

EXPLORATIONS OF TEACHER LABOR MARKETS

By

Seth L. Gershenson

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Economics

2011

ABSTRACT

EXPLORATIONS OF TEACHER LABOR MARKETS

By

Seth L. Gershenson

My dissertation is comprised of three chapters that analyze various aspects of teacher labor supply. The first two chapters use the same primary dataset and similar empirical strategies to investigate substitute-teacher labor supply in an intermediate school district in Michigan. The data comes from an automated calling system used to offer jobs to available substitute teachers and is notable in two respects. First, both accepted and rejected offers are observed, which facilitates the estimation a sequential binary-choice model of substitute teachers' job-offer acceptance decisions. Second, the calling system makes offers in a conditionally random order, which generates exogenous variation in offer quality across substitute teachers. This exogenous variation is exploited to identify the causal effects of a variety of job attributes on substitute teachers' labor-supply decisions.

Substitute teachers are an important, but often overlooked, source of instruction in U.S. public schools. Chapter 1 investigates substitute teachers' preferences for several non-wage job characteristics and their potential implications for education policy. I find that important determinants of the offer-acceptance decision include the offer's arrival time, commute time, day of week, classroom type, school type, and school quality. Interestingly, conditional on school quality student demographics do not significantly influence substitutes' decisions. Longer and higher paying full-day jobs are preferred to half-day jobs, although conditional on daily pay, job length does not significantly impact daily labor-supply decisions. Preferences for several job

characteristics are found to vary with substitutes' regular-teacher certification status. Policy implications of these findings are discussed.

Chapter 2 estimates the causal effect of commute time on daily labor supply. The substitute-teacher labor market is an ideal environment in which to answer this question because workers are subject to daily exogenous variation in commute time and are free to adjust labor supply on a daily basis. The main result is an estimated offer-acceptance elasticity (with respect to commute time) of about -0.4, which suggests that commute time plays an important role in labor supply decisions. The effect of commute time on labor supply is significantly larger on mornings when the temperature is below 20 degrees Fahrenheit, but fuel prices and rain do not significantly alter the effect of commuting. There is no statistically significant difference in the overall aversion to commuting between men and women, however women are particularly averse to commuting the cold weather and are significantly more responsive to fuel prices than men.

Chapter 3 investigates the impact of an increase in the stakes of mandatory testing created by the 2001 *No Child Left Behind Act* (NCLB) on teacher quality in California. NCLB simultaneously created strong incentives for schools to improve student achievement and increased the stress and pressure on teachers. The empirics use a difference-in-differences identification strategy that compares teachers in tested second-grade classrooms to those in non-tested first-grade classrooms. I find that the probability of second-grade teachers holding a graduate degree significantly decreased in response to (and in anticipation of) NCLB, a small decrease in average years of experience in tested versus non-tested classrooms, and no effect on teacher certification.

Copyright by
SETH L. GERSHENSON
2011

Dedicated to the memory of my mother, Shona Zangari Gershenson

ACKNOWLEDGEMENTS

I am incredibly lucky to have had an advisor, Professor Steven Haider, who provided a tremendous amount of support, encouragement, and constructive criticism over the past several years. I will be forever grateful for the countless hours that he spent reading some *very* preliminary drafts of my work, discussing potential improvements, and teaching me how to read, write, and present economic research. Professor Haider was a fantastic mentor in every sense of the word. Professor Gary Solon also deserves a huge amount of thanks for the many important contributions he made to my dissertation and the constant friendly encouragement that he provided; in many ways Professor Solon acted as a second advisor. I also thank the other members of my dissertation committee, Professors Cassie Guarino and Stephen Woodbury, for their encouragement and support throughout.

I have thoroughly enjoyed my time at MSU. During the course of my graduate studies I have benefitted from conversations with fellow students, faculty members, and seminar participants too numerous to mention. I received excellent training at MSU and sincerely appreciate the time and effort that Professors John Giles, Steven Haider, Gary Solon, and Jeff Wooldridge put into their graduate courses. I am also lucky to have had some terrific classmates; Brian McNamara, Brian Moore, and Nick Sly have become good friends that helped me to cope with the ups and downs of graduate school and to prepare for my comprehensive exams. I am also thankful for the financial support and travel grants that I have received from MSU's Department of Economics, Graduate School, College of Social Science, and Council of Graduate Students. I would also like to acknowledge financial support from the American Education Finance Association's Pre-Doctoral New Scholar Award, which I received in

recognition of the first two chapters of my dissertation. I thank Cassie Guarino for encouraging me to apply for the award.

Finally, I am blessed to have a wonderfully supportive father. His blind faith that I would succeed helped me to persevere in the first year of graduate school and his occasional “loans” made things much less stressful. Thanks Dad, I cannot begin to tell you how much I appreciate everything that you have done for me over the years.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER 1: How do Substitute Teachers Substitute?	
1.1 Introduction.....	2
1.2 Background and Literature	3
1.3 Institutional Details and Data.....	5
1.3.1 Data	7
1.3.2 Descriptive Statistics.....	9
1.4 Econometric Model.....	14
1.4.1 Substitutes' Optimal Decision Rule.....	14
1.4.2 Estimation	17
1.5 Results.....	20
1.6 Conclusions.....	28
Appendix 1.1 Tables	32
Appendix 1.2 Figures.....	43
Appendix 1.3 Average Partial Effect (APE) Definitions.....	47
Appendix 1.4 MP Probit Coefficients.....	49
References	54
CHAPTER 2: Going the Extra Mile	
2.1 Introduction.....	58
2.2 Literature Review.....	60
2.3 Labor Market Environment & Data	63
2.3.1 The Intermediate School District	63
2.3.2 Data	64
2.3.3 Descriptive Statistics.....	65
2.4 Econometric Model.....	67
2.4.1 Optimal Decision Rule.....	67
2.4.2 Estimation	70
2.5 Results.....	72

2.5.1 Main Results	72
2.5.2 Sensitivity Analysis	76
2.6 Conclusions.....	79
Appendix 2.1 Tables	83
Appendix 2.2 Figures.....	88
Appendix 2.3 Average Partial Effects, Elasticities, & Interaction Effects	92
Appendix 2.4 Baseline RE Probit Coefficients.....	95
References.....	99

CHAPTER 3: The Effect of High-Stakes Testing on Teacher Quality - Evidence from California

3.1 Introduction.....	103
3.2 Literature Review.....	107
3.3 Institutional Details & Data	109
3.3.1 Pre-NCLB Education Policy in California	109
3.3.2 NCLB's Impact in California.....	111
3.3.3 Data	112
3.4 Empirical Model and Estimation	114
3.5 Results.....	116
3.5.1 DD Estimates	116
3.5.2 Event History Estimates.....	117
3.5.3 Sensitivity Analysis	119
3.6 Conclusion & Discussion.....	120
Appendix 3.1 Tables	124
Appendix 3.2 Figures.....	129
Appendix 3.3 FE Logit Coefficients.....	131
References.....	133

LIST OF TABLES

Table 1.1: Mean Job Characteristics	33
Table 1.2: Mean Offer Characteristics	34
Table 1.3: Daily Offers Received and Daily Selectivity of Substitutes.....	35
Table 1.4: RE-Probit Coefficients.....	36
Table 1.5: Average Partial Effects	39
Table 1.6: Mass-point Probit APE	42
Table A1: Mass-point Probit Coefficients	50
Table 2.1: Mean Offer Characteristics	84
Table 2.2: RE-Probit Results	85
Table 2.3: Linear Probability Model (LPM) Estimates	87
Table A2: Baseline RE-Probit Coefficients	96
Table 3.1: PAIF Data Description	127
Table 3.2: Standard DD Estimates.....	128
Table 3.3: Event History Estimates (time-varying NCLB effects).....	129
Table 3.4: Sensitivity Analysis of Graduate-degree Results	130
Table A3: FE Logit Coefficients.....	134

LIST OF FIGURES

Figure 1.1: Job-Length Distribution	44
Figure 1.2a: Day Of-Offer Time Distribution.....	45
Figure 1.2b: Day Before-Offer Time Distribution.....	45
Figure 1.3: Offer Time-Acceptance Probability Gradient	46
Figure 2.1: Commute-Time Distributions.....	89
Figure 2.2: Daily Weather Conditions	90
Figure 2.3: County-Level Average Daily Fuel Prices	91
Figure 3.1: Grade-Specific Trends in Average Teacher Characteristics	132

CHAPTER 1

HOW DO SUBSTITUTE TEACHERS SUBSTITUTE?

1.1 Introduction

The quality of public education in the U.S. is important due to its relationship with economic growth (Hanushek & Woessmann, 2008) and individual labor market outcomes (Card & Krueger, 1992). Instruction is a primary input of the education production function and an extensive literature studies the principal purveyors of instruction: regular teachers (Dolton, 2006; Hanushek & Rivkin, 2006). Regular-teacher absence rates are between five and ten percent and teacher absences are typically covered by substitute teachers (Roza, 2007). Little is known about this secondary source of instruction, however, and the present paper begins to fill this gap in the education literature by analyzing daily substitute-teacher labor supply.

Understanding the preferences of substitute teachers, particularly those certified as regular teachers, is potentially important for several reasons. First, many schools have trouble satisfying their demand for substitute teachers (Henderson et al., 2002; Rogers, 2001; Dorward et al., 2000). When a substitute teacher cannot be found, regular teachers and school administrators work overtime to cover their colleague's absence (Rogers, 2001). This increased workload likely decreases the covering teachers' effectiveness throughout the day. Second, recent work documenting the negative effect of teacher absences on student achievement suggests that absences covered by certified substitutes are sometimes less harmful than absences covered by non-certified substitutes (Clotfelter et al., 2009), suggesting that substitute-teacher quality may influence student achievement. Third, poor and low-achieving schools have higher regular-teacher absence rates (Clotfelter et al., 2009; Miller et al., 2008a, 2008b) and are more likely to lose their regular teachers to wealthier and higher-achieving schools (Hanushek et al., 2004). If substitute teachers similarly avoid low-achieving schools, the problems associated with the availability and quality of substitute teachers discussed above are concentrated among the

schools and students that can least afford them. Finally, understanding the preferences of substitute teachers might allow the design of a pay system that minimizes expenditures on substitutes or that increases efficiency or equity by altering the distribution of substitutes or substitute quality across schools.

I estimate a sequential binary-choice model based on an expected utility-maximizing optimal decision rule that is hypothesized to govern substitutes' job-offer acceptance decisions. The empirics utilize data on the job offers, both accepted and rejected, made by an automated calling system to substitute teachers. The offers are made in a conditionally random order that creates exogenous variation in offer quality across substitute teachers.

Several non-wage offer characteristics are found to play an important role in substitutes' daily labor supply decisions, including commute time, school type, school quality, and time of offer. Friday jobs are significantly less likely to be accepted and certified substitutes are more likely to accept offers than non-certified substitutes. Interestingly, conditional on achievement, a school's demographic composition does not influence substitutes' daily decisions, nor does job length conditional on daily pay. Substitutes do, however, systematically prefer longer and higher paying full-day jobs to half-day jobs.

1.2 Background and Literature

Substitute teachers have recently received attention from both policy makers and the popular media. For example, in 2007 H.R. 3345 (*The Substitute Teacher Improvement Act*) was introduced in Congress and in 2010 a *New York Times* editorial lamented the difficulties of substitute teaching (Bucior, 2010). Despite the apparent interest in substitute teachers, however, they have been neglected by economists and education-policy researchers. A possible

explanation for the lack of rigorous research on substitute-teacher labor supply is the dearth of data on substitute teachers in large, nationally representative, data sets like the National Center for Education Statistics' Schools and Staffing Survey.

Existing studies of the substitute-teacher labor market come mainly from outside of economics. The contingent-labor literature, for instance, contains two case studies of substitute teaching. Rogers (2001) found that substitutes in a Pennsylvania school district felt underpaid and underemployed. A sociological study found that both substitutes and regular teachers preferred arranging jobs personally to using an automated call system (Coverdill & Oulevey, 2007).

Strauss (2003) was primarily interested in the demand for substitute teachers in the Pittsburgh area, but did ask some qualitative questions of Pittsburgh-area substitutes. Over 40% cited daily pay as the most important job characteristic. Overall, 98.4% of surveyed substitutes said that daily pay was either "very important" or "somewhat important." Other commonly mentioned important job characteristics were "advance professional career," "discipline in school," "safety of school," and "proximity to residence." Dorward et al. (2000) surveyed a random sample of 500 U.S. school districts on "issues related to substitute teaching." The authors report that 86% of school districts claimed to have a "problem" or "serious problem" with substitute availability and that 7% of districts deemed their substitutes "below average." The average daily pay in their sample was \$65 per 6 hour day and ranged from \$35 to \$180.

What, if any, findings from the regular-teacher literature might apply to substitute teachers? Substitute teachers operate on a daily margin, and regular teachers choose daily labor supply by being absent. Roza (2007) finds that regular teachers are absent about ten times per school year, accounting for about 5% of school days, while comparable professionals take only

three sick days during an equivalent time period. While this difference may result from teachers being sick more often as a result of their close contact with children, a significant number of teacher absences appear to be discretionary: Ehrenberg et al. (1991) found that annual teacher absences are responsive to district-level policies and Jacobson (1988) found that a small cash bonus for perfect attendance caused a significant drop in absence rates and a large increase in perfect attendance.

In reviewing the literature on teacher quality, Hanushek and Rivkin (2006) generally find that certification standards and advanced degrees have little to no effect on student achievement. Absence rates, however, have been shown to negatively impact student achievement in a variety of settings: Clotfelter et al. (2009) in North Carolina, Miller et al. (2008a, b) in a large urban U.S. school district, and Das et al. (2007) in Zambia. Miller et al. (2008a) suggest that the negative effect of teacher absences may partially result from the low quality of substitutes. Substitute teachers are subject to significantly less-stringent requirements than regular teachers (Henderson et al., 2002). Clotfelter et al. (2009) provide evidence that substitute quality matters: absences in primary-school reading classes covered by certified substitutes are marginally less harmful than absences covered by non-certified substitutes.

1.3 Institutional Details and Data

This paper analyzes the daily labor supply of substitute teachers in a consortium of ten adjacent and autonomous Michigan school districts that contains more than 70 schools. The consortium's members enjoy economies of scale in a variety of administrative duties. For example, districts share the fixed costs of recruiting, training, and maintaining a large pool of substitute teachers and of running an automated calling system used to offer jobs to substitutes. The requirements

to substitute teach in the consortium include passing a criminal background check, at least three years of credits from an accredited college or university, completion of a four-hour orientation program, and either a valid Michigan teaching certificate or a Michigan substitute-teaching license. The latter costs \$25 and must be renewed annually.

Regular teachers in the consortium requested about 20,000 substitutes during the 2006-07 school year. About half of these requests were fulfilled via personal arrangements between regular teachers and substitutes. All remaining jobs were filled by the automated calling system. When using the automated calling system a regular teacher may request a specific substitute by name; this accounts for less than 10% of call-system requests. The subsequent analysis is restricted to the approximately 9,000 requests (jobs) that were filled by the calling system but did not specify a substitute teacher by name.

At any time prior to the start of a job, regular teachers can request a substitute by phone. The request must specify the job's characteristics, including start and end time, subject or grade level, location, and (optionally) a voicemail containing special instructions. The calling system then repeatedly offers the job to available substitutes until it is either accepted or the job begins. Offers (phone calls) are made between 4:00 p.m. and 11:00 p.m. one or more days in advance of the job and beginning at 5:00 a.m. on the morning of the job.

The automated calling system makes offers in a random order, conditional on two observed characteristics: substitutes' regular-teacher certification status (in Michigan) and substitutes' offer-specific "preferred-list status." From the call system's perspective certification is a binary variable; it does not take the job's subject or the substitute's area of certification into account. Each regular teacher, school, and district maintains a fluid list of "preferred" substitute teachers. All substitutes are included on the district list, which is a substitute's default status.

Because “list status” is offer specific, and some substitutes might be on more lists than others, I create a substitute-specific variable equal to the percentage of “preferred-list” offers.¹

Substitutes are not penalized by the system for rejecting offers and continue to receive offers after rejecting one. Nor are substitutes penalized for reneging on an acceptance in advance of the job’s start time, in which case the job simply reenters the calling system’s queue. After accepting a job, however, substitutes cease receiving offers that conflict with the accepted job. Returning to previously rejected offers is also prohibited. The model developed in section 4 accurately portrays the functioning of the automated calling system.

Upon answering a phone call from the automated system, a substitute learns the job’s start and end time, regular teacher’s name, subject, and school. The wage is not explicitly stated because it is a function of job length. Daily pay for all substitutes at all consortium schools is binary; half days pay \$40 and full days pay \$75. Full-day jobs are those longer than four hours and twenty minutes. Variation in job length within half and full days is created by differences in school schedules, class schedules, and regular teachers’ discretionary choices.

1.3.1 Data

Job offers (phone calls) are the primary unit of observation. About 5% of the roughly 100,000 offers made by the automated calling system during the 2006-07 school year are dropped from the analysis because they concerned “alternative schools” for which school-level information is unavailable. The time and date of each offer, along with characteristics of the job being offered and the offer recipient, are observed. Job characteristics include start and end time, subject, school, and a unique job identifier. Recipient characteristics include substitutes’ certification

¹ While the lists are updated throughout the school year, status changes are relatively rare and tend to happen early in the school year.

status, preferred-list status, gender, home zip code, and a unique substitute identifier. Measures of commute time were constructed for each substitute zip code-school address pair using MapQuest.com.²

The calling-system data is augmented with school-level data from two additional sources. First, total enrollment, student-teacher ratio, and the percentage of black, Hispanic, and lunch program-eligible students at each school was taken from the 2006-07 Common Core of Data.³ Second, the school grades assigned by the Michigan Department of Education (MDE) in 2006-07 were taken as a measure of school achievement (quality). The MDE annually publishes a School Report Card that assigns a letter grade (A, B, C, D, or F) to each school in the state.

Published grades are the average of three distinct grades for achievement status (test scores in levels), achievement change (first-differenced test scores), and implementation of “best practices” (self-reported usage of 40 specified instructional methods). The first two grades are based on Michigan Education Assessment Program (MEAP) standardized-test scores and are adjusted to account for variation across schools in average student socioeconomic status and to emphasize scores at the low end of the distribution.⁴ Finally, overall grades are subject to two potential modifications based on the school’s Adequate Yearly Progress (AYP) status: schools

² MapQuest uses geocoding technology to assign approximate latitude-longitude coordinates to each school’s address and each substitute’s zip-code centroid. An algorithm that favors higher posted speed limits and fewer turns and intersections searches for an optimal route. Approximate driving distance and travel time are then estimated using posted speed limits, average stop-time at each intersection, and average time it takes to make each left turn along the route. For additional details and references see Layton (2005). A trivial number of substitutes were assigned a non-Michigan zip code; these substitutes were dropped from the analysis.

³ Lunch-programs provide low-income students with free or reduced-price lunches. Eligibility for such programs is a commonly-used indicator of student poverty. The Common Core of Data is publicly provided by the National Center for Education Statistics: <http://nces.ed.gov/ccd/>.

⁴ For additional MEAP information see http://www.michigan.gov/mde/0,1607,7-140-22709_31168---,00.html.

that make AYP and earn a D will have their grade improved to a C while schools that fail to make AYP and earn an A will have their grade lowered to a B.⁵ The formulas used to compute school grades are provided in MDE (2007).⁶

1.3.2 Descriptive Statistics

Table 1.1 describes the jobs offered by the automated calling system. Column 1 reports the average characteristics of all 8,566 unique jobs, 98% of which were ultimately accepted. The average job was offered about eleven times before being accepted and the initial offer was made on the day of the job about one third of the time. Slightly more than one third of jobs were half-days, but given the way that daily pay is determined the overall average job length and hourly wage are not particularly interesting. Jobs were roughly evenly distributed between elementary, middle, and high schools. Only 4% of jobs were in charter schools. The majority of jobs were in well-performing schools; only 1% of jobs were in D schools and less than 10% of jobs were in C schools. The average job was in a school in which 19% of students were eligible for lunch assistance, 10% were black, and 4% were Hispanic. Finally, jobs were roughly evenly distributed across days of the week, with a slightly higher percentage of jobs occurring on Fridays. Not reported in table 1.1 is that jobs were evenly distributed across months, with the exception of fewer jobs in September and June.

The mean characteristics of the 3,126 never-accepted jobs are given in column 2. Two notable differences between columns 1 and 2 emerge. First, nearly three quarters of the never-

⁵ AYP is binary. It is computed using the percentage of “non-proficient” students and attendance rates (or graduation rates in high schools).

⁶ The most recent School Report Cards and accompanying documentation are publicly available at <https://oeaa.state.mi.us/ayp/>. Past School Report Cards and documentation are available from the MDE upon request.

accepted jobs were first offered on the day of the job. Second, the never-accepted jobs overwhelmingly fall on Fridays.

Columns 3 and 4 of table 1.1 investigate the differences between full and half-day jobs. Despite being significantly shorter on average, half days pay about \$1.40 more per hour than full-day jobs. Half-days are less likely to be in rural districts and more likely to be in elementary schools. Full-day jobs are more likely to be in high schools. Half-day jobs are somewhat more likely to be in A schools while full-day jobs are more likely to be in B, C, and D schools. The remaining job characteristics do not systematically vary with half-day status.

Figure 1.1 shows that the job-length distribution is a bimodal mixture of two distributions centered on the half-day and full-day means. The full-day distribution is tightly centered around the full-day mean, while the half-day distribution exhibits more variation. One potential explanation for this is that full-day absences result from regular teachers being unavailable for the entire school day, rather than being unavailable for a period of time greater than four hours and 20 minutes but less than the length of the school day, while half-day absences result from commitments shorter than four hours and 20 minutes. Note, however, the non-zero mass to the immediate right of the full-half cutoff. These potentially interesting jobs pay a high hourly wage and are investigated in column 5 of table 1.1.

There are 139 “short” full-day jobs that are less than five hours but pay the full-day wage. At first blush it is surprising that so many of these high-hourly wage jobs were never accepted, but this phenomenon is at least partly explained by the high percentage of “short” full-day jobs that were initially offered on the day of the job. For example, a regular teacher may have gone to school, unexpectedly needed to leave, and called a substitute for the remainder of the day. Thus

the job was in the four to five hour range because the teacher had been in school for an hour or two before requesting a substitute but the calling system had insufficient time to find a substitute.

The “short” full-day jobs paid over \$16 per hour, several dollars more than average, and were less likely to be in rural districts, elementary schools, and high schools. A potential concern is that teachers in unobservably bad classrooms took it upon themselves to attract substitute teachers by purposely choosing a job length to the immediate right of the half-day cutoff with the intention of creating a high hourly wage. Column 5 provides two pieces of evidence against this hypothesis, however. First, these jobs are no less likely to be in A-graded schools. Second, they are actually less likely to be on Fridays, which are the days on which it is most difficult to find a substitute. Thus it appears that teachers are not paying compensating wage differentials based on observable measures of quality. It is reasonable to assume, therefore, that regular teachers are not behaving this way based on unobservable job characteristics either.

Table 1.2 reports mean offer characteristics. Focusing on offers rather than jobs introduces two new dimensions to the data: the offer’s recipient and timing. Column 1 summarizes the 94,106 offers made by the automated system. The average offer was made 1.6 days prior to the job’s start, half of offers were made on the day of the job, and 36% were made on the day before the job. The average one-way commute was about 20 minutes (15 miles).

Regarding offer-recipient characteristics, 25% of offers went to certified substitutes and 93% went to substitutes who accepted at least one call-system offer during the year. Just 1% of offers went to a substitute on a teacher’s preferred list and 6% went to substitutes on a school’s preferred list. The remaining school characteristics have offer-averages similar to the corresponding job averages in column 1 of table 1.1.

Columns 2 and 3 report offer means separately for half and full days, respectively. The average acceptance rate for full-day jobs is slightly higher. Most of the other offer-specific characteristics are similar across half and full days with one exception: certified substitutes receive a larger share of half-day offers.

Columns 4 and 5 summarize offer characteristics by recipients' certification status. The acceptance rate of certified substitutes is three percentage points higher than that of non-certified substitutes. This could indicate that certified substitutes have stronger tastes for substitute teaching or that certified substitutes are more likely to accept offers because they receive higher quality offers as a result of the calling system's preference for certified substitutes. As expected, on average certified substitutes receive offers one full day earlier than non-certified substitutes and are less likely to receive day-of offers.

Average commutes are about the same for certified and non-certified substitutes, but school characteristics are not. Certified substitutes are more than twice as likely to receive elementary-school offers and less than half as likely to receive high-school offers as their non-certified counterparts. Certified substitutes are also more likely to receive offers from A schools, which could result from jobs being offered to certified substitutes first, who quickly accept the A offers leaving fewer A jobs to be offered to non-certified substitutes.

Over 85% of offers were made either on the day of or day before the job. Figure 1.2 provides day-of and day-before offer-time histograms that examine the precise timing of these offers. Not surprisingly, the likelihood of receiving a day-of offer decreases monotonically with time. Offers are made until relatively late in the school day because jobs, particularly half-day jobs, can start at any time. The day-before offer-time distribution follows a U-shaped pattern: the probability of receiving an offer decreases from 4:30 to 6:00 p.m., remains relatively flat

from 6:00 to 8:00 p.m., and then increases between 8:00 and 10:00 p.m. before slowly decreasing again. One explanation of this pattern is that many substitute requests are made during and immediately after the school day, then there is a lull during dinner time before another batch of requests are made later in the evening. These qualitative patterns remain when looking at the offer-time distributions separately for certified and non-certified substitutes.

Table 1.3 investigates the daily selectiveness of substitutes. For each of the 32,338 “sub-days” for which a substitute received at least one offer to work on a given day, the total number of sub-day offers received is tabulated separately for substitutes who worked on the day in question and those that did not. The average non-worker received 3.16 offers to work on the day in question while the average worker received 2.05 such offers. This difference is a result of the fact that offers to work on a given day essentially stop arriving once the substitute has accepted an offer to work on that day.

Two striking features of table 1.3 will be revisited when discussing the empirical specification. First, of the non-workers on a given day, nearly 16% rejected six or more offers and over 40% of total offers went to these “multiple rejecters.” It is unlikely that all of these substitutes received a series of unlucky draws from the job distribution; instead, these figures suggest that a nontrivial fraction of substitutes had a prohibitively large opportunity cost of substitute teaching on the day in question. Second, of the substitutes who worked on a given day, nearly 60% accepted the first offer that they received. One explanation of the high percentage of first-offer acceptances is that many substitutes have a low opportunity cost of subbing on a given day. Alternatively, these quick-to-accept substitutes may be extremely risk averse or worried that another offer may not arrive. This is not to say that there is no variation in

the number of offers received before accepting, however, as 20% of substitutes sampled two offers, 9% sampled three offers, 5% sampled four offers, and over 3% sampled 7 or more offers.

1.4 Econometric Model

1.4.1 *Substitutes' Optimal Decision Rule*

The functioning of the automated calling system in this labor market is remarkably similar to a finite-horizon job-search model with no recall and no on-the-job search (Mortensen, 1986).

Supposing that substitute teachers maximize expected utility when making daily labor supply decisions, the optimal strategy can be defined in terms of a reservation-utility decision rule:

accept an offer if and only if the utility of accepting (U^A) exceeds the expected utility of rejecting (U^R). The former depends on the individual's tastes for substitute teaching and on the offer's characteristics. The latter depends on both the individual's non-subbing alternative (U^N), or opportunity cost of substitute teaching, and expectations regarding future offers.

Let T represent the end of the school day, at which point the probability of receiving an offer becomes zero. Rejecting an offer at time T is therefore equivalent to choosing the non-subbing alternative, so $U_T^R = U^N$. If offers arrive at time t with probability π_t and only one offer can be received per period, the expected utility of rejecting at all t less than T is

$$U_t^R = \pi_{t+1} E \left[\max \left(U_{t+1}^A, U_{t+1}^R \right) \right] + (1 - \pi_{t+1}) E \left[U_{t+1}^R \right]. \quad (1.1)$$

Because U_t^R is decreasing in t and $U_T^R = U^N$, U_t^R can be approximated as

$$U_t^R = U^N + b(t), \quad (1.2)$$

where $b(t)$ is nonnegative, monotonically decreasing in t , and equals zero at time T . Empirically, I will employ a flexible piecewise-linear approximation of the $b(t)$ function that allows the first derivative to vary with the number of days in advance that the call is made.

I assume that substitute teachers' daily preferences are represented by the same utility function regardless of where, or if, they work. Daily utility is a function of non-labor income (Y), labor income (M), hours worked (H), commute time (h), and observed and unobserved individual, day, and non-wage job characteristics (ψ). Specifically, let daily utility be separable in income and leisure, taking the form

$$U = f(Y) + \beta M - g(H, h) + \psi. \quad (1.3)$$

The functions f and g are both increasing. M is valued linearly because small changes to lifetime earnings have approximately no income effect (Goette et al., 2004).

The empirics will take a linear approximation of g , so the utility accruing to substitute s of accepting an offer to work on day d at time t is

$$U_{sdt}^A = f(Y_{sd}) + \gamma^A \mathbf{x}_{sdt} + \lambda^A \mathbf{z}_s + \mathbf{b}^A \mathbf{j}_{sdt} + \delta^A \mathbf{r}_d + \omega_{sd}^A + \varepsilon_{sdt}, \quad (1.4)$$

where \mathbf{x}_{sdt} is a vector of observed job characteristics including M , H , and h ; \mathbf{z}_s is a vector of observed individual characteristics including gender, certification status, and preferred-list status;

\mathbf{r}_d is a vector of day-of-job variables including day of week and month; \mathbf{j}_{sdt} is a vector of time-of-call variables; ω_{sd}^A is the substitute's unobserved day-specific taste for substitute teaching; and ε_{sdt} is an offer-specific error term capturing unobserved offer characteristics and distractions to the substitute at the time of offer.⁷

A substitute's daily non-subbing utility depends on non-labor income and varies with observable individual and day characteristics. The characteristics of the non-subbing activity are unobserved and subsumed in an unobserved sub-day term ω_{sd}^N . Therefore, the utility of the non-subbing alternative is

$$U_{sd}^N = f(Y_{sd}) + \lambda^N \mathbf{z}_s + \delta^N \mathbf{r}_d + \omega_{sd}^N. \quad (1.5)$$

Combining equations (1.2), (1.4), and (1.5) with the optimal decision rule discussed above yields the probability that an offer will be accepted conditional on it being received. Formally,

$$\begin{aligned} & \Pr\left(A_{sdt} = 1 \mid p_{sdt} = 1, \mathbf{x}_{sdt}, \mathbf{z}_s, \mathbf{r}_d, \mathbf{j}_{sdt}, \omega_{sd}^A, \omega_{sd}^N\right) \\ &= \Pr\left(\gamma^A \mathbf{x}_{sdt} + \lambda \mathbf{z}_s + \delta \mathbf{r}_d + \mathbf{b} \mathbf{j}_{sdt} + \omega_{sd} + \varepsilon_{sdt} > 0 \mid p_{sdt} = 1\right), \end{aligned} \quad (1.6)$$

⁷ Because daily pay is binary, M is replaced by a half-day dummy in the empirics. The day-of-week and month dummies enter equation (4) because they contain information on job quality. For example, students may be systematically rowdier on Fridays and in June because they are excited for the weekend and summer vacation, respectively. Time of call enters equation (4) because it proxies for job quality to the extent that the unobserved job-quality distribution changes over time and because it provides a measure of the substitutes' preparation time for the job.

where A_{sdt} is a binary indicator of offer acceptance, p_{sdt} is a binary indicator of having received an offer, and parameters lacking a superscript represent net effects that are defined as follows: $\lambda = \lambda^A - \lambda^N$, $\delta = \delta^A - \delta^N$, $\mathbf{b} = \mathbf{b}^A - \mathbf{b}^N$, and $\omega_{sd} = \omega_{sd}^A - \omega_{sd}^N$. With the exception of γ^A , the primary object of interest in this study, only net effects of the model's covariates are identified because the same covariates enter equations (1.4) and (1.5). Finally, note that non-labor income was differenced out of (1.6) because it is valued identically in both U^A and U^N .⁸

1.4.2 Estimation

The observance of all rejected offers made by the call system distinguishes this dataset from those typically used to estimate job-search models (Devine & Kiefer, 1991, p. 8). This is important because the usual sample-selection problem associated with observing only accepted offers is avoided and equation (1.6) can be estimated in a straightforward binary-response framework. A different sample-selection problem remains, however, because of the no on-the-job search rule followed by the automated calling system: substitutes who work on day d will, on average, receive fewer offers and have higher values of ω_{sd} than day- d non-workers. Because offer-acceptance decisions are only observed when an offer was made, the data can be viewed as a selected sample where p_{sdt} serves as the selection indicator. Thus pooled estimators of (1.6), which leave ω_{sd} in the error term, are inconsistent because ω_{sd} is negatively correlated with p_{sdt} .

⁸ Intuitively, this is a result of consumption smoothing over the lifecycle and preferences that are separable in consumption and leisure. The assumption that non-labor income is valued differently on subbing and non-subbing days can be relaxed entirely by noting that any difference in utility would be sub-day specific and hence captured in ω_{sd} .

Conditional on \mathbf{z}_s and ω_{sd} , however, the offer-specific error term ε_{sdt} is independent of the selection indicator. This is a direct result of the call system's randomness. Accordingly, conditional on \mathbf{z}_s and ω_{sd} , time periods in which no offer is received can be considered “missing at random” (Wooldridge, 2010, p. 795) and the sample-selection problem can be safely ignored. This solution to an unbalanced-panels problem in a nonlinear model is similar in spirit to Kiefer and Neumann (1981).

I assume that $\varepsilon_{sdt} \mid \mathbf{x}_{sdt}, \mathbf{z}_s, \mathbf{r}_d, \mathbf{j}_{sdt}, \omega_{sd} \sim N(0,1)$, so (1.6) can be rewritten as

$$\begin{aligned} & \Pr\left(\gamma^A \mathbf{x}_{sdt} + \lambda \mathbf{z}_s + \delta \mathbf{r}_d + \mathbf{b} \mathbf{j}_{sdt} + \omega_{sd} + \varepsilon_{sdt} > 0 \mid p_{sdt} = 1\right) \\ &= \Phi\left(\gamma^A \mathbf{x}_{sdt} + \lambda \mathbf{z}_s + \delta \mathbf{r}_d + \mathbf{b} \mathbf{j}_{sdt} + \omega_{sd}\right). \end{aligned} \tag{1.7}$$

Assuming that $\omega_{sd} \mid \mathbf{z}_s, \mathbf{r}_d \sim N(0, \sigma_\omega)$, equation (1.7) can be estimated using the random effects (RE) probit estimator of Butler and Moffitt (1982). I will treat the RE-probit model as the baseline model.

The RE-probit model makes several strong assumptions, so I consider alternative estimators as well to verify the robustness of the results. First, consider relaxing the distributional assumption made on the unobserved sub-day effect ω_{sd} . As seen in table 1.4, the raw data suggests that on any given day a nontrivial number of substitutes have an extremely low opportunity cost (or high level of risk aversion) and another subset of substitutes have a prohibitively high opportunity cost. This suggests that ω_{sd} may not be normally distributed. An alternative is a nonparametric “mass point” distribution of ω_{sd} (Heckman & Singer, 1984). An

additional benefit of the mass-point (MP) model is that the proportion of sub-days located at each mass point and mass point-specific marginal effects can be estimated. Both the number and preferences of substitutes “at the margin” of accepting a day- d offer may be of particular interest to policy makers because these are the substitutes who are likely to be influenced by small policy changes.

Second, I relax the assumption made by both the RE-probit and MP-probit models that ω_{sd} is conditionally independent of the offer characteristics by using the linear fixed-effects (FE) estimator to estimate a linear probability model (LPM). Comparing the linear-FE estimates to linear-RE estimates provides an approximate test of the call system’s conditional randomness.⁹ The FE-logit estimator is not an attractive option in the present case because the majority of observations would be dropped from the analysis because there is no variation in the offer-acceptance decisions for the majority of sub-days.

The assumption that ε_{sdt} is independent of the offer characteristics is more contentious and less testable than the independence of ω_{sd} . This is because regular teachers may take it upon themselves to pay compensating wage differentials. For example, teachers in unobservably bad classrooms might systematically offer shorter assignments, causing ε_{sdt} to be correlated with H_{sdt} . Recall that the discussion of table 1.1 in section 1.3, however, suggests that regular teachers are not paying compensating wage differentials based on observable job characteristics,

⁹ The test is approximate because the linear-RE estimator is inconsistent as a result of the unbalanced-panel problem discussed above (Wooldridge, 2010, p. 831).

making it unlikely that they do pay compensating wage differentials based on unobservables.

Furthermore, this type of behavior is unlikely to be problematic for a number of reasons.¹⁰

1.5 Results

Table 1.4 reports estimated RE-probit coefficients and their standard errors clustered at the substitute level for four alternative specifications. Columns 1 and 2 estimate the baseline RE-probit model using all offers (observations). The only difference between the two is the presence of a half day-hours interaction term in the former, which is negative and statistically significant at the 5% level. Surprisingly, the column 1 coefficient on hours is positive and statistically significant at 5%. The half day-hours coefficient is larger in magnitude, however, implying that the marginal effect of hours is negative for half-day jobs and positive for full-day jobs. The model estimated in column 2 assumes that the marginal effect of hours is identical for both full-day and half-day jobs and precisely estimates a zero hours coefficient. The hours coefficients in column 1 are quite small as well, suggesting that the true effect of job length on the labor supply decisions of substitute teachers is quite small. The half-day dummy coefficient cannot be directly interpreted because half-day status cannot change while holding hours constant; coherent average partial effects (APE) will be discussed shortly.

There are essentially no differences in the remaining coefficient estimates between columns 1 and 2. Several offer characteristics have relatively large and statistically significant

¹⁰ First, if teachers do behave this way, it is only problematic if substitutes are aware of each job's unobserved quality. Considering the large number of substitutes and regular teachers working in the consortium it is unlikely that many substitutes, especially those accepting offers from the randomized call system, are aware of the intricacies of each specific classroom. Second, as seen in table 1.1, 98% of jobs are eventually accepted. With such a high fill rate the threat of not finding a substitute is quite low, which significantly lowers the incentive to implement such a strategy. Finally, concerned teachers have the more effective option of specifically requesting a substitute or compiling a list of teacher-preferred substitutes.

coefficients, including commute time, school type, MDE-assigned school grades, special education, certification, preferred list status, and the time-of-call variables. Two sets of covariates included in the models are excluded from table 1.4 in the interest of brevity. First is a set of day-of-job indicators that omits Wednesday; only the Friday coefficient is statistically significant and reported in table 1.4. Second is a set of month-of-job indicators that omits October as the reference point. The fall and early winter month coefficients are statistically insignificant. The coefficients on the spring months (March, April, May, and June) are all about -0.30 and statistically significant at the 1% level.

Columns 3 and 4 of table 1.4 estimate the baseline specification of column 1 separately for certified and non-certified substitutes, respectively. A likelihood ratio test strongly rejects that the model's parameters are identical for certified and non-certified substitutes.¹¹ There are several noticeable differences in coefficients, particularly commute time, where the certified coefficient is more than three times larger than that for non-certified substitutes. Other large differences are found for school type, school quality, and subjects including foreign language and special education.

Columns 1 – 3 of table 1.5 report APE for the RE-probit models estimated in columns 1, 3 and 4 of table 1.4. These APE are comparable to the LPM coefficients reported in columns 4 and 5 of table 1.5 and were computed following Wooldridge (2010, p. 613), exploiting the fact that the conditional expectation of (1.7) can be written as

¹¹ The LR test statistic was formed by taking the log likelihood of the unrestricted model to be the sum of the log likelihoods from columns 3 and 4. The resulting LR statistic has a p-value significantly less than 0.0001.

$$E(A_{sdt} | p_{sdt}=1, \mathbf{x}_{sdt}, \mathbf{z}_s, \mathbf{r}_d, \mathbf{j}_{sdt}, \omega_{sd}) = \Phi \left(\frac{\gamma^A \mathbf{x}_{sdt} + \lambda \mathbf{z}_s + \delta \mathbf{r}_d + \mathbf{b} \mathbf{j}_{sdt}}{(1 + \sigma_\omega^2)^{0.5}} \right). \quad (1.8)$$

The value of the RHS of (1.8), averaged across all offers, provides an estimate of the predicted acceptance probability and is reported at the bottom of table 1.5.¹² Precise definitions of the estimated APE are provided in appendix 1.3. The APE standard errors were computed by taking the standard deviation of 50 bootstrapped APE estimates. The bootstrap procedure resampled with replacement at the substitute level, utilizing all observations from the chosen substitute. Resampling at the substitute level produces standard errors that are robust to substitute-level clustering and that are asymptotically equivalent to the usual robust “sandwich” standard error estimates (Cameron & Trivedi, 2005).

The half-day APE reported in column 1 of table 1.5 indicates that half-day jobs are 1.8 percentage points less likely to be accepted than full-day jobs. The preference for full-day jobs is about one percentage point higher among certified substitutes than non-certified substitutes. The overall APE of half-day job hours on the acceptance probability is -0.003 and does not vary with certification status. It is not significantly different from zero and the standard error of 0.003 suggests that this is a precisely estimated “zero effect” of hours on the acceptance decision. The full day-hours APE is larger in magnitude, positive, and even more precisely estimated.

¹² APE were computed by averaging across all offers (observations). It is worth pointing out that different APE might be considered, however. For example, we might average across jobs because low-quality (frequently rejected) jobs are over represented in the sample. Similarly, averaging across substitutes might be useful to the extent that high-opportunity cost substitutes (frequent rejecters) are overrepresented in the sample. An alternative to computing APE at all is to simply scale the probit coefficients reported in table 1.4 by values ranging from zero to 0.4 (the range of possible values of the normal pdf).

Although statistically different from zero, at 0.007, a half-hour (one standard deviation) increase in job length only raises the acceptance probability by one third of one percentage point, suggesting that job length has no economically significant impact on the offer-acceptance decisions made by substitute teachers.

Commute time is measured in one-way hours, so a fifteen-minute increase lowers the acceptance rate by about 1.4 percentage points. This is a substantial effect given that the overall predicted acceptance rate is 0.12. The effect of commute time on certified substitutes is substantially larger than for non-certified substitutes, where a mere 15 minute increase in one-way commute time lowers the predicted acceptance rate by over four percentage points. A likely explanation for this difference is that the certified substitutes' opportunity cost of time is greater.

Regarding school type, overall, high-school jobs are two percentage points more likely to be accepted than elementary and middle-school jobs. This effect is statistically significant at the 1% level. Among certified substitutes, elementary-school jobs are nearly three percentage points less likely to be accepted, while non-certified substitutes react similarly to offers from elementary schools and middle schools. The preference for high schools is driven entirely by non-certified substitutes, however, who are 2.7 percentage points more likely to accept high-school jobs. Both types of substitutes are significantly less likely to accept jobs in charter schools, with a certified APE of about -0.05 and a non-certified APE of about -0.03. The charter-school result is interesting, especially because student demographics within the schools are being controlled for. If it is not the students that make these jobs less desirable, one possibility is that these jobs are more structured and require greater effort from the substitute teacher; another is that these jobs provide substitutes with fewer networking opportunities. Finally, it is important when interpreting these results to remember that the certified substitutes

are called first and are therefore choosing from a different job-quality distribution than are non-certified substitutes.

Relative to the reference group of highest-achieving A schools, both B and C schools are about 1.5 percentage points less likely to be accepted, and this effect is statistically significant at the 1% level. The lowest-graded schools observed in the consortium, those earning a D, are 6.6 percentage points less likely to be accepted than A schools. Magnitude wise, the effects are slightly larger among non-certified substitutes. The school-grade effects for certified substitutes are similar to the overall APE, but imprecisely estimated. Having controlled for schools' achievement levels, it is interesting to note that the student-demographic variables are mostly insignificant. There also appear to be some subtle differences between certified and non-certified preferences for student type. For example, the APE of percent black suggests that a 10% increase in black enrollment lowers the probability that a certified substitute will accept the offer by one percentage point but raise the probability that a non-certified substitute will accept the offer by a little more than half a percentage point, and these effects are marginally statistically significant. Again, part of this might due to the fact that certified substitutes are called first.

The raw call-system data contains about 70 unique subject descriptions that I aggregated into 14 broad groups. The 13 indicator variables are strongly jointly significant, with English/reading serving as the omitted reference group because these subjects generally require less subject-specific knowledge than math or science, for example. Only a few subjects are individually statistically significant, however. Large, negative, statistically significant effects were found on the art/gym/music, special education, and "other" indicators. These results suggest that substitutes generally preferred academic to non-academic subjects, but had little

subject-specific preference within these broad groups.¹³ The overall negative effect of each of these seemingly diverse groups of subjects ranges from 0.02 to 0.03. These subjects share some common characteristics, however, notably that they each require some type of specific training and increased attentiveness on the part of the substitute. The latter is particularly true of gym and special education classes. The large and significant negative effect of foreign language classrooms for certified substitutes, coupled with a larger APE of the art/gym/music group, suggests that certified substitutes are particularly averse to accepting jobs in subjects that are foreign to them.

Certified substitutes are about four percentage points more likely to accept an offer, according to the APE of certified in column 1, and this is precisely the difference in average predicted acceptance rates for certified and non-certified substitutes found at the bottom of columns 2 and 3. The APE of the list-status indicators, at two to three times the average predicted acceptance rate, are quite large. While it is tempting to infer from this that forging personal relationships between substitute and regular teachers is a surefire way to increase offer-acceptance rates, remember that part of this positive effect results from the call order's dependence on list status, which inflates the estimated effect. The same applies to the certified dummy because of certification's role in the call order.

Friday jobs are about two percentage points less likely to be accepted than jobs on any other day of the week, and this is true for both certified and non-certified substitutes. This is likely the result of some combination of increased demand for substitutes on Fridays, higher opportunity costs of substitutes on Fridays, and poorer student behavior on Fridays.

¹³ Similar effects were found on art, gym, and music when each was assigned its own indicator variable. The "other" category includes agriculture, English as a second language, family life, home economics, life skills, and speech therapy.

Finally, recall the statistical significance and large magnitudes of the time-of-call coefficients in table 1.4. The day-of and time-of call interactions complicate the calculation of scalar APE. Instead, I plot the average acceptance rate as a function of time of call in figure 1.3, which is easier to interpret. Three separate curves are plotted, representing the overall, certified, and non-certified response to time of call. The offer time-acceptance gradient is essentially flat for all substitutes for offers made more than one day in advance, which is intuitive because this is too early to worry about not receiving additional offers and also because the average offer quality is unlikely to vary across time within days this far in advance of the job. The gradient is upward sloping on the day before, suggesting that the reservation utility is decreasing with time, as predicted by the search model. This effect is greater for non-certified substitutes, which again is intuitive because the non-certified subs are called last and thus might worry more than certified subs about receiving additional future offers.

For day-of offers, however, two things change. First, the gradient becomes downward sloping, which suggests that offer quality decreases rapidly with time on the morning of and that this negative effect dominates the search effect observed for day-before offers. One reason for this is that not only have the jobs been thoroughly picked over by the morning of, but also that offers made on the morning of are more likely to have been placed by the regular teacher on the morning of and thus these jobs are less desirable because the teacher has not left a lesson plan and the students have not been prepared for the absence. Second, the certified gradient is now much steeper than the non-certified gradient, perhaps because certified substitutes have a stronger aversion to low-quality jobs.

Columns 4 and 5 of table 1.5 report RE- and FE-LPM estimates, respectively. The LPM estimates, for the most part, are similar in sign and magnitude to the probit APE in column 1.

Additionally, the linear RE and FE estimates are themselves quite similar, which suggests that the calling system is conditionally random. Finally, an advantage of the linear models is that standard errors robust to two-way clustering are easily computed (Cameron et al., 2006). The probit standard errors are one-way clustered at the substitute level, which is important because the unobserved sub-day tastes for subbing and opportunity costs of subbing are likely correlated across days, within substitutes. But if there are important unobserved job characteristics, proper inference must account for this second source of clustering. The two-way standard errors reported in square brackets are quite similar to the one-way clustered standard errors in parentheses for the linear models. In the RE estimates, on average, the two-way standard errors are 12% larger than the one-way standard errors. The similarity between the one-way and two-way clustered standard errors in the linear model is reassuring for the interpretation of the one-way probit-model standard errors.

The overall mass-point APE, reported in column 1 of table 1.6, is the weighted average of the mass point-specific APE reported in columns 2 – 4, which are arranged in descending order of predicted acceptance probability. Equation (1.8) is unnecessary for the calculation of the mass-point APE because the estimated mass-point locations can be plugged directly in to the RHS of equation (1.7). There is no practical difference between the RE-probit APE in column 1 of table 1.5 and the MP-probit APE in column 1 of table 1.6, which indicates that the results are robust to the assumed heterogeneity distribution.

The mass-point locations and corresponding location probabilities are provided at the bottom of table 1.6, where we see that 12% of offers go to substitutes with a 59% chance of accepting the offer, 48% go to a substitute with a 9% chance of accepting, and a remarkable 40% go to substitutes who will almost certainly reject any offer they receive. This last result was

foreshadowed in table 1.3, where 43% of calls were made to substitutes who rejected six or more day- d offers on the way to not working on day d . With nearly half of offers going to substitutes who have no intention of accepting or even listening to the offer, the overall APE is effectively biased towards zero.

The mass-point APE are useful, then, because they provide the preferences of substitutes “at the margin” of accepting an offer. The APE in columns 2 and 3 are about four and two times larger than the overall APE, respectively, suggesting that substitutes “at the margin” are significantly more responsive to offer characteristics than implied by the overall APE discussed in table 1.5. Policy makers seeking to redistribute substitutes or substitute quality across schools may be particularly interested in the APE of column 2, because these are the substitutes who are significantly “at risk” of working on a given day.

1.6 Conclusions

This paper has used data on accepted and rejected job offers to estimate a sequential binary-choice model of substitute teachers’ daily labor supply. A variety of non-wage job characteristics were found to significantly affect the offer-acceptance probability, including commute time, school type, school quality, subject, day of job, and time of offer. Higher-paying longer jobs were preferred to lower-paying shorter jobs. Job length, conditional on daily pay, was a notable non-factor in substitutes’ decision making as were student demographics conditional on school achievement level. Future work might probe the wage elasticity by experimentally varying daily pay or rigorously analyzing the impact of a wage change.

The basic results of the paper are of general interest for at least three reasons. First, the research potential of pseudo-random automated calling systems is displayed, both as a source of

exogenous variation and as a collector of high-quality data. Second, economists study how individuals make decisions and this paper has provided a unique glimpse into the determinants of a fundamentally important decision: when and where to work. Indeed, the analysis highlights the multitude of factors in addition to hours and wages that enter the decision-making process.

Finally, the results may contribute to the regular-teacher literature more generally. Exogenous variation in important factors such as commute time, student achievement, and student demographics is typically nonexistent in studies of teacher attrition and sorting across schools. The main results suggest several potentially welfare-enhancing substitute-teacher policies. First, the call-order algorithm might be adjusted to offer jobs to nearby substitutes first. This policy would decrease the number of calls made by the call system and increase substitutes' preparation time for jobs. More importantly, if schools routinely call the same set of substitutes first, these substitutes will repeatedly work in the same schools. Doing so will provide these substitutes with specific human capital with regards to schools' policies, layout, and individual students' needs. Similarly, substitutes would accumulate social capital with the administration, faculty, and students. A lack of both types of capital is often seen as a challenge to successful substitute teaching (Coverdill & Oulevey, 2007).

Second, a variety of methods might be implemented to attract certified substitutes to underperforming schools and improve the equity of the distribution of substitute-teacher quality. One solution is to pay compensating wage differentials in the low-achieving schools that stand to benefit the most from attracting higher-quality substitutes. Such a policy can be budget neutral, and even decrease expenditures, if the compensating wage differential is created by decreasing daily wages of preferred jobs. Similarly, the observed preferences for commute time and time of offer can be exploited in the calling-system algorithm to direct substitutes, or a subset of

substitutes, towards particular schools or classrooms. There are, however, potentially complicated general equilibrium effects of such policy changes that must be considered.

Finally, regarding the external validity of these findings, the estimated partial effects on the acceptance probability are likely to be small relative to national or overall effects. This is because over 98% of substitute requests were ultimately filled and substitute teachers nevertheless exhibited strong preferences over a variety of job characteristics. It stands to reason, then, that in labor markets with excess demand (substitute-teacher shortages) substitutes would be even choosier when accepting jobs. Still, it would be useful to employ a similar empirical approach to call-system data in other substitute-teacher labor markets to verify the robustness of these results.

CHAPTER 1 APPENDICIES

APPENDIX 1.1

CHAPTER 1 TABLES

Table 1.1: Mean Job Characteristics

	All Jobs 1	Never- Accepted 2	Half-day Jobs 3	Full-day Jobs 4	Full-Day, Hours < 5 5
Never accepted	0.02	1	0.03	0.02	0.09
Times offered	10.99 (15.56)	47.21 (46.25)	11.40 (15.04)	10.75 (15.85)	18.53 (26.76)
First offer was day of	0.32	0.74	0.32	0.32	0.38
Half day	0.36	0.40	1	0	0
Hours	5.76 (1.84)	5.41 (1.86)	3.43 (0.52)	7.09 (0.53)	4.63 (0.22)
Hourly wage	\$11.16 (2.19)	\$11.89 (3.23)	\$12.05 (3.14)	\$10.66 (1.09)	\$16.23 (0.75)
Rural district	0.35	0.36	0.28	0.40	0.22
<u>School Characteristics</u>					
Elementary	0.36	0.31	0.46	0.30	0.29
Middle	0.28	0.36	0.25	0.30	0.49
High	0.36	0.33	0.29	0.40	0.22
Charter	0.04	0.07	0.02	0.04	0.06
Grade A	0.51	0.42	0.57	0.47	0.52
Grade B	0.39	0.42	0.36	0.41	0.34
Grade C	0.09	0.14	0.07	0.11	0.14
Grade D	0.01	0.02	0.00	0.01	0.00
% lunch	0.19 (0.15)	0.23 (0.18)	0.19 (0.15)	0.19 (0.16)	0.25 (0.15)
% black	0.10 (0.13)	0.13 (0.18)	0.10 (0.12)	0.10 (0.14)	0.12 (0.13)
% Hispanic	0.04 (0.04)	0.05 (0.04)	0.05 (0.03)	0.04 (0.04)	0.05 (0.04)
<u>Day of job</u>					
Monday	0.19	0.08	0.18	0.20	0.08
Tuesday	0.19	0.11	0.17	0.20	0.14
Wednesday	0.19	0.07	0.23	0.17	0.45
Thursday	0.20	0.10	0.20	0.20	0.14
Friday	0.23	0.64	0.21	0.24	0.20
N	8,566	214	3,126	5,440	139

Notes: Standard deviations of non-binary variables are given in parentheses. Time of first offer is measured in days prior to the job beginning. For example, 0 means the offer was made on the morning of the job, 1 the day before, and so on.

Table 1.2: Mean Offer Characteristics

	All Offers 1	Job Length		Certification Status	
		Half 2	Full 3	Certified 4	Non-cert. 5
Accepted	0.08	0.07	0.08	0.10	0.07
Half day	0.38	1	0	0.45	0.35
Hours	5.67	3.43	7.03	5.47	5.74
	(1.85)	(0.52)	(0.66)	(1.91)	(1.83)
Hourly wage	\$11.27	\$12.07	\$10.79	\$11.18	\$11.31
	(2.41)	(3.36)	(1.36)	(1.82)	(2.58)
Lead days	1.61	1.73	1.53	2.37	1.35
	(4.88)	(5.05)	(4.78)	(6.06)	(4.38)
Day-of job	0.49	0.47	0.50	0.35	0.54
Day-before job	0.36	0.35	0.36	0.42	0.34
One-way commute min.	19.41	18.89	19.73	19.24	19.47
	(14.75)	(14.88)	(14.66)	(12.60)	(15.41)
One-way commute miles	14.62	14.07	14.95	14.67	14.60
	(14.50)	(14.54)	(14.47)	(12.26)	(15.19)
Certified	0.25	0.30	0.22	1	0
Worked at least once	0.93	0.93	0.93	0.94	0.93
Teacher's list	0.01	0.00	0.01	0.01	0.00
School's list	0.06	0.06	0.05	0.05	0.06
Rural district	0.34	0.27	0.38	0.26	0.37
Elementary	0.38	0.46	0.33	0.65	0.29
Middle	0.30	0.27	0.32	0.21	0.33
High	0.32	0.27	0.35	0.14	0.38
Charter	0.07	0.06	0.08	0.06	0.08
Grade A	0.47	0.51	0.44	0.60	0.42
Grade B	0.38	0.38	0.39	0.29	0.41
Grade C	0.14	0.10	0.16	0.10	0.15
Grade D	0.01	0.01	0.01	0.01	0.02
% lunch	0.23	0.22	0.23	0.23	0.22
	(0.19)	(0.18)	(0.19)	(0.18)	(0.19)
% black	0.13	0.13	0.13	0.12	0.13
	(0.17)	(0.17)	(0.18)	(0.16)	(0.18)
% Hispanic	0.05	0.05	0.05	0.05	0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Monday	0.18	0.17	0.19	0.18	0.18
Tuesday	0.16	0.15	0.17	0.17	0.16
Wednesday	0.18	0.23	0.14	0.19	0.17
Thursday	0.19	0.20	0.19	0.21	0.19
Friday	0.29	0.25	0.31	0.26	0.30
N	94,106	35,645	58,461	23,823	70,283

Notes: Standard deviations of non-binary variables are given in parentheses. Lead days measure the days prior to the job that the offer is made. For example, 0 means the offer was made on the morning of the job, 1 the day before, and so on.

Table 1.3: Daily Offers Received and Daily Selectivity of Substitutes

Offers	Day- <i>d</i> Non-workers			Day- <i>d</i> Workers		
	Sub-days	% sub-days	% of offers	Sub-days	% sub-days	% of offers
1	9,552	37.87	11.98	4,027	57.46	28.01
2	4,952	19.64	12.42	1,408	20.09	19.59
3	3,055	12.11	11.5	646	9.22	13.48
4	2,036	8.07	10.22	366	5.22	10.18
5	1,662	6.59	10.42	179	2.55	6.22
6	1,159	4.6	8.72	131	1.87	5.47
7	791	3.14	6.95	79	1.13	3.85
8	576	2.28	5.78	53	0.76	2.95
9	410	1.63	4.63	38	0.54	2.38
10	251	1	3.15	16	0.23	1.11
11-20	695	2.76	11.69	57	0.81	5.42
> 20	81	0.3	2.55	8	0.09	1.37
Total	25,220	100	100	7,008	100	100

Notes: Only 7,008 sub-days are observed for working substitutes because jobs accepted by substitutes with non-Michigan zip codes are dropped from the sample. There were also 82 cases in which a substitute worked two non-overlapping half-day jobs on the same day.

Table 1.4: RE-Probit Coefficients

	All 1	All 2	Certified 3	Non- Certified 4
Half day	0.3963 (0.2160)*	-0.0596 (0.0833)	0.6465 (0.4873)	0.3767 (0.2420)
Half-day*Hours	-0.0981 (0.0475)**	.	-0.1325 (0.1022)	-0.0943 (0.0540)*
Hours	0.0646 (0.0280)**	0.0322 (0.0248)	0.1055 (0.0540)*	0.0604 (0.0339)*
One-way commute	-0.5356 (0.2159)**	-0.5361 (0.2162)**	-1.2762 (0.4519)***	-0.3615 (0.2258)
Rural district	-0.0055 (0.0550)	-0.0014 (0.0550)	0.0224 (0.1015)	0.0039 (0.0624)
Elementary school	-0.0455 (0.0551)	-0.0431 (0.0552)	-0.2197 (0.1245)*	-0.0179 (0.0601)
High school	0.1980 (0.0640)***	0.2004 (0.0641)***	-0.0830 (0.1472)	0.2943 (0.0752)***
Charter school	-0.3536 (0.1217)***	-0.3459 (0.1212)***	-0.4384 (0.2363)*	-0.3526 (0.1303)***
Student-Teacher ratio	0.0169 (0.0087)*	0.0172 (0.0087)**	0.0002 (0.0113)	0.0159 (0.0129)
Grade B	-0.1451 (0.0498)***	-0.1468 (0.0498)***	-0.0956 (0.0839)	-0.2064 (0.0588)***
Grade C	-0.1459 (0.0773)*	-0.1500 (0.0773)*	-0.1126 (0.1442)	-0.2392 (0.0872)***
Grade D	-0.8619 (0.3779)**	-0.8902 (0.3775)**	-0.2353 (0.6443)	-1.1220 (0.4619)**
Percent lunch program	0.0266 (0.2568)	0.0285 (0.2567)	-0.0297 (0.5128)	0.3270 (0.2810)
Percent black	0.2949 (0.3242)	0.3216 (0.3242)	-0.8487 (0.6502)	0.5996 (0.3965)
Percent Hispanic	-1.2323 (0.9328)	-1.2631 (0.9318)	-1.9631 (1.7965)	-0.7467 (1.1357)
Pre-K/Kindergarten	-0.0345 (0.1048)	-0.0358 (0.1048)	0.1288 (0.2570)	-0.0816 (0.1167)
First/Second Grade	0.0771 (0.0902)	0.0790 (0.0903)	0.2596 (0.2416)	-0.0755 (0.1075)
Third/Fourth Grade	0.0367 (0.0934)	0.0404 (0.0934)	0.2249 (0.2463)	-0.1383 (0.0989)
Fifth/Sixth Grade	0.0428 (0.0848)	0.0465 (0.0847)	0.1237 (0.2376)	-0.0753 (0.0844)
Seventh/Eighth Grade	0.1296 (0.1194)	0.1327 (0.1192)	0.3029 (0.2704)	-0.0635 (0.1368)
Math	0.0445 (0.0863)	0.0418 (0.0865)	0.3477 (0.3232)	-0.0166 (0.0729)

Table 1.4, Continued

Science	0.0818 (0.0775)	0.0824 (0.0775)	0.1917 (0.2855)	0.0565 (0.0736)
Social Studies	-0.0203 (0.0885)	-0.0198 (0.0887)	-0.2560 (0.3097)	0.0307 (0.0818)
Art/Gym/Music	-0.2644 (0.0755)***	-0.2580 (0.0753)***	-0.4309 (0.3039)	-0.2454 (0.0700)***
Business/Technology	-0.0752 (0.0779)	-0.0740 (0.0780)	0.0089 (0.3937)	-0.0833 (0.0707)
Foreign Language	-0.0278 (0.0893)	-0.0302 (0.0895)	-0.6902 (0.3135)**	0.0097 (0.0866)
Special Education	-0.1760 (0.0796)**	-0.1749 (0.0796)**	-0.0892 (0.3139)	-0.1787 (0.0768)**
Other	-0.1955 (0.1052)*	-0.1938 (0.1053)*	-0.2030 (0.3351)	-0.2086 (0.1101)*
Certified	0.3692 (0.1343)***	0.3703 (0.1344)***	Yes	No
Male	0.0583 (0.1212)	0.0587 (0.1213)	0.3112 (0.2379)	0.0133 (0.1392)
Teacher's List	2.0205 (0.1832)***	2.0284 (0.1831)***	1.6924 (0.3370)***	2.2188 (0.2241)***
School's List	1.3836 (0.1134)***	1.3887 (0.1136)***	0.8878 (0.1849)***	1.5859 (0.1434)***
% Teacher's Lists	-0.5716 (1.9553)	-0.5771 (1.9540)	-2.1378 (2.7874)	0.4711 (2.4342)
% School's Lists	0.1853 (0.5722)	0.1821 (0.5729)	-1.4233 (1.3497)	0.3614 (0.6441)
Friday job	-0.2119 (0.0505)***	-0.2101 (0.0504)***	-0.1748 (0.0850)**	-0.2141 (0.0630)***
Time of call ($T - t$)	-0.0007 (0.0001)***	-0.0007 (0.0001)***	-0.0006 (0.0002)***	-0.0006 (0.0002)***
Day-of offer	-1.1566 (0.1393)***	-1.1500 (0.1395)***	-2.0669 (0.2447)***	-0.7591 (0.1702)***
Day-before offer	0.9938 (0.2232)***	1.0054 (0.2234)***	0.4098 (0.3566)	1.2544 (0.2951)***
Day of*($T - t$)	0.1074 (0.0141)***	0.1069 (0.0141)***	0.1962 (0.0261)***	0.0734 (0.0165)***
Day before*($T - t$)	-0.0420 (0.0097)***	-0.0425 (0.0098)***	-0.0201 (0.0156)	-0.0504 (0.0129)***
Observations	94,106	94,106	23,823	70,283
Sub-days (RE)	32,228	32,228	9,037	23,191
Substitutes (clusters)	771	771	195	576
Log likelihood	-21,587	-21,590	-6,751	-14,667
rho	0.685	0.686	0.647	0.697

Table 1.4, Continued

Notes: Standard errors, in parentheses, are robust to clustering at the substitute level. ***, **, and * indicate statistical significance at 1, 5, and 10 percent. Coefficients for all covariates are reported with the exception of month dummies and day-of-week dummies other than Friday. The rho statistic is the percentage of unobserved variation due to the sub-day random effect. The RE probits were fit using 12-point adaptive quadrature, which is the preferred approximation method when rho is relatively large (Rabe-Hesketh et al., 2005). The coefficient estimates are stable when the number of quadrature points is increased.

Table 1.5: Average Partial Effects

	RE Probits			RE-LPM	FE-LPM
	All 1	Certified 2	Non-Cert. 3	All 4	All 5
Half day	-0.0176 (0.0031)***	-0.0247 (0.0074)***	-0.0149 (0.0036)***	-0.0037 (0.0122) [0.0161]	-0.0095 (0.0126) [0.0133]
Half-day hours	-0.0032 (0.0028)	-0.0032 (0.0072)	-0.0030 (0.0034)	-0.0007 (0.0030) [0.0033]	0.0004 (0.0032) [0.0032]
Full-day hours	0.0068 (0.0019)***	0.0142 (0.0054)**	0.0058 (0.0016)***	0.0014 (0.0014) [0.0020]	0.0009 (0.0014) [0.0015]
One-way	-0.0544 (0.0112)***	-0.1611 (0.0335)***	-0.0326 (0.0134)***	-0.0568 (0.0139)*** [0.0140]***	-0.0677 (0.0139)*** [0.0139]***
Rural district	-0.0006 (0.0034)	0.0028 (0.0086)	0.0004 (0.0033)	0.0019 (0.0032) [0.0037]	0.0018 (0.0028) [0.0029]
Elementary	-0.0046 (0.0048)	-0.0283 (0.0089)***	-0.0017 (0.0038)	-0.0022 (0.0030) [0.0037]	-0.0005 (0.0030) [0.0031]
High school	0.0208 (0.0044)***	-0.0101 (0.0109)	0.0278 (0.0050)***	0.0146 (0.0040)*** [0.0045]***	0.0130 (0.0038)*** [0.0039]***
Charter school	-0.0325 (0.0084)***	-0.0495 (0.0201)***	-0.0292 (0.0099)***	-0.0233 (0.0076)*** [0.0087]***	-0.0221 (0.0080)*** [0.0083]***
Student/teach.	0.0017 (0.0005)***	0.00003 (0.0010)	0.0015 (0.0009)*	0.0009 (0.0005)* [0.0006]	0.0002 (0.0005) [0.0005]
Grade B school	-0.0147 (0.0024)***	-0.0120 (0.0069)*	-0.0187 (0.0031)***	-0.0098 (0.0028)*** [0.0033]***	-0.0079 (0.0025)*** [0.0026]***
Grade C school	-0.0145 (0.0041)***	-0.0140 (0.0109)	-0.0208 (0.0043)***	-0.0139 (0.0045)*** [0.0051]***	-0.0133 (0.0043)*** [0.0044]***
Grade D school	-0.0663 (0.0173)***	-0.0296 (0.0482)	-0.0712 (0.0213)***	-0.0506 (0.0185)*** [0.0218]**	-0.0348 (0.0172)** [0.0178]*
% lunch program	0.0024 (0.0137)	-0.0038 (0.0386)	0.0294 (0.0146)**	0.0141 (0.0150) [0.0172]	0.0149 (0.0136) [0.0144]
% black	0.0304 (0.0202)	-0.1058 (0.0591)*	0.0560 (0.0254)**	0.0196 (0.0186) [0.0213]	0.0015 (0.0172) [0.0176]
% Hispanic	-0.1262 (0.0610)**	-0.2527 (0.1747)*	-0.0691 (0.0732)	-0.0755 (0.0471) [0.0548]	-0.0327 (0.0451) [0.0461]

Table 1.5, Continued

Math	0.0048 (0.0043)	0.0487 (0.0218)**	-0.0011 (0.0035)	-0.0034 (0.0052) [0.0064]	-0.0082 (0.0045)* [0.0049]*
Science	0.0087 (0.0046)*	0.0258 (0.0268)	0.0056 (0.0054)	0.0028 (0.0051) [0.0061]	-0.0001 (0.0050) [0.0052]
Social Studies	-0.0021 (0.0053)	-0.0303 (0.0203)	0.0032 (0.0053)	0.0003 (0.0061) [0.0071]	0.0026 (0.0060) [0.0062]
Art/Gym/Music	-0.0251 (0.0030)***	-0.0482 (0.0189)**	-0.0212 (0.0048)***	-0.0192 (0.0043)*** [0.0052]***	-0.0169 (0.0041)*** [0.0043]***
Business/Tech.	-0.0073 (0.0052)	0.0017 (0.0364)	-0.0073 (0.0044)*	-0.0075 (0.0048) [0.0061]	-0.0076 (0.0047) [0.0050]
Foreign Language	-0.0024 (0.0049)	-0.0705 (0.0188)***	0.0014 (0.0052)	-0.0055 (0.0055) [0.0069]	-0.0074 (0.0052) [0.0054]
Special Education	-0.0171 (0.0042)***	-0.0112 (0.0155)	-0.0157 (0.0053)***	-0.0134 (0.0048)*** [0.0056]**	-0.0129 (0.0047)*** [0.0048]***
Other	-0.0184 (0.0062)***	-0.0238 (0.0221)	-0.0178 (0.0060)***	-0.0156 (0.0058)*** [0.0075]**	-0.0167 (0.0059)*** [0.0060]***
Certified	0.0398 (0.0076)***	Yes	No	0.0542 (0.0167)*** [0.0167]***	.
Male	0.0057 (0.0076)	0.0414 (0.0309)	0.0012 (0.0071)	0.0151 (0.0139) [0.0139]	.
Teacher's List	0.3327 (0.0335)***	0.3091 (0.0548)***	0.3541 (0.0355)***	0.3103 (0.0397)*** [0.0413]***	0.2674 (0.0501)*** [0.0511]***
School's List	0.1999 (0.0145)***	0.1393 (0.0229)***	0.2217 (0.0161)***	0.1545 (0.0168)*** [0.0184]***	0.1151 (0.0146)*** [0.0154]***
Friday job	-0.0208 (0.0031)***	-0.0218 (0.0068)***	-0.0188 (0.0031)***	-0.0134 (0.0060)** [0.0066]**	.
Average	0.115 (0.005)***	0.14 (0.01)***	0.10 (0.005)***	.	.
Observations	94,106	23,823	70,283	94,106	94,106
Sub days	32,228	9,037	23,191	32,228	32,228
Substitutes	771	195	576	771	771
Jobs	8,566	5,796	7,343	8,566	8,566

Table 1.5, Continued

Notes: Substitute-clustered standard errors are reported in parentheses. The probit APE standard errors are based on 50 bootstrap replications. The square brackets in columns 4 and 5 contain two-way substitute-job clustered standard errors (Cameron et al., 2006). ***, **, and * indicate statistical significance at 1, 5, and 10 percent.

Table 1.6: Mass-point Probit APE

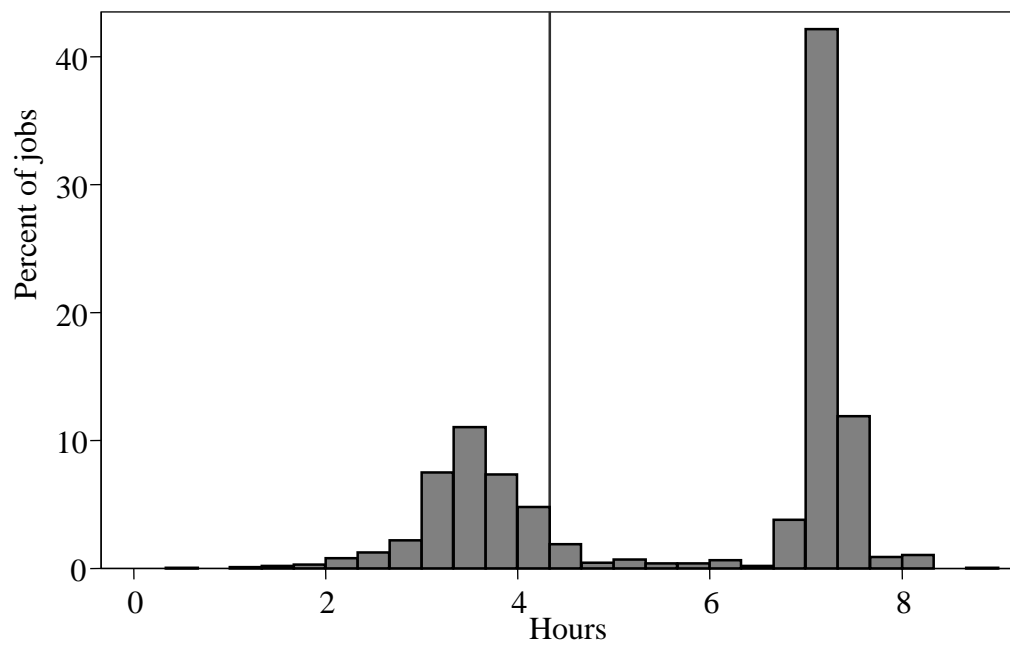
	APE 1	APE MP 1 2	APE MP 2 3	APE MP 3 4
Half day	-0.0168 (0.0153)	-0.0567 (0.0594)	-0.0205 (0.0279)	-0.0000 (0.0017)
One-way commute	-0.0558 (0.0208)***	-0.1859 (0.0795)**	-0.0684 (0.0651)	-0.0001 (0.0035)
Elementary school	-0.0052 (0.0046)	-0.0174 (0.0177)	-0.0064 (0.0092)	-0.0000 (0.0005)
High school	0.0190 (0.0071)***	0.0619 (0.0264)**	0.0238 (0.0224)	0.0000 (0.0014)
Charter school	-0.0315 (0.0104)***	-0.1177 (0.0342)***	-0.0354 (0.0479)	-0.0001 (0.0012)
Grade B school	-0.0131 (0.0043)***	-0.0443 (0.0166)***	-0.0160 (0.0159)	-0.0000 (0.0008)
Grade C school	-0.0132 (0.0081)*	-0.0460 (0.0273)*	-0.0158 (0.0306)	-0.0000 (0.0009)
Grade D school	-0.0625 (0.0259)***	-0.2649 (0.0680)***	-0.0620 (0.1651)	-0.0001 (0.0018)
Percent lunch program	0.0006 (0.0266)	0.0019 (0.1045)	0.0007 (0.0494)	0.0000 (0.0026)
Percent black	0.0254 (0.0317)	0.0845 (0.1257)	0.0311 (0.0563)	0.0001 (0.0035)
Percent Hispanic	-0.1110 (0.0858)	-0.3701 (0.3308)	-0.1362 (0.1623)	-0.0003 (0.0088)
Math	0.0036 (0.0048)	0.0120 (0.0192)	0.0045 (0.0102)	0.0000 (0.0005)
Science	0.0088 (0.0048)	0.0286 (0.0185)	0.0111 (0.0094)	0.0000 (0.0004)
Social Studies	-0.0043 (0.0055)	-0.0144 (0.0222)	-0.0052 (0.0091)	-0.0000 (0.0005)
Art/Gym/Music	-0.0244 (0.0060)***	-0.0876 (0.0211)***	-0.0284 (0.0335)	-0.0000 (0.0012)
Special education	-0.0172 (0.0065)***	-0.0599 (0.0229)***	-0.0204 (0.0260)	-0.0000 (0.0010)
Teacher's list	0.3660 (0.0358)***	0.4019 (0.0374)***	0.6502 (0.0480)***	0.0144 (0.0552)
School's list	0.2077 (0.0170)***	0.3563 (0.0416)***	0.3415 (0.0474)***	0.0012 (0.0183)
Friday job	-0.0217 (0.0050)***	-0.0752 (0.0185)***	-0.0259 (0.0133)*	-0.0000 (0.0005)
MP Location	3 MP	0.12	-1.74	-5.06
Location Probability	.	0.12	0.48	0.40
Avg. Predicted A	0.11 (0.0098)***	0.59 (0.0461)***	0.09 (0.0485)*	0.0001 (0.0020)

Notes: The model's coefficients are reported in table A1. Standard errors are based on 50 bootstrap replications. ***, **, and * indicate statistical significance at 1, 5, and 10 percent. The three mass-point probit model was fit using the GLLAMM Stata package (Rabe-Hesketh et al., 2002). The likelihood function of a four mass-point model did not converge.

APPENDIX 1.2

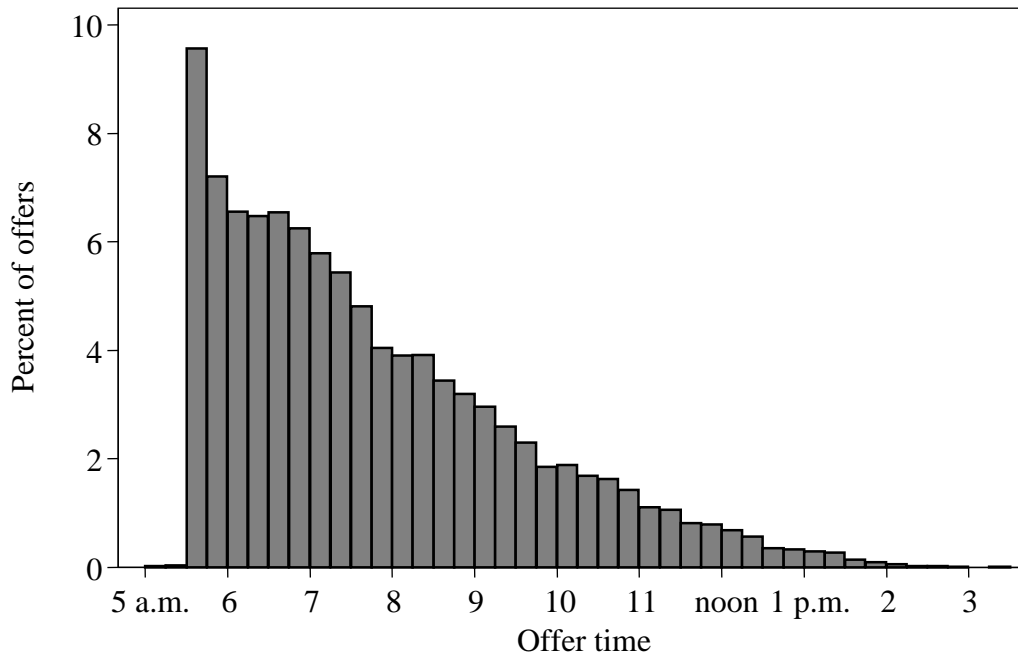
CHAPTER 1 FIGURES

Figure 1.1: Job-length distribution



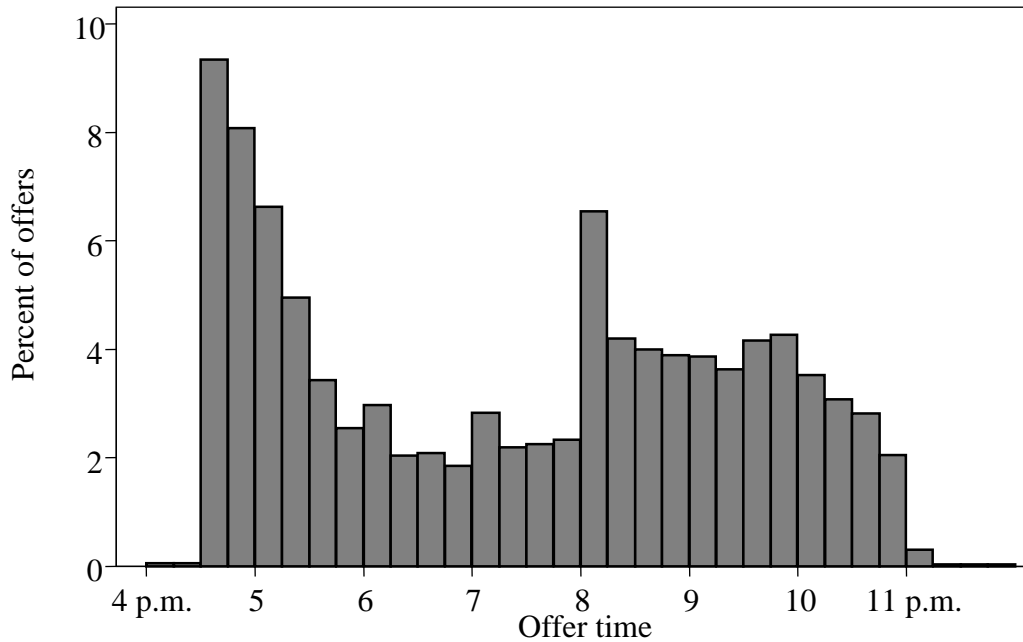
Notes: Bins are 20 minutes wide. The vertical line indicates the half-full cutoff.

Figure 1.2a: Day of-offer time distribution

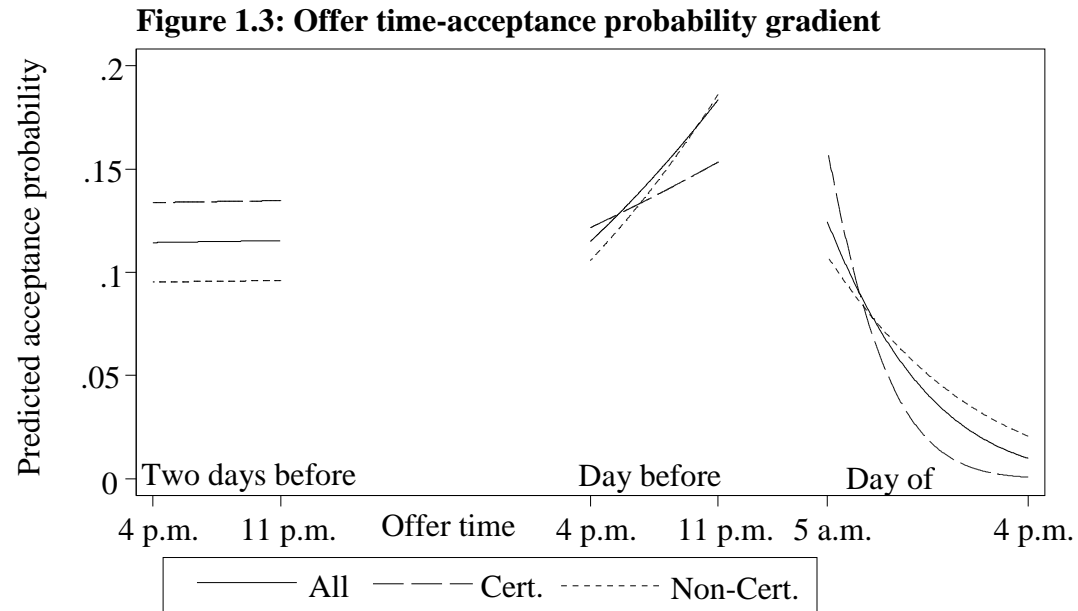


Note: Bins are 15 minutes wide.

Figure 1.2b: Day before-offer time distribution



Note: Bins are 15 minutes wide.



Note: The gradients were computed using the RE-probit coefficients reported in columns 1, 3, and 4 of table 1.4.

APPENDIX 1.3

AVERAGE PARTIAL EFFECT (APE) DEFINITIONS

Let N represent the total number of offers (observations). In the baseline model, which assumes that the sub-day random effect ω_{sd} is $\sim \text{Normal}(0, \sigma_\omega^2)$, the APE of a continuous variable k is

$$\text{APE}_k = \frac{\gamma_k}{N(1 + \sigma_\omega^2)^{0.5}} \sum_{sdt=1}^N \phi \left(\frac{\gamma \mathbf{x}_{sdt} + \lambda \mathbf{z}_s + \delta \mathbf{d}_d + \mathbf{b} \mathbf{t}_{sdt}}{(1 + \sigma_\omega^2)^{0.5}} \right) \quad (\text{A1.1})$$

and the APE of a binary variable k is

$$\text{APE}_k = N^{-1} \sum_{sdt=1}^N \left\{ \Phi \left(\frac{\gamma \mathbf{x}_{sdt} + \gamma_k (1 - x_{sdt}^k) + \lambda \mathbf{z}_s + \delta \mathbf{d}_d + \mathbf{b} \mathbf{t}_{sdt}}{(1 + \sigma_\omega^2)^{0.5}} \right) - \Phi \left(\frac{\gamma \mathbf{x}_{sdt} - \gamma_k x_{sdt}^k + \lambda \mathbf{z}_s + \delta \mathbf{d}_d + \mathbf{b} \mathbf{t}_{sdt}}{(1 + \sigma_\omega^2)^{0.5}} \right) \right\}. \quad (\text{A1.2})$$

For the half-day APE, equation (A1.2) is modified as follows: For the first CDF in A1.2, if the offered job was a half-day, the vector \mathbf{x}_{sdt} is left as is. If a full-day job was offered, in addition to adding γ_{half} , H_{sdt} is changed to 3.4, the mean half-day job length. Similarly, for the second CDF in (A1.2), if a full-day job was offered the vector \mathbf{x}_{sdt} is left as is. If a half-day job was offered, γ_{half} is subtracted, and H_{sdt} is changed to the full-day mean job length (7.2).

In the three-mass point model, ω_{sd} takes the values ω_1 , ω_2 , and ω_3 with probabilities π_1 , π_2 , and π_3 , respectively. The APE of a continuous variable k , at mass point j , is

$$\text{APE}_{k,j} = N^{-1} \gamma_k \sum_{sdt=1}^N \phi \left(\gamma \mathbf{x}_{sdt} + \lambda \mathbf{z}_s + \delta \mathbf{d}_d + \mathbf{b} \mathbf{t}_{sdt} + \omega_j \right). \quad (\text{A1.3})$$

The APE of binary variables, and the adjustment for the half-day APE, are computed in similar fashion. The overall APE of variable k is simply the weighted average of the mass point-specific APE:

$$\text{APE}_k = \sum_{j=1}^3 \pi_j \text{APE}_{k,j}. \quad (\text{A1.4})$$

APPENDIX 1.4

MASS-POINT PROBIT COEFFICIENTS

Table A1: Mass-point Probit Coefficients

Half day	0.3248 (0.2102)
Half-day*Hours	-0.0817 (0.0451)*
Hours	0.0581 (0.0274)**
One-way commute <i>h</i>	-0.5414 (0.1941)***
Rural district	0.0041 (0.0515)
Elementary school	-0.0507 (0.0538)
High school	0.1812 (0.0619)***
Charter school	-0.3356 (0.1171)***
Student-Teacher ratio	0.0165 (0.0084)*
Grade B	-0.1291 (0.0476)***
Grade C	-0.1328 (0.0743)*
Grade D	-0.7756 (0.3461)**
% lunch program	0.0055 (0.2472)
% black	0.2460 (0.3081)
% Hispanic	-1.0780 (0.8961)
Pre-K/Kindergarten	-0.0285 (0.1031)
First/Second Grade	0.0858 (0.0853)
Third/Fourth Grade	0.0432 (0.0889)
Fifth/Sixth Grade	0.0310 (0.0812)
Sev./Eighth Grade	0.1569 (0.1193)
Math	0.0350 (0.0825)

Table A1, Continued

Science	0.0840 (0.0759)
Social Studies	-0.0419 (0.0874)
Art/Gym/Music	-0.2515 (0.0721)***
Business/Technology	-0.0779 (0.0752)
Foreign Language	-0.0465 (0.0884)
Special Education	-0.1729 (0.0777)**
Other	-0.2036 (0.1008)**
Certified	0.3629 (0.1221)***
Male	0.0453 (0.1112)
Teacher's List	2.3162 (0.2312)***
School's List	1.4084 (0.1193)***
% Teacher's Lists	-0.5942 (1.4940)
% School's Lists	-0.0819 (0.5900)
Friday job	-0.2168 (0.0462)***
Time of call ($T - t$)	-0.0006 (0.0001)***
Day-of offer	-1.1166 (0.1236)***
Day-before offer	0.8010 (0.2185)***
Day of*($T - t$)	0.0992 (0.0122)***
Day before*($T - t$)	-0.0340 (0.0095)***
Observations	94,106
Sub-days (RE)	32,228
Substitutes (clusters)	771
Log likelihood	-21,587

Table A1, Continued

Notes: Standard errors, in parentheses, are robust to clustering at the substitute level. ***, **, and * indicate statistical significance at 1, 5, and 10 percent. Coefficients for all covariates are reported with the exception of month dummies and day-of-week dummies other than Friday. The model was fit using the GLLAMM package (Rabe-Hesketh et al., 2002).

CHAPTER 1 REFERENCES

CHAPTER 1 REFERENCES

- Bucior, C. 2010. The Replacements. *New York Times*, January 2.
- Butler, J.S., and R. Moffitt. 1982. A computationally efficient quadrature procedure for the One Factor Multinomial Probit Model. *Econometrica* 50(3): 761-764.
- Cameron, A.C., J.B. Gelbach, and D.L. Miller. 2006. Robust inference with multi-way clustering. NBER Technical Working Paper No. 327.
- Cameron, A.C., and P.K. Trivedi. 2005. *Microeconometrics: Methods and Applications*, New York, NY: Cambridge Univ. Press.
- Card, D., and A. 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): 1-40.
- Clotfelter, C. F., H. Ladd, and J. Vigdor. 2009. Are teacher absences worth worrying about in the U.S.? *Education Finance and Policy* 4(2): 115-149.
- Coverdill, J., and P. Oulevey. 2007. Getting contingent work: Insights into on-call work, matching processes, and staffing technology from a study of substitute teachers. *Sociological Quarterly* 48(3): 533-557.
- Das, J., S. Dercon, J. Habyarimana, and P. Krishnan. 2007. Teacher shocks and student learning: Evidence from Zambia. *Journal of Human Resources* 42(4): 820-862.
- Devine, T. J., and N. M. Kiefer. 1991. *Empirical Labor Economics: The Search Approach*. Oxford: Oxford Univ. Press.
- Dolton, P. 2006. Teacher supply. In *Handbook of the Economics of Education*, vol. 2, ed. E. Hanushek and F. Welch, 1079-1161. Amsterdam: North-Holland.
- Dorward, J., A. Hawkins, and G. Smith. 2000. Substitute teacher availability, pay, and influence on teacher professional development: A national survey. *ERS Spectrum* Summer: 40-46.
- Ehrenberg, R. G., R. A. Ehrenberg, D. Rees, and E. Ehrenberg. 1991. School district leave policies, teacher absenteeism, and student achievement. *Journal of Human Resources* 26(1): 72-105.
- Goette, L., D. Huffman, and E. Fehr. 2004. Loss aversion and labor supply. *Journal of the European Economic Association* 2(2-3): 216-228.

- Hanushek, E., J. Kain, and S.G. Rivkin. 2004. Why public schools lose teachers? *Journal of Human Resources* 39(2): 326-354.
- Hanushek, E., and S.G. Rivkin. 2006. Teacher quality. In *Handbook of the Economics of Education*, vol. 2, ed. E. Hanushek and F. Welch., 1019-1078. Amsterdam: North-Holland.
- Hanushek, E., and L. Woessmann. 2008. The role of cognitive skills in economic development. *Journal of Economic Literature* 46(3): 607-668.
- Heckman, J. J., and B. Singer. 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52(2): 271-320.
- Henderson, E., N. Protheroe, and S. Porch. 2002. *Developing an Effective Substitute Teacher Program*. Arlington, VA: Educational Research Service.
- Jacobson, S. 1988. The effects of pay incentives on teacher absenteeism. *Journal of Human Resources* 24(2): 280-287.
- Kiefer, N., and G. Neumann. 1981. Individual effects in a nonlinear model: Explicit treatment of heterogeneity in the empirical job-search model. *Econometrica* 49(4): 965-979.
- Layton, J. 2005. How MapQuest works. *HowStuffWorks.com*.
<http://www.howstuffworks.com/mapquest.htm> (accessed March 27, 2011).
- MDE. See Michigan Department of Education.
- Michigan Department of Education: Office of Educational Assessment and Accountability. 2007. *Guide to Reading the Michigan School Report Cards (2007 Edition)*.
- Miller, R., R. Murnane, and J. Willett. 2008a. Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. *Educational Evaluation and Policy Analysis* 30(2): 181-200.
- . 2008b. Do worker absences affect productivity? The case of teachers. *International Labour Review* 147(1): 71-89.
- Mortensen, D. 1986. Job search and labor market analysis. In *Handbook of Labor Economics*, vol. 2, ed. O. Ashenfelter and R. Layard, 849-919. Amsterdam: North-Holland.
- Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2002. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* 2: 1-21.
- . 2005. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 128(2): 301-323.
- Rogers, J. 2001. There's no substitute. *Work and Occupations* 28(1): 64-90.

Roza, M. 2007. Frozen assets: Rethinking teacher contracts could free billions for school reform. January 2007 Education Sector Reports.

Strauss, R. 2003. *The Market for Substitute Classroom Teachers in South-West Pennsylvania in 2001-2002*. Pittsburgh, PA: The Pittsburgh Foundation.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd Ed. Cambridge, MA: MIT Press.

CHAPTER 2

GOING THE EXTRA MILE

2.1 Introduction

The cost of commuting influences a variety of economic decisions. It is a fundamental parameter in urban-economic spatial models of firm and household location (Muth, 1969) and is central to cost-benefit analyses of proposed transportation-infrastructure investments (Small & Verhoef, 2007, p.181).¹⁴ A typical goal of the latter is to spur economic development in suburban and rural areas by decreasing the commute time to jobs in neighboring cities (So et al., 2001). The effect of commuting on labor supply enters firms' hiring decisions as well by shaping the optimal "spatial search radius" over which to recruit (Russo et al., 1996). Finally, both explicit and implicit commuting costs are fixed costs of working that may influence labor-force participation (Cogan, 1981). Black et al. (2010), for example, find significantly lower labor-force participation rates among married women in cities that have longer-than-average commutes. Identifying the effect of commuting on labor supply is complicated by a fundamental endogeneity problem, however: commute time is jointly determined by individuals' job and residence choices.¹⁵

I estimate the causal effect of commute time on daily labor supply by studying a unique labor market in which workers are subject to daily exogenous variation in commute time and are not constrained in their daily labor supply decisions.¹⁶ To motivate my approach, consider how

¹⁴ Muth (1969) is a classic text on urban-economic spatial models, in which the marginal cost of commuting determines the size and shape of cities (p. 90-93), the wage-commute gradient, and the housing price-commute gradient (p. 71). The spatial model may be out of equilibrium, however; Stutzer and Frey (2008) find that individuals with longer commutes systematically report being unhappier than those with shorter commutes.

¹⁵ The existence of two-worker households increases the problem's complexity.

¹⁶ This approach is very much in the spirit of the work on intertemporal labor supply that, for similar reasons, focuses on the cab-driver and stadium-vendor labor markets (Camerer et al., 1997; Oettinger, 1999).

the causal effect of commuting on daily labor supply would be identified experimentally. We would begin by holding each subject's residential location and transportation mode fixed. Then, on a daily basis, individuals would be asked to choose between accepting a job at a randomly determined location and not working. If all jobs are identical aside from location, it is straightforward to use the observed decisions to estimate the marginal effect of commute time on daily labor supply.

One particular substitute-teacher labor market is similar to this hypothetical ideal experiment. Each day, a consortium comprised of ten Michigan school districts and over 75 schools makes hundreds of take-it-or-leave-it job offers to substitute teachers via an automated calling system. Importantly, the call system makes offers in a conditionally random order, which generates exogenous variation in offer quality and commute time across substitute teachers. The call-system's randomness solves the usual endogeneity problem of commute times being correlated with individuals' unobserved tastes. Unlike in the ideal experiment, however, all jobs and schools (locations) are not identical. These confounding factors must be "partialled out" by controlling for a variety of job characteristics and school fixed effects.

The empirics are based on an optimal decision rule that is motivated by a job-search model of substitute teachers' expected utility maximization. I use data on accepted and rejected offers to estimate sequential binary-choice models of substitutes' offer acceptance decisions. The main results suggest that a fifteen-minute increase in one-way commute time decreases the acceptance probability by three percentage points and that the elasticity of the acceptance probability with respect to commute time is about -0.4.

I also investigate whether certain day and individual characteristics influence the disutility of commuting. On average, the negative effect of commute time is about 36% larger

when the 6:00 a.m. temperature is below 20 degrees Fahrenheit but rainfall over the past 24 hours is not associated with commuting preferences. Fuel prices appear to increase the cost of commuting, but this effect is imprecisely estimated. Similarly, both women and substitutes who are certified as regular teachers tend to have a larger, but imprecisely estimated, aversion to commuting. Estimating the model separately for men and women, however, yields the interesting results that women are significantly more averse to commuting in cold temperatures and are significantly more responsive to changes in fuel prices than men: the negative effect of commute time for women is over 50% larger on frigid days and a one-dollar increase in the price per gallon of fuel increases women's negative effect of commute time by 44%.

2.2 Literature Review

There are both explicit and implicit private costs of commuting.¹⁷ There are two types of explicit commuting costs. The first is monetary: the American Automobile Association (AAA) reports average vehicle costs of \$0.42 to \$0.66 per mile, about \$0.10 of which is for fuel (AAA, 2009). The second includes potential physical and mental-health costs of commuting (Koslowsky et al. 1995). The primary implicit cost is forgone time: the average one-way commute in the US in 2004 was about 25 minutes, up from about 20 minutes in 1980 (Pisarski, 2006).

Two recent studies have directly investigated the effect of commuting on labor supply. Using panel-data methods, Gutiérrez-i-Puigarnau and Van Ommeren (2010) find that long-commute German workers work fewer but longer days per week than individuals with shorter commutes, but find no difference between the groups in total weekly hours. Similarly, applying

¹⁷ Though not relevant here, there are also public (external) costs of commuting (e.g. Lemp & Kockelman, 2008).

an instrumental-variables procedure to Spanish time-use data, Gimenez-Nadal and Molina (2011) find that an extra hour of commute time leads to a 35 minute increase in the length of the workday. Both of these studies are subject to the criticism that workers are likely to be constrained in their labor supply choices, however (Dickens & Lundberg, 1993; Kahn & Lang, 1991).

Rather than estimating the effect of commute time on outcomes such as labor supply, the early empirical commuting literature used discrete-choice models of commuters' transportation-mode choices to estimate the willingness to pay per hour of commute time (WTP). On average, these studies find the cost of an hour of commute time to be about 50% of the hourly wage (Small & Verhoef, 2007, p. 52). A well-documented problem with this method is the implicit assumption that time spent travelling in one mode (e.g., a car) is equivalent to time spent traveling in another (e.g., a bus).

Stated-preference survey data, in which respondents rank or choose from a hypothetical set of commute-wage bundles, has been proposed as a solution to the "comparability" problem inherent in the transportation-mode choice literature mentioned above. Calfee et al. (2001) and Calfee and Winston (1988) evaluate stated-preference data using various econometric methods and find a significantly lower WTP of about 20 percent of the hourly wage. However, experimentalists have repeatedly found a "hypothetical bias" in answers to subjective and hypothetical questions, questioning the validity of estimates based on stated-preference data (Harrison, 2006).

A third approach to estimating WTP employs structural job-search models that treat job offers as wage-commute bundles. For example, Van Ommeren et al. (2000) find a WTP of about 50% of the hourly wage. Van Ommeren and Fosgerau (2009) extend the basic search model to

incorporate job-switching behavior and find that the total commuting cost, including both time and monetary costs, is 200% of the hourly wage.

Commuting costs, particularly the opportunity cost of time, potentially vary across both observed and unobserved individual attributes. So et al. (2001) find that commuters, relative to non-commuters, earn higher wages, are younger, and are disproportionately male. One possible explanation of the latter is that women must stay closer to home because they are active in home production. This view is supported by Van den Berg and Gorter (1997) who find that women with children have significantly higher WTP than those without children, but no significant difference in WTP between men and childless women. Inclement weather might also increase the marginal cost of commuting for a variety of reasons. Snow, for example, has been shown to decrease commuters' welfare by decreasing travel speed (Sabir et al., 2010).

Despite not estimating a formal WTP, I contribute to the existing literature on commuting preferences in several ways. First, to the best of my knowledge this is the first paper to exploit arguably exogenous variation in the actual commute times faced by workers making labor-supply decisions in real time. Thus I am able to estimate the effect of commute time on labor supply without making the stronger assumptions required by the fixed-effects and instrumental-variables estimators used in previous work. Second, my analysis is immune to the criticism that commute time-labor supply elasticities are attenuated because substitute teachers are unconstrained in their daily labor supply decisions.¹⁸ Third, I am able to make inroads on the longstanding question of whether there are gender-specific commuting preferences because the call system's randomness eliminates the confounding dual worker-household problem. Finally,

¹⁸ Again, in this regard, the substitute-teacher labor market is similar to the cab-driver and stadium-vendor labor markets studied by Camerer et al. (1997) and Oettinger (1999).

the presence of daily variation in commute times allows me to investigate the role that cost shifters such as fuel prices and inclement weather play in commuting decisions.

2.3 Labor-Market Environment and Data

2.3.1 The Intermediate School District

I investigate the daily labor supply decisions of substitute teachers in a consortium of ten adjacent and autonomous school districts in Michigan. The consortium consists of over 75 schools located across approximately 600 square miles. Substitute teachers live both within and outside the consortium. Membership in the consortium enables districts to enjoy economies of scale in training substitute teachers and in operating an automated calling system. The calling system is used to satisfy regular teachers' requests for substitute teachers who were not filled personally. The subsequent analysis focuses solely on jobs filled by the automated calling system, accounting for about half of the consortium's annual teacher absences.

At any time prior to the start of a job a regular teacher may request a substitute through the automated calling system. After the regular teacher has specified the job's characteristics the calling system sequentially offers the job to available substitute teachers until either the job is accepted or it begins. Job offers are made over the phone for same-day jobs beginning at 5:00 a.m. and for jobs one or more days in the future between 4:00 p.m. and 11:00 p.m. each day of the week. Substitutes are not penalized for rejecting an offer, but once a substitute accepts a job he or she will not receive any conflicting offers. Additionally, substitute teachers are prohibited from returning to previously-rejected offers.

Substitutes receive offers from all consortium schools and are called in a conditionally random order. There are two conditioning variables: the substitutes' regular-teacher certification

status and “preferred-list status.” Each regular teacher and school enters a list of “preferred substitutes” in the system. The phone calls made to available substitute teachers state the start and end time, teacher name, subject, and school of the job being offered. The job’s wage is not explicitly stated because it is a known function of job length. Daily pay in this labor market, for all substitutes and for all schools, is binary: half days pay \$40 and full days pay \$75, where half days are jobs lasting less than four hours and 21 minutes. Job length is ultimately at the discretion of the regular teacher making the request but also influenced by school- and subject-specific schedules.

2.3.2 *Data*

The primary labor-supply data comes directly from the automated calling system’s computer and includes every offer made during the 2006-07 school year. Over 100,000 offers regarding nearly 9,000 jobs (unique substitute requests) were made. Remarkably, 98% of these jobs were successfully filled. In addition to the job attributes mentioned above, I observe the day and time at which each offer was made, whether or not the offer was accepted, and a unique identifier of each substitute along with his or her certification status, preferred-list status, gender, and home zip code.¹⁹

Measures of commute time and distance from the center of each substitute’s home zip code to each school’s street address were computed using MapQuest.com.²⁰ MapQuest uses geocoding technology to assign approximate latitude-longitude coordinates to each school’s

¹⁹ Age is only observed for about 70% of the substitutes, so is not used in the analysis.

²⁰ The use of centroids was necessitated by privacy requirements that prevent access to the substitutes’ home addresses. Commutes, therefore, are measured with error because substitutes can live anywhere within the zip code. Implications of this measurement error are discussed in section 2.5.2.

street address and to the centroid of each five-digit zip code. An algorithm then searches a database of roadmaps and evaluates potential routes. The algorithm chooses an optimal route based on driving distance, posted speed limits, the number of left-hand turns, and the number of intersections. The travel distance (in miles) and estimated travel time for the optimal route are then reported. See Layton (2005) for additional details and references.

I further augment the call-system data with information on daily weather and fuel-prices, both of which potentially influence the marginal cost of commuting. To account for inclement weather I use daily measures of rainfall and temperature from the U.S. National Climatic Data Center's "Land Surface Data."²¹ While snowfall is likely the most important weather-related shifter of commute costs for the general labor force (Sabir et al., 2010), it is of little interest in the present context because schools in the consortium close when winter weather creates hazardous driving conditions. To control for fuel costs, which represent as much as one quarter of per-mile vehicle costs (AAA, 2009), I use county-level daily average fuel prices that are based on daily samplings of about 100 gas stations located in the consortium's MSA.²²

2.3.3 *Descriptive Statistics*

Table 2.1 provides summary statistics of one-way commutes and some other offer characteristics. The average offered one-way commute was about 13 miles or 18 minutes, which is slightly shorter than the U.S. national average of about 25 minutes (Pisarski, 2006). About 16% of offers were made to substitutes residing within the offering district. Of the offer

²¹ The "Land Surface Data" is collected daily at 6 a.m. by a CO-OP station in the center of the consortium and is publicly available from the National Climatic Data Center at <http://www.ncdc.noaa.gov/oa/climate/stationlocator.html>.

²² This is proprietary data that was purchased from the private market-research firm Oil Price Information Service (OPIS).

recipients, 34% were male and 24% were certified as regular teachers. The average accepted commute was about 2.5 miles and 3 minutes shorter than the average rejected commute, suggesting that longer commutes were less likely to be accepted regardless of whether commutes are measured in time or distance. In the subsequent analysis I focus only on commute times because the time and distance measures in my data are highly correlated and produce nearly identical estimates of the elasticity of the offer-acceptance probability with respect to commute length.²³ Similarly, on average substitutes residing within the offering district are eight percentage points more likely to accept.²⁴

Figure 2.1 depicts the distributions of offered, accepted, and rejected commute times. The majority of offered commutes are shorter than 30 minutes. Comparing the kernel density estimates of the accepted-offer and rejected-offer distributions again suggests that accepted offers tend to be associated with shorter commutes.

Some substitutes' home zip codes provided in the data suggest one-way commute times longer than two hours, raising the concern that some zip codes are incorrect. I drop suspect zip codes from the subsequent analysis as follows. First, using a zip-code map of the area, I retain all zip codes contiguous to the consortium. Second, I retain all zip codes containing at least one active substitute that are contiguous to the area defined in step 1. I repeat step 2, retaining all zip codes that are contiguous to the area defined in the previous step *and* contain at least one active substitute teacher, until the region is encapsulated by a ring of zip codes containing no active

²³ Van Ommeren and Fosgerau (2009) note that commute time and commute distance are typically not equivalent measures and discuss the relative merits of each. The correlation coefficient is 0.96 in my data, however, which likely results from numerous accessible highways and a general lack of traffic congestion in the consortium.

²⁴ This may have to do with preferences for the neighborhood school rather than commute time, however, a possibility that is investigated in the sensitivity analysis of section 2.5.2.

substitutes. The result is a contiguous region of 48 zip codes, 11 of which are located within the consortium.

Figure 2.2 plots rainfall over the 24-hour period ending at 6:00 a.m. and the daily temperature at 6:00 a.m. for each school day. Rainfall is only reported for days when at least one consortium school was open. Figure 2.3 plots the county-level average daily price per gallon of regular unleaded fuel over the course of the school year. Fuel prices decrease in the month of September and remain relatively stable until the end of December. Fuel prices then fall below \$2.00 in January before steadily increasing over the remainder of the school year.

2.4 Econometric Model

2.4.1 Optimal Decision Rule

This section draws heavily upon the model developed in section 1.4 of this volume. For additional details, the interested reader is referred to section 1.4. I assume that substitute teachers maximize expected utility when deciding whether to accept or reject an offer, which in this case is accomplished by following a reservation-utility decision rule: accept if and only if the utility of accepting (U^A) exceeds the expected utility of rejecting (U^R).²⁵ U^A is a function of the offer's and recipient's characteristics. U^R is a function of both the recipient's non-subbing alternative (U^N) and expectations of future offers.

Let T be the last time that an offer to work on a particular day can be made. Rejecting an offer at time T , therefore, is equivalent to choosing the non-subbing alternative on that day; this

²⁵ The functioning of the automated calling system is essentially a finite-horizon job-search model with no recall and no on-the-job search, a la Mortensen (1986).

implies that $U_T^R = U^N$. For all t less than T , U_t^R can be approximated by the sum of U^N and a nonnegative, monotonically-decreasing function of offer time.

Substitute teachers' daily utility is assumed to take the same functional form whether substitute teaching, working elsewhere, or not working at all. Daily utility is a function of non-labor income (Y), labor income (M), hours worked (H), commuting costs (C), and a variety of individual, day, and non-wage job characteristics that are both observed and unobserved (ψ). Formally, let daily utility take the form

$$U = f(Y) + \beta M - g(H) - C + \psi. \quad (2.1)$$

Taking a first-order approximation of g , so H enters the utility function linearly as an observed job characteristic, the utility accruing to substitute s of accepting job j on day d at time t is

$$U_{sdjt}^A = f(Y_{sd}) + \gamma^A \mathbf{x}_j - C_{sdjt} + \lambda^A \mathbf{z}_s + \mathbf{b}^A \mathbf{w}_{sdjt} + \delta^A \mathbf{r}_d + \omega_{sd}^A + \varepsilon_{sdjt}, \quad (2.2)$$

where \mathbf{x}_j is a vector of observed job characteristics including job length, daily pay, and full sets of subject and school dummies; \mathbf{z}_s is a vector of observed substitute characteristics including gender, certification status, and preferred-list status; \mathbf{w}_{sdjt} is a vector of offer-time variables that is a piecewise linear function in $(T - t)$; \mathbf{r}_d is a vector of day-of-job variables including rainfall,

²⁶ Daily pay (M) is valued linearly because there is approximately no income effect of a small change to lifetime earnings (Goette et al., 2004).

temperature, average fuel price, and full sets of day-of-week and month dummies; ω_{sd}^A is the substitute's unobserved day-specific taste for substitute teaching; and ε_{sdjt} is an offer-specific error term that captures job attributes that are unobserved by the econometrician (i.e. within-school, within-subject variation in classroom quality), substitutes' mood and attention level at the time of offer, and substitutes' preferences for specific jobs or schools.²⁷

The non-subbing utility U_{sd}^N , which can be interpreted as a substitute's opportunity cost of subbing on day d , depends on observed individual and day characteristics as well as unobserved sub-day specific non-subbing opportunities (ω_{sd}^N). Formally,

$$U_{sd}^N = f(Y_{sd}) + \lambda^N \mathbf{z}_s + \delta^N \mathbf{r}_d + \omega_{sd}^N. \quad (2.3)$$

Combining equations (2.2) and (2.3) with the reservation-utility decision rule yields the probability that an offer will be accepted conditional on it being received:

²⁷ Because daily pay is binary, M will be replaced by a half-day indicator. Offer time enters U^A because it might proxy for offer quality in several ways. First, the distribution of offers might worsen over time. Second, a late-arriving offer might indicate that the regular teacher made the request late in the morning and therefore did not have time to prepare a lesson plan for the substitute teacher or prepare students for the absence. Third, the amount of time the substitute has to prepare for the job might be an important measure of offer quality. Day-of-week variables enter U^A because they may contain information on job quality via student behavior. For example, students may behave differently on rainy or warm days, Fridays, and towards the end of the school year. The implications of school-specific tastes among substitutes are considered and tested for in the sensitivity analysis of section 2.5.2.

$$\begin{aligned}
& \Pr(A_{sdjt} = 1 | p_{sdt} = 1, \mathbf{x}_j, C_{sdjt}, \mathbf{z}_s, \mathbf{w}_{sdjt}, \mathbf{r}_d, \omega_{sd}) \\
& = \Pr(\gamma^A \mathbf{x}_j - C_{sdjt} + \boldsymbol{\lambda} \mathbf{z}_s + \mathbf{b} \mathbf{w}_{sdjt} + \boldsymbol{\delta} \mathbf{r}_d + \omega_{sd} + \varepsilon_{sdjt} > 0 | p_{sdt} = 1),
\end{aligned} \tag{2.4}$$

where A_{sdjt} is a binary indicator of offer acceptance and p_{sdt} is a binary indicator of substitute s having received a day- d offer at time t . The lack of superscripts on the coefficients of \mathbf{z} , \mathbf{w} , and \mathbf{r} and on the unobserved sub-day effect is notational, indicating that only the net effect of these variables on the acceptance probability are identified; specifically, $\boldsymbol{\lambda} = \boldsymbol{\lambda}^A - \boldsymbol{\lambda}^N$, $\mathbf{b} = \mathbf{b}^A - \mathbf{b}^R$, $\boldsymbol{\delta} = \boldsymbol{\delta}^A - \boldsymbol{\delta}^N$, and $\omega = \omega^A - \omega^N$. The cost of commuting, C_{sdjt} , will be approximated in the empirics by both linear and quadratic functions of commute time and commute time interacted with elements of \mathbf{z} and \mathbf{r} . A final comment regarding equation (2.4) is that non-labor income was differenced out because it is valued identically in both U^A and U^N .²⁸

2.4.2 Estimation

The typical sample-selection problem caused by a lack of data on rejected offers is absent here because all offers, accepted and rejected, are observed. However, the call-system data is a selected sample in the sense that offer-acceptance decisions are only observed when an offer was made (i.e., when $p_{sdt} = 1$). Substitutes who work on day d will, on average, receive fewer offers and have higher values of ω_{sd} than those that do not work on day d because substitutes do not

²⁸ Intuitively, this is a result of consumption smoothing over the lifecycle and preferences that are separable in consumption and leisure. The assumption that non-labor income is valued differently on subbing and non-subbing days can be relaxed entirely by noting that any difference in utility would be sub-day specific and hence incorporated in ω_{sd} .

receive offers that conflict with previously accepted jobs. The resulting negative correlation between ω_{sd} and p_{sdt} implies that pooled estimators of (2.4) that fail to account for the presence of ω_{sd} are inconsistent.

Conditional on \mathbf{z}_s and ω_{sd} , however, the offer-specific error term ε_{sdjt} is independent of the selection indicator. This is a direct result of the call system's randomness. Accordingly, conditional on \mathbf{z}_s and ω_{sd} , missing observations (time periods in which no offer is received) can be considered "missing at random" (Cameron & Trivedi, 2005, p. 926) and the conditioning on an offer being received can be removed from the RHS of equation (2.4). This solution to the problem of unbalanced panels in a nonlinear model is similar in spirit to Kiefer and Neumann (1981).

The baseline model imposes the following assumptions:

$$\left(\varepsilon_{sdjt} \mid p_{sdt} = 1, \mathbf{x}_j, C_{sdjt}, \mathbf{z}_s, \mathbf{w}_{sdjt}, \mathbf{r}_d, \omega_{sd} \right) = \left(\varepsilon_{sdjt} \mid \mathbf{x}_j, C_{sdjt}, \mathbf{z}_s, \mathbf{w}_{sdjt}, \mathbf{r}_d, \omega_{sd} \right) \sim N(0, 1) \quad (2.5a)$$

$$\text{and} \quad \left(\omega_{sd} \mid p_{sdt} = 1, \mathbf{x}_j, C_{sdjt}, \mathbf{z}_s, \mathbf{w}_{sdjt}, \mathbf{r}_d \right) = \left(\omega_{sd} \mid p_{sdt} = 1, \mathbf{z}_s, \mathbf{r}_d \right) \sim N\left(0, \sigma_\omega^2\right). \quad (2.5b)$$

Assumption (2.5a) reflects that ε_{sdjt} is independent of p_{sdt} , conditional on ω_{sd} , as discussed above. Assumption (2.5b) is a direct result of the call system's conditional randomness. Under these assumptions equation (2.4) can be rewritten as

$$\begin{aligned}
& \Pr\left(\gamma^A \mathbf{x}_j - C_{sdjt} + \lambda \mathbf{z}_s + \mathbf{b} \mathbf{w}_{sdjt} + \delta \mathbf{r}_d + \omega_{sd} + \varepsilon_{sdjt} > 0 \mid p_{sdt} = 1\right) \\
& = \Phi\left(\gamma^A \mathbf{x}_j - C_{sdjt} + \lambda \mathbf{z}_s + \mathbf{b} \mathbf{w}_{sdjt} + \delta \mathbf{r}_d + \omega_{sd}\right),
\end{aligned} \tag{2.6}$$

and estimated using the random-effects (RE) probit procedure of Butler and Moffitt (1982).

2.5 Results

2.5.1 Main Results

Table 2.2 reports the estimated RE-probit coefficients, average partial effects (APE), and elasticities of commute time. APE and elasticities for the RE-probit model are defined in appendix 2.3. The APE and elasticity standard errors were computed by taking the standard deviation of the estimates from 50 bootstrap replications. The bootstrap procedure resampled with replacement at the substitute level, utilizing all observations from the chosen substitute. Resampling at the substitute level produces standard errors that are robust to substitute-level clustering and that are asymptotically equivalent to the usual robust “sandwich” standard error estimates (Cameron & Trivedi, 2002). Clustering at the substitute level allows for individuals’ opportunity costs (ω_{sd}) to be correlated across days. Omitted from table 2.2, but included in its regressions, are the substitute, day, and job characteristics described in section 2.4; the full set of coefficient estimates is reported in table A2.

Column 1 assumes that commute time enters the model linearly. The commute-time coefficient is negative and strongly statistically significant. The APE indicates that a fifteen-minute increase in one-way commute time lowers the acceptance probability by about three percentage points. In elasticity terms, a ten percent increase in commute time lowers the acceptance probability by about four percent. Allowing for a quadratic in commute time does

not lead to a meaningful change in the estimated APE or elasticity of commute time. Because the quadratic-term coefficient is statistically insignificant and a likelihood-ratio test fails to reject the null that the quadratic model does not provide a better fit of the data, I subsequently treat column 1 as the baseline model.

In column 3, the baseline model is expanded to include several commute time-interaction terms that allow the effect of commute time to vary with day, job, and individual characteristics.²⁹ The first interactions are weather variables. The frigid interaction uses a dummy variable equal to one when the temperature at 6:00 a.m. on the morning of the job was below 20 degrees Fahrenheit. Over 35% of job offers are made on the day before, and exactly 50% are made on the morning of the job, suggesting that the majority of offers are received at a time when substitutes have an expectation of the morning-of-job temperature. The frigid-interaction effect is statistically significant at five percent confidence and magnifies the effect of commuting by 0.045 (36%). One possible explanation for the aversion to driving in the cold is that it is physically uncomfortable; another is the presence of safety concerns over icy roads.

Rainfall over the past 24 hours, measured at 6:00 a.m. on the day of the job, is an imperfect measure because the rain may have ended the previous day. Nonetheless, this is the best measure available and there are at least two reasons to believe that this noisy measure of rainfall contains useful information. First, the timing is less of an issue for the 35% of offers made the day before because despite the presence of forecasts, the precise ending time of the rain is uncertain. Second, even when the majority of the rain fell during the previous day, it is possible that roads were still wet at 6:00 a.m. the following day. Regardless, the interaction

²⁹ I do not report the interaction coefficients because neither the sign nor the statistical significance of the interaction coefficients is interpretable (Ai & Norton, 2003).

effect is a precisely estimated zero, suggesting that rainfall does not significantly influence substitutes' commuting preferences.

Fuel prices are the next determinant of the cost of commuting to be considered in column 3. The interaction term was constructed using the county-level daily (day-of-job) average per-gallon price of regular unleaded gasoline. I use the day-of-job fuel price because past fuel purchases are a sunk cost that should not enter in today's decisions and for forward-looking substitutes today's fuel price is likely to be the best predictor of tomorrow's. The estimated fuel-price interaction effect suggests that a one-dollar increase in the per-gallon price of gasoline increases the negative effect of commute time by 0.027 (about 22%). This fairly large effect is imprecisely estimated, however, and not statistically significant at traditional confidence levels.³⁰

The next two interaction terms in column 3 are job-length (in hours) and a half-day dummy. Intuitively, because the marginal opportunity cost of being away from home is presumably increasing with time, we might expect to see a greater aversion to commuting on longer days. This is confirmed in column 3, where a one-hour increase in job length, all else equal, increases the negative effect of commute time by 0.008 (6%), but the effect is not statistically significant. The job-length interaction effect on commuting preferences only captures the increasing marginal cost of being away from home because the half day-commute time interaction is holding the effect of daily pay constant. The half-day interaction effect is negative but also imprecisely estimated.

³⁰ As with rainfall, there is some question as to whether I am using the correct measure of fuel price: for instance, lagged fuel prices may influence substitutes' decision making. The qualitative result of a negative but statistically insignificant effect of fuel price is robust to instead using lagged daily fuel prices or lagged one- or two-week moving averages.

It is well established that, on average, women have shorter commutes than men. Explaining this stylized fact is difficult, however, and complicated by the fact that many women live in two-worker households. Including a male-commute interaction term allows for a simple test of whether women's aversion to commuting is significantly larger than that of men when commute lengths are exogenously determined and not confounded by a joint residential decision. A disproportionate number of female substitutes in the sample are certified as regular teachers, however, so a certified-commute interaction is also included to disentangle differences in gender preferences for commuting from those of certified-teachers. This is important if, as is likely, certified substitutes have a higher opportunity cost of time than their non-certified counterparts. Both the male and certified interaction effects are of the expected sign, but imprecisely estimated. The lack of a statistically significant difference between men and women suggests that the shorter commutes frequently observed among women are not due to inherent differences between the sexes in commuting preferences.

Given the commuting literature's longstanding interest in the differential between male and female commute times, I take this opportunity to further examine the differences in commuting preferences between genders by estimating the interaction model of column separately for men and women. These results are reported in columns 4 and 5 of table 2.2. A likelihood ratio test strongly rejects the null that the parameter values of the interaction model are the same for both men and women.³¹ A few striking results emerge when comparing columns 4 and 5. First, it appears that the entire aversion to commuting in cold temperatures was driven by women: the negative effect of commute time was about 50% larger for women on frigid mornings while there was virtually no temperature effect among men. Second, females

³¹ The log likelihood of the unrestricted model was computed by summing the log likelihoods of the male-only and female-only models.

have a significantly larger aversion to commuting when fuel prices are high: a one-dollar increase in fuel price increases the partial effect of commute time by about 44%.

2.5.2 *Sensitivity Analysis*

The measurement error in commute times that results from the use of substitutes' zip-code centroids rather than home addresses is a potential cause for concern. If centroid-based commute times are independent of the measurement error, however, linear probability model (LPM) estimates are consistent (Deaton, 1997, p. 101). Similarly, in the probit model, the presence of a normally distributed measurement-error term that is independent of the model's covariates creates an attenuation bias in the estimated probit coefficients, but not in the estimated APE.³² One reason that the measurement error might be independent of the centroid-based commute time is if centroid-based commute times represent average commute times of substitutes living within the zip code (Deaton, 1997, p. 101).

Another potential concern is that individuals' commute times are negatively correlated with unobserved tastes for specific jobs if substitutes prefer to work in nearby schools for reasons unrelated to commuting. In terms of the model, the concern is that

$$\varepsilon_{sdjt} = \eta_{sj} + \nu_{sdjt} \text{ and } \text{cov}(\eta_{sj}, C_{sdjt}) < 0, \quad (2.7)$$

where η_{sj} is an unobserved substitute-job-specific match effect. If equation (2.7) is true, perhaps because substitutes prefer to work in the schools that their children attend or in the schools that

³² This is similar to the “neglected heterogeneity” problem discussed in Wooldridge (2010, p. 582-4).

their neighbors work in, the baseline estimates discussed above would overstate the aversion to commuting. I show below that this is not the case.

I test for the presence of confounding “neighborhood-school preferences” by estimating the baseline model on two restricted samples: one that excludes within-zip code offers and offered commutes of less than ten minutes and a second that excludes offers from the school that the substitute worked in most frequently (the substitute’s modal school). These results are reported in columns 6 and 7 of table 2.2. In neither case does the estimated commute APE change in a meaningful way. The estimated coefficient and APE of commute time in column 6 are actually slightly larger in magnitude than their counterparts in the baseline model of column 1, suggesting that “neighborhood-school preferences” are not driving the results.³³ Together, the results in columns 6 and 7 suggest that the baseline results in column 1 are not driven by substitute-school matching effects or by measurement error in commute times.

To this point the discussion has centered on probit-model estimates. Linear probability models (LPM) are useful too, however, for a number of reasons. First, given that analytic results on CME-induced attenuation bias only exist for linear models, the robustness of the main results to the sample restriction imposed in column 6 of table 2.2 should be verified in the linear model.

Linear sub-day random-effect estimates on the full sample are provided in column 1 for comparison’s sake, but are inconsistent due to the endogenously-unbalanced nature of the

³³ The slight increase in APE observed in the restricted-sample estimates could be caused by the presence of nonlinearities in the effect of commute time or by the presence of classical measurement error (CME) in commute times. Whether substitutes live on the near or far side of a zip-code’s centroid is arguably random. It is well known that CME in linear models causes an attenuation bias, but there are no similar analytic results for non-linear models. Monte Carlo studies, however, have shown that the coefficients in binary-response models are attenuated and that the magnitude of the bias is negatively correlated with the signal-to-noise ratio (Cameron & Trivedi, 2005, p. 919). The signal-to-noise ratio is smaller when the substitute and school are located in the same zip code. The restricted sample, therefore, produces estimates that are less susceptible to CME-induced attenuation bias.

substitute-day panels (Wooldridge, 2010, p. 831).³⁴ Instead, I take the linear sub-day fixed-effects estimates in column 2 to be the baseline LPM estimates. The LPM estimates are strongly statistically significant and similar in magnitude to the probit APE, albeit slightly smaller. The linear estimates on the restricted sample, reported in column 3, are remarkably similar to the baseline LPM estimates in column 2 and are actually slightly larger, as observed in the RE-probit results discussed above. Again, there is no evidence that the results are being driven by measurement error or preferences for neighborhood schools.

A second advantage of the LPM is that it is straightforward to compute standard errors that are robust to two-way clustering (Cameron et al., 2006). Two-way clustering might be important if, in addition to the presence of a substitute-specific taste for subbing, unobserved job effects (ζ_j) enter the model. Unobserved job effects might exist because the offers state the regular teacher's name, but this information is not available in the data. The LPM estimates in table 2.3 report one-way substitute-clustered standard errors in parentheses and two-way substitute-job clustered standard errors in brackets. In each case, the one-way and two-way standard errors are nearly identical, suggesting that failing to compute two-way robust standard errors of the RE-probit estimates in table 2.2 does not significantly impact statistical inference.

Unobserved job effects would create a more serious problem if ζ_j is correlated with \mathbf{x}_j . While the randomness of the call system implies that unobserved job quality is not correlated with commute time itself, it is conceivable that job length, which is under the control of the regular teacher, is correlated with unobserved job quality. For example, the regular teacher in a classroom full of unusually difficult students might systematically make shorter substitute

³⁴ The difference is that the linear-RE model, unlike the RE-probit, does not condition on random effect.

requests to ensure that the position gets accepted. If regular teachers routinely follow this “compensating wage differential strategy,” the unobserved job effect will be positively correlated with job length and potentially bias the estimated effect of commute time.

In this case, two-way sub day-job fixed effects are required for consistency, but can only be implemented in the linear model. Job fixed effects can be included in the model because within-job variation in commute time is created when the same job is offered multiple times to substitutes living in different zip codes. As discussed in Abowd et al. (1999), the usual approach of applying OLS to mean-differenced data is infeasible due to the unbalanced nature of the panels and the high dimensionality of the problem (there are about 9,000 jobs and 32,000 sub-days). Instead, I use the two way-FE estimator of Abowd et al. (2002), the results of which are reported in column 4 of table 2.3.³⁵ The estimated effect of commute time is slightly smaller than the baseline sub-day FE estimate in column 2, but still strongly statistically significant. It is worth noting, however, that including job fixed effects sweeps away a substantial portion of variation in the data: specifically, 1,705 (about 20%) jobs have no variation in commute time because they are only offered to substitutes residing within a single zip code. Furthermore, the LPM in a panel-data setting makes restrictive assumptions of its own on the range of values that the FE can take (Wooldridge, 2010, p. 608).

2.6 Conclusion

I used data on the job offers made to substitute teachers by an automated calling system to estimate the causal effect of commute time on labor supply. Substitute teaching is an ideal labor market in which to answer this question because substitutes are both free to make daily labor

³⁵ Abowd et al. (2002) use the iterative conjugate gradient method and sparse matrixes to develop the exact two-way FE estimator. I use Ouazad’s (2008) A2REG Stata module.

supply decisions and are subject to daily exogenous variation in commute time. The main result is characterized by an offer-acceptance elasticity with respect to commute time of about -0.4. Interestingly, on average, no statistically-significant effects of rainfall or fuel prices on the disutility of commuting were found, although women's commute preferences were found to vary with fuel prices. Extremely low temperatures do increase the cost of commuting, however, and again this effect is particularly strong among women. Interestingly, while much has been made of the typically shorter commutes of women, I find no evidence that women are inherently more averse to commuting than men.

Because 98 percent of jobs were eventually accepted, there is no substitute-teacher shortage in this particular labor market. Were there a shortage, substitutes would likely be more selective when considering job offers and exhibit even stronger preferences over commute time. In this sense, the effect of commute time found in this paper can be considered a lower bound. While the generalizability of substitute teachers' preferences to the U.S. workforce is an open question, these results are potentially particularly relevant to two important labor markets: regular teacher labor markets and contingent labor markets.

Ideally the WTP for commute time would be computed in addition to the reported estimates of the causal effect of commute time on daily labor supply by taking the ratio of the marginal disutility of commuting to the marginal utility of daily pay (i.e. the marginal rate of substitution). In terms of the empirical model, this would simply be the ratio of the coefficients on daily pay and commute time. I cannot do this convincingly, however, because the positive baseline-model coefficient on job hours reported in table A2 obfuscates the interpretation of the daily-pay coefficient (half-day dummy). In other words, I am unable to disentangle the effect of

job hours from the effect of daily pay and thus cannot compute the marginal rate of substitution between commute time and daily pay.

Generally, the finding that commute time plays an important role in labor supply decisions suggests that employment policies and studies of labor supply ought to seriously consider time spent commuting in addition to hours worked. Similarly, firms ought to take potential employees' locations seriously in the hiring and recruiting processes. From an education-policy perspective, schools might be advised to actively seek nearby residents to work as substitute teachers, compensate regular teachers who make long commutes to schools in less desirable neighborhoods, or even subsidize housing for regular teachers who choose to live nearby less desirable neighborhoods' schools.

CHAPTER 2 APPENDICIES

APPENDIX 2.1

CHAPTER 2 TABLES

Table 2.1: Mean Offer Characteristics

	Offers		
	All	Rejected	Accepted
Acceptance rate	0.07	0	1
Half day	0.37 (0.48)	0.37 (0.48)	0.36 (0.48)
Hours	5.69 (1.85)	5.68 (1.85)	5.79 (1.83)
Wage	\$11.29 (2.43)	\$11.30 (2.45)	\$11.12 (2.15)
One-way miles	12.63 (8.71)	12.81 (8.74)	10.35 (7.99)
One-way minutes	17.49 (9.29)	17.70 (9.29)	14.87 (8.89)
<u>Offer Recipient</u>			
Same town	0.16	0.16	0.24
Male	0.33	0.33	0.34
Certified	0.24	0.24	0.32
N	97,205	90,040	7,165

Notes: Standard deviations are provided in parentheses for non-binary variables.

Table 2.2: RE-Probit Results

	Linear <i>C</i> 1	Quadratic <i>C</i> 2	Interactions 3	Men 4	Women 5	Drop Short 6	Drop Modal 7
<hr/>							
Coefficients							
One-way hours	-1.1943 (0.2530)***	-1.2661 (0.2543)***	-1.1333 (0.4049)***	-0.7869 (1.01)	-1.1535 (0.4229)***	-1.4571 (0.3444)***	-1.1806 (0.2853)***
One-way hours ²	.	-0.9542 (1.101)
<hr/>							
Average Effects							
Commute APE	-0.1255 (0.0133)***	-0.1206 (0.0123)***	-0.1247 (0.012)***	-0.0980 (0.0208)***	-0.1367 (0.0178)***	-0.1380 (0.0179)***	-0.1068 (0.0122)***
Commute Elast.	-0.3886 (0.0378)***	-0.4450 (0.0424)***	.	.	.	-0.5887 (0.0666)***	-0.4224 (0.0425)***
Hours*Frigid	.	.	-0.0450 (0.0185)**	0.0221 (0.0352)	-0.0698 (0.0224)***	.	.
Hours*Rain	.	.	0.0001 (0.0002)	-0.0004 (0.0006)	0.0003 (0.0002)	.	.
Hours*Gas	.	.	-0.027 (0.0201)	0.0589 (0.0359)	-0.0607 (0.0275)**	.	.
Hours*Cert.	.	.	-0.0282 (0.0329)	-0.0858 (0.0761)	-0.0267 (0.0334)	.	.
Hours*Half-day	.	.	-0.0464 (0.0371)	-0.0558 (0.0907)	-0.0291 (0.0426)	.	.
Hours*Job length	.	.	-0.0076 (0.0098)	-0.0127 (0.0202)	-0.0018 (0.0117)	.	.
Hours*Male	.	.	0.0222 (0.0258)
Predicted A	0.111 (0.005)***	0.111 (0.005)***	0.111 (0.005)***	0.117 (0.0075)***	0.107 (0.0062)***	0.0953 (0.0053)***	0.086 (0.0049)***
<hr/>							
Observations	97,205	97,205	97,205	32,435	64,770	74,383	84,191
Subs (clusters)	763	763	763	217	546	667	740

Table 2.2, Continued

Sub-days (RE)	32,057	32,057	32,057	9,827	22, 230	25,347	27,928
Log Likelihood	-22,007	-22,004	-22,000	-7,205	-14,637	-15,530	-16,708
Rho	0.63	0.63	0.63	0.67	0.62	0.63	0.63

Notes: Standard errors, reported in parentheses, are based on 50 bootstrap replications and robust to substitute-level clustering.

Bootstraps for column 7 are in progress. In column 6 offered commutes shorter than 15 minutes and within-zip offers are dropped from the analysis, while in column 7 offers from each substitute's most-frequently-worked-in school are dropped. Definitions of the partial effects are reported in appendix 2.3. All regressions include the full set of control variables described in the text. The full set of estimated coefficients in the baseline (column 1) model are reported in table A2.

Table 2.3: Linear Probability Model (LPM) Estimates

	All 1	All 2	Drop Short 3	All 4
Commute Coefficient	-0.0968 (0.0156)*** [0.0156]***	-0.0854 (0.0152)*** [0.0153]***	-0.0889 (0.0193)*** [0.0193]***	-0.0601 (0.0098)*** .
Sub-day effect	Random	Fixed	Fixed	Fixed
Job effects	None	None	None	Fixed
Observations	97,205	97,205	74,383	97,205
Substitutes	763	763	667	763
Sub-days	32,057	32,057	25,347	32,057
Jobs	8,950	8,950	8,123	8,950

Notes: The standard errors reported in parentheses are robust to clustering at the substitute level. The standard errors in square brackets are robust to two-way substitute-job clustering. All regressions include the full set of covariates discussed in the text. Column 3 makes the same sample restriction as column 6 in table 2.2. The two-way FE model in column 4 was estimated using the A2REG Stata package (Ouazad, 2008).

APPENDIX 2.2

CHAPTER 2 FIGURES

Figure 2.1: Commute-time Distributions

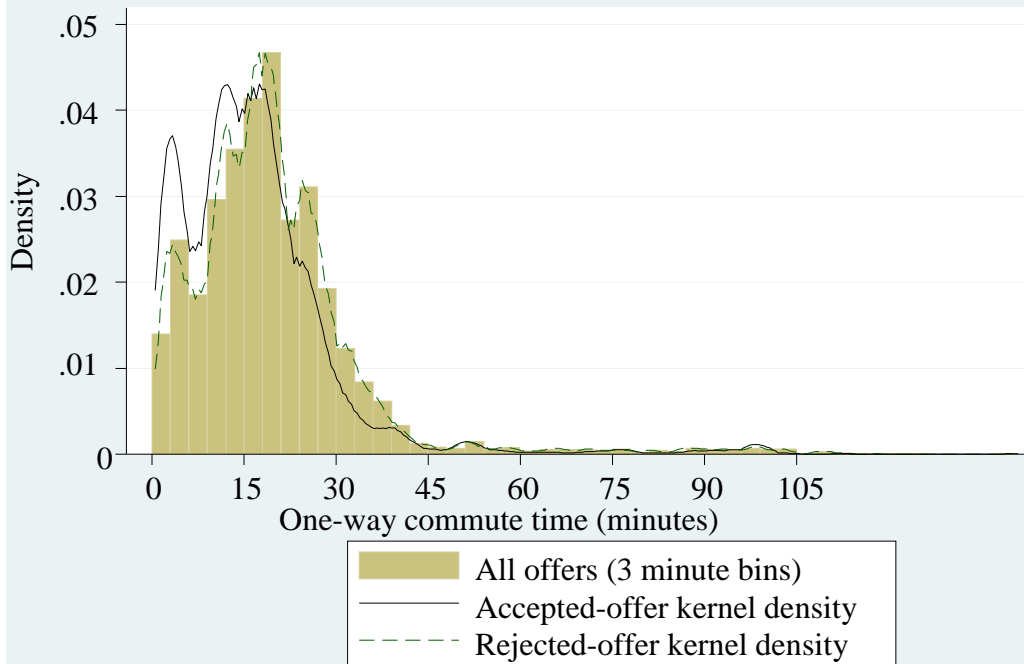


Figure 2.2: Daily Weather Conditions

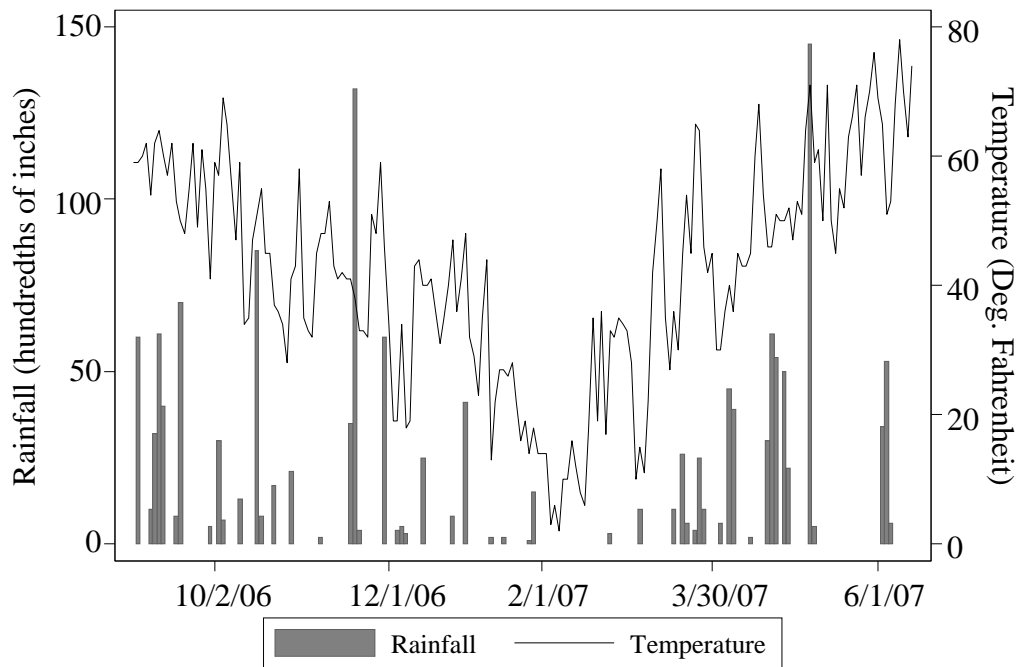
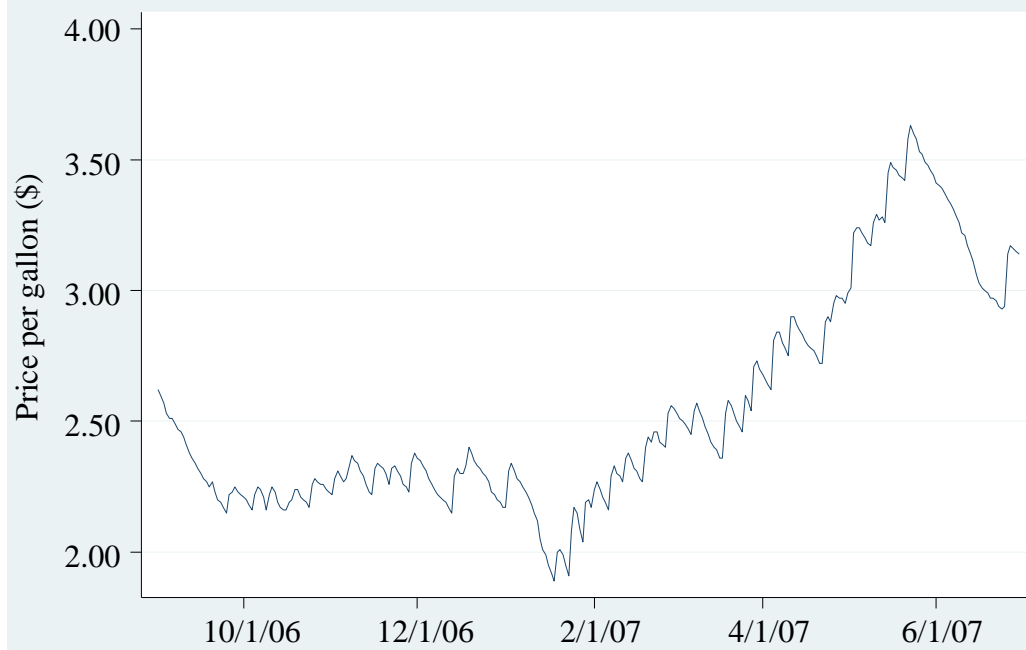


Figure 2.3: County-level Average Daily Fuel Prices



Source: OPIS daily average retail gasoline prices.

APPENDIX 2.3

AVERAGE PARTIAL EFFECTS, ELASTICITIES, AND INTERACTION EFFECTS

Following Wooldridge (2010, p. 613) the conditional expectation of the acceptance probability is

$$E(A_{sdjt} | p_{sdt} = 1, \mathbf{x}_{sdjt}, C_{sdjt}, \mathbf{z}_s, \mathbf{w}_{sdjt}, \mathbf{r}_d, \omega_{sd}) = \Phi \left(\frac{\gamma^A \mathbf{x}_{sdjt} - C_{sdjt} + \lambda \mathbf{z}_s + \mathbf{b} \mathbf{w}_{sdjt} + \delta \mathbf{r}_d}{(1 + \sigma_\omega^2)^{0.5}} \right). \quad (\text{A2.1})$$

Let N represent the total number of offers (observations), $\mathbf{x}\boldsymbol{\beta}$ represent the numerator of (A1), and s represent the denominator of (A1). The APE of a continuous variable x_k is

$$\text{APE}_k = \frac{\beta_k}{Ns} \sum_{i=1}^N \phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{s} \right) \quad (\text{A2.2})$$

and the APE of a binary variable x_k is

$$\text{APE}_k = N^{-1} \sum_{i=1}^N \left\{ \Phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta} + \beta_k (1 - x_i^k)}{s} \right) - \Phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta} - \beta_k x_i^k}{s} \right) \right\}. \quad (\text{A2.3})$$

The average elasticity (E) of continuous variable x_k with respect to the acceptance probability is

$$E_k = \frac{\beta_k}{Ns} \sum_{i=1}^N x_i^k \phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{s} \right) / \Phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{s} \right). \quad (\text{A2.4})$$

The “average partial interaction effect” (APIE) of $x_j x_k$ when x_j and x_k are continuous is based on the cross derivative $\frac{\partial^2 E(A_i | \mathbf{x}_i)}{\partial x_j \partial x_k}$, which equals

$$\begin{aligned} \text{APIE}_{jk} &= N^{-1} \sum_{i=1}^N \left(\frac{(\beta_j + \beta_{jk} x_i^k)(\beta_k + \beta_{jk} x_i^j)}{s^2} \phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{s} \right) + \frac{\beta_{jk}}{s} \phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{s} \right) \right) \\ &= N^{-1} \sum_{i=1}^N \left(\phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{s} \right) \left[\frac{\beta_{jk}}{s} - \frac{\mathbf{x}_i \boldsymbol{\beta} (\beta_j + \beta_{jk} x_i^k)(\beta_k + \beta_{jk} x_i^j)}{s^3} \right] \right). \end{aligned} \quad (\text{A2.5})$$

The APIE of $x_j x_k$ when x_j is continuous and x_k is binary is

$$\text{APIE}_{jk} = N^{-1} \sum_{i=1}^N \left(\frac{\partial E(A_i \mid \mathbf{x}_i, x_k = 1)}{\partial x_j} \frac{\partial E(A_i \mid \mathbf{x}_i, x_k = 0)}{\partial x_j} \right). \quad (\text{A2.6})$$

APPENDIX 2.4

BASELINE RE-PROBIT COEFFICIENTS

Table A2: Baseline RE-Probit Coefficients

Variable	Coefficient	Variable	Coefficient
Commute hours	-1.1943 (0.2530)***	Monday	-0.0088 (0.0703)
Male	0.1064 (0.1470)	Tuesday	0.0268 (0.0495)
Certified	0.4176 (0.2045)**	Thursday	-0.0125 (0.0589)
Half-day	0.0782 (0.1175)	Friday	-0.2408 (0.0629)***
Frigid	0.1141 (0.0602)*	September	-0.0563 (0.1315)
Gas price	0.0426 (0.1742)	November	-0.1255 (0.0756)*
Rain	-0.0002 (0.0008)	December	-0.3056 (0.1161)***
Hours	0.0743 (0.0353)**	January	-0.1129 (0.1364)
Pre-K/Kindergarten	-0.0268 (0.1402)	February	-0.3054 (0.1286)**
First/Second Grade	0.0166 (0.1208)	March	-0.2996 (0.1354)**
Third/Fourth Grade	-0.0017 (0.1380)	April	-0.3237 (0.1702)*
Fifth/Sixth Grade	0.0228 (0.1165)	May	-0.3693 (0.2138)*
Sev./Eighth Grade	0.0765 (0.1795)	June	-0.3605 (0.2815)
Math	0.0501 (0.1230)	On Teacher's List	1.4046 (0.2175)***
Science	0.0980 (0.0832)	On School's List	1.4469 (0.1208)***
Social studies	-0.0018 (0.1043)	Total Teacher Lists	3.5683 (1.8846)*
Art/Music/Gym	-0.2047 (0.0872)**	Total School Lists	1.2762 (0.9410)
Tech./Computers	-0.0214 (0.0917)	Time of call (T-t)	-0.0006 (0.0002)***
Foreign Language	-0.0718 (0.1064)	Day of (do)	-0.6458 (0.2505)***
Special Education	-0.1686 (0.1059)	Day before (db)	0.8865 (0.4526)*
Other	0.0262 (0.1054)	db*(T-t)	-0.0373 (0.0183)**
		do*(T-t)	0.0510 (0.0194)***

Table A2, Continued

Notes: These are the variables included in the baseline RE-probit model discussed in column 1 of table 2.2. The school-dummy coefficients are not reported, but are strongly jointly significant. Standard errors are robust to substitute-level clustering.

CHAPTER 2 REFERENCES

CHAPTER 2 REFERENCES

AAA. *See* American Automobile Association.

Abowd, J. M., R. H. Creecy, and F. Kramarz. 2002. Computing person and firm effects using linked longitudinal employer-employee data. Technical Paper No. 2002-06, Longitudinal Employer-Household Dynamics, Center for Economic Studies, U.S. Census Bureau.

Abowd, J. M., F. Kramarz, and D. N. Margolis. 1999. High wage workers and high wage firms. *Econometrica* 67(2): 251-333.

Ai, C., and E. C. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80(1): 123-129.

American Automobile Association. 2009. *Your Driving Costs*. Heathrow, FL: AAA.

Black, D., N. Kolesnikova, and L. Taylor. 2010. Why do so few women work in New York (and so many in Minneapolis)? Labor supply of married women across U.S. cities. Federal Reserve Bank of St. Louis Working Paper Series: Working Paper 2007-043E.

Butler, J.S., and R. Moffitt. 1982. A computationally efficient quadrature procedure for the One Factor Multinomial Probit Model. *Econometrica* 50(3): 761-764.

Calfee, J., and C. Winston. 1998. The value of automobile congestion time: Implications for congestion policy. *Journal of Public Economics* 69(1): 83-102.

Calfee, J., C. Winston, and R. Stempiski. 2001. Econometric issues in estimating consumer preferences from stated preference data: A case study of the value of automobile travel time. *The Review of Economics and Statistics* 83(4): 699-707.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler. 1997. Labor supply of New York City cabdrivers: One day at a time. *Quarterly Journal of Economics* 112(2): 407-441.

Cameron, A.C., J.B. Gelbach, and D.L. Miller. 2006. Robust inference with multi-way clustering. NBER Technical Working Paper No. 327.

Cameron, A.C., and P.K. Trivedi. 2005. *Microeconometrics: Methods and Applications*, New York, NY: Cambridge Univ. Press.

Cogan, J. 1981. Fixed costs and labor supply. *Econometrica* 49(4): 945-964.

Dickens, W., and S. Lundberg. 1993. Hours restrictions and labor supply. *International Economic Review* 34(1): 169-92.

- Gimnez-Nadal, J. I., and J. A. Molina. 2011. Commuting time and labour supply: A causal effect? IZA Discussion Paper No. 5529.
- Goette, L., D. Huffman, and E. Fehr. 2004. Loss aversion and labor supply. *Journal of the European Economic Association* 2(2-3): 216-228.
- Gutiérrez-i-Puigarnau, E., and J.N. van Ommeren. 2010. Labour supply and commuting. *Journal of Urban Economics* 68(1): 82-89.
- Harrison, G. W. 2006. Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics* 34(1): 125-162.
- Kahn, S., and K. Lang. 1991. The effect of hours constraints on labor supply estimates. *Review of Economics and Statistics* 73(4): 605-611.
- Kiefer, N., and G. Neumann. 1981. Individual effects in a nonlinear model: Explicit treatment of heterogeneity in the empirical job-search model. *Econometrica* 49(4): 965-979.
- Koslowsky, M., A. N. Kluger, and M. Reich. 1995. *Commuting Stress: Causes, Effects, and Methods of Coping*. New York: Plenum Press.
- Layton, J. 2005. How MapQuest works. *HowStuffWorks.com*.
<http://www.howstuffworks.com/mapquest.htm> (accessed March 27, 2011).
- Lemp, J. D., and K. M. Kockelman. 2008. Quantifying the external costs of vehicle use: Evidence from America's top-selling light-duty models. *Transportation Research Part D: Transport and Environment* 13(8): 491-504.
- Mortensen, D. 1986. Job search and labor market analysis. In *Handbook of Labor Economics*, vol. 2, ed. O. Ashenfelter and R. Layard, 849-919. Amsterdam: North-Holland.
- Muth, R. 1969. *Cities and Housing*. Chicago: University of Chicago Press.
- Oettinger, G. 1999. An empirical analysis of the daily labor supply of stadium vendors. *Journal of Political Economy* 107(2): 360-392.
- Ouazad, A. 2008. A2REG: Stata module to estimate models with two fixed effects.
<http://econpapers.repec.org/RePEc:boc:bocode:s456942> (accessed March 27, 2011).
- Pisarski, A. 2006. *Commuting in America III: The Third National Report on Commuting Patterns and Trends*. Washington, DC: Transportation Research Board.
- Russo G., P. Rietveld, P. Nijkamp, and C. Gorter. 1996. Spatial aspects of recruitment behaviour of firms: An empirical investigation. *Environment and Planning A* 28(6): 1077-1093.

- Sabir, M., J. S. van Ommeren, M. Koetse, and P. Rietveld. 2010. Adverse weather and commuting speed. *Networks and Spatial Economics*: 1-12.
- Small, K. A., and E. T. Verhoef. 2007. *The Economics of Urban Transportation*. Abingdon, UK: Routledge.
- So, S., P. Orazem, and M. Otto. 2001. The effects of housing prices, wages, and commuting time on joint residential and job location choices. *American Journal of Agricultural Economics* 83(4): 1036-1048.
- Stutzer, A., and B.S. Frey. 2008. Stress that doesn't pay: The commuting paradox. *Scandinavian Journal of Economics* 110(2): 339-366.
- Van den Berg, G.J., and Gorter, C. 1997. Job search and commuting time. *Journal of Business & Economic Statistics* 15(2): 269-281.
- Van Ommeren J. S., and M. Fosgerau. 2009. Workers' marginal costs of commuting. *Journal of Urban Economics* 65(1): 38-47.
- Van Ommeren, J. S., G.J. van den Berg, and C. Gorter. 2000. Estimating the marginal willingness to pay for commuting. *Journal of Regional Science* 40(3): 541-563.
- Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd Ed. Cambridge, MA: MIT Press.

CHAPTER 3

The Effect of High-Stakes Testing on Teacher Quality: Evidence from California

3.1 Introduction

The 2001 passage of the *No Child Left Behind Act* (NCLB) represents one of the U.S. Federal Government's largest forays into education policy since the 1965 *Elementary and Secondary Education Act* (ESEA) and the beginning of an era of increasing national attention to public-school quality. NCLB mandated, among other things, the publication of school report cards and sanctions on underperforming schools and districts that strengthened existing incentives for schools to improve student achievement and strengthened incentives to increase school quality more broadly. However, this pressure to improve student achievement can also potentially result in unintended consequences that undermine the policy's objectives, so understanding how both schools and teachers respond to evidence-based accountability programs is of the utmost importance to policy makers tasked with improving future iterations of education policy.³⁶

This paper uses teacher-level data from California to investigate what effect, if any, increasing the stakes of standardized testing had on teacher quality as measured by education, experience, and certification status. The effect's direction is theoretically ambiguous because both schools and teachers are presented with potentially competing incentives.³⁷ The increase in

³⁶ Hamilton et al. (2008) provide a thorough review of both the history and existing literature regarding standards-based education policy. Dee and Jacob (2010) do the same with a specific focus on NCLB. Carnoy and Loeb (2002) exploit cross-sectional variation in accountability strength across states and find larger achievement gains in "strong accountability" states. Strategic responses of schools have been found to include re-classifying predicted low scorers as non-tested special-education students (Cullen & Reback, 2006; Figlio & Getzler, 2006), suspending predicted low scorers on test days (Figlio, 2006), reassigning low-performing teachers to non-tested grades and subjects (Chingos & West, 2011), and offering high-calorie school lunches on test days (Figlio & Winicki, 2005). Similarly, teachers have been found to "game the system" by explicitly cheating (Jacob & Levitt, 2003) and by "teaching to the test" (Jacob, 2005). Hannaway and Hamilton (2008) review the literature on how accountability influences teachers' instructional strategies.

³⁷ The empirical evidence on the effect of evidence-based accountability policy on teacher attrition, for example, is mixed: Boyd et al. (2008) show that attrition of New York fourth-grade

tests' stakes provided schools with strong incentives to increase achievement as measured by standardized test scores. Schools, for example, might respond by upgrading teacher quality in all grades and subjects, or by attempting to increase test scores at the expense of general learning and learning in non-tested grades and subjects by reallocating low-performing teachers from tested to non-tested grades (Chingos & West, 2011). The latter approach is short sighted and potentially harmful, of course, because decreasing educational quality in non-tested early grades may delay students' development and have long-lasting consequences (Chetty et al., 2011; Heckman et al., 2010; Heckman & Masterov, 2007).

Some teachers may relish the opportunity to make a difference in children's lives and welcome accountability programs. Not all teachers necessarily feel this way, however, and more generally the ability of schools to respond to the incentives created by accountability programs might be limited by a number of factors including budget constraints, teacher shortages, and the preferences of teachers and teachers' unions. Specifically, high-stakes testing might increase teacher attrition and exacerbate teacher shortages by decreasing teachers' autonomy in the classroom (Luna & Turner, 2001), decreasing teachers' sense of job security (Reback et al., 2011), or increasing teachers' stress levels (Daly & Chrispeels, 2005). Highly-educated teachers might be those most at risk of leaving if they have access to better non-teaching job opportunities.

The existing literature on the response of teacher quality to evidence-based accountability programs largely overlooks the potential for grade-specific effects of high-stakes testing on teacher quality (e.g., Clotfelter et al., 2004; Lee & Young, 2004), which is problematic if accountability programs are primarily affecting tested grades because estimates will be weighted

teachers decreased in response to new high-stakes fourth-grade tests while Clotfelter et al. (2004) find that attrition increased in North Carolina in response to a new accountability system.

averages of the effect in tested-grades effects and non-effects in non-tested grades. Two exceptions to this critique include Boyd et al. (2008), who find that fourth-grade attrition rates decreased in response to New York's introduction of a fourth-grade test, and Phillips and Flashman (2007), who use pre-NCLB nationally-representative data to compare inputs in tested versus non-tested grades across "strong-" and "weak-accountability" states; they find small difference in class size but no difference in teacher quality.

I contribute to this literature by estimating the effect of the NCLB-induced increase in testing stakes on teacher quality in California using a difference-in-differences (DD) approach that compares teacher quality in non-tested first-grade to tested second-grade classrooms. The empirical strategy is based on the fact that beginning in the 1997/98 school year all second-through eleventh-grade students in California took mandatory standardized tests that were used by California's pre-NCLB accountability system, but NCLB substantially increased the tests' stakes beginning in the 2002/03 school year. I also undertake two additional analyses. First, I fully interact the DD model's covariates with a Title 1-school indicator to test for a differential effect of NCLB in Title 1 schools, which were subject to significantly stronger sanctions under NCLB. Second, I consider an event-history specification that allows for year-specific policy effects, which is useful for understanding the timing of schools' and teachers' response to the policy and pre-existing differences in teacher quality across grades.

Two caveats of the analysis are worth stressing at the outset. First, this is not a study of the effect of NCLB, which is itself a bundle of several policies. Rather, I examine the effect of increasing the stakes of standardized testing on the distribution of teacher quality across tested and non-tested grades. Prior to NCLB, California had an accountability program that provided schools with two types of incentives to improve student achievement. One was a rewards

program that awarded high-achieving schools monetary rewards and public praise. The other program, which low-achieving schools could voluntarily enter, provided financial assistance in conjunction with a threat of district or state intervention if the schools did not subsequently improve. NCLB raised the stakes of California's standardized tests by raising the bar of acceptable performance, mandating that the state provide school report cards indicating schools' performance, and imposing strong sanctions on underperforming Title 1 schools.

A second caveat is the possibility of confounding spill-over effects of the increased testing stakes to the control group (non-tested grades) caused by school administrators seeking to improve second-grade test scores by increasing first-grade teacher quality.³⁸ Similarly, first-grade teachers might worry about being blamed for low second-grade test scores. The direction of the spill-over bias is ambiguous, again because of the potentially competing school and teacher reactions.

The strongest findings regard teacher education: the fraction of second-grade teachers holding a graduate degree decreased relative to that of first-grade teachers by about 1.4 percentage points in the year preceding NCLB and in each subsequent year through 2005/06. However, the fact that an "effect" is found in the year before the testing stakes were raised obfuscates the interpretation of the results. One possible interpretation of this anomalous finding is that there was a preemptive movement of highly-educated teachers out of the tested second grade in anticipation of the changes to come due to the fact that NCLB was publicly debated for one full year before being passed in January of 2002. Alternatively, this might indicate the presence of a pre-existing differential trend between first- and second-grade classrooms.

³⁸ This would be a greater concern if the tests were administered early in the school year, but California administers its standardized tests in March when about 85% of the school year is complete.

Regardless of the cause of the difference, however, it is interesting to note that there does appear to be a difference in the prevalence of highly-educated teachers in first and second grades.

A much smaller and shorter-lived decrease in second-grade teachers' average years of experience is found and no statistically-significant effects are found on teacher-certification measures. Nor do Title 1 schools appear to behave any differently than their non-Title 1 counterparts, which is surprising given the stronger incentives and greater pressure placed on Title 1 schools by NCLB. Policy implications of these findings are discussed in the conclusion.

3.2 Literature Review

Lee and Young (2004) use nationally representative teacher-level data from the 1990's to investigate the effect of state-level accountability strength on a variety of teacher and school outcomes. Specifically, the authors find no effect of accountability strength on class size or teacher quality, the latter of which they measure by in-field teaching. Clotfelter et al. (2004) investigate the impact of the implementation of an accountability program in North Carolina. The authors find that the accountability program increased teacher attrition, decreased teacher experience, and had no effect on teacher quality (measured by teachers' undergraduate institution's selectivity). Neither paper, however, allows for grade-specific effects of accountability. If responses are concentrated in tested grades as hypothesized in the introduction, however, studies that fail to account for this by averaging across all grades will produce attenuated policy-effect estimates.

More recently, a handful of papers have examined the potential for differential impacts of accountability policies on tested versus non-tested grades. Reback et al. (2011) use state-level variation in the definition of adequate yearly progress (AYP) to find that NCLB led teachers in

tested grades to work more hours per week and to be more concerned about job security than their peers in non-tested grades. Chingos and West (2011) use post-NCLB administrative data from Florida to find that low-value added teachers are more likely to both move to low-stakes (non-tested) positions within their current school and to exit teaching.

The most methodologically similar paper to my approach is Boyd et al. (2008), who use administrative data from New York State to investigate the effect of a newly-implemented fourth-grade testing requirement on fourth-grade teacher attrition and find the counter-intuitive result that testing led to a decrease in attrition in the newly tested fourth grade. The teachers who did leave the fourth grade in response to the new test were more likely to be experienced and less likely to have graduated from a highly-selective college. Post-test entrants to fourth grade were less likely to be new teachers and more likely to have graduated from a highly-selective college. Similarly, Phillips and Flashman (2007) use nationally representative pre-NCLB data from 1993 and 1999 to examine the difference in several teacher and classroom-level characteristics between tested and non-tested grades in “strong-” and “weak-accountability” states. The authors find marginally smaller class sizes in strong-accountability states, but no significant difference in teachers’ highest degree obtained, years of experience, licensure status, or college quality.

I contribute to this literature in several ways. First, understanding the teacher labor market in as large and diverse a state as California is important in its own right. Second, I examine the effect of strengthening existing accountability policies rather than the effect of implementing new policies when none existed before (e.g., Clotfelter et al., 2004; Boyd et al., 2008). As federal and state policies evolve, the former question is arguably of more interest.

3.3 Institutional Details and Data

3.3.1 Pre-NCLB Education Policy in California

The federal *Improving America's Schools Act of 1994* encouraged states to create or expand standards-based accountability programs. California responded by passing the *Public Schools Accountability Act of 1999* (PSAA), which was comprised of three interrelated programs that sought to motivate schools to improve student achievement.³⁹ The PSAA's primary innovation was the creation of an Academic Performance Index (API), which is an annual school-level achievement score based on student and demographic-group performance on California's *Standardized Testing and Reporting Program* (STAR).⁴⁰

From STAR's inception in 1997/98, all students in second through eighth grade were tested annually in a minimum of two subjects: math and English.⁴¹ The tests are administered within a ten-day window of the date on which 85% of the school year is complete; testing dates typically fall in early to mid-March. API is scored on a scale from 200 to 1,000, with 800 being the target for proficiency. Under PSAA, schools must score better than 800 or make annual gains of at least five percent of their distance from 800 to remain in good standing. Schools that met these requirements and met a threshold test-participation rate became eligible for monetary

³⁹ A thorough independent review of PSAA, commissioned by the California Department of Education, was conducted by the American Institutes for Research (AIR, 2003).

⁴⁰ API is computed as follows. First, student scores are sorted into five performance categories. The scores are then weighted by category-specific weights, with the higher-achieving categories receiving larger weights. Finally, all of the weighted individual test scores within a school are summed. The resulting number is the school's API. For additional details see <http://www.ed-data.k12.ca.us/articles/Article.asp?title=understanding%20the%20API>.

⁴¹ For additional information on the history of STAR, exemptions, test formats, and results see <http://star.cde.ca.gov/>.

prizes via the Governor's Performance Award Program, which was the second component of PSAA.

Schools that failed to meet these guidelines became eligible to enter the Immediate Intervention / Underperforming Schools Program (II/USP), which was the third component of PSAA. Each year, schools that scored below a preordained API percentile became eligible to apply to II/USP.⁴² Despite the voluntary nature of participation in II/USP from the state's perspective, many districts in California effectively forced their eligible schools to apply (AIR, 2003). Because applications typically outnumbered available funding, 430 applicant schools were randomly selected each year to receive the modest funding increase.⁴³ Analyses by Goe (2006) and AIR (2005) find no evidence of a significant effect of II/USP funding on student achievement, however, and it is unlikely that a slight increase in funding would impact the outcomes of interest in the present paper.⁴⁴ II/USP schools that fail to improve in the two to three years after entering the program were technically subject to state-level interventions, but most teachers and principals did not consider this a credible threat (AIR, 2003).

A final relevant pre-NCLB policy in California is the *Grades K-3 Class Size Reduction Program* (CSR), which is distinct from PSAA and provides schools and districts with financial

⁴² The median was the initial cutoff, but in subsequent years it was reduced to the lower quartile.

⁴³ The initial II/USP lasted three years. It has since been reinstituted under various names.

⁴⁴ Two potential explanations for the program's apparent lack of success have been put forth. First, principals may have been unsure of how to spend the funds; indeed, principals complained that state administrators provided no guidance in this regard. Second, the comparison group used in the empirical analyses may have been contaminated by some combination of districts responding endogenously by cutting funding to II/USP schools and redirecting it to similarly low-performing schools that did not receive II/USP funding and non-II/USP schools receiving money from similar federal programs (e.g., the *Comprehensive School Reform Demonstration Program* (CSRD)).

incentives to reduce class size.⁴⁵ Two key features of CSR are crucial for proceeding in the absence of detailed school-level CSR participation data. First, CSR schools must reduce class sizes in the following order: first grade, second grade, kindergarten *or* third grade, grade level not chosen in previous step. Second, by the 1997/98 school year virtually all first and second-grade classrooms were participating in CSR (Carroll et al., 2000). For these reasons, the empirics will focus on comparisons between first and second grade classrooms. Still, a valid concern is that NCLB encouraged some schools to enter CSR for first grade only or expand CSR from first to second grade, differentially affecting first- and second-grade classrooms.

3.3.2 *NCLB's Impact in California*

NCLB increased the stakes of California's STAR tests in three fundamental ways. First, it required that all schools make adequate yearly progress (AYP), which was stricter than existing API-score requirements because in addition to reaching API growth targets, AYP requires that schools meet percent proficient, attendance, and test participation thresholds. Second, NCLB mandated that states publish "school report cards" containing information on schools' performance levels and AYP status. And third, NCLB threatened stronger sanctions on schools that receive Title 1 funds but fail to make AYP.

Title 1 was a component of the original ESEA that was reinstituted by NCLB, which provides federal funds to schools in proportion to the number of low-income students attending the school. This money can be used to cover the cost of tutoring, after-school, and summer

⁴⁵ CSR was launched in 1996 and is still active today. Schools must have no more than 20 students per class. For additional information see <http://www.cde.ca.gov/ls/cs/k3/>.

programs that reinforce the school's standard curriculum.⁴⁶ Under NCLB, when a Title-1 school fails to make AYP for two consecutive years it must enter Program Improvement (PI), which is a five-year process of steadily increasing consequences that culminates with the drastic restructuring of the school (e.g., the school is reinvented as a charter, taken over by the state, or replaces a majority of the staff).⁴⁷ To leave PI and avoid restructuring a school must make AYP in two consecutive years.

3.3.3 Data

The teacher data analyzed in this paper comes from the California Basic Educational Data System (CBEDS) Professional Assignment Information Form (PAIF).⁴⁸ I use PAIF data on all fulltime first- and second-grade teachers in self-contained classrooms from 1998/99 through 2005/06 because schools' Title-1 status is unavailable prior to the 1998/99 school year, nearly all first- and second-grade classrooms were participating in CSR by 1998/99, and I am primarily interested in schools' and teachers' immediate responses to the NCLB-induced increase in testing stakes that took effect in 2002/03. I augment the PAIF data with information on schools' Title 1 status and demographic composition from the National Center for Education Statistics' Common Core of Data (CCD). Both data sources are publicly available.⁴⁹

Table 3.1 provides an overview of the schools containing first- and second-grade classrooms. The black share of enrollment fell slightly during the study's time frame from 8.8%

⁴⁶ For more information on Title 1 see <http://www2.ed.gov/programs/titleiparta/index.html>.

⁴⁷ For the precise PI timeline see <http://www.cde.ca.gov/ta/ac/ti/nclbpireq.asp>.

⁴⁸ A copy of a PAIF is available here: www.cde.ca.gov/ds/dc/cb/documents/paif08.doc.

⁴⁹ CBEDS data is provided by the California Department of Education at <http://www.cde.ca.gov/ds/sd/df/>. The CCD is available at <http://nces.ed.gov/ccd/>.

to 7.3% while the Hispanic share of enrollment increased from 45.8% to 51.5%. Charter schools became more popular during this time period and about 70% of schools received Title 1 funds. The PAIF sample contains about 45,000 self-contained classroom teachers each year that are evenly split between first and second grade. About 750 school districts and 5,000 schools are represented in each year.

In figure 3.1, as a prelude to the empirical analysis, I show how four teacher-quality measures varied by grade between 1998/99 and 2005/06. Figure 3.1A examines trends in the fraction of teachers holding a graduate degree (masters or doctorate). The pre- and post-NCLB trends are quite similar. Second-grade classrooms are about three percentage points more likely to be taught by a teacher who holds a graduate degree, but this gap narrows somewhat over time. Similarly, figure 3.1B shows that second-grade teachers tend to have about one more year of experience than their first-grade counterparts and that the trends in average experience are similar for both grades, before and after NCLB.

Figure 3.1C defines inexperienced teachers as those with either zero or one year of prior teaching experience. First-grade teachers are more likely to be inexperienced throughout, but the pre- and post-NCLB trends differ for both grades. Specifically, the post-NCLB second-grade gradient flattens out while the corresponding first-grade gradient increases slightly. Finally, figure 3.1D shows the fraction of self-contained first and second-grade teachers who are fully credentialed each year. There is a small gap between first and second-grade credential rates initially, but over time the two trends converge to about 97% fully credentialed. In sum, figure 3.1 does not suggest any obvious or large effects of NCLB, although to definitively answer this question a multivariate analysis that controls for school characteristics is necessary.

3.4 Empirical Model and Estimation

NCLB increased the stakes of California's STAR tests beginning in 2002/03, but this did not affect all grades equally because kindergarten and first-grade classrooms remained untested. I restrict the analysis to only first- and second-grade classrooms, however, for two reasons. First, as discussed in section 3.3.1, the CSR program requires that schools begin by reducing first- and second-grade class sizes and kindergarten and third-grade classes are differentially affected by CSR. Most first- and second-grade classrooms in California were participating in CSR by 1997/98, although kindergarten and third-grade participation potentially varied across schools. The DD estimator maintains the assumption that the treatment and control groups were not differentially affected by any other policy interventions, so given the presence of the CSR program, this assumption is most plausible when comparing first- to second-grade classrooms.

Second, apart from testing status, first and second grade are similarly structured. Kindergarten is fundamentally different in that it is a half-day in many school districts and many kindergarten teachers teach two classes per day. Similarly, higher elementary grades are more likely to compartmentalize (have one teacher teach only math, one teacher teach only science, etc.) or track students (group students of like ability).

The standard DD estimator compares differences in outcomes between the treated and non-treated groups, before and after the policy change, and assumes that the treatment effect of the policy is constant across years.⁵⁰ I assume that the outcome of interest (y) is determined by

$$y_{ist} = \lambda \text{Second}_{ist} + \delta \text{NCLB}_t \cdot \text{Second}_{ist} + \theta_{st} + \varepsilon_{ist}, \quad (3.1)$$

⁵⁰ For a textbook treatment of DD estimators see Wooldridge (2010, pp. 147-51).

where i indexes classrooms, s indexes schools, t indexes years, $Second_{ist}$ is a dummy variable equal to one if classroom i is a second-grade classroom, $NCLB_t$ is a binary indicator of NCLB being in effect in year t , θ_{st} is a school-year fixed effect that controls for both observed and unobserved school-year attributes such as principal quality, student type, school size, and participation in programs such as CSR, II/USP, and CSRD, and ε_{ist} is an idiosyncratic error term that captures the effect of unobserved classroom-specific determinants of teacher quality.

The school-year fixed effects (FE) subsume the year dummies typically included in a DD regression and are included in all subsequent models in order to partial out school-specific time trends. Conditioning on the school-year FE means that the results are identified by within-school year differences between first- and second-grade classrooms.

NCLB was being publicly discussed in congress during the year before it was implemented and was a prominent component of George W. Bush's 2000 Presidential Campaign before that. Given the high profile and controversial nature of NCLB and the strong incentives that it provided schools to make AYP, schools could very well have reacted in anticipation of the passage of NCLB. Similarly, teachers who wanted to avoid the pressures of NCLB might have preemptively exited the profession or tested classrooms. Alternatively, there may have been a pre-existing difference in the characteristics of first- and second-grade teachers. For these reasons, along with the fact that some schools and teachers might be slow in responding to NCLB, I also estimate event-history models that allow the treatment effect to vary across years.

These richer models replace the $NCLB_t$ indicator seen in equation (3.1) with a full set of year dummies, yielding the estimating equation

$$y_{ist} = \lambda Second_{ist} + \sum_{j=1}^8 \delta^j Year_t^j Second_{ist} + \theta_{st} + \varepsilon_{ist}, \quad (3.2)$$

where $Year_t^j = 1$ when $j = t$, and 0 otherwise.

Equations (3.1) and (3.2) implicitly assume that all schools responded to NCLB in the same way, regardless of Title-1 status. This is a strong assumption, of course, because only schools that received Title 1 funds were required to enter PI after failing to make AYP in two consecutive years. The hypothesis that NCLB had a differential effect on Title 1 versus non-Title 1 schools is easily tested by fully interacting the covariates in equation (3.1) or (3.2) with a Title-1 dummy and testing the significance of the year-second grade-Title 1 triple interaction terms.

Equations (3.1), (3.2), and the Title 1-interacted analogue of equation (3.1) will be estimated by the standard linear-FE estimator. Because school-years are nested in schools that are nested in districts, standard errors are made robust to clustering at the district level in all models. Doing so makes inference robust to the presence of district-wide and school-wide initiatives, programs, demographic effects, and superintendent effects.⁵¹

3.5 Results

3.5.1 *Difference-in-differences Estimates*

Table 3.2 reports school-year fixed effects (FE) linear-model estimates of the simple DD regressions described by equation (3.1) for each of the following teacher characteristics: holds a

⁵¹ Clustering at the highest level, in this case the district, is advocated by Angrist and Pischke (2009, p. 319).

graduate degree, years of experience, is inexperienced (binary indicator of < 2 years of experience), holds a full California teaching credential, and holds a general elementary license. The top panel, which assumes no heterogeneity in schools' responses to NCLB, finds small but marginally statistically significant decreases of about one percentage point in both the probability that second-grade teachers hold a graduate degree and the probability that second-grade teachers hold a full California teaching credential. The negative sign on the NCLB-second grade interaction term suggests that teacher quality decreased in the tested second grade relative to the non-tested first grade in response to NCLB. Precise null effects of NCLB are estimated for each of the other three measures of teacher quality.

The second panel of table 3.2 extends the baseline specification to allow for a differential effect in Title 1 schools. The Title 1-second grade-NCLB triple interaction term is not statistically significant for any of the five outcomes of interest, however, suggesting that the response of Title 1 schools to NCLB did not systematically differ from that of non-Title-1 schools. The lack of a differential response is not evidence of a lack of Title 1 schools' *desire* to respond, of course, as teacher shortages, resistance from teachers, and schools' budget constraints could all prevent schools from enacting desired changes.

3.5.2 *Event History Estimates*

As discussed in section 3.4, the results of the simple DD estimates reported in table 3.2 are misleading if there were delayed effects of the policy change, schools and teachers altered their behavior in anticipation of NCLB, or if there were pre-existing differences between first and second grades in teacher quality. To accommodate these possibilities, table 3.3 estimates the event-history models described by equation 3.2 for each of the five measures of teacher quality.

The estimates in columns 1 and 2 of table 3.3 are problematic: there were fairly large and statistically significant decreases in both the probability of holding a graduate degree and in years of experience in 2001/02, the year before NCLB took effect. By treating the 2001/02 school year as pre-NCLB, the simple DD effect on the probability of holding a graduate degree of -0.007 reported in column 1 of table 3.2 is only half the size of the effect reported in the event-history specification of column 1, table 3.3. It is interesting to note that the decrease in the likelihood of second-grade teachers holding a graduate degree persisted in each year following the passage of NCLB.

The effect on teacher experience was short-lived, dying out in subsequent years. Column 2 of table 3.3 shows that teachers' average experience fell by about 0.2 years in both the year preceding NCLB and the first year of NCLB, which cancelled each other out in the simple pre/post comparison of table 3.2. The point estimates of the interaction terms remain negative in subsequent years, but decrease in magnitude and are not statistically significant at traditional confidence levels. In column 3 of table 3.3 there appears to be a slight increase in the probability that second-grade classrooms are staffed by an inexperienced teacher in 2002/03, the first year of NCLB, but essentially no effect in subsequent years. This finding fits with the estimates in column 2, perhaps suggesting that a small subset of experienced teachers left second-grade classrooms immediately after the passage of NCLB and were replaced by inexperienced teachers. Finally, as seen in table 3.2, NCLB does not appear to have had a strong effect on either fully-credentialed or elementary-licensed teachers.

3.5.3 *Sensitivity Analysis*

Table 3.4 examines the robustness of the graduate-degree results to the choice of a linear probability model (LPM) and performs two falsification exercises that provide further evidence that the results discussed above are attributable to the increase in testing stakes brought about by NCLB. I focus on graduate degree because this was the only outcome for which strong effects were found. The linear estimates were taken as the preferred baseline estimates because the FE logit estimator precludes the calculation of precise partial effects and requires dropping observations from school-years that contain no variation in the dependent variable.⁵²

Column 1 repeats the baseline estimates from column 1 of table 3.3 to facilitate comparisons with the alternative specifications reported in columns 2 and 3. Column 2 estimates the LPM on a restricted sample that excludes observations from school years that did not experience any variation in the dependent variable (i.e., either all first and second-grade teachers in the given school-year held graduate degree, or none did). This is the sample restriction imposed by the conditional FE logit estimator reported in column 3, so it is reassuring to see that the LPM estimates in columns 1 and 2 are quite similar.

The FE-logit coefficients reported in column 3 follow the same sign and statistical-significance patterns as the LPM, suggesting that the results are robust to the linear functional form assumed by the LPM. The logit coefficients cannot be directly compared to the LPM coefficients, but scaled coefficients that are comparable to the LPM partial effects can be computed using the product of the sample average probability of holding a graduate degree

⁵² For a textbook treatment of the FE logit estimator, see Wooldridge (2010, pp. 621-2).

(0.25) and one minus this probability as an approximate scaling factor.⁵³ This scaling factor is approximate because we are effectively computing the “partial effect at the average” rather than the “average partial effect” and the year-grade interactions are binary variables. Nonetheless, it is reassuring that the resulting scaling factor of 0.1875 produces partial-effect estimates of about -0.02, which are in line with the LPM estimates reported in column 2. FE-logit estimates for the other binary measures of teacher quality are reported in table A3.

Columns 4 and 5 of table 3.4 are falsification exercises that look for a differential effect on the probability of the classroom teacher holding a graduate degree between two tested grades and between two non-tested grades, respectively. Column 4 shows no statistically significant “effect” of NCLB on third-grade teachers relative to second-grade teachers in any year before or after NCLB; this was to be expected, given that NCLB did not differentially effect the two grades. In column 5, which compares kindergarten to first grade, larger, positive, and marginally statistically significant effects are found that suggest that NCLB increased the probability that first-grade classrooms were taught by highly-educated teachers. One potential explanation of this finding is that some graduate-degree holding teachers moved from second- to first-grade classrooms in response to the increased testing stakes.

3.6 Conclusions and Discussion

This paper uses teacher-level data to analyze the effect of an increase in the stakes of standardized testing created by NCLB on teacher quality in California’s second-grade classrooms. The identification strategy compares teachers in the non-tested first grade to those

⁵³ It is impossible to estimate proper “average partial effects” because the distribution of the school-year fixed effect θ_{st} is unknown (Wooldridge, 2010, pp. 620-1).

in the tested second grade. Fairly large, statistically significant, and persistent decreases in the probability that second-grade teachers held a graduate degree were found. A small, marginally significant decrease in teachers' experience was found in the years preceding and immediately following NCLB, but died out in later years. No effect was found on teachers' certification. Surprisingly, NCLB's effect did not vary with schools' Title-1 status.

One possible explanation of the finding that teacher quality decreased in tested grades despite schools' short-run incentives to increase teacher quality is that highly-educated teachers were leaving tested "high-stakes" grades of their own accord in response to the increased pressure and decreased autonomy created by NCLB. That the exit of highly-educated teachers continued for several years after the initial passage of NCLB might indicate that each year cumulative accountability pressures induced a new group of highly-educated teachers to leave tested classrooms. Highly-educated teachers may be particularly vulnerable to leaving the teaching profession when external factors such as high-stakes testing increase the stresses associated with the job because their education affords them viable alternative careers.

However, alternative explanations of the results exist. Specifically, while the statistically significant effect in the year prior to NCLB might indicate anticipatory behavior, it could just as easily be evidence of a pre-existing trend towards a difference in teacher quality between first and second grades. Furthermore, the results should be interpreted cautiously because the effects are relatively small and only statistically significant for one of the five investigated outcomes. Of course, as mentioned previously, the non-findings may be the result of spillover effects to the first grade. Simply put, more evidence is necessary to definitively answer the questions posed in this paper. Future work applying a similar methodology to administrative data from other states

and nationally representative data such as the NCES Schools and Staffing Survey will prove invaluable, as will the use of panel data that allows researchers to follow teachers over time.

Taken at face value, the finding that high-educated teachers became less likely to be in tested grades is a troubling unintended consequence of NCLB, but one that might be relatively easy to correct from a policy perspective. Standard labor-economic theory suggests that if jobs in tested grades are significantly more stressful, these jobs should pay higher wages in order to compensate teachers for coping with that stress.⁵⁴ Compensation could be non-pecuniary as well, provided via additional resources such as additional planning periods, teaching aids, or professional development. Finally, even if the difference in educational attainment between first- and second-grade teachers is not driven by the presence of high-stakes testing, this research has identified grade-specific differences in teacher quality that are interesting in their own right. Are the differences driven by the supply- or demand-side of the teacher labor market? A better understanding the causes of such differences has the potential to improve teacher hiring and training practices, for example.

⁵⁴ This is known as the theory of compensating wage differentials (Borjas, Chapter 6, 2008).

CHAPTER 3 APPENDICIES

APPENDIX 3.1

Chapter 3 TABLES

Table 3.1: PAIF Data Description

Year	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	Total
<u>Full sample (weighted) average characteristics</u>									
Second grade	49.7%	49.6%	50.4%	50.3%	50.1%	50.0%	49.9%	49.8%	50.0%
% black	8.8%	8.6%	8.4%	8.2%	8.0%	7.7%	7.5%	7.3%	8.1%
% Hispanic	45.8%	46.9%	48.4%	49.3%	49.8%	50.4%	51.0%	51.5%	49.1%
% free lunch	48.3%	48.4%	47.5%	47.2%	46.5%	46.6%	46.5%	45.4%	47.0%
Charter	0.8%	1.4%	1.5%	1.6%	1.7%	1.6%	1.8%	2.1%	1.5%
Title-1 eligible	67.8%	61.8%	71.1%	72.0%	73.7%	73.2%	67.2%	69.2%	69.5%
N (Teachers)	45,425	45,971	46,893	47,240	47,137	46,234	45,960	45,761	370,621
<u>School-level average characteristics</u>									
% black	8.8%	8.6%	8.3%	8.2%	8.1%	8.0%	7.8%	7.5%	.
% Hispanic	39.7%	40.8%	41.8%	43.0%	43.8%	44.8%	46.0%	47.1%	.
% free lunch	44.2%	43.8%	42.6%	42.6%	42.3%	43.1%	43.5%	42.5%	.
Charter	1.0%	1.8%	2.0%	2.3%	2.6%	2.6%	2.9%	3.3%	.
Title-1 eligible	64.2%	58.4%	67.6%	69.0%	71.2%	71.2%	64.9%	67.3%	.
N (districts)	746	752	760	756	762	765	767	767	.
N (schools)	4,802	4,857	4,924	4,973	5,038	5,087	5,145	5,241	.

Notes: The school-characteristic data comes from the Common Core of Data (CCD). Grade-specific means of the dependent variables are provided in the main results table alongside each regression and separately by year in figure 3.1.

Table 3.2: Standard DD Estimates

Dependent var.	Graduate degree 1	Years teaching 2	New teacher 3	Full credential 4	Elem. license 5
Baseline Specification					
Second grade	0.023 (0.004)***	1.079 (0.097)***	-0.012 (0.002)***	0.013 (0.006)**	-0.001 (0.001)
NCLB*second	-0.007 (0.003)**	-0.025 (0.066)	0.001 (0.002)	-0.008 (0.005)*	0.001 (0.001)
Constant	0.241 (0.002)***	11.428 (0.045)***	0.067 (0.001)***	0.904 (0.002)***	0.971 (0.000)***
Differential Title-1 effect					
Second grade	0.021 (0.006)***	1.321 (0.168)***	-0.013 (0.002)***	0.006 (0.003)*	0.000 (0.001)
Title-1*second	0.003 (0.006)	-0.356 (0.174)**	0.002 (0.003)	0.010 (0.006)*	-0.002 (0.002)
NCLB*second	-0.011 (0.006)*	-0.043 (0.141)	0.001 (0.003)	-0.003 (0.004)	0.000 (0.002)
Title-1*NCLB*second	0.005 (0.007)	0.037 (0.183)	0.000 (0.004)	-0.007 (0.004)	0.001 (0.002)
Constant	0.241 (0.002)***	11.427 (0.047)***	0.067 (0.001)***	0.904 (0.002)***	0.971 (0.000)***
First grade mean	0.24	11.4	0.07	0.90	0.97
Second grade mean	0.26	12.5	0.06	0.91	0.97
Observations	370,621	370,479	370,479	370,621	370,621
School-year FE	40,067	40,066	40,066	40,067	40,067
District clusters	823	823	823	823	823

Notes: All models estimated in this table include school-by-year fixed effects (FE). The standard errors reported in parentheses are robust to district-level clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. “New teacher” was defined as teachers with less than two years of experience. Sample sizes vary because experience was not available for all teachers.

Table 3.3: Event History Estimates (time-varying NCLB effects)

Dependent var.	Graduate degree 1	Years teaching 2	New teacher 3	Full credential 4	Elem. license 5
Second grade	0.028 (0.004)***	1.153 (0.120)***	-0.015 (0.003)***	0.014 (0.008)*	-0.001 (0.002)
Second*99/00	-0.002 (0.004)	0.041 (0.065)	0.006 (0.003)*	0.002 (0.002)	0.001 (0.002)
Second*00/01	-0.006 (0.004)	-0.094 (0.103)	0.004 (0.004)	-0.002 (0.005)	-0.000 (0.002)
Second*01/02	-0.013 (0.005)***	-0.236 (0.096)**	0.002 (0.003)	-0.002 (0.004)	0.000 (0.002)
Second*02/03	-0.010 (0.005)**	-0.192 (0.102)*	0.007 (0.003)**	-0.004 (0.006)	0.001 (0.002)
Second*03/04	-0.014 (0.005)***	-0.114 (0.102)	0.003 (0.003)	-0.008 (0.006)	0.000 (0.002)
Second*04/05	-0.014 (0.005)***	-0.035 (0.112)	0.003 (0.005)	-0.012 (0.007)	0.002 (0.002)
Second*05/06	-0.014 (0.006)**	-0.055 (0.111)	0.004 (0.004)	-0.011 (0.007)	0.001 (0.002)
Constant	0.241 (0.002)***	11.428 (0.045)***	0.067 (0.001)***	0.904 (0.002)***	0.971 (0.000)***
First grade mean	0.24	11.4	0.07	0.90	0.97
Second grade mean	0.26	12.5	0.06	0.91	0.97
Observations	370,621	370,479	370,479	370,621	370,621
School-year FE	40,067	40,066	40,066	40,067	40,067
District clusters	823	823	823	823	823

Notes: All models estimated in this table include school-by-year fixed effects (FE). The standard errors reported in parentheses are robust to district-level clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. “New teacher” was defined as teachers with less than two years of experience. Sample sizes vary because experience was not available for all teachers.

Table 3.4: Sensitivity Analysis of Graduate-degree Results

Grades used:	1 st and 2 nd	1 st and 2 nd	1 st and 2 nd	2 nd and 3 rd	K and 1 st
Specification:	Baseline LPM 1	Restricted LPM 2	FE Logit Coeff. 3	LPM 4	LPM 5
High-grade	0.028 (0.004)***	0.033 (0.005)***	0.178 (0.026)***	0.009 (0.004)**	-0.020 (0.004)***
High-grade*99/00	-0.002 (0.004)	-0.002 (0.005)	-0.016 (0.024)	-0.002 (0.003)	0.003 (0.006)
High-grade*00/01	-0.006 (0.004)	-0.008 (0.005)	-0.044 (0.024)*	0.006 (0.004)	0.003 (0.005)
High-grade*01/02	-0.013 (0.005)***	-0.015 (0.005)***	-0.087 (0.028)***	0.006 (0.004)	0.010 (0.006)*
High-grade*02/03	-0.010 (0.005)**	-0.011 (0.005)**	-0.069 (0.029)**	0.006 (0.005)	0.005 (0.005)
High-grade*03/04	-0.014 (0.005)***	-0.017 (0.005)***	-0.098 (0.028)***	0.002 (0.005)	0.011 (0.006)*
High-grade*04/05	-0.014 (0.005)***	-0.017 (0.006)***	-0.098 (0.029)***	0.005 (0.005)	0.014 (0.006)**
High-grade*05/06	-0.014 (0.006)**	-0.016 (0.007)**	-0.098 (0.032)***	0.001 (0.005)	0.022 (0.005)***
Constant	0.241 (0.002)***	0.278 (0.002)***	.	0.260 (0.001)***	0.250 (0.002)***
High grade mean	0.24	0.28	0.28	0.26	0.25
Low grade mean	0.26	0.30	0.30	0.27	0.24
Observations	370,621	315,804	315,804	368,933	367,643
School-year FE	40,067	31,044	31,044	40,372	40,610
District clusters	823	667	667	824	845

Notes: Column 1 is identical to column 1 in table 3.3; it is repeated to facilitate comparison with the restricted-sample LPM in column 2. The sample restriction in column 2 mimics that of the FE-logit, which drops observations from school-years in which there was no variation in the dependent variable. This is why the sample sizes in this table vary. All models estimated in this table include school-by-year fixed effects (FE). The standard errors reported in parentheses are robust to district-level clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

APPENDIX 3.2

CHAPTER 3 FIGURES

Figure 3.1A: Fraction Grad. Degree

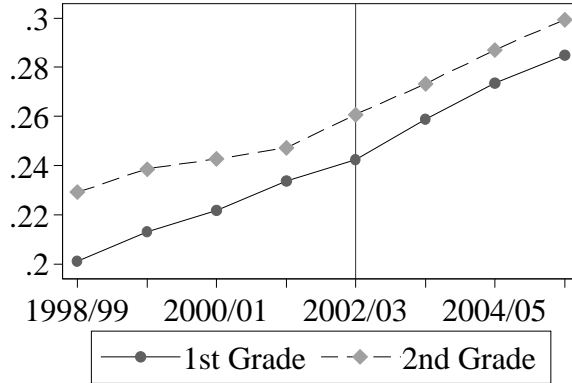


Figure 3.1B: Average Experience

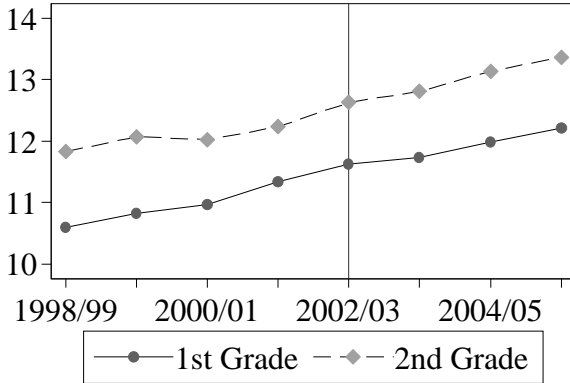


Figure 3.1C: Fraction Inexperienced

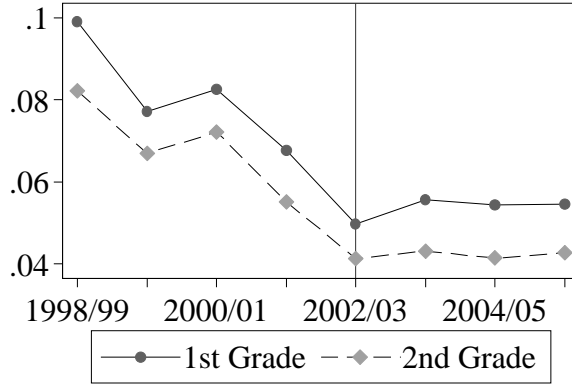
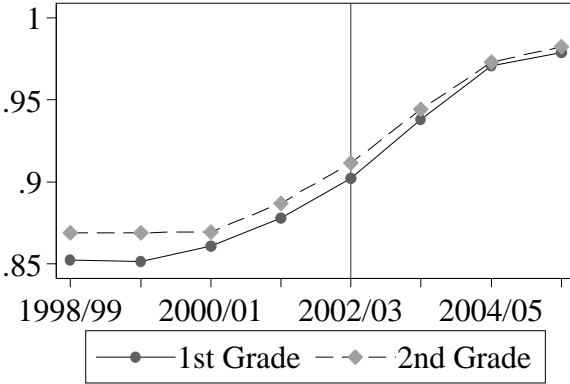


Figure 3.1D: Fraction Full Credential



APPENDIX 3.3

FE LOGIT COEFFICIENTS

Table A3: FE Logit Coefficients

VARIABLES	Graduate 1	New teacher 2	Full Cred. 3	Elem. Lic. 4
Second	0.178 (0.026)***	-0.194 (0.035)***	0.139 (0.062)**	-0.042 (0.063)
Second*99/00	-0.016 (0.024)	0.059 (0.050)	0.015 (0.025)	0.047 (0.070)
Second*00/01	-0.044 (0.024)*	0.031 (0.051)	-0.012 (0.044)	-0.032 (0.071)
Second*01/02	-0.087 (0.028)***	-0.034 (0.045)	-0.002 (0.044)	-0.036 (0.092)
Second*02/03	-0.069 (0.029)**	0.007 (0.060)	-0.004 (0.057)	0.040 (0.080)
Second*03/04	-0.098 (0.028)***	-0.102 (0.055)*	-0.013 (0.064)	-0.023 (0.106)
Second*04/05	-0.098 (0.029)***	-0.093 (0.070)	-0.079 (0.066)	0.104 (0.096)
Second*05/06	-0.098 (0.032)***	-0.053 (0.057)	0.011 (0.073)	0.043 (0.086)
Observations	315,804	154,879	153,582	52,634
School-Years	31,044	13,682	13,055	4,901
Districts	667	658	520	366
Pseudo R ²	0.001	0.002	0.001	0.0001
Log Likelihood	-129294	-44643	-55881	-15149

Notes: All models estimated in this table include school-by-year fixed effects (FE). The standard errors reported in parentheses are robust to district-level clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. “New teacher” was defined as teachers with less than two years of experience. Sample sizes vary both because experience was not available for all teachers and because the FE-logit estimator drops observations from school-years that did not experience any variation in the dependent variable (i.e., all 0 or all 1).

CHAPTER 3 REFERENCES

CHAPTER 3 REFERENCES

AIR. *See* American Institutes for Research.

American Institutes for Research. 2003. *Evaluation Study of the Immediate Intervention/Underperforming Schools Program and the High Achieving/Improving Schools Program of the Public Schools Accountability Act of 1999*. Washington DC: American Institutes for Research.

———. 2005. *Evaluation Study of the Immediate Intervention/Underperforming Schools Program of the Public Schools Accountability Act of 1999*. Washington DC: American Institutes for Research.

Angrist, J., and S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricists' Companion*. Princeton, NJ: Princeton Univ. Press.

Borjas, G. 2008. *Labor Economics*. 4th ed. New York, NY: McGraw-Hill.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff. 2008. The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences? *Public Finance Review* January 36(1): 88-111.

Carnoy, M., and S. Loeb. 2002. Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis* 24(4): 305-331.

Carroll, S., R. Reichardt, and C. Guarino. 2000. *The distribution of teachers among California's school districts and schools*. Santa Monica, CA: RAND Corporation.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. 2011. How does your kindergarten classroom affect your earnings? Evidence from Project Star. NBER Working Paper No. 16381.

Chingos, M. M., and M. R. West. 2011. Promotion and reassignment in public school districts: How do schools respond to differences in teacher effectiveness? *Economics of Education Review* 30(3): 419-433.

Clotfelter, C.T., H. F. Ladd, J. L. Vigdor, and R.A. Diaz. 2004. Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management* 23(2): 251-271.

Cullen J. B., and R. Reback. 2006. Tinkering toward accolades: School gaming under a performance accountability system. In *Improving School Accountability: Check-ups or choice?* (*Advances in Applied Microeconomics, Vol. 14*), ed. T. Gronberg and D. Jansen, 1-34. Amsterdam: JAI Press.

- Daly, A. J., and J. Chrispeels. 2005. From problem to possibility: Leadership for implementing and deepening the process of effective schools. *Journal for Effective Schools* 4(1): 7-25.
- Dee, T. S., and B. Jacob. 2010. Impact of NCLB on students, teachers, and schools. Fall 2010 Brookings Papers on Economic Activity.
- Figlio, D. N. 2006. Testing, crime, and punishment. *Journal of Public Economics* 90(4-5): 837-851.
- Figlio, D.N., and L. S. Getzler. 2006. Accountability, ability and disability: Gaming the system? In *Improving School Accountability: Check-ups or choice? (Advances in Applied Microeconomics, Vol. 14)*, ed. T. Gronberg and D. Jansen, 35-49. Amsterdam: JAI Press.
- Figlio, D. N., and J. Winicki. 2005. Food for thought: The effects of school accountability plans on school nutrition. *Journal of Public Economics* 89(2-3): 381-394.
- Goe, L. 2006. Evaluating a state-sponsored school improvement program through an improved school finance lens. *Journal of Education Finance* 31(4): 395-419.
- Hamilton, L.S., B. M. Stecher, and K. Yuan. 2008. *Standards-based reform in the United States: History, research, and future directions*. RAND Corporation Report Number 1384. Santa Monica, CA: RAND.
- Hannaway, J., and Hamilton, L. 2008. *Effects of Accountability Policies on Classroom Practices*. Washington, DC: The Urban Institute.
- Heckman, J. J., and D. V. Masterov. 2007. The productivity argument for investing in young children. *Review of Agricultural Economics* 29(3): 446-493.
- Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz. 2010. The rate of the return to the HighScope Perry Preschool Program. *Journal of Public Economics* 94(1-2): 114-128.
- Jacob, B.A. 2005. Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics* 89(5-6): 761-796.
- Jacob, B. A., and S. Levitt. 2003. Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics* 118(3): 843-877.
- Lee, J., and K. K. Wong. 2004. The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal* 41(4): 797-832.
- Luna, C., and C.L. Turner. 2001. The impact of the MCAS: Teachers talk about high-stakes testing. *The English Journal* 91(1): 79-87.

Phillips, M., and J. Flashman. 2007. How did the statewide assessment and accountability policies of the 1990's affect instructional quality in low-income elementary schools? In *Standards-Based Reform and the Poverty Gap*, ed. A. Gamoran, 47-90. Washington, DC: Brookings Institution Press.

Reback, R., J. Rockoff, and H. L. Schwartz. 2011. Under pressure: Job security, resource allocation, and productivity in schools under NCLB. NBER Working Paper No. 16745.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd Ed. Cambridge, MA: MIT Press.