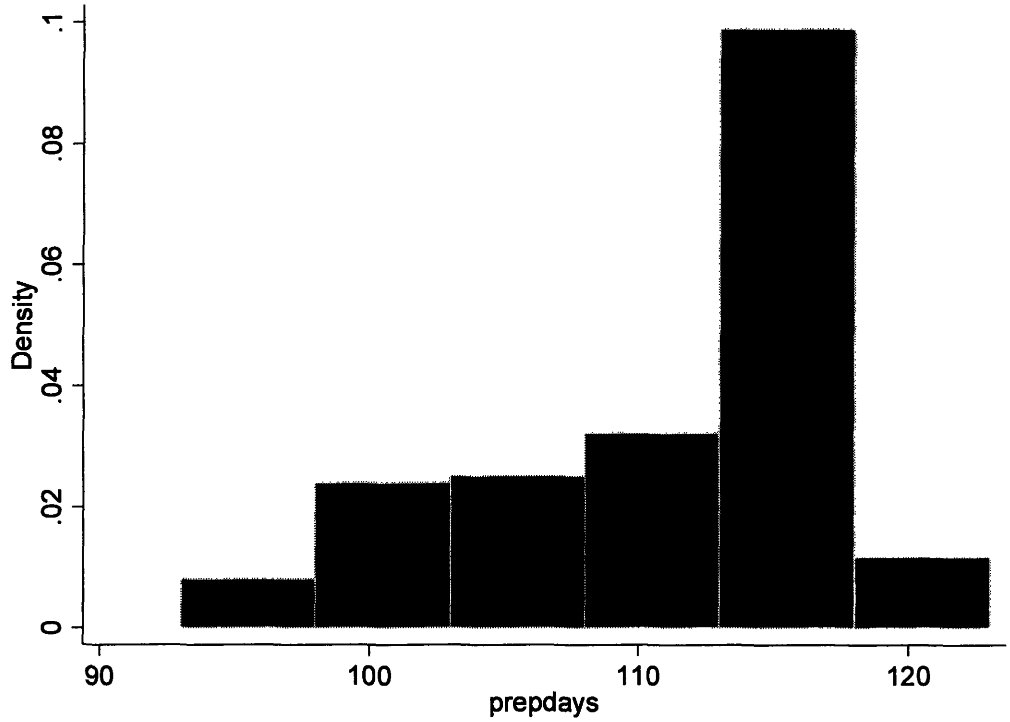


Figure 3a: School Days Preceding High Stakes Testing in Wisconsin 1998-2001.

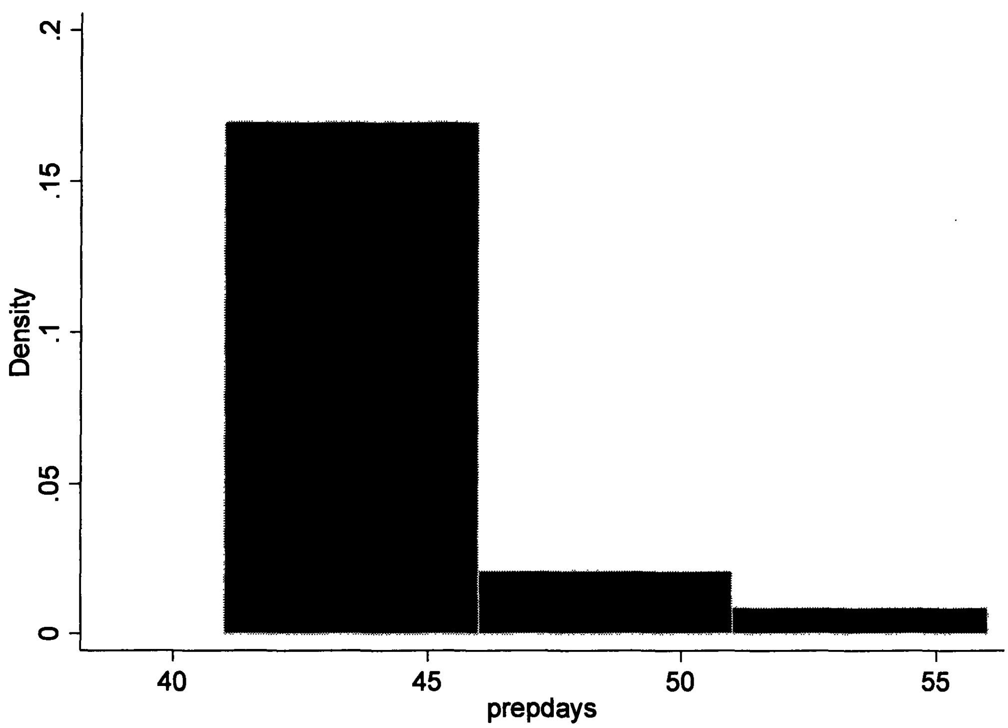


Figure 3b: School Days Preceding High Stakes Testing in Wisconsin 2002.
(Source: Author's calculations on data from Wisconsin DPI)

Table 1 - Descriptive Statistics

School Length Variables	
Early School Days	5.576 (3.803)
School Days Before Test	97.010 (27.229)
School Resource Variables	
Pupil Teacher Ratio	12.364 (1.670)
District Spending (thousands of dollars)	8.744 (1.451)
Teacher Experience	15.214 (2.231)
Teacher Salary (thousands of dollars)	40.645 (4.051)
Percent Minority Students	6.573 (10.301)
Attendance Rate	95.033 (1.445)

Standard errors are listed below means in parentheses. Each observation is a school district year.

Table 1a - Descriptive Statistics (continued)

Grade	3rd Grade	4th Grade	8th Grade
Test Score Measures			
WRCT Minimal	3.788		
	(5.236)		
WRCT Basic	12.632		
	(7.814)		
WRCT Proficient	51.330		
	(8.930)		
WRCT Advanced	28.015		
	(11.074)		
Math Percentile Rank		68.464	70.682
		(8.912)	(9.030)
Math Mean Scaled Score		641.245	
		(11.599)	
Language Percentile Rank		64.560	66.985
		(7.884)	(8.311)
Language Mean Scaled Score		652.229	
		(9.888)	
Reading Mean Scaled Score		654.527	
		(10.683)	
n=	2092	2092	1704

Standard errors are listed below means in parentheses. The WRCT is the Wisconsin Reading Comprehension Test, administered annually to all third graders. All other test scores are from the Wisconsin Knowledge and Concepts Examination. Percentile scores are from the years 1997-2001, other test scores are from 1997-2002. Percentile scores represent how a hypothetical average district student would rank nationally.

Table 2 - The Effect of Math Scores on Future District School Start Dates

	Dependant Variable: Number of School Days Before Labor Day						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Prior Test Scores							
4th Grade Math Scores (t-1)			0.042		0.037	0.037	0.036
			(0.011)		(0.011)	(0.011)	(0.011)
4th Grade Math Scores (t-2)	-0.075		-0.095		-0.079	-0.058	-0.055
	(0.009)		(0.011)		(0.012)	(0.011)	(0.011)
8th Grade Math Scores (t-1)				0.028	0.022	0.021	0.022
				(0.013)	(0.013)	(0.013)	(0.012)
8th Grade Math Scores (t-2)		-0.053		-0.070	-0.054	-0.047	-0.046
		(0.011)		(0.014)	(0.013)	(0.013)	(0.012)
Resource Controls							
Pupil Teacher Ratio						-0.036	
						(0.031)	
District Spending Per Pupil (in thousands)						-0.117	
						(0.077)	
Average Teacher Experience						0.195	
						(0.046)	
Average Teacher Salary (in thousands)						-0.221	
						(0.027)	
Percent Minority Students						-0.024	
						(0.944)	
Higher Order Controls	no	no	no	no	no	no	yes

Heteroskedasticity robust standard errors in parentheses. Higher order controls include cubics of school resources and time effects. All higher order controls except for percentage minorities fail statistical tests for inclusion in the model. Test scores are measured as the National Percentile Scaled Ranking of the hypothetical district average student. N=1245

Table 3 - The Effect of Language Scores on Future District School Start Dates

	Dependant Variable: Number of School Days Before Labor Day						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Prior Test Scores							
4th Grade Language Scores (t-1)			-0.023		-0.011	0.001	0.004
			(0.014)		(0.013)	(0.013)	(0.013)
4th Grade Language Scores (t-2)	-0.025		-0.011		-0.010	-0.008	-0.006
	(0.012)		(0.014)		(0.014)	(0.014)	(0.013)
8th Grade Language Scores (t-1)				0.063	0.062	0.049	0.048
				(0.012)	(0.013)	(0.013)	(0.013)
8th Grade Language Scores (t-2)		-0.053		-0.109	-0.112	-0.096	-0.091
		(0.011)		(0.012)	(0.012)	(0.012)	(0.012)
Resource Controls							
Pupil Teacher Ratio						-0.036	
						(0.031)	
District Spending Per Pupil (in thousands)						-0.139	
						(0.074)	
Average Teacher Experience						0.175	
						(0.045)	
Average Teacher Salary (in thousands)						-0.211	
						(0.027)	
Percent Minority Students						-0.028	
						(0.921)	
Higher Order Controls	no	no	no	no	no	no	yes

Heteroskedasticity robust standard errors in parentheses. Higher order controls include cubics of school resources and time effects. All higher order controls except for percentage minorities fail statistical tests for inclusion in the model. Test scores are measured as the National Percentile Scaled Ranking of the hypothetical district average student. N=1245

Table 4 - The Effect of School Time on 4th Grade Math Achievement

	(1)	(2)	(3)	(4)
School Time				
School Days before Test	0.147	0.113	0.039	0.041
	(0.009)	(0.008)	(0.008)	(0.008)
Effect Size (1 week increase)	0.066	0.051	0.018	0.019
Lagged Test Scores				
Math Score (t-1)		0.529	0.259	0.258
		(0.021)	(0.025)	(0.025)
Math Score (t-2)			0.387	0.388
			(0.027)	(0.028)
Resource Controls				
Pupil Teacher Ratio	-0.219	-0.227	-0.408	
	(0.140)	(0.119)	(0.119)	
District Spending (in thousands)	0.380	0.332	0.521	
	(0.203)	(0.174)	(0.182)	
Teacher Experience	-0.461	-0.229	-0.367	
	(0.126)	(0.108)	(0.112)	
Teacher Salary (in thousands)	0.572	0.261	0.265	
	(0.072)	(0.063)	(0.066)	
Minority Students	-19.679	-8.662	-8.043	
	(2.418)	(2.111)	(2.175)	
higher order controls	no	no	no	yes
n=	1643	1642	1227	1227

Robust standard errors in parentheses. Higher order controls include cubics for school resources and year effects. Test scores measures are mean scaled math scores. Calculated effect sizes are relative to between observation test score standard deviation.

Table 5 - The Effect of Class Time on 4th Grade Achievement – First Differenced Estimates

Dependent Variable	math scores		language scores		reading scores	
	all years	law change	all years	law change	all years	law change
	(1)	(2)	(3)	(4)	(5)	(6)
Δ School Days before Test	0.170	0.209	0.013	0.032	0.029	0.069
	(0.073)	(0.114)	(0.068)	(0.096)	(0.071)	(0.104)
Effect Size (1 week increase)	0.073	0.097	0.007	0.016	0.014	0.033
Test Score gain (t-1)	-0.399	-0.383	-0.452	-0.413	-0.444	-0.418
	(0.036)	(0.063)	(0.033)	(-0.413)	(0.036)	(0.066)
n=	1227	398	1227	398	1227	398

Higher order controls include cubics of school resources and time effects. All variables except time effects are in differenced form. Test scores measures mean scaled scores on each test. The Law change sample in includes only the 2001 and 2002 school years. Calculated effect sizes are relative to between observation test score standard deviation. Test score gain(t-1) represents the difference between student scores in period (t-1) and period (t-2).

Table 6 - Test Score Changes 2001 - 2002

	2001 Scores	2002 Scores	Difference	σ
Reading	656.84	647.45	-9.39	-0.91
Language	655.41	647.34	-8.07	-0.77
Math	641.15	633.16	-7.99	-0.75

Test scores are mean district scores scaled to be comparable across districts and years. The σ column gives the score change in between district standard deviation units.

Table 7 - The Effect of Early School Start on 3rd Grade Writing Achievement

Dependent Variable: Percent of Students Passing				
	(1)	(2)	(3)	(4)
School Time Measure				
School Days before Labor Day	-0.035	-0.052	-0.123	-0.126
	(0.039)	(0.039)	(0.042)	(0.042)
Resource Controls				
Pupil Teacher Ratio	0.025		0.127	
	(0.051)		(0.073)	
District Spending Per Pupil (in thousands)	0.271		-0.046	
	(0.088)		(0.121)	
Average Teacher Experience	-0.041		0.038	
	(0.052)		(0.078)	
Average Teacher Salary (in thousands)	-0.041		-0.013	
	(0.028)		(0.046)	
Percent Minority Students	-0.081		-0.113	
	(.012)		(.122)	
Estimation Method	ols	ols	differenced	differenced
higher order controls	no	yes	no	yes
year/time controls	yes	yes	yes	yes
n=	1658	1658	1242	1242

Heteroskedasticity robust standard errors in parentheses. Higher order controls include time effects and cubic resource controls. Columns (3) – (4) have all variables first differenced.

Table 8 – First Differenced Estimates of Early School Start Effects on Categorical Writing Achievement

Competency level	minimal	basic	proficient	advanced
	(1)	(2)	(3)	(4)
School Time Measure				
Δ Early School Days	-0.033	-0.018	-0.286	0.215
	(0.038)	(0.072)	(0.099)	(0.100)
Resource Controls				
Δ Pupil Teacher Ratio	0.033	0.199	-0.266	0.162
	(0.037)	(0.072)	(0.130)	(0.187)
Δ District Spending (in thousands)	-0.124	-0.308	-0.053	0.440
	(0.094)	(0.195)	(0.285)	(0.274)
Δ Teacher Experience	0.053	-0.195	-0.133	0.313
	(0.079)	(0.145)	(0.229)	(0.261)
Δ Teacher Salary (in thousands)	-0.044	-0.036	0.378	-0.311
	(0.035)	(0.063)	(0.115)	(0.136)
Δ Minority Students	-0.072	-0.035	0.527	-0.533
	(0.082)	(0.159)	(0.241)	(0.251)

Heteroskedasticity robust standard errors in parentheses. Not reported are linear controls for resources and time effects. All variables differenced except period effects. Analogous regressions with higher order controls yield similar estimates. n=1242.

Table 9 – Differenced Estimates of the Effect of Early School Start on Attendance Rates

sample	all years	law change
	(1)	(2)
School Time Measure		
Δ Early School Days	-0.027	-0.024
	(0.011)	(0.012)
Resource Controls	-0.069	-0.051
Δ Pupil Teacher Ratio	(0.029)	(0.046)
Δ District Spending	0.049	0.085
(in thousands)	(0.032)	(0.050)
Δ Teacher Experience	0.005	-0.022
	(0.031)	(0.019)
Δ Teacher Salary	0.004	0.005
(in thousands)	(0.011)	(0.013)
Δ Minority Students	-0.187	-1.619
	(2.893)	(1.837)
n=	1278	426

Heteroskedasticity robust standard errors in parentheses. Time effects included but not reported. All variables are differenced except period effects. Regressions weighted by district enrollment average over 5 year study period.