Performance Pay and Teacher Selection: Do performance pay programs attract higher-ability teachers?

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Matthew D. Hendricks

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Morris M. Kleiner, Adviser

July 2011

I would like to thank the members of my committee for their help during this project and the numerous teachers, mentors, and family members that have helped me reach this point in my career.

For my parents, Stephen and Cheryl.

Abstract

Many public school districts are beginning to implement performance pay programs that provide teachers the opportunity to earn pay bonuses based on measures of teaching performance. Despite the growing number of districts offering these programs, we know little about their effects. Theory suggests that performance pay programs may provide an incentive for teachers to work harder (an effort effect). In addition, districts offering performance pay may attract teachers with higher average ability (a selection effect). Empirical work investigating the existence of these effects is mixed in the case of the effort effect and nonexistent in the case of the selection effect. This study is the first to attempt to empirically test for the existence of a teacher selection effect resulting from performance pay programs. I show that the existence of a selection effect may be revealed in differences in total pay and bonus probabilities between teachers who self-select into performance pay programs and teachers who are exogenously assigned. If self-selectors earn a higher expected total pay or are more likely to earn a bonus than an exogenously assigned teacher, I show that this implies that self-selectors are higher ability teachers on average. I test for this difference in cross-sectional data from the Schools and Staffing Survey (SASS) and in longitudinal data from a single performance pay district in Minnesota. In each case, I fail to find evidence of a selection effect. In the cross-sectional analysis, I find that while self-selectors earn a higher expected total pay, they are also less likely to earn a performance bonus. In the longitudinal analysis, I find that teachers who joined a district in the few years prior to its adoption of performance pay are not measurably different from teachers who joined in the years after adoption. Post performance pay joiners are not measurably different in terms of their education, experience, or likelihood of earning a performance bonus. While I fail to find evidence of a selection effect, that should not be taken as proof that performance pay programs, in general, do not produce their advertised benefits. The SASS analysis relies on several strong assumptions that potentially undermine the credibility of the results. For the analysis of the performance pay district in Minnesota, this study's inability to find evidence of a selection effect is likely a result of the district's high relative base compensation and nearly guaranteed bonus award.

Contents

List of Tables

Chapter 1

Introduction

Single salary or "steps and lanes" teacher pay schedules have served as the primary method of teacher pay determination in U.S. public school districts since at least the early 1950s (Podgursky and Springer, 2007). In the year 2000, Podgursky (2007) reports that roughly 96 percent of public school districts paid teachers according to a single salary pay schedule, in which teacher pay is entirely determined by education and experience. In the "step and lanes" pay schedule, within a school district, teachers with the same education and experience earn the same salary. High performers earn the same as low performers. Teachers in minority schools earn the same salary as those in affluent schools. Teachers in hard to staff subjects earn the same salary as those in easily staffed subjects.

In recent years, proponents of education reform have advocated policies aimed at improving the incentive structure for teachers. Critics of the single salary schedule argue that it restricts districts from offering pay incentives for teaching performance, and it prohibits a district from offering pay incentives to attract teachers to under-achieving schools and hard-to-fill subject areas. Viewed by advocates as a way to provide stronger incentives and attract higher ability teachers, teacher performance pay programs have gained political momentum.

The number of public school districts adopting teacher performance pay policies in the United States is substantial and growing. Funding and support of performance pay programs has come from nonprofit organizations and from local, state, and federal governments. Minnesota, Texas, and Florida have implemented some of the largest state-level performance pay initiatives. In Minnesota, 15% of the state's regular school districts (50 out of 333 districts) participated in the state's performance pay program, Q-Comp, during the 2010-2011 school year. In Texas, as of 2010-11, approximately $197 million in annual supplemental state funding is going to over 200 of the state's 1041 districts participating in the state performance pay program, DATE (Springer et al., 2010). The state initiatives in Minnesota and Texas combined with Florida's MAP program provide over $550 million annually in teacher pay incentives (Taylor et al., 2009).

The federal government is also promoting performance pay, particularly for high-needs schools. Through the Teacher Incentive Fund (TIF), the federal government provides five-year grants to aid states, districts, and charter schools in developing and implementing performance pay programs for teachers and principals. From 2006-2010, the federal government has awarded roughly $665 million through 92 grants.

Many of the TIF grants provide supplemental funding for large ongoing local performance pay initiatives. For example, TIF grants helped to fund district-level initiatives in Houston (the ASPIRE program), in Denver (ProComp), and in New York City (The Urban Excellence Initiative). In recent years, the prevalence of local performance pay initiatives has increased, in large part, because of the increased availability of funding and support for program development and implementation. In particular, in the 1990s, the National Institute for Excellence in Teaching (NIET) developed a model for teacher compensation reform - the Teacher Advancement Program (TAP). Since its development, NIET has helped implement TAP in over 66 public school districts and charter schools across the United States[1].

It is difficult to obtain an accurate number of districts in the United States that have adopted performance pay. However, using Schools and Staffing Survey data, the National Center for Education Statistics (NCES) estimates that 7.9% of public school districts in 2003 and 10.2% of public school districts in 2007 used pay incentives, such as bonuses, to reward excellence in teaching (NCES, 2010). These numbers suggest that the incidence of performance pay in public schools is growing and that the number of public schools employing performance pay is nontrivial, amounting to possibly more than 11% of districts in 2010.

Simply put, performance pay programs seek to link teacher pay to measures of teaching performance. In most performance pay programs, teacher pay is determined by a combination of student performance on standardized tests and professional evaluations. For example, some of the largest and well-known performance pay programs in the United States provide a framework for using subjective performance evaluations or student achievement measures to determine teacher pay: the Q-Comp program in Minnesota, the MAP program in Florida, and the DATE program in Texas.

While performance pay programs have gained support in recent years, they are still highly controversial. Opponents of performance pay argue that performance pay programs have several negative consequences for teaching and ultimately long-term student learning. Opponents argue that performance pay programs, especially those that reward bonuses based on relative teacher performance, erode cooperation among teachers. Also, critics argue that performance pay based on student test scores encourages teachers to "teach to the test", which they argue only improves students' test taking skills and reduces teacher effort devoted to important aspects of learning that are not tested[2]. Other prominent issues raised by opponents include

[1] For more information about TAP, go to http://www.tapsystem.org/

[2] This argument has theoretical support in Holmstrom and Milgrom (1991), which shows that performance pay contracts can distort incentives in multitasking environments.

concerns that performance pay will decrease equity and fairness in teacher pay, it will encourage racial or gender discrimination in teacher evaluation and pay, and it will erode a teacher's intrinsic motivation.

Proponents of performance pay argue that it will improve public education in two important ways. First, they argue that performance pay will encourage greater teacher effort and improve student outcomes. Linking performance pay bonuses to improved student achievement measures, they argue, will increase teacher effort and align teacher incentives with the goals of public education. Second, advocates of performance pay argue that by altering the teacher pay schedule and incentive structure, the profession may attract a higher quality teacher pool. That is, by rewarding good teachers with higher pay increases and greater advancement opportunity, the profession may encourage high talent individuals to pursue teaching careers who might have otherwise worked in higher paying jobs outside of education. The first potential effect mentioned above is often called the incentive or effort effect, while the second potential benefit of performance pay is often called a selection or sorting effect.

While advocates of performance pay cite these possible outcomes, evidence supporting these assertions is weak. Of the recent studies of performance pay programs, most focus on measuring effort effects and none, that I am aware of, attempt to estimate selection effects of performance pay. Evidence from other labor markets suggest that focusing on the effort effects alone may exclude as much as half of the potential benefit of performance pay. In addition to ignoring a major potential benefit of teacher performance pay programs, the literature has provided less than conclusive evidence of the size and existence of effort effects. While most studies have found positive effects of performance pay on measured student achievement, econometric concerns and evidence of teachers "teaching to the test" raise doubts as to whether this is valid evidence of increased teacher effectiveness. Clearly, more research in this area is warranted, specifically in the area of teacher selection effects. Obtaining more comprehensive measures of the effects of teacher performance pay programs will significantly improve our assessments of current performance pay programs, and it will aid policy makers in crafting and installing effective policies in the future.

This study begins to fill a major void in the literature by attempting to measure the effect of performance pay programs on teacher selection. That is, this research attempts to answer the following question: do performance pay programs attract a disproportionate number of high ability teachers? This research is motivated by several related studies that suggest the answer to this question may be yes. Studies of the effects of performance pay outside of the teaching profession suggest that when firms link pay to performance, existing workers become more productive and the firm eventually attracts a higher quality workforce. This study will attempt to apply and test prior theory regarding selection effects of performance pay as it relates to the current round of teacher pay reform. Specifically, this study uses variation in performance pay policies across districts and within a district over time to attempt to identify a causal relationship between performance pay and the type of teachers that choose to join the district. I intend to measure a difference in average teacher ability between teachers who self-select into performance pay districts and teachers who choose traditional school districts.

This work is organized as follows: chapter 2 begins with a literature review; chapter 3 provides the theoretical groundwork for this study; chapters 4 through 7 provide cross-sectional and longitudinal evidence; chapter 8 concludes.

Chapter 2

Literature Review

2.1 Incentive contracts and worker performance

There is a large body of literature examining the effect of performance pay on worker productivity in professions outside of teaching. Many studies have focused on the effects of piece rate pay schemes (pay that depends on worker output or performance measures) relative to salary pay (pay that is dependent on inputs of time and independent of worker output) in primarily blue collar jobs, where individual worker output is easily observed. In an influential study, Lazear (1986) develops a theory of the effects of piece rate pay schemes relative to time-rates. Lazear's theory highlights two potential sources of increased worker productivity in piece rate pay schemes: an effort effect and a selection effect.

The effort effect refers to worker productivity gains that result from an increase in the worker's optimal effort choice. Under time-rates, pay does not directly reward worker effort; therefore, the optimal behavior of a salaried worker is to exert the minimal amount of effort required to maintain employment. Lazear shows that in theory a worker's optimal choice of effort increases as the reward associated with effort rises. Piece rates or performance pay can increase the reward for effort, and with it, likely increase average worker effort and productivity.

The selection effect of piece rate pay refers to the firm's added ability to hire and retain high ability workers as result of the incentive pay scheme. Lazear argues that "The best workers select firms where performance has a payoff" (Lazear, 1986, p. 413). That is, piece rate pay policies may attract workers that are more productive on average. For example, workers who select into firms offering piece rates may be more willing to exert high effort (have a lower disutility of effort) or may have a high unobserved ability, which may not be fully rewarded in a firm that pays a time-rate based on observable characteristics.

While theory predicts how workers will respond to incentive contracts, the empirical literature provides evidence about how workers actually respond. A key question in the incentive contract literature is whether performance incentives matter. That is, do incentive contracts, such as piece rates or performance pay, influence worker performance? This question has been explored in several empirical studies, most of which confirm that workers generally do respond to incentives. For example, all of the following studies suggest that workers perform better under performance pay relative to pay regimes based on fixed payments or time-wages: Lazear (2000), Shearer (2004), Adams and Ferreira (2008), Banker et al. (1996), Fernie and Metcalf (1999), and Foster and Rosenzweig (1994).

Lazear (2000) measures a large productivity increase in an auto glass company after it switched from an hourly wage pay regime to piece rate pay. He finds that the switch to piece rates increased average output per worker by about 44%, and he attributes this to two effects: effort and selection. Lazear reports that as much as half of the productivity gain is attributable to the selection effect, which occurred primarily because the firm was able to hire a larger proportion of high quality workers under the piece rate regime.

Shearer (2004) finds that piece rates are associated with higher worker productivity in a Canadian tree planting firm. Shearer employs an experimental design, in which nine individual tree planters within the firm were randomly assigned to work under piece rates or fixed wages. Each person in the study was observed under both pay regimes. Shearer finds that workers were, on average, 20% more productive under piece rates relative to fixed wages. It is important to note that the design of this study prohibits Shearer from estimating any selection effects, since the workers were randomly assigned to each pay regime. Therefore, the productivity increase measured in this study is attributable primarily to an effort effect.

Banker et al. (1996) find that sales at retail stores are higher when the store pays it sales consultants according to a performance pay plan. Banker et al. use data from a major national retail firm that operates 34 department stores. Of the firm's 34 stores, 15 implemented a performance pay plan, which allowed individual sales personnel to earn a bonus payment (in addition to base salary) if he or she met a predetermined sales target. In this case, the bonus amounts represented about 20% of a worker's base salary. Also, prior to the performance pay plan, Banker et al. report that compensation for sales consultants within the firm was based primarily on seniority and not performance. Banker et al. compare sales in stores that implemented performance pay to stores that maintained the prior compensation scheme. They find that performance pay is associated with a roughly 10% increase in average store sales, and this difference in sales increased the longer the performance pay plan was in effect - suggesting a possible selection effect.

Fernie and Metcalf (1999) test the impact of performance incentives on professional horse racing jockeys in the United Kingdom. They examine whether jockey performance is influenced by the jockey's type of employment contract: a performance contingent contract or a fixed retainer unrelated to performance. Fernie and Metcalf find that individual jockey performance decreases when the jockey switches from a performance contingent contract to a large fixed retainer.

Foster and Rosenzweig (1994) examine the impact of time-wage versus piece rate pay schemes in rural farming households in the Philippines. They find evidence that, consistent with the findings above, workers exert more effort under a piece rate pay regime. The Foster and Rosenzweig study is unique in that they use more direct measures of worker effort as the outcome variable: namely caloric intake and body mass. They find that farmers under a piece rate contract lose 10% more body mass, conditional on caloric intake, than workers under a time-wage arrangement. Also, workers under the piece rate consumed more calories than workers in the time-wage arrangement. Both of these results suggest that Filipino farmers paid with piece rates exert higher energy or effort than farmers that are paid time-wages.

Adams and Ferreira (2008) find evidence that people respond to financial incentives even if the incentive payment is small and the agent claims she is "not in it for the money". Using data from corporate boards of directors from S&P 1,500 firms, Adams and Ferreira find that when directors are paid a higher fee to attend corporate board meetings, attendance at meetings improves. This result has two significant implications. First, the average payment for board meeting attendance is roughly $1,000. This financial incentive is minuscule compared to the income of the average director, who is typically a CEO or top executive. This suggests that even small incentives may have an impact on individual behavior. Second, while corporate directors often claim that they are not in their role for the money, this result suggests that they still respond to monetary incentives.

The incentive contract literature offers theoretical implications for the likely effects performance pay will have on teacher quality. The studies summarized above all show that, relative to firms that offer time-wages, firms that link pay directly to measures of worker output will potentially increase worker effort and attract higher quality workers. The results of these studies may be relevant to teacher performance pay, where teacher salary is determined by performance rather than experience and education alone. If performance pay affects teachers in similar ways found in the studies summarized above, we can expect that when a school district switches from a "steps and lanes" pay scale to performance pay that it may affect the quality of its teachers in two ways. First, by offering performance pay, the district may increase teacher motivation and effort. Second, the district may attract higher ability teachers after it adopts a performance pay regime.

In the next section, I describe the traditional "steps and lanes" pay schedule, which establishes little or no link between teacher pay and performance. In section 2.3 I describe some of the recent teacher performance pay programs in the U.S. and contrast these with the "steps and lanes" pay schedule. In section 2.4 I summarize the literature and evidence regarding the effect of performance pay programs on teacher effectiveness or quality.

2.2 The "steps and lanes" teacher pay schedule

In U.S. public schools, teacher pay is almost universally determined by a district level single salary or "steps and lanes" pay schedule, where teacher pay is determined almost entirely by years of teaching experience and education level. Responses to the 1999-2000 Schools and Staffing Surveys (SASS), a nationally representative sample of school districts and teachers in the United States, show that 96.3 percent of public school districts employ a single salary or "steps and lanes" teacher salary schedule, (Podgursky, 2007)[1].

This pay schedule gets its name from the manner in which teacher pay is determined. Teachers are placed on a pay grid where the step (row) is determined by their experience and the lane (column) is determined by the teacher's level of education. Each teacher in the district earns the same salary based on their corresponding cell

[1] In contrast to the 96.3% use of a salary schedule in public schools, Podgursky (2007) reports that 65.9% of private schools and 45.1% of nonreligious private schools employ a single salary pay schedule.

Table 2.1: Anoka-Hennepin district 11 “steps and lanes” pay schedule

2008-09 TEACHERS SALARY SCHEDULE

LANE	3	4	5	6	7	8	9	10	11
STEP	BA	BA+15	BA+30	BA+45	BA+60	MA	MA+15	MA+30	MA+45
1	37,231	38,076	38,832	39,702	NONE	42,142	42,933	44,045	44,780
2	38,271	39,327	40,179	41,112	NONE	43,951	44,985	46,065	47,155
3 & 4	40,144	41,277	42,328	43,380	NONE	46,217	47,469	48,800	49,907
5 & 6	41,587	42,836	43,909	45,035	NONE	48,659	49,868	51,493	52,634

Source: “Working Agreement by and Between Anoka-Hennepin Independent School District No. 11 School Board and Anoka-Hennepin Education Minnesota.” July 1, 2007-June 30, 2009.

assignment in the pay grid. The “steps and lanes” pay schedule was born out of the desire to make teacher pay equitable and has been a main stay in public education for over a half century. Sharpes (1987) reports that 97% of all schools had adopted a single salary schedule by 1950, which is remarkably similar to modern estimates of the prevalence of the steps and lanes pay schedule in public schools.

As an example of a typical single salary pay schedule, table 2.1 shows the first 6 steps of the 2008-2009 pay schedule negotiated between the Anoka-Hennepin School District 11 in Minnesota and the local teacher’s union affiliate, Anoka-Hennepin Education Minnesota.

A first year teacher in this district with a bachelor’s degree would be placed in step 1 and lane 3, earning an annual salary of $37,231. Similarly a teacher with 4 years of experience and a master’s degree plus 15 additional master’s credits would be placed in step 3&4 and lane 9, and earn an annual salary of $47,469. Each additional year of experience a teacher earns will place him or her in a higher step. Similarly, additional graduate level credits place teachers in higher lanes. It is important to note that this district salary contract makes no distinction between teachers of different subjects or grade. All teachers in the district, whether high school math or first grade art, are paid according to their cell assignment on the salary schedule.

Anoka-Hennepin’s pay schedule is typical of the “steps and lanes” system of pay employed by approximately 96% all school districts in the country. By construction, this system of pay is completely determined by education and experience and is not linked to any measure of teacher performance. Teachers earn raises with additional college degrees or credits and each year of experience they acquire. There is, however, a clause in the collective bargaining agreement for Anoka-Hennepin that requires satisfactory teacher performance to gain step increases. The collective bargaining agreement for Anoka-Hennepin school district states that "the annual (step) increment shall be contingent upon satisfactory service and evidence of growth on the part of staff members. The School Board may, upon administrative recommendation, withhold increases in salary if work is not satisfactory". There is, however, likely a significant cost of such action by the school board; therefore, a district's withholding of a step increase would likely occur in only the most obvious cases of poor teacher performance. Further, in these cases the teacher may be more likely terminated than simply barred a step increase.

This minimal performance standard is the only way in which teacher pay is dependent on performance in the "steps and lanes" system. While this standard appears similar to that faced by a typical salaried worker in the private sector, for teachers, the tenure system tends to dilute even this minimal performance standard. Many teachers are awarded tenure after two to five years of employment, which significantly lowers the teacher's probability of losing his or her job for poor performance[2]. Since teachers often obtain tenure after only a few years of experience, it is likely that teachers almost always face weaker external performance incentives than comparable workers in other professions.

Because of its perceived lack of incentives, the traditional single salary pay schedule has been a magnet for criticism. Critics argue that the "steps and lanes" pay

[2]Bridges and Groves (1984) report that teacher dismissals resulting from incompetency or poor performance are relatively rare. Tenured teachers are more likely dismissed for insubordination or illegal or immoral behavior.

schedule prohibits the district from rewarding high quality teachers with increased pay and restricts the district's ability to award higher salaries for positions that are consistently difficult to staff. In short, opponents maintain that the "steps and lanes" pay schedule lacks incentives and could cause teacher shortages, especially of high quality teachers with high earnings potential outside of teaching. As a result of these criticisms, many school districts and state governments are exploring ways to implement performance pay in an effort to improve incentives for teachers and improve the quality of the teacher pool.

2.3 Teacher performance pay programs

Teacher performance pay encompasses a wide range of policies that all have one thing in common – teacher pay is linked to a measure of performance. The are two crucial aspects to any teacher performance pay program. First, to link pay to teacher performance, designers must identify or create an accurate measure of teacher performance. Second, program designers must determine how to link the performance measures to pay rewards. In this section I describe how several recent performance pay programs have addressed each of these issues.

Accurately and objectively measuring teacher performance is difficult. This fact may explain why so few performance pay programs exist. It is relatively easy to observe the performance of an auto glass worker, a jockey, a tree planter, or a salesman. The output from these professions is readily observable. One can simply count the number of windshields fixed, trees planted, sales made, or wins. It is not, however, easy to observe or measure a teacher's output or performance.

Despite this difficulty, districts have adopted a range of methods designed to measure teacher performance. The most common performance measures employed in performance pay programs use one or both of the following: student test scores and principal or peer evaluations. The teacher's own students' achievement levels or gains on standardized tests is a popular performance measure employed in performance pay programs. Maybe the most popular teacher quality measurements rely on subjective assessments of teaching performance.

The use of student test scores in teacher performance evaluations is particularly controversial. By overemphasizing test scores, critics argue, teachers might devote less effort to equally important aspects of student learning that are not measured in standardized tests. This is the so called "teaching to the test" critique. Opponents of the use of test scores also argue that test scores are often not an accurate reflection of teacher quality since student achievement is influenced by factors that are outside of the teacher's control. Teacher performance assessments based on test scores are potentially distorted by nonrandom assignment of students to teachers and by random variability.

Proponents of the use of test scores highlight the need for objective measures of teacher performance. They argue that teacher performance measures based on subjective evaluations are potentially contaminated by error related to noisy observed signals, and the subjective nature of these evaluations can encourage outright bias or discrimination. Furthermore, Neal (2011) suggests that performance pay programs that rely too heavily on subjective evaluations tend to morph into teacher base pay increases that are not accompanied by improvements in performance. Arguments against the use of test scores are less relevant today with the development of detailed student level longitudinal data. These detailed data systems allow analysts to estimate teacher value-added models, which control for student characteristics and more accurately link student achievement gains to the performance of individual teachers.

For example, the baseline teacher value-added model is estimated with the following procedure. First, one estimates a child's expected annual test score gain, given her characteristics. Second, one estimates each student's residual achievement gain by subtracting her expected test score gain from her actual test score gain. Finally, a simple teacher value-added measure is obtained by averaging these residuals for the teacher's own students. If the average residual for a teacher's students is positive, then that suggests the teacher is contributing more to student learning than the average teacher would contribute to the same students. The advantage of value-added measures is that they control for determinants of student achievement that are beyond the teacher's control, including effects of socioeconomic status and ability. In addition, value-added measures are may not be highly sensitive to omitted student characteristics (Ballou et al., 2004) or nonrandom student assignment to teachers (Kane and Staiger, 2008).

While value-added methods are widely used in academic studies, few districts are employing these measures in high-stakes performance assessments. Instead, in current performance pay programs, teacher performance measures come primarily from student test score levels or gains at the teacher, school, or district level and subjective principal or peer evaluations.

Translating teacher performance measures into teacher pay is another point of contention for districts considering performance pay. How should teachers be rewarded for performance: through one time bonuses or permanent base salary increases? Should performance payments be offered at multiple levels of performance or awarded only if teacher performance exceeds a predetermined threshold? According to Podgursky and Springer (2007), most recent performance pay programs maintain the steps and lanes pay schedule and offer teachers the opportunity to earn additional one-time bonuses that are conditioned on annual performance measures. In what follows, I provide several examples of recent performance pay programs, all of which offer teachers a base salary and the opportunity to earn bonuses based on annual performance measures [3].

[3]For a state-by-state listing of current performance pay programs, see the National Center on Performance Incentives website: http://www.performanceincentives.org/statebystate_

In 2004, Denver Public Schools implemented the Professional Compensation System for Teachers (ProComp). The program awards teacher pay incentives through annual bonus awards in four components: knowledge and skills, professional evaluation, student growth, and market incentives. In 2010- 2011 teachers in ProComp were eligible for nine separate bonus awards in these components, some of which are permanent awards (base salary building).

In the knowledge and skills component, teachers can earn base salary increases for participating in professional development programs and obtaining advanced degrees. Within this component, teachers can also earn up to $4,000 in annual tuition and student loan reimbursement. In the second component, professional evaluation, non-probationary teachers are eligible for annual base pay increases of $1,127 for satisfactory annual subjective performance evaluations. In the market incentives component, teachers are eligible for two types of annual bonus awards of $2,403. A teacher may earn one or both of these annual bonus awards if he or she teaches in a hard to staff school or in a hard to staff position within a school.

Finally, teachers in ProComp are eligible for four bonus awards, one of which is a base pay increment, for elements within the student growth component. A teacher can earn a $376 permanent annual base pay increase if she meets two of her predetermined student growth objectives within an academic year. In addition, teachers may earn three annual bonus awards equal to $2,203, each with a separate individual or school-level performance requirement. A teacher in ProComp receives a $2,203 bonus for each of the following conditions: the teacher's own students' standardized test scores exceed district expectations, the teacher works in a school designated by Denver Public Schools as "top-performing", and the teacher works in a school designated by Denver Public Schools as "high-growth". In total, teachers in ProComp are eligible for bonus incentives of about $3,700 annually for individual performance and an

resources/index.aspx

additional $4,400 for school-wide performance.

In 2005, Minnesota adopted Quality Compensation for Teachers or Q-Comp. The Q-Comp program is a state-wide voluntary performance pay program that rewards participating school districts with increased funding up to $260 per student. To participate in Q-Comp, the district must design and implement a plan that provides teacher bonuses that link to student performance and subjective performance evaluations. In 2009, Q Comp awarded more than $49 million in supplemental state funding to participating districts and charter schools (Minnesota Office of the Legislative Auditor, 2009) and has awarded over $200 million since its inception (Sojourner et al., 2011). During the 2010-2011 school year, 50 of Minnesota's regular school districts (out of 333) and 54 charter schools participated in Q-Comp.

Q-Comp provides flexibility in how districts design their performance pay plans, but each district participant is required to align at least 60 percent of teacher bonuses to measures of student academic achievement and progress. The average Q-Comp plan offers teachers the opportunity to earn over $2,000 in annual bonuses if the teacher meets specific performance targets (Sojourner et al., 2011). Typically, Q-Comp districts award three types of performance bonuses: bonuses based on subjective performance evaluations, bonuses based on the teacher's own students' performance, and bonuses based on school-wide student performance measures. Current Q-Comp plans typically attach the highest bonus awards to subjective performance evaluations. For example, Sojourner et al. (2011) find that, in the average Q-Comp plan, subjective performance evaluations account for the largest bonus award, $1,100 on average. By meeting school-wide performance goals or teacher level performance goals, Sojourner et al. find that on average Q-Comp teachers can earn additional bonuses of $247 and $872 respectively.

In Texas, as of the 2010-2011 academic year, over $197 million in annual supplemental state funding is going to over 200 of the state's 1041 districts (over 19%)

participating in the state performance pay program, District Awards for Teacher Excellence (DATE) (Springer et al., 2010, B). Like Minnesota's Q-Comp program, DATE is a district level voluntary performance pay program available to all of the state's school districts. Districts choose whether to participate and which schools within the district will be eligible for the program. The DATE program provides two types of funds to participating districts. Part 1 funds (60% of the total district grant) are to be used for teacher incentives or bonus awards that link to student achievement. The state guidelines for DATE also suggest that Part 1 should be distributed based on quantifiable and objective measures of student achievement (Springer et al., 2010, B).The other 40% of district grants (Part 2 funds) can be used for additional incentive awards, professional growth activities, and data development. In the first year of DATE (2008-09), 42% of teachers in districts with district-wide incentive plans earned a Part 1 award and the average award was $1,361 (Springer et al., 2010, B). Springer et al. (2010, B) report that most districts designed plans that awarded teacher incentives using a combination of individual and group-based performance measures. Performance measures were typically based on student achievement levels or growth on state standardized assessments (Springer et al., 2010, C).

In 2007-2008, Florida implemented the Merit Award Program (MAP). MAP is a district-level voluntary performance pay program, like Q-Comp and DATE. Districts are free to join the program and develop a performance pay program that fits within the MAP parameters. Unlike Q-Comp, MAP requires that performance measures for teachers in tested subjects come from objective metrics, such as students' achievement on state, national, or locally produced tests. The MAP program requires that districts base 60 percent of teacher performance measures on the performance of students assigned to the teacher or students within the teacher's school or instructional team. The state also requires that districts award all of its top performing personnel with pay bonuses in the range of 5%-10% of the average district salary (Florida Department of Education, 2007).

Houston is home to one of the most innovative performance pay programs to date. The Accelerating Student Progress, Increasing Results and Expectations program (ASPIRE) began in 2007-08 and operates in 130 of Houston's "high-needs" or economically disadvantaged schools. Project ASPIRE is remarkable in that the program uses primarily value-added measures of teacher performance. The ASPIRE program utilizes a multivariate value-added model to objectively evaluate teacher and school performance, and it attaches large bonus amounts to these measures. Teachers can earn bonus awards of up to $10,300 in the program. ASPIRE is on the cutting-edge in terms of its use of advanced teacher performance assessments and may be a model for future performance pay programs that wish to rely more heavily on objective performance measures.

These programs provide a sense of the nature of performance pay initiatives that are developing around the country. While the programs differ in many respects, most offer pay incentives based on a mix of student performance measures and teacher evaluations. It is difficult to know the percentage of teachers that are currently employed under performance pay policies. However, using survey data from the SASS, we can get a sense of its prevalence. The SASS district survey includes the following question: "Does the district (school) use any pay incentives such as cash bonuses, salary increases, or different steps on the salary schedule to reward excellence in teaching". Podgursky (2007) reports that 5.5% of traditional public school districts and 42.9% of non-religious private schools indicated that they offer pay incentives for excellence in teaching during the 1999-2000 school year. More recent estimates from the National Center for Education Statistics (NCES) (2010) suggest that the prevalence of performance pay in public schools increased in just a few years. According to NCES estimates from the 2003-04 and 2007-08 SASS, about 7.9 percent of public school districts in 2003 and 10.2 percent of public school districts in 2007 offered pay incentives for excellence in teaching. Taylor et al. (2009) also suggest that the size of performance bonuses in the 2003-04 SASS are nontrivial, equaling \$2,005 (4.6% of base salary) on average.

The previous sections highlight the differences between the traditional teacher pay schedule and performance pay reforms. The traditional pay schedule rewards teachers' experience and education levels which have been found to have little or no relationship with student performance (Hanushek, 2003)[4]. In contrast, current performance pay initiatives seek to align teacher incentives with student performance measures. These initiatives intend to pay teachers according to a measure of their production or quality, a compensation plan that is similar to piece rate pay. The contrast in incentives between traditional teacher pay and performance pay initiatives is analogous to the contrast in incentives between piece rates and time-wages. This suggests that the results found in the piece rates versus time-wages studies may be relevant to teachers in "steps and lanes" districts versus performance pay districts. That is, teacher quality may be higher in districts offering performance pay as a result of sorting and effort effects.

2.4 The effect of performance pay on teacher quality

A common way of defining teacher quality is described in Hanushek (2003), where he states, "High quality teachers are ones who consistently obtain higher than expected gains in student performance, while low quality teachers are ones who consistently obtain lower than expected gains" (p. F90). According to Hanushek, measuring teacher quality is equivalent to measuring the teacher's value added to student performance.

[4] Hanushek (2003) summarizes all econometric estimates of education production functions published before 1995 that include controls for family inputs. He shows that of 170 estimates of the effect of teacher education on student performance, only 9% show a positive effect and 5% show negative effects; the rest show statistically insignificant results. Similarly, of 206 estimates of the effect of teacher experience on student performance, 29% show a positive effect and 5% show a negative effect. Further, many of these estimates may suffer from upward bias as a result of more experienced and highly educated teachers self-selecting into better schools.

Using this definition of teacher quality, many studies have attempted measure the effect of performance pay policies on teacher quality using student test scores or other student achievement measures. Below I discuss several of these studies and highlight the strengths and weaknesses of each in measuring the potential causal effects of performance pay policies on teacher quality. Specifically, I highlight their ability to identify selection and effort effects of performance pay. Most studies show a positive relationship between individual teacher pay incentives and targeted student achievement measures. However, not all studies find a relationship. In particular, a recent experimental study fails to find a relationship between student achievement and large teacher bonuses that were contingent on the same student achievement measures.

A growing number of studies have attempted to measure the causal relationship between student achievement measures and performance pay programs in the United States. Figlio and Kenny (2007) use cross-sectional data from the National Education Longitudinal Survey (NELS) and their own data collection of school personnel practices in 2000 to estimate the relationship between performance pay and student test scores. They find that in districts offering performance pay, student test scores are higher. However, the authors acknowledge that this study's reliance on cross-sectional estimates make it difficult to infer whether performance pay is increasing teacher quality or if higher quality schools are those that tend to adopt performance pay schemes. Without the use of other estimation techniques we cannot be sure that these estimates are capturing effort and sorting effects of performance pay or if they are capturing the sum of these effects and positive bias. For example, positive bias is introduced when schools or districts that are already performing above average are also more likely to implement performance pay plans.

Eberts et al. (2002) find evidence that teacher incentives aimed at specific student achievement measures are effective. In a simple difference and difference comparison of two Michigan high schools, Eberts et al. (2002) find that a teacher performance pay plan which rewarded teachers for increasing student retention achieved the desired goal. The school that offered the performance pay incentive increased student retention relative to a comparable school which maintained traditional teacher pay. This study suggests that performance pay may influence teachers by focusing their effort toward achieving outcomes that are directly rewarded by the performance pay policy. However, the small sample size (two schools) severely limits our ability to draw broad conclusions from this study's results.

Barnett et al. (2007) examined the effect of a Little Rock, Arkansas performance pay program (the Academic Achievement Pilot Project (ACPP)) on student achievement. Active in 5 Little Rock elementary schools during the 2006-07 school year, the ACPP provided teacher financial incentives based solely on student one year gains on the Stanford Achievement Test Version-9 (SAT-9) (Barnett et al., 2007). The teacher pay incentives in the ACPP were awarded as bonuses contingent on the number of students achieving gains in their SAT-9 score and the size of the gains. They found that one-year average test score gains for students in ACPP schools were greater than the average test score gains of students in comparable traditional pay schools. While this study shows evidence of a positive teacher effort effect, the study is small in scope, covering only 89 elementary students in ACPP schools. Such a small sample makes it difficult to accurately infer the effects of ACPP type policies in other schools or states. Also, the study does not attempt to measure any of the potential long-term teacher selection effects of performance pay.

Ladd (1999) finds that a performance pay program introduced in Dallas in 1991 had positive effects on student test scores and reduced dropout rates. The Dallas program was a group level performance pay program, which rewarded all teachers of the same school pay bonuses based on the school's success in increasing student test score measures beyond what would be expected, given the school's student body characteristics (race, gender, and socioeconomic status). Ladd (1999) reports that students in the performance pay program had a higher pass rate on the Texas Assessment of Academic Skills test and lower high school drop-out rates relative to students in comparable schools not in the program. Again a major disadvantage of this study is its small scope, which limits the study's policy implications. Also, while Ladd examines the effects of the performance pay program over a five year period (a time length long enough to observe teacher turnover), she makes no attempt to decompose the positive effect on student outcomes into teacher sorting and incentive effects.

One of the most recent and ambitious studies of performance pay in the United States found less than encouraging evidence regarding the incentive effects of performance pay. Springer et al. (2010, A) report results from a three-year controlled experiment of performance pay in Nashville, Tennessee - the Project on Incentives in Teaching (POINT). The purpose of the POINT experiment was to measure whether offering large financial incentives to math teachers improves student performance on standardized mathematics exams. The POINT study followed a controlled experimental design, in which teachers were randomly assigned to treatment and control groups. Individual teachers in the treatment group were eligible for bonuses up to $15,000 if the teacher's students performed above a predetermined performance standard[5]. Teachers in the control group were not eligible for bonuses. The results of the experiment failed to find evidence of an incentive effect. Students assigned to teachers in the treatment group failed to outperform students assigned to control group teachers.

The POINT study also failed to find evidence of a selection effect. Although teachers were randomly assigned to the treatment group in the first year of the experiment, attrition from the experiment was high. Half of the participants had exited

[5]Performance standards in POINT were set by historical teacher performance. For example, teachers were eligible for the $15,000 bonus if annual student performance was at a level that had historically been achieved by only the top 5% of teachers. Similarly, teachers were eligible for bonuses of $5,000 and $10,000 if annual student performance was at a level that had historically been achieved by only the top 20% and 10% of teachers respectively.

the experiment by the third year. This attrition provides an opportunity to estimate a selection effect if teachers who remained in the treatment group appear to be of higher quality. That is, more effective teachers may be less likely to exit the treatment group, which would be consistent with a selection effect associated with performance incentives. While the potential to measure a selection effect existed in the POINT study, Springer et al. fail to find evidence of differential attrition from the treatment group. Throughout the three-year study, differences in observable characteristics and measures of prior teacher performance between teachers in the control and treatment groups remain constant, contrary to what would be expected given a selection effect.

More experimental evidence of the effects of performance pay on teacher quality has come from studies outside of the United States. Two important studies used recent performance pay initiatives in Israel to identify the effects of teacher incentives on student achievement.Lavy (2002) and Lavy (2009) suggest that performance pay incentives for teachers and schools in Israel led to gains in several student outcomes, including higher average test scores and lower dropout rates[6]. Lavy (2002) estimates the effects school-based incentives after the first two years of the implementation of a performance pay policy. Since two years is long enough to have some teacher turnover, it is possible that these estimates are capturing the positive effects of both teacher sorting and effort effects, but it is likely that most of the measured effect is a result of increased teacher effort. Lavy (2002) does not investigate the mechanism by which teacher incentives have produced higher student achievement. Lavy (2009) does investigate this mechanism and finds that individual teacher incentives raised student achievement largely as a result of increased teacher effort, reflected in overtime devoted to student instruction after the regular school day. However, Lavy (2009) does not attempt to measure any potential long-term sorting effects resulting from individual teacher incentives.

[6] Lavy (2002) investigates the effects of school level teacher pay incentives and Lavy (2009) examines the effect of individual teacher pay incentives.

Muralidharan and Sundararaman (2009) present evidence from an experimental study of performance pay in India. In their study, a representative sample of Indian schools in the state of Andhra Pradesh were randomly assigned to a teacher performance pay program that awarded teacher bonus payments based on average improvement in student test scores in math and language. Muralidharan and Sundararaman (2009) find that in schools with the incentive program, student scores were significantly higher in the incentivized math and language tests, but also higher in science and social studies which suggests there may be spillover effects. Based on teacher interviews, Muralidharan and Sundararaman suggest that much of the increase in student achievement can be attributed to higher teacher effort in the schools participating in the incentive program. They report that teachers in these schools were more likely to assign additional homework and class work, provide practice tests, and provide after school classes.

Like the Muralidharan and Sundararaman study, Glewwe et al. (2003) report the results of a randomly assigned incentive pay program that rewarded teachers with bonuses based on student test scores. They present evidence that performance pay in Kenya raised student test scores, but they argue that much of this gain was a result of increased teacher effort to raise short-run test scores rather than to increase long-term learning. They find that teachers in performance pay schools were more likely to offer test preparation sessions, but did not significantly alter other teaching methods. Also, while students that attended performance pay schools scored higher on exams during the program, Glewwe et al. (2003) find that after the program ended these students performed no better than students in the control schools.

From the studies presented above, we have some evidence that teacher pay incentives are associated with higher teacher quality, reflected in higher student achievement. However, the literature is incomplete and is far from conclusive. The studies of recent reforms in the U.S. either suffer from potential econometric problems or are focused on local programs, which are likely too small in scope to accurately represent the effects that performance pay policies may have on a national scale.

Furthermore, no prior studies I am aware of attempt to measure the potential sorting effects of performance pay, instead most focus on measuring incentive or effort effects. Some of these studies, the experimental studies in particular, are too short-lived and follow the same population of teachers before and after the adoption of performance pay programs. This research design eliminates the possibility of measuring any teacher selection effects. Some of the best designed experimental studies focus only on measuring effort effects of performance pay, while other studies have the potential to measure both effort and sorting effects, but for some reason fail to estimate them separately. Clearly, more evidence of the effect performance pay policies is needed, especially for the performance pay programs that are currently being adopted within the United States. Measuring these effects would greatly improve our understanding of the short and long-term mechanisms by which performance pay policies might affect teacher quality. As highlighted in related literature on piece-rate pay, the literature's focus on only incentive effects could be missing up to half of the overall positive effects of performance pay on teacher quality. This study contributes to the existing literature by attempting to measure teacher selection effects of performance pay programs.

Chapter 3

Theory

A major challenge of any empirical study measuring the effect of performance pay is to find an accurate measure of teacher quality. In prior studies of performance pay, the most popular measures of teacher quality rely on student outcome measures: standardized test scores, drop-out rates, and others. This study breaks with tradition and employs a unique method of measuring teacher quality: total pay or bonus probabilities. As described above, performance pay policies seek to link teacher pay to teacher quality. Assuming that performance pay policies are successful in this regard, we can use a teacher's pay in a performance pay district as a measure of relative teacher quality. That is, if schools are successful in both measuring teacher quality and awarding higher pay to higher quality teachers, then a teacher's performance pay salary should be an accurate reflection of the school's own teacher quality measurement. Given this assumption, it follows that teachers who are more likely to earn a performance bonus also higher quality teachers.

The use of a teacher's performance pay salary (or bonus probability) as a measure of teacher quality has some advantages over using student test scores. A teacher's total pay in a performance pay district is likely a more accurate measure of teacher quality than student test scores alone, because it is a reflection of the school's own evaluation of teacher quality, which usually depends on a combination of student test score gains and professional evaluations.

There is also an econometric advantage of using total pay over using only student test scores. Relative to student test score models, models that use teacher total pay as a measure of teacher quality are not as dependent on student body and family characteristics, which are often missing in large representative samples. Unobservables at the student level, such as family upbringing, intelligence level, and motivation, are crucial in estimating student test scores. However, these covariates are potentially not as crucial to an analysis of a teacher's total pay in a performance pay district, since it is often awarded after providing some account for student body characteristics.

For example, most performance pay programs base a majority of teacher pay on on principal evaluations. Principals are likely to take into account the characteristics of the children in the classroom when evaluating teacher performance. As a result, principal evaluations of teacher performance are likely to be less dependent on student characteristics. Also, some performance pay districts are employing value-added measures in test score based evaluations of teacher performance, which provide some account for student characteristics[1]. For these reasons, models that use total pay as a measure of teacher quality are potentially less exposed to omitted variable bias than are measures of teacher quality using a cross-section of student test scores.

Because of the wide availability of teacher pay data and the advantages described above, this study uses total pay in performance pay districts to attempt to measure the effect of performance pay programs on teacher sorting. The foundation of this study is that total pay or bonus probabilities provide a measure or indicator of relative teacher quality. For example, if teachers in group A are more likely to earn a performance bonus than teachers in group B then, assuming that performance bonuses are more likely awarded to higher quality teachers, it follows that group A teachers are higher

[1] See for example the ASPIRE program in Houston, Texas.

quality on average than teachers in group B.

This type of comparison is the basic research design of this study. What if we could compare the performance pay salary of teachers who self-select into performance pay districts to that of a random sample of teachers? Suppose we find that the total pay for teachers who self-select is higher, on average, than the expected total pay of a random sample. Then, given the assumption that total pay is higher for higher quality teachers, we could conclude that teacher quality among self-selectors is higher, on average, than the teacher quality of a random sample of teachers. This conclusion implies a sorting effect. This research will investigate whether this sorting effect exists in the data by attempting to measure differences in total pay and bonus probabilities across self-selectors and exogenously assigned teachers.

While this research design seems simple in theory, implementing it is complicated by the fact that we only observe a teacher's total pay in a performance pay district if she actually taught in a performance pay district. We do not observe total pay for those teachers in traditional "steps and lanes" districts. This complication makes it difficult, but not impossible, to obtain estimates of the expected total pay of a random sample or exogenously assigned group of teachers. I delay discussion of the techniques that allow one to implement this research design to chapter 4. In the next section, I formalize the theory behind this design.

In the theory that follows, I develop the conditions under which performance pay programs attract only high ability teachers. I develop a basic model of teacher behavior assuming teachers are risk averse, teachers have full knowledge of their own ability, and districts have an imperfect measure of teacher quality. In sections 3.1 through 3.4 I describe the model's assumptions, characterize the teacher's optimal effort and district choice, and describe how these decisions are related to teacher ability. In section, 3.5 I define the selection effect and describe the conditions under which a selection effect exists within this framework. Finally, in section 3.6 I highlight the implications of this theoretical model for performance pay program design. This model shows that under some conditions, districts offering performance pay attract only teachers from the upper tail of the ability distribution. Further, this framework suggests a unique empirical technique that can be used to test for the existence of a selection effect, which I discuss in chapter 4.

3.1 Model assumptions

I assume teacher quality is a function of teacher i's effort (e_i) and ability (a_i): $Q = Q(e_i, a_i)$. This quality function is the same across all teachers, subjects, and schools. Further, I assume that teacher quality is increasing in both effort and ability and that for higher levels of teacher effort, teacher quality increases at a lower rate. This assumption captures the intuition that, for teachers that are exerting low effort, a little extra effort goes a long way in increasing teacher quality, but for teachers that are already exerting high effort, a small increase in effort only slightly increases teacher quality. Also, ability and effort can be either complements or substitutes in teacher quality. In mathematical language these assumptions imply the following results hold for the first and second partial derivatives of $Q : Q_e > 0$, $Q_a > 0$, $Q_{ae} \geq 0$ and $Q_{ee} < 0$.

I assume neither effort nor ability are observable to the school or district at any cost but districts are able to obtain a noisy teacher quality measure (Q^m) and perfectly determine whether an individual's teacher quality is above a minimum standard ($\underline{S}$). That is, I assume that districts can perfectly identify teachers whose quality level has dropped below some minimal standard ($\underline{S}$), but the district has a more difficult time assessing teacher quality at higher levels. Districts must rely on stochastic teacher quality assessments to measure higher levels of teacher quality. The noisy teacher quality measure is defined as the following: $Q^m = Q(e_i, a_i) + \mu$, where $Q(e_i, a_i)$ is the teacher's true teaching quality and μ is a random error term. I assume that μ has cumulative distribution function $G()$ and probability density function $g()$.

There are two types of school districts, performance pay districts and steps and lanes districts. These district types award teacher compensation differently. Steps and lanes districts award teachers a fixed pay if the teacher meets a minimal teacher quality standard, which is determined without error. Performance pay districts also offer teachers a base salary if they meet the minimal teacher quality standard. However, in performance pay districts, teachers have the opportunity to earn a pay bonus (B) if a noisy teacher quality measure (Q^m) is above some high performance standard (S).

Teacher pay in each district type is a function of teacher quality, which is a function of teacher ability and effort. As such, teacher compensation in each district type depends on teacher effort and ability. Denote teacher i's pay in steps and lanes district j as $Y_{ij}^s(e_i, a_i)$, and denote teacher i's pay in performance pay district j as $Y_{ij}^p(e_i, a_i)$. These compensation functions are the following:

$$Y_{ij}^s(e_i, a_i) \equiv y_{ij} = y(\mathbf{w}_i, \mathbf{v}_j) \text{ if } Q(e_i, a_i) \geq \underline{S}$$
$$Y_{ij}^p(e_i, a_i) \equiv \begin{cases} y(\mathbf{w}_i, \mathbf{v}_j) + B & \text{if } Q^m \geq S \text{ and } Q(e_i, a_i) \geq \underline{S} \\ y(\mathbf{w}_i, \mathbf{v}_j) & \text{if } Q^m < S \text{ and } Q(e_i, a_i) \geq \underline{S} \end{cases}$$

As shown above, pay in the steps and lanes district is determined by the function $y(\mathbf{w}_i, \mathbf{v}_j)$, and teachers receive this compensation only if their true teacher quality ($Q(e_i, a_i)$) is at least meeting the minimal quality standard ($\underline{S}$). I assume that the minimal teacher quality standard is identical across districts. The function $y(\mathbf{w}_i, \mathbf{v}_j)$ is a base compensation function. Teacher i's base compensation in district j is determined by a two variable vector of teacher i's education and experience level ($\mathbf{w}_i$) and a vector of district j's characteristics ($\mathbf{v}_j$).

Compensation in performance pay district j is slightly different than compensation in a steps and lanes district. In the performance pay district a teacher must still meet the minimal quality standard in order to earn any compensation ($Q(e_i, a_i) \geq \underline{S}$). However, in the performance pay district, teachers can earn a bonus (B) in addition to their base compensation ($y(\mathbf{w}_i, \mathbf{v}_j)$) if a noisy teacher quality measure (Q^m) is at or above some high performance standard (S). Again, I assume that the high quality standard (S) is identical across performance pay districts.

In this model, teachers maximize their expected utility choosing their effort level and district type. I assume that teachers choose their effort level with full knowledge of their own ability level (a_i). A teacher's job satisfaction, or utility, in district j is determined by the amount of money she earns in district j, the characteristics of district j, and the amount of effort she exerts on the the job (e_i). Specifically, teacher utility is determined by the following function:

$$U(Y_{ij}^t(e_i, a_i), \mathbf{v}_j, e_i) \equiv u(Y_{ij}^t(e_i, a_i)) + f(\mathbf{v}_j) - C(e_i) \text{ where } t = s \text{ or } p$$

This utility function has three components: utility derived from teacher pay, utility derived from the characteristics of the district in which teacher i works, and disutility derived from effort exerted on the job. The three components have three associated functions: $u()$, $f()$, and $C()$. I assume that the functions $u()$ and $C()$ are strictly increasing, such that their first derivatives are positive. Agents in this model are risk averse so that the function $u()$ is strictly concave (ie. $u''() < 0$). Also, I assume that the function $C()$, which represents the cost of effort, is an increasing, strictly convex function such that $C' > 0$ and $C'' > 0$.

Given these assumptions about teacher utility and compensation, the teacher's problem is to choose the district type and effort level that maximizes her utility. That is, the individual chooses a performance pay district or steps and lanes district and an optimal effort level. This decision can be broken into two stages. First, each individual chooses his or her optimal effort in each district. Second, the teacher chooses the district type that gives her the highest utility evaluated at the optimal effort choice.

3.2 The teacher's optimal effort

In this model, a teacher's optimal effort depends on the district type under consideration. As I have characterized the compensation scheme and individual utility function, a teacher is not rewarded for high effort in a steps and lanes type district. In a steps and lanes type district, a teacher has no incentive to exert more effort than is necessary to achieve the minimal quality standard $\underline{S}$. Therefore, teacher i's optimal effort in a steps and lanes type district is $\underline{e}_i$, which is the effort level for individual i that sets his or her teacher quality to the minimum standard. That is, the minimal effort level $\underline{e}_i$ solves the following: $Q(\underline{e}_i, a_i) = \underline{S}$.

In any arbitrary performance pay district j, the teacher's optimal effort is the effort level that maximizes his or her expected utility in district j:[2]

$$\max_{e} \left\{ E\left[u\left(Y_{ij}^{p}(e_i, a_i) \right) \right] + f(\mathbf{v}_j) - C\left(e_i\right) \right\} \text{ subject to } Q(e_i, a_i) \geq \underline{S}$$

This problem has two types of solutions. A teacher may choose to exert low effort $\underline{e}_i$, as defined above, or choose a higher effort level $e_{ij}^{*} > \underline{e}_i$ and increase the probability that she will receive a performance bonus. In the performance pay district, teachers are rewarded for high effort by an increased likelihood of receiving a bonus. The harder a teacher works in the performance pay district, the greater the probability that the teacher will receive a performance bonus. The question is, what determines whether the teacher exerts high or low effort in the performance pay district?

[2]Given the previously stated assumptions about the teacher quality function and cost of effort, this objective function can be shown to be strictly concave in effort.

The teacher chooses high effort (e^*_{ij}) if her expected utility under high effort is greater than her expected utility under low effort. In other words, if the teacher's marginal expected utility is positive when evaluated at the minimal effort level ($\underline{e}_i$), then the teacher will choose high effort. That is, the teacher chooses high effort (e^*_{ij}) if:

$$g\left(S - \underline{S}\right)\left[u\left(y_{ij} + B\right) - u\left(y_{ij}\right)\right] Q_e\left(\underline{e}_i, a_i\right) - C'(\underline{e}_i) > 0 \tag{3.1}$$

This result shows that the decision to choose high or low effort is determined by the distribution of the error term (μ), the teacher's increase in utility given that she receives the bonus, the marginal return to effort in the teacher quality function, and the marginal cost of effort. This decision rule says that if a teacher is exerting the minimal effort and could increase her expected utility by increasing effort, then the teacher should choose high effort e^*_{ij}. Finally, e^*_{ij} is defined as the solution to the following:

$$g\left(S - Q\left(e^*_{ij}, a_i\right)\right)\left[u\left(y_{ij} + B\right) - u\left(y_{ij}\right)\right] Q_e\left(e^*_{ij}, a_i\right) = C'(e^*_{ij})$$

That is, if the teacher chooses high effort in performance pay district j, then she should increase her effort level until e^*_{ij} sets the marginal benefit of effort equal to the marginal cost of effort. From this equation it is clear that e^*_{ij} is a function of the teacher's ability level (a_i), the amount of the performance bonus B, and the teacher's base compensation y_{ij}, which is a function of experience and education levels ($\mathbf{w}_i$) and performance pay district $j's$ characteristics ($\mathbf{v}_j$).

3.2.1 ability and the effort choice

This model allows one to characterize the effort choice completely in terms of teacher ability levels. It can be shown that there is threshold level of ability for each individual and performance pay district (a'_{ij}) such that teachers with ability levels less than or equal to a'_{ij} always choose low effort in performance pay district j. Similarly, teachers with ability levels greater than a'_{ij} choose to exert high effort in performance pay district j.

Above, I showed that teacher i chooses high effort in performance pay district j if the inequality in expression 3.1 holds. Given the previous assumptions about the cost of effort and teacher quality function, it can be shown that the left hand side of expression 3.1 is strictly increasing in ability (a_i). Therefore, there is some a'_{ij} such that:

$$g\left(S-\underline{S}\right)\left[u\left(y_{ij}+B\right)-u\left(y_{ij}\right)\right]Q_e\left(\underline{e}_i,a'_{ij}\right)-C'(\underline{e}_i)=0 \tag{3.2}$$

If teacher i has ability equal to the threshold value ($a_i = a'_{ij}$), then she is indifferent between exerting high effort or low effort in performance pay district j. Since the left hand side of this equation is strictly increasing in ability, if teacher i's ability is greater than a'_{ij}, then she chooses high effort (e^*_{ij}) in performance pay district j. Likewise, if $a_i \leq a'_{ij}$, then teacher i chooses low effort ($\underline{e}_i$) in performance pay district j. From equation 3.2 it is clear that a'_{ij} varies with the bonus award (B) and the teacher's base compensation (y_{ij}), which is a function of individual experience and education levels ($\mathbf{w}_i$) and performance pay district $j's$ characteristics ($\mathbf{v}_j$).

This model predicts that teachers with ability levels higher than the threshold value (a'_{ij}) will choose to exert high effort in performance pay district j. Low ability teachers choose to exert low effort. The intuition behind this result is simple. Relative to low ability teachers, teachers with high ability exert relatively low effort in order to meet the low quality standard. Since high ability teachers are already exerting relatively low effort, increasing their effort adds relatively less disutility of effort. High ability teachers can therefore increase their probability of earning a performance bonus with little added disutility of effort, relative to low ability teachers. This result implies that performance pay district j induces greater effort from teachers with sufficiently high ability: teachers with $a_i > a'_{ij}$.

3.3 The teacher's district selection

Now that we have characterized the teacher's effort choice in each district type, we can now characterize the individual's optimal district choice. Suppose that each teacher is endowed with a best steps and lanes district option. That is, assume that district s is teacher i's best available steps and lanes district option. Specifically, district s gives teacher i the greatest expected utility among all steps and lanes district options under her optimal low effort choice.

An individual's choice of district type (performance pay or steps and lanes) can be summarized by the teacher's choice between the teacher's best steps and lanes district option (district s), and an alternative performance pay district (district j). Under what circumstances does a teacher select performance pay district j over her best available steps and lanes district option? In this section, I characterize the teacher's choice of district type and show how teacher ability levels are related to the teacher's optimal district selection.

The teacher's choice of district is determined by which district gives the teacher the highest expected utility. Teacher i chooses performance pay district j if the teacher's expected utility in district j is greater than her utility in her best available steps and lanes district option, district s. That is, teacher i chooses performance pay district j if either of the following two cases are true:

$$\text{case I:}\;\; E\left[U(Y^p_{ij}(e^*_{ij}, a_i), \mathbf{v}_j, e^*_{ij})\right] \geq U(Y^s_{is}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i) \text{ and } a_i > a'_{ij}$$

or

$$\text{case II:}\;\; E\left[U(Y^p_{ij}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] \geq U(Y^s_{is}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$$

In case I, the teacher chooses performance pay district j because her expected utility in the performance pay district under high effort (e^*_{ij}) is greater than the utility she can earn in the steps and lanes district under the optimal low effort choice. In case II, the teacher chooses the performance pay district because the performance pay district offers a higher expected utility than the steps and lanes district under the low effort choice in both districts. In case II, teachers choose the performance pay district because the performance pay district offers a large enough base salary so that even under low effort the teacher is better-off selecting into the performance pay district.

3.3.1 ability and district selection

As with the effort decision, we can fully characterize the teacher's district selection decision in terms of teacher ability levels. Above, I described two cases. I assume that conditional on education and experience levels, ability is independent of the pay and characteristics of a teacher's best available steps and lanes district option. This assumption implies that the district selection decision in case I depends on teacher ability while decision in case II does not. Below, I formalize this assumption and describe how teacher ability is related to district selection.

Before introducing a major assumption regarding the teachers best available steps and lanes district and teacher ability, it is useful to define some additional notation. Define indicator variable s_{ij} to denote whether case II is false for teacher i with regard to performance pay district j. That is $s_{ij} = 1$ if teacher i has lower expected utility in performance pay district j under low effort relative to her best available steps and lanes district option: $E\left[U(Y^p_{ij}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] < U(Y^s_{is}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$. Also, $s_{ij} = 0$ if teacher i has higher expected utility in performance pay district j under low effort relative to her best available steps and lanes district option: $E\left[U(Y^p_{ij}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] \geq U(Y^s_{is}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$. Given this definition, I provide a formal definition of the assumption that conditional on education and experience levels, ability is independent of the pay and characteristics of a teacher's best available steps and lanes district option.

Assumption 1: *Conditional on individual and district characteristics* $(\mathbf{w}_i, \mathbf{v}_j)$, s_{ij} *and teacher ability* (a_i) *are independent. That is,* $P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j, a_i) = P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j)$ *and* $P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j, a_i) = P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)$.

This assumption implies that the characteristics and pay in a teacher's best available steps and lanes district are unrelated to the teacher's ability after controlling for the teacher's education and experience and performance pay district $j's$ characteristics. In other words, the teacher's ability level is not related to the likelihood that the her best available steps and lanes district will offer a higher total compensation than is offered by performance pay district j under the low effort choice. Intuitively, this assumption implies that steps and lanes districts are unable to identify or measure teacher ability in the hiring process, beyond that which is correlated with experience and education.

The decision rule in case II implies that teacher i will select into performance pay district j if the following relation holds:

$$u\left(y_{ij} + B\right)\left[1 - G\left(S - \underline{S}\right)\right] + u\left(y_{ij}\right)G\left(S - \underline{S}\right) + f(\mathbf{v}_j) \geq u\left(y_{is}\right) + f(\mathbf{v}_s)$$

Assumption 1 implies that individual ability levels are unrelated to pay and district attributes in the teacher's best available steps and lanes district s; therefore, this selection decision is unrelated to teacher ability conditional on the teacher's characteristics and performance pay district $j's$ characteristics.

Case II says that teachers will always choose the performance pay district if it offers a higher expected utility under the low effort choice. Given assumption 1, conditional on teacher education and experience levels, case II may be true for teachers with all levels of ability. Therefore, in the cases where performance pay districts offer a relatively high expected base compensation, the performance pay district could potentially attract teachers of all ability levels.

The selection decision in case I, however, is related to teacher ability. In case I, teachers select into performance pay district j if their ability is greater than a'_{ij} and the following relation holds:

$$u(y_{ij}+B)\left[1-G\left(S-Q(e^*_{ij},a_i)\right)\right]+u\left(y_{ij}\right)G\left(S-Q(e^*_{ij},a_i)\right)+f\left(\mathbf{v}_j\right)-C\left(e^*_{ij}\right)$$
$$-u(y_{is})-f(\mathbf{v}_s)+C\left(\underline{e}_i\right)\geq 0 \quad (3.3)$$

Given assumption 1, it can be shown that the left hand side of expression 3.3 is strictly increasing in teacher ability, conditional on the teacher's education and experience[3]. Hence, for each combination of teacher experience and education and each performance pay district j there is a threshold ability level (a^*_{ij}) such that the relation in 3.3 becomes an equality. Furthermore, it can be shown that if the relation is case II does not hold, then $a^*_{ij} > a'_{ij}$. Tteachers with high ability levels higher than this threshold $\left(a_i \geq a^*_{ij}\right)$ self-select into performance pay district j. It should be noted that a^*_{ij} is a function of individual experience and education levels $(\mathbf{w}_i)$, performance pay district $j's$ characteristics $(\mathbf{v}_j)$, and the characteristics of the teacher's best available steps and lanes district $(\mathbf{v}_s)$.

Combining the results derived in this section, we can now summarize teacher selection decisions across the ability distribution. Based on the model presented here,

[3]This result relies on the prior assumptions about the cost of effort function, teacher quality function, and the assumption that ability is unrelated to the district attributes in the teacher's best available steps and lanes type district (Assumption 1).

teachers select into performance pay district j if either of the following are true:

$$a_i \geq a_{ij}^*$$

or

$$E\left[U(Y_{ij}^p(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] \geq U(Y_{is}^s(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$$

Teachers select into performance pay districts either because they are high ability or because the performance pay district offers a high base compensation. If neither of these cases are true, the teacher selects the steps and lanes district.

3.4 Effort and district selection across the ability distribution

In the previous sections, I characterized the teacher's effort and district choice. I have also summarized the way in which these decisions are related to teacher ability. In this section I combine the results on effort and district selection and summarize them for the entire distribution of teacher ability.

Table 3.1 summarizes the teacher's district selection, teacher quality in performance pay district j, and expected teacher total pay in district j for all possible cases that may arise in the model described above.

In column (1) the table reports the teacher's district selection decision in terms of an indicator variable, d_{ij}, where $d_{ij} = 1$ if teacher i selects into performance pay district (j) and $d_{ij} = 0$ otherwise.

Column 2 reports the teacher's actual or potential teaching quality in performance pay district j (Q_{ij}^p). As described earlier, teacher quality depends on ability and effort in the following function: $Q(e_i, a_i)$. If the teacher exerts minimal effort in district j ($\underline{e}_i$), then her teacher quality in district j is $\underline{S} = Q(\underline{e}_i, a_i)$. If the teacher chooses high effort in district j, $e^*_{ij} > \underline{e}_i$, then the teacher's quality is higher than the minimal standard. I denote teacher quality in district j under high effort as: $S^*_{ij} = Q(e^*_{ij}, a_i)$, and $S^*_{ij} > \underline{S}$.

Finally, column 3 reports the teacher's actual or potential expected total pay in performance pay district j, conditioned on the teacher's education and experience and the teacher's ability level: $\left(E\left[Y^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, a_i\right]\right)$. If the teacher chooses low effort in district j, then her teaching quality is $\underline{S}$, and her conditional expected total pay is: $y_{ij} + G\left(\underline{S} - S\right)B$. That is, her expected total pay is the sum of base pay in the performance pay district (p) and her expected bonus given low effort ($G\left(\underline{S} - S\right)B$). If the teacher chooses high effort (e^*_{ij}) in performance pay district j, then her teaching quality is S^*_{ij}, and her conditional expected total pay is: $y_{ip} + G\left(S^*_{ij} - S\right)B$. It is worth noting that since $G()$ is strictly increasing and $S^*_{ij} > \underline{S}$, the teacher's expected total pay in district j is always greater under high effort relative to low effort. The more effort an individual teacher exerts, the greater the probability that the teacher will earn a performance bonus.

Table 3.1, summarizes teacher i's decision to select performance pay district j, the observed and potential teacher quality of teacher i in district j, and the teachers observed or potential expected total pay in district j:

Table 3.1: The individual's optimal district and effort choice

	(1)	(2)	(3)
Case:	d_{ij}	Q^p_{ij}	$E\left[Y^p_{ij}\|\mathbf{w}_i, \mathbf{v}_j, a_i\right]$
$E\left[U(Y^p_{ij}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] \geq U(Y^s_{is}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$			
1. and $a_i \leq a'_{ij}$	1	$\underline{S}$	$y_{ij} + G\left(\underline{S} - S\right)B$
2. and $a_i > a'_{ij}$	1	S^*_{ij}	$y_{ij} + G\left(S^*_{ij} - S\right)B$
$E\left[U(Y^p_{ij}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] < U(Y^s_{is}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$			
3. and $a_i \leq a'_{ij} < a^*_{ij}$	0	$\underline{S}$	$y_{ij} + G\left(\underline{S} - S\right)B$
4. and $a'_{ij} < a_i < a^*_{ij}$	0	S^*_{ij}	$y_{ij} + G\left(S^*_{ij} - S\right)B$
5. and $a'_{ij} < a^*_{ij} \leq a_i$	1	S^*_{ij}	$y_{ij} + G\left(S^*_{ij} - S\right)B$

There are five possible cases that may arise in the model, which depend on the teacher's ability level and the expected utility the teacher may earn in her best available steps and lanes district. In the first case, teachers are low ability ($a_i \leq a'_{ij}$), but the teacher's best available steps and lanes district offers a lower utility than the expected utility she can earn under low effort in performance pay district j. Therefore, in the first case, it is always optimal for the teacher to choose district j, despite the fact that she is low ability. These teachers select into district j, exert low effort, and always produce the minimal teacher quality in district j.

As in case one, in case two the teachers' best steps and lanes district option offers a lower expected utility than district j, under the low effort choice. Therefore, teachers in case two will always select into performance pay district j. However, what makes case two different than case one is that teachers in case two have higher ability ($a_i > a'_{ij}$). These higher ability teachers can achieve a higher expected utility in district j under high effort (e^*_{ij}). Therefore, teachers in case two select into performance pay district j and exert high effort. Conditional on experience and education, teachers in case two exert higher effort and earn a higher expected total pay in district j than teachers in case one.

In cases three through five, the teacher's expected utility in district j under low effort is less than the utility the teacher can earn in her best available steps and lanes district. Therefore, teachers in cases three through five will never choose performance pay district j and low effort. Teachers in these cases will only choose the performance pay district if under the high effort choice it provides a higher utility than the best available steps and lanes district.

Teachers in case three are low ability ($a_i \leq a'_{ij} < a^*_{ij}$) and do not select into the performance pay district. The teacher's best available steps and lanes district offers these teachers a higher utility since they are low ability and cannot substantially increase their chances of earning a performance bonus in district j without a large effort cost. Teachers in case three choose the steps and lanes district and exert low effort. Also, teachers in case three would not exert high effort in performance pay district j since their ability level is below the threshold value a'_{ij}. As a result, these teachers choose low effort in performance pay district j and earn the following potential expected total pay: $y_{ij} + G\left(\underline{S} - S\right)B$.

In case four, teachers have moderate ability $\left(a'_{ij} < a_i < a^*_{ij}\right)$. These teachers choose not to enter performance pay district j, but they would choose high effort in district j. Teachers in case four have ability levels that are high enough so that their cost of exerting high effort is less than the added benefit of obtaining a higher expected bonus in district j. However, even under the high effort choice in performance pay district j, teachers in case four are better off choosing their best available steps and lanes district. Therefore, these teachers do not self select into performance pay district j, but they would exert high effort if forced to work in district j. The potential conditional expected total pay for these teachers in district j is $y_{ij} + G\left(S^*_{ij} - S\right)B$.

Finally, in case five, teachers are high ability. These teachers self-select into performance pay district j, choose high effort, and achieve the following expected total pay: $y_{ij} + G\left(S^*_{ij} - S\right)B$.

3.5 The definition of the selection effect

To this point, I have shown that a teacher performance pay program can be characterized by five cases within a simple model of teacher behavior. In the model, teachers with relatively high ability levels are always attracted to performance pay, and in some cases low ability teachers prefer not to join a performance pay district. However, in the model (specifically case 1) we have seen that even low ability teachers might prefer to teach in a performance pay district if the base compensation in those districts is high enough. The key question is: how do these results relate to the selection and effort effect? That is, how do we decompose teacher quality effects that result from differences in effort exerted and teacher quality effects that are a result of selection on ability. Also, under what circumstances do performance pay programs induce a selection effect? To answer these questions we need to define the effort and selection effects.

I define selection and effort effects in terms of conditional differences in teachers' quality in performance pay district j: (Q^p_{ij}). Specifically, I decompose the difference in conditional observed teacher quality among those teachers who self-select into performance pay district j $(d_{ij} = 1)$ from those teachers who select their best available steps and lanes district $(d_{ij} = 0)$ over performance pay district j. Conditional on teacher characteristics and district j's characteristics, the difference in expected teacher quality in performance pay district j between teachers who self-select into performance pay district j and those who select the "steps and lanes" district is the following:

$$\text{selection effect}_j = E\left[Q^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] - E\left[Q^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0\right] \tag{3.4}$$

Equation (3.4) formalizes the idea that the selection effect for performance pay district j is defined as the difference in teacher quality between teachers who self select into performance pay district j and teachers who choose a steps and lanes districts over performance pay district j. This difference captures quality effects that stem from differing ability levels in the two groups of teachers. Specifically, given the assumptions in the previous sections, it can be shown that a positive selection effect implies that the average ability of teachers who self-select into performance pay district j is greater than the average ability of those who choose a steps and lanes district. This result is summarized by the following proposition:

Proposition 1. *Given the assumptions in sections 3.1-3.4, a positive selection effect exists (as defined in equation 3.4) if and only if the conditional expected ability* *for teachers who self-select into performance pay district j is greater than the conditional expected ability of teachers who select into steps and lanes districts over district j* [4]: $E\left[Q^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[Q^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0\right] \iff E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0\right]$

Proposition 1 states that if the average teacher quality is higher among teachers who choose performance pay district j over their best available steps and lanes district, then the average ability of teachers who self-select into performance pay district j is greater than the average ability of teachers who choose the steps and lanes district. That is, a positive selection effect for district j implies that district j is attracting a disproportionate number of teachers from the upper tail of the ability distribution. Assuming that district j, without a performance pay program, would have attracted a representative sample of teachers from the ability distribution (consistent with assumption 1), then a positive selection effect for performance pay district j implies that the performance pay program is the force behind this differential selection on ability.

3.6 The existence of a selection effect and implications for program design

Now that we have formally defined the selection effect, we can examine the circumstances under which a selection effect exists within this theoretical model. Using this information, I highlight the performance pay program characteristics and relevant policy variables that are most likely to induce a selection effect.

The conditions necessary for a performance pay program to induce a selection effect are illustrated in the proof of proposition 1, part I[5]. According to the assumptions

[4] For a proof of this statement, see appendix section 9.1.5

[5] See appendix section 9.1.5

in sections 3.1 - 3.4, a selection effect exists if and only if at least one teacher is in case 5 of table 3.1. That is, performance pay district j's performance pay program is attracting teachers with higher average ability if and only if:

$$P\left(s_{ij} = 1 \cap a_i \geq a_{ij}^{*} | \mathbf{w}_i, \mathbf{v}_j\right) > 0 \tag{3.5}$$

Intuitively, this result says that performance pay district j induces a selection effect if at least two requirements are true of at least one teacher. First, at least one teacher must be in the cases where district j offers a lower expected base compensation than the teacher's best available steps and lanes district. Second, this teacher must have high enough ability such that through high effort she can achieve a higher expected utility in the performance pay district than she can achieve in her best available steps and lanes district.

If a district wishes to adopt a performance pay program that induces a selection effect, the condition in relation 3.5 suggests that the district should work to maintain a sufficiently large proportion of teachers with $s_{ij} = 1$ and $a_i \geq a_{ij}^{*}$. That is, in order for a performance pay district to increase its likelihood of inducing a selection effect, it should use its influence to increase the probability that $s_{ij} = 1$ and $a_i \geq a_{ij}^{*}$. There are several policy instruments available to the district that influence each of these probabilities.

First, a performance pay district might increase the probability that $s_{ij} = 1$ by offering a low base compensation relative to other comparable districts. District administers might achieve this policy aim by simply offering a base salary schedule that is strictly lower than all other district salary schedules within its surrounding region. The lower the district sets its base salary schedule, the less attractive the district becomes to low ability teachers who would choose low effort in the performance pay program. As an alternative policy, a performance pay district might increase the probability that $s_{ij} = 1$ by offering teachers less in terms of benefits or amenities relative to comparable districts. Once again, this type of policy reduces the probability that teachers will self-select into the performance pay district simply because the performance pay district offers a relatively high base compensation.

Second, a performance pay district that wishes to induce a positive selection effect, must ensure that $a_i \geq a_{ij}^*$ for a sufficiently large number of teachers. To increase the proportion of teacher for which $a_i \geq a_{ij}^*$, the district might pursue policies that decrease the threshold value a_{ij}^*. Recall from above that a_{ij}^* is defined as the ability level that sets the inequality in relation 3.3 equal to zero.

From this equation, it is clear that a_{ij}^* is a function of two variables that are at least partially under the district's control when designing a performance pay program: the bonus size (B) and the distribution of the error in the teacher quality measurement μ. Assuming that the error term on the performance measure (μ) has mean zero (such that it is unbiased) and is symmetric such that $g(\mu) = g(-\mu)$, the relationship between a_{ij}^* and the bonus size B is negative. That is, as the district increases the bonus size, the proportion of teachers with $a_i > a_{ij}^*$ is likely to increase. This implies that a performance pay district that offers a larger performance bonus is more likely to induce a selection effect given that there are some teachers with $s_{ij} = 1$.

The model presented above implies there are ways in which districts may design performance pay programs that are more likely to attract teachers from only the upper tail of the ability distribution. The model suggests that for a performance pay district to attract only high ability teachers it must offer a low base compensation relative to comparable competing districts and offer a large performance bonus. The theory predicts that a performance pay program of this type will simultaneously repel low ability teachers and attract only high ability teachers to the district.

Chapter 4

Cross-sectional Methods and Procedures

Measuring the quality difference between teachers in performance pay districts and traditional pay districts has been the focus of several empirical studies. As I have mentioned above, most of these studies use student test scores to measure teacher quality, and almost all focus on measuring only the effort effect. In this section, I present an alternative method that uses teacher total pay and bonus probabilities to identify the existence of a selection effect. In section 4.1, I lay the theoretical foundation for this method. In section 4.2, I describe an instrumental variable procedures that may be used to estimate the teacher selection effect in cross-sectional data.

4.1 Potential total pay differences and the selection effect

Since we are usually unable to observe data on teacher quality, but data on teacher pay is readily available, one might wonder if we can infer anything about teacher quality by using data on teacher total pay in performance pay districts and bonus probabilities.

That is, can we learn about the expected ability difference between teachers in performance pay and traditional districts ($E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1\right] > E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=0\right]$) using some difference in teacher pay? It turns out that we can. In this section, I detail a method that identifies a selection effect from differences in teacher total pay and bonus probabilities.

Consider the following difference in conditional teacher total pay:

$$E\left[Y_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1\right] - E\left[Y_{ij}^p|\mathbf{w}_i, \mathbf{v}_j\right] \tag{4.1}$$

where as above, $\mathbf{w}_i' \equiv \{exp_i, educ_i\}$ is a 1x2 row vector containing the education and experience level of teacher i. This is the difference in total pay between teachers who self-select into performance pay district j and a random sample of teachers, conditioned on education level, experience, and district j's characteristics. Given the assumptions of the previous sections, if the difference defined in 4.1 is positive, then that implies performance pay district j is inducing a positive selection effect and attracting a teacher workforce that has a higher average ability. This result is formalized in the following proposition.

Proposition 2. *Given the assumptions in sections 3.1-3.4, the difference in expectation in equation (4.1) is strictly positive, if and only if district j's performance pay program is inducing a positive selection effect. In mathematical terms, the result is* [1]*:*
$E\left[Y_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1\right] > E\left[Y_{ij}^p|\mathbf{w}_i, \mathbf{v}_j\right] \Longleftrightarrow E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1\right] > E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=0\right].$

Proposition 2 states that if teachers who self-select into performance pay district j earn more than a random sample of teachers, then district j's performance pay program is inducing a positive selection effect. In chapter 4, I describe a method that allows one to estimate the difference in expression (4.1). If we estimate this difference and find that it is positive, then theory suggests that we are measuring a positive

[1] For a proof of this statement, see appendix section 9.1.6. This statement also applies to differences in conditional bonus award probabilities.

selection effect of district j's performance pay program.

4.2 Estimating conditional expectations

As I argued in the previous section, by measuring

$$E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1\right] - E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j\right] > 0,$$

we are indirectly measuring a positive teacher selection effect of district j's performance pay program. We can estimate these conditional expectations using a well known selection correction technique developed by Heckman (1979): the Heckman two-step estimator.

Heckman's two-step model specifies a selection equation, which determines the compensation type we observe for each individual, and a equation relating the dependent variable of interest to observable covariates. We have already defined the selection criteria in the previous section. Teachers select into performance pay district j if either of the following are true:

$$\begin{aligned} & a_i \geq a_{ij}^{*} \\ \text{or} & \\ & E\left[U(Y_{ij}^{p}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] \geq U(Y_{is}^{s}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i) \Longleftrightarrow s_{ij}=0 \end{aligned} \tag{4.2}$$

Recall from the theory that a_{ij}^{*} depends on the amount of the performance bonus (B), the teacher quality standards ($S, \underline{S}$), the functional form of the cost of effort, the functional form of the teacher quality function, the functional form of the teacher's utility function, and the relative compensation in performance pay districts (pecuniary and nonpecuniary). For the empirical portion of this analysis, we need to make simplifying assumptions about the nature of the selection decision. In particular, I assume that teacher i chooses the performance pay district j if:

$$\theta_{ij} > -\mathbf{x}'\boldsymbol{\gamma_1} - \mathbf{z}'\boldsymbol{\gamma_2} \tag{4.3}$$

In relation 4.3, $\boldsymbol{\gamma_1}$ is a vector of parameters to be estimated and $\mathbf{x}$ is a 4x1 vector defined as $\mathbf{x}' = \left\{1, educ_i, exp_i, \mathbf{v}_j'\right\}$. Also, $\boldsymbol{\gamma_2}$ is an mx1 vector of parameters to be estimated and $\mathbf{z}$ is an mx1 vector of instruments to be defined later. The individual level variable θ_{ij} is a function of the components of the teacher decision rule in 4.2. That is θ_{ij} is a function of a^*_{ij} and the indicator s_{ij}.

As above, the indicator variable d_{ij} is equal to one if we observe a teacher's total pay (Y^p_{ij}) in performance pay district j. That is, $d_{ij} = 1$ when $\theta_{ij} > -\mathbf{x}'\boldsymbol{\gamma_1} - \mathbf{z}'\boldsymbol{\gamma_2}$. We observe a teacher's steps and lanes salary and $d_{ij} = 0$ when $\theta_{ij} \leq -\mathbf{x}'\boldsymbol{\gamma_1} - \mathbf{z}'\boldsymbol{\gamma_2}$.

Now I specify the equation of interest. I assume that total pay in district j is related to education, experience, and the performance pay district's characteristics as in the following equation:

$$Y^p_{ij} = \mathbf{x}'\boldsymbol{\beta} + \varepsilon_{ij} \tag{4.4}$$

where $\boldsymbol{\beta}$ is a 4x1 vector and I assume the error term ε_{ij} satisfies $E\left[\varepsilon_{ij}|\mathbf{x}\right] = 0$[2]. Given this assumption, the conditional expectation of teacher total pay under a random assignment of teachers into performance pay district j is:

$$E\left[Y^p_{ij}|\mathbf{x}\right] = \mathbf{x}'\boldsymbol{\beta} \tag{4.5}$$

The other conditional expectation we are interested in estimating is $E\left[Y^p_{ij}|\mathbf{x}, d_{ij} = 1\right]$,

[2]It is important to note here that we are not trying to estimate a causal relationship between the covariates and total pay. Instead, in this case we want to estimate a conditional expectation, which may or may not have a causal interpretation. It is a well known result in econometrics that linear regression models such as that defined in equation (4.4) provide an accurate approximation of the conditional expectation. Angrist and Pischke (2009) make this point nicely in their theorem 3.1.6: The function $\mathbf{x}'\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is defined such that ε_{ij} is uncorrelated with $\mathbf{x}$ in equation (4.4), provides the minimum mean square error (MMSE) linear approximation to the conditional expectation $E\left[Y^p_{ij}|\mathbf{x}\right]$. Given this result, we could drop the assumption that $E\left[\varepsilon_{ij}|\mathbf{x}\right] = 0$. Instead, we could define ε_{ij} such that it is uncorrelated with $\mathbf{x}$.

which is teacher i's expected conditional total pay, given that the teacher self-selected into performance pay district j. In the cross-sectional analysis that follows, I assume that there are no unobservable differences among teachers who have self-selected into at least one performance pay district. To be more precise about this assumption, it is useful to define more notation. Define the indicator $d_i = 1$ if a teacher has self-selected into at least one performance pay district. Assuming that there are no unobservable differences among teachers who self-select into at least one performance pay district is equivalent to assuming the following:

$$E\left[Y_{ij}^{p}|\mathbf{x}, d_{ij} = 1\right] = E\left[Y_{ij}^{p}|\mathbf{x}, d_i = 1\right]$$

Or equivalently this assumption is the following:

$$E\left[\varepsilon_{ij}|\mathbf{x}, d_{ij} = 1\right] = E\left[\varepsilon_{ij}|\mathbf{x}, d_i = 1\right]$$

Given equation (4.4) and this assumption, the conditional expectation $E\left[Y_{ij}^{p}|\mathbf{x}, d_{ij} = 1\right]$ is the following:

$$E\left[Y_{ij}^{p}|\mathbf{x}, d_{ij} = 1\right] = E\left[Y_{ij}^{p}|\mathbf{x}, d_i = 1\right] = \mathbf{x}'\boldsymbol{\beta} + E\left[\varepsilon_{ij}|\mathbf{x}, d_i = 1\right] \tag{4.6}$$

We can estimate the conditional expectations in equations (4.5) and (4.6) given the following assumptions:

1. θ_{ij} and ε_{ij} are independent of $\mathbf{x}$ and $\mathbf{z}$ with zero mean.

2. $\theta_{ij} \sim N(0, 1)$

3. $E\left[\varepsilon_{ij}|\theta_{ij}, \mathbf{x}\right] = E\left[\varepsilon_{ij}|\theta_{ij}\right] = \varphi\theta_{ij}$ where φ is a constant.

Given these assumptions, we can rewrite equation (4.6):

$E\left[\varepsilon_{ij}|\mathbf{x}, d_i = 1\right] = E\left[\varepsilon_{ij}|\mathbf{x}, \theta_{ij} > -\mathbf{x}'\boldsymbol{\gamma_1} - \mathbf{z}'\boldsymbol{\gamma_2}\right] = E\left[\varepsilon_{ij}|\theta_{ij} > -\mathbf{x}'\boldsymbol{\gamma_1} - \mathbf{z}'\boldsymbol{\gamma_2}\right]$. By the law of iterated expectations,

$E\left[\varepsilon_{ij}|\theta_{ij} > -\mathbf{x}^{'}\boldsymbol{\gamma_1} - \mathbf{z}^{'}\boldsymbol{\gamma_2}\right] = E\left[E\left[\varepsilon_{ij}|\theta_{ij}\right]|\theta_{ij} > -\mathbf{x}^{'}\boldsymbol{\gamma_1} - \mathbf{z}^{'}\boldsymbol{\gamma_2}\right]$ applying assumption 3 above, we have: $E\left[\varepsilon_{ij}|\mathbf{x}, d_i = 1\right] = E\left[\varphi\theta_{ij}|\theta_{ij} > -\mathbf{x}^{'}\boldsymbol{\gamma_1} - \mathbf{z}^{'}\boldsymbol{\gamma_2}\right]$ which is equal to $\varphi E\left[\theta_{ij}|\theta_{ij} > -\mathbf{x}^{'}\boldsymbol{\gamma_1} - \mathbf{z}^{'}\boldsymbol{\gamma_2}\right]$.

By assumption 2, $\theta_{ij} \sim N(0,1)$ which implies,

$E\left[\theta_{ij}|\theta_{ij} > -\mathbf{x}^{'}\boldsymbol{\gamma_1} - \mathbf{z}^{'}\boldsymbol{\gamma_2}\right] = \frac{\phi\left(-\mathbf{x}^{'}\boldsymbol{\gamma_1} - \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)}{1-\Phi\left(-\mathbf{x}^{'}\boldsymbol{\gamma_1} - \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)} = \frac{\phi\left(\mathbf{x}^{'}\boldsymbol{\gamma_1} + \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)}{\Phi\left(\mathbf{x}^{'}\boldsymbol{\gamma_1} + \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)}$ which is the inverse mills ratio, where $\phi()$ is the probability density function of a standard normal distribution and $\Phi()$ is the cumulative distribution function of a standard normal. Now we can see that $E\left[\varepsilon_{ij}|\mathbf{x}, d_i = 1\right] = \varphi\frac{\phi\left(\mathbf{x}^{'}\boldsymbol{\gamma_1} + \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)}{\Phi\left(\mathbf{x}^{'}\boldsymbol{\gamma_1} + \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)}$. Therefore, equation (4.6) becomes:

$$E\left[Y_{ij}^{p}|\mathbf{x}, d_i = 1\right] = \mathbf{x}^{'}\boldsymbol{\beta} + \varphi\frac{\phi\left(\mathbf{x}^{'}\boldsymbol{\gamma_1} + \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)}{\Phi\left(\mathbf{x}^{'}\boldsymbol{\gamma_1} + \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)} \tag{4.7}$$

From these assumption we can estimate the parameters in equations (4.5) and (4.7) using a two-step procedure described in Heckman (1979). The first step of the procedure is to obtain probit estimates of $\boldsymbol{\gamma_1}$,$\boldsymbol{\gamma_2}$ from the following model:

$$prob\left(d_i = 1|\mathbf{x}, \mathbf{z}\right) = prob\left(\theta_{ij} > -\mathbf{x}^{'}\boldsymbol{\gamma_1} - \mathbf{z}^{'}\boldsymbol{\gamma_2}|\mathbf{x}, \mathbf{z}\right) = \Phi\left(\mathbf{x}^{'}\boldsymbol{\gamma_1} + \mathbf{z}^{'}\boldsymbol{\gamma_2}\right)$$

After obtaining $\hat{\boldsymbol{\gamma}}_{\mathbf{1}}, \hat{\boldsymbol{\gamma}}_{\mathbf{2}}$ from the probit model, I use these estimates to calculate the inverse Mills ratio for all individuals where Y_{ij}^{p} is observed to get $\frac{\phi\left(\mathbf{x}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{1}} + \mathbf{z}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)}{\Phi\left(\mathbf{x}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{1}} + \mathbf{z}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)}$ for all individuals who have self-selected into a performance pay district. Finally, in the second step, I obtain $\hat{\boldsymbol{\beta}}, \hat{\varphi}$ from the OLS regression of Y_{ij}^{p} on $\boldsymbol{x}$ and $\frac{\phi\left(\mathbf{x}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{1}} + \mathbf{z}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)}{\Phi\left(\mathbf{x}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{1}} + \mathbf{z}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)}$ using only the sample of teachers in performance pay districts. Using these parameter estimates, we can obtain estimates of the conditional expectations in (4.5),(4.7). Finally, for this model, the estimated difference in conditional expected total pay is:

$$E\left[Y_{ij}^{p}|\mathbf{x}, d_i = 1\right] - E\left[Y_{ij}^{p}|\mathbf{x}\right] = \hat{\varphi}\frac{\phi\left(\mathbf{x}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{1}} + \mathbf{z}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)}{\Phi\left(\mathbf{x}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{1}} + \mathbf{z}^{'}\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)} \tag{4.8}$$

This suggests, a simple test for the existence of a selection effect. It is clear that this difference in conditional expected pay is positive if and only if $\varphi > 0$. We can test this hypothesis with a simple t-test using the parameter estimate $\hat{\varphi}$. A positive and statistically significant estimate of φ would suggest, according to the results presented in propositions 1 and 2, that performance pay districts attract higher ability teachers. In other words, a positive φ suggests a teacher selection effect exists.

Chapter 5

Cross-sectional Analysis

5.1 Cross-sectional data

The cross-sectional method described above requires data regarding district performance pay and teacher, district, and regional characteristics. These data are available in the Schools and Staffing Survey (SASS) and the Common Core of Data (CCD), collected by the National Center on Education Statistics (NCES). In this section, I describe the SASS and CCD data and provide descriptive statistics.

The SASS is the nation's largest sample survey of teacher, school, and district characteristics. It has been administered six times: during the 1987-88, 1990-91, 1993-94, 1999-2000, 2003-04, and 2007-08 school years. In this analysis I use the two most recent years of the survey: 2003-04 and 2007-2008. The 2003-04 SASS provides a state representative sample of 43,244 public school teachers from 4,421 public school districts. The most recent year of the SASS (2007-08) provides a state representative sample of about 38,240 public school teachers from 4,601 public school districts.

The SASS consists of 5 separate surveys: the district survey, the school survey, the teacher survey, the principal survey, and the library/media center survey. The sample design for the SASS adhered to the following procedure. First, NCES obtain a representative sample of schools within each state. NCES asked these schools to complete the school, principal, and library surveys. Second, schools were linked to their districts, and each district was asked to complete the district level survey. Finally, the sampled schools were asked to provide a list of all teachers employed in the school to NCES. NCES then randomly sampled teachers from this list at a rate of at least one teacher per school and no more than twenty per school. The teachers selected in this sample were asked to complete the teacher survey. The response rates for the SASS are fairly high. For the 2007-08 SASS, of the 9,800 public schools that were sampled, 80% completed the school survey. Of 5,250 public school districts in the sample, 88% completed the district survey. Finally, of 47,600 public school teachers sampled, about 80% completed the teacher survey.

The district, school, and teacher surveys of the SASS provide all of the relevant information needed to implement the cross-sectional methods described in section 4.2. The SASS teacher survey provides detailed individual teacher information including education level, experience, base salary, and bonus salary. The SASS district survey includes teacher and student demographics aggregated at the district level, such as: the percent of students on free or reduced price lunch, percent of students belonging to a racial minority, total number of schools, total number of teachers, total number of students, district locale, and more. The district survey also includes information on district level collective bargaining, hiring practices, and the salary schedule. The restricted use version of the SASS provides the opportunity to link individual teachers to their schools, schools to their districts, and districts to their regions and states, a requirement that is crucial for this analysis. The restricted use file of the SASS can also be merged to the CCD, which provides even more detail about district characteristics. The CCD provides information on total district revenue and district demographic information available in the Census 2000 school district tabulation, including the following distinct level demographics: median family income, the proportion of children in poverty, and the proportion of the population aged 25 and over with highest educational attainment equal to high school or college.

One of the key variables for this study, which is included in the SASS district survey, is whether the district offers performance pay. The district survey includes the following question: "Does the district use any pay incentives such as cash bonuses, salary increases, or different steps on the salary schedule to reward excellence in teaching?". About 10% of districts responded yes to this question. This is the key measure of the incidence of performance pay in the SASS data.

5.1.1 instrument candidate

The cross-sectional methods described earlier require at least one valid instrumental variable. In this section, I describe the requirements for a valid instrument, and I suggest a candidate instrument that is available in the data and may satisfy these requirements.

In the Heckman model, a valid instrumental variable (z) must satisfy two conditions. First, it must be related to a teacher's decision to enter a performance pay district. That is, z must be related to a_{ij}^* or s_{ij}. Second, it must be unrelated to unobservables in the model that determine a teacher's total pay in a performance pay district.

Recall from above that a_{ij}^* is dependent on the shape of an individual's cost of effort function, the teacher quality function, the bonus, the performance standards, the shape of the individual's utility function, and the distribution of the measurement error in the teacher quality measurement. While variables that are related to these components of a_{ij}^* would satisfy the first assumption of a valid instrument, any variable related to these components would also likely violate the second assumption, since each of these components is likely related to total teacher pay.

Another determinant of the selection decision, which may provide the opportunity to find a valid instrument is the indicator s_{ij}. Recall that $s_{ij} = 1$ if teacher i has lower expected utility in performance pay district j under low effort relative to her best available steps and lanes district option: $E\left[U(Y_{ij}^{p}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] < U(Y_{is}^{s}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$. Also, $s_{ij} = 0$ if teacher i has higher expected utility in performance pay district j under low effort relative to her best available steps and lanes district option: $E\left[U(Y_{ij}^{p}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] \geq U(Y_{is}^{s}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$. If a teacher has $s_{ij} = 0$, then the teacher selects into performance pay district j, regardless of her ability level. Therefore, s_{ij} is related to an individual's selection decision, but is not related to the individual's ability level or conditional total pay.

Since s_{ij} is related to the selection decision but is assumed to be unrelated to teacher ability, variables that determine s_{ij} may satisfy the conditions of a valid instrument in this case. One variable that may determine s_{ij} and is available in the SASS is a measure of the distance between a new teacher's residence and her best available performance pay district. For example, a teacher who lives in an area where there are a large number of performance pay districts and relatively few steps and lanes type districts may be more likely to have $s_{ij} = 0$. That is, teachers who reside in an area with a high concentration of performance pay districts may have fewer high quality steps and lanes district options. These teachers are potentially more likely to select into a performance pay district for reasons that are unrelated to their teaching ability.

The SASS provides the opportunity to construct a measure of the concentration of performance pay districts in the state where the teacher attended college. This variable may satisfy the two conditions of a valid instrument. With regard to the first criterion, teachers who attended a college in a state where performance pay is more prevalent may be more likely to work in a performance pay district for several reasons. First, teachers might prefer to stay close to where they grew-up and therefore choose to attend college and work close to home. If the teacher's home state happens to have a high proportion of performance pay districts, then the teacher will be more likely to join a performance pay district than teachers from other states. Second, even if teachers are not more likely to attend college in their home state, teachers may be more likely to find a teaching job in the state where he or she received training. This stem from several factors: the teacher may prefer to stay in the state after graduation, it may be easier to obtain a teaching license in the state where the university is located, and teacher labor markets may be localized, in that districts tend to hire and recruit teaching candidates that are nearby. Given that teachers tend to work in the state where they received their undergraduate training, if that state happens to have a high proportion of districts that offer performance pay, then the teacher will also be more likely to select into a performance pay district.

To satisfy the second criterion of an instrumental variable, it must be true that the instrument is not related to teacher pay or ability. The measure of the proportion of performance pay teaching jobs in the state where a teacher attended college likely satisfies these criteria. I can think of no reason why high ability undergraduates may be more attracted to universities in states where there is a higher incidence of performance pay. Also, districts do not offer teachers differential pay based on where the teacher received his or her bachelor's degree.

This instrument candidate is much like a measure of distance from a performance pay district. The measure of distance from where a program or intervention was administered has has been used as a valid instrument in several selection correction studies. For example, Card (1993) uses college proximity as an instrument to estimate the returns to education and Neal (1997) uses the number of catholic schools per square mile within a region as an instrument to estimate the effects of catholic secondary school enrollment. The instrument that I am proposing here is similar to the instruments used in these studies. Teachers who attended college in a state with a high concentration of performance pay are, on average, closer to performance pay districts, and they are therefore potentially more likely to join those districts. At the same time, there is no reason to believe that this measure of distance is related to a teacher's ability level or pay (after controlling for individual and district covariates).

In this section, I have described the conditions that are necessary for a valid instrumental variable and described a promising candidate. A valid instrumental variable must be related to the selection decision, but unrelated to a teacher's ability and pay. I argued that a measure of the concentration of performance pay jobs in the state where the teacher attended college may be a valid instrument. I argued that this instrument candidate is much like a measure of "distance to treatment" used in prior studies. In the analysis that follows, I verify that this candidate satisfies the first criterion of a valid instrument and proceed to estimate the Heckman two-step model proposed in section 4.2.

5.2 Sample restrictions, variable definitions, and descriptive statistics

As mentioned above, this study's cross-sectional analysis employs the following data sources: the 03-04 SASS, the 07-08 SASS, the 03-04 CCD, and the 07-08 CCD. Analyzing this data required several data merges and sample restrictions. First, I merged the SASS teacher survey sample with each teacher's SASS school and district survey responses. Second, I merged the CCD School District Universe Survey and CCD School District Finance Survey with the SASS data. The final raw sample contained teacher-level observations for each year of the SASS, matched with a large number of school and district variables from the SASS school and district surveys and the CCD. The 03-04 raw sample contained 43,244 teachers and the 07-08 sample contained 38,240 teachers. In this section I describe the final sample of teachers that I use for the cross-sectional analysis. Specifically, I describe sample restrictions and variable definitions, and I provide descriptive statistics.

I imposed two restrictions to each year's raw sample that were related to missing or inconsistent district identifiers. In some observations, district identifiers were missing or inconsistent with the CCD district universe. I dropped observations where the teacher's district did not respond to the SASS district survey: 6,637 observations dropped from 03-04 sample, and 2,887 dropped from 07-08 sample. I also dropped observations where the teacher's district identifier did not exist in the corresponding year of the CCD: 20 observations dropped in the 03-04 sample and 217 dropped in the 07-08 sample.

I imposed three sample restrictions related to teacher characteristics and school type. I limited the sample to only teachers who identified themselves as a "regular full-time teacher". This eliminated an additional 3,250 observations from the 03-04 sample and an additional 3,047 observations in the 07-08 sample. I also kept only teachers who were employed by a "regular elementary or secondary school". This eliminated an additional 3,169 observations from the 03-04 sample and an additional 2,561 observations from the 07-08 sample. I dropped teachers who did not work for a school that is part of a school district or local education agency; this eliminated an additional 2,237 observations from the 07-08 sample and 2,083 from the 03-04 sample.

I imposed three sample restrictions related to the teacher's district characteristics. First, I kept only teachers employed by districts that offer at least grades 1-12. This eliminated an additional 1,348 observations from the 07-08 sample and 1,019 observations from the 03-04 sample. Second, I kept only districts that employ a single salary or "steps and lanes" pay schedule, which eliminated an additional 229 observations from the 03-04 sample and 248 observations from the 07-08 sample. Finally, I kept only teachers from districts that award a high school diploma. This eliminated an additional 189 observations for the 03-04 sample and 88 observations for the 07-08 sample.

Finally, I imposed four restrictions based on values of variables in the data. First, I dropped observations where the teacher reported a base salary that was more than \$1,000 greater than the maximum base salary reported by the district. Second, I dropped observations where the teacher reported a base salary that was less than the lowest base salary reported by the district minus \$1,000. Third, I dropped teachers who had less than a Bachelor's degree. This restriction eliminated only 265 additional teachers from the 03-04 sample and 193 from the 07-08 sample. Finally, I dropped teachers with age greater than 70 years.

After imposing all of the restrictions described above, the 07-08 sample contained 23,122 teacher observations and the 03-04 sample contained 24,608 observations. Any deviation from these sample sizes in the analysis or descriptive statistics tables are a result of missing values in the SASS or CCD data. For all of this analysis, I assume that missing values occur at random.

The final sample contains teachers who belong to either a traditional public school district or a district that offers a performance pay program. Table 5.1 reports the number and proportion of teachers belonging to performance pay districts by year. As mentioned earlier, I identify performance pay districts by the district's response to the following question: "Does the district use any pay incentives such as cash bonuses, salary increases, or different steps on the salary schedule to reward excellence in teaching?" As is shown in the table, about 9% of teachers in the sample belong to districts that responded yes to this question. The proportion of teachers belonging to performance pay districts is slightly higher in the 2007-08 sample relative to the 2003-04 sample. Each sample year provides over 2,000 observations of teachers who belong to performance pay districts.

The key question I intend to answer in this analysis is whether teachers who select performance pay districts earn higher then expected total pay on average, relative to a representative teacher. I define total pay as the sum of base salary and any

Table 5.1: Performance pay district participation by year

	2003-04 Sample		2007-08 Sample		Total	
	Number	%	Number	%	Number	%
No	22,331	90.75%	20,926	90.50%	43,257	90.63%
Yes	2,277	9.25%	2,196	9.50%	4,473	9.37%
Total	24,608	100.00%	23,122	100.00%	47,730	100.00%

performance bonus a teacher received. Only teachers in performance pay districts can receive a performance bonus, but for every teacher in the population I estimate an expected total pay. Teachers who self-select may earn higher than average total pay through three avenues: by earning a higher than average base-salary, by receiving larger than average performance bonuses, or by earning a performance bonus more frequently than a representative teacher. In the analysis that follows, I investigate all three of these avenues that might explain any measured difference in teacher total pay. Therefore, the key dependent variables for this analysis are: total pay, base-salary in a performance pay district, performance bonus size, and a bonus indicator variable. Table 5.2 provides descriptive statistics for these dependent variables.

Table 5.2: Descriptive statistics: dependent variables

Variable	N	Mean	SD	Min	Max
Base Teacher Salary					
full sample	47730	43,683.69	11,678.03	17,000	121,000
if in a performance pay district	4473	42,827.85	11,355.81	22,600	95,500
if in a traditional district	43257	43,772.19	11,707.40	17,000	121,000
Teacher's Total Pay	4464	43,310.81	11,469.77	22,600	95,500
Teacher Received a Performance Pay Bonus	4464	0.266	0.44	0	1
Teacher's Performance Pay Bonus					
full sample	4464	490.93	1,241.14	0	13,474
if received a bonus	1187	1,846.25	1,814.49	50	13,474

Table 5.2 reveals a few interesting details about teacher pay in the pooled sample. First, average base salary in traditional districts is $1,000 higher than the average base salary for teachers in performance pay districts. However, when base salary is combined with bonus pay for teachers in performance pay districts, average base

salary for teachers in traditional districts is comparable to the average total pay teachers receive in performance pay districts ($43,772 vs. $43,311). Also, of the teachers in performance pay districts, about 27% earn a bonus and the average bonus size is about $1,846. However, bonus sizes are highly variable, with a standard deviation of $1,814. The largest bonus in the final sample was $13,474.

In chapter 4 I described a cross-sectional estimation method that requires several individual and district level covariates. Table 5.3 displays descriptive statistics of the teacher and district level covariates that I use in this analysis.

Table 5.3: Descriptive statistics: teacher and district characteristics

Variable	N	Mean	SD	Min	Max
Individual Characteristics:					
Education: Bachelor's Degree	47730	0.51	0.5	0	1
Education: Master's Degree	47730	0.42	0.49	0	1
Education: PhD, Prof. Degree, Education Specialist	47730	0.07	0.25	0	1
Years of Teaching Experience	47730	13.24	10.15	0	53
District Characteristics:					
Total Number of K-12 Students Enrolled in the District	47730	29971	94215.4	2	1100000
Percent of Students on Free or Reduced Price Lunch	47513	41.59	21.86	0	100
Perc. of K-12 Students Belonging to a Racial Minority	47730	31.91	28.47	0	100
Num. of Full-Time Equiv. (FTE) Teachers in District	47730	1933.5	6383.8	2.9	89322
Number of Students Per FTE Teacher	47579	15.64	11.13	3.32	174
Number of Schools in the District	47730	43.19	110.05	1	1289
Collective Bargaining: Yes, Meet-and-Confer	47730	0.11	0.31	0	1
Collective Bargaining: Yes	47730	0.63	0.48	0	1
Collective Bargaining: No	47730	0.27	0.44	0	1
District Offers Teachers Tuition Reimbursement	47730	0.43	0.5	0	1
District Offers Magnet Program	47730	0.27	0.44	0	1
Dist. Req. Student Community Service for HS Diploma	47730	0.13	0.33	0	1
District Locale: Large City	47730	0.07	0.25	0	1
District Locale: Mid-Size City	47730	0.14	0.35	0	1
District Locale: Urban Fringe of a Large City	47730	0.2	0.4	0	1
District Locale: Urban Fringe of a Mid-Size City	47730	0.14	0.35	0	1
District Locale: Large Town	47730	0.02	0.15	0	1
District Locale: Small Town	47730	0.15	0.36	0	1
District Locale: Rural, outside CBSA	47730	0.17	0.37	0	1
District Locale: Rural, inside CBSA	47730	0.11	0.31	0	1
CCD: District Total Revenue per Student	47423	10245	9712.07	3090.91	440000
District Characteristics from 2000 Census:					
Prop. of District Pop. 16 and Over in Labor Force	47031	0.64	0.07	0.32	0.86
Prop. of Children Living in District in Poverty	47031	0.16	0.1	0	0.62
Prop. aged 25 and over with Highest Educ. = HS	47031	0.4	0.09	0.04	0.69
Prop. aged 25 and over with Highest Educ. = College	47031	0.36	0.15	0.06	0.93
Median Family Income in District	47031	48320	14035	16163	162000

For individual teacher characteristics, the SASS samples provide measures of teaching experience and education level, which are the primary determinants of teacher pay within a district. In the combined sample, about half of the 47,430 teachers have a Bachelor's degree and the other half has some type of advanced degree. Also, average total teaching experience, as measured by the sum of years as a full-time or part-time teacher in either a public or private school, is about 13 years.

Table 5.4 displays the mean values of individual teacher characteristics by district type. The table shows that traditional districts employ a larger proportion of teachers with Master's degrees and lower proportion of teachers with Bachelor's degrees relative to performance pay districts. However, there is not a statistically significant difference between the two district types in their employment of teachers with more advanced degrees (PhD, professional degree, or Education Specialist). There is also a statistically significant difference in teacher experience between the district types. Teachers in traditional districts have, on average, 0.73 additional years of teaching experience relative to teachers in performance pay districts, and this difference is statistically significant at the 1% level.

Table 5.4: Descriptive statistics: individual characteristics by district type

	Full Sample	**Traditional District**	**Performance Pay Dist.**	**difference**
Variable	**Mean**	**Mean**	**Mean**	**ppay-trad**
Education: Bachelor's Degree	0.51	0.51	0.53	0.02***
Education: Master's Degree	0.42	0.43	0.40	-0.03***
Education: PhD, Prof. Degree, Ed. Spec.	0.07	0.07	0.06	-0.01
Years of Teaching Experience	13.24	13.31	12.58	-0.73***

Note: N=47,730 for Full sample, N=43,257 for Traditional, and N=4,473 for Perf. Pay Dist; ***p<0.01, **p<0.05, *p<0.1

In the final sample I am able to merge the 47,730 teachers in the SASS teacher survey with a large number of district level covariates. As discussed in section 4.2, the proposed empirical analysis requires variables in $\mathbf{v}_j$, which is a district level vector of covariates that influence a district's base teacher pay and the utility level of teachers in the district. That is, $\mathbf{v}_j$ is a vector of district attributes related to the district's pecuniary and nonpecuniary attributes. In table 5.3, I provide a list of all of the district level covariates that may be elements of $\mathbf{v}_j$ and that I later include in the Heckman model estimation. Table 5.3 also provides a listing of the descriptive statistics of these district covariates.

The district level covariates that I include in this analysis are intended to measure district characteristics in seven categories: district size, student characteristics, classroom environment, district financials, benefits offered, district location, and district population demographics. For each category, I include one or more district variables in the analysis. For the district size category, I include measures of the number of students, teachers, and schools in the district. To capture district attributes related to student characteristics, I include a measure of the proportion of students belonging to a racial minority and the proportion of students qualifying for free or reduced price lunch. To capture district attributes related to the classroom environment, I include a variable measuring the proportion of students to teachers in the district to get a crude measure of average class size. I also include indicators for the existence of a magnet program and whether the district requires community service as a graduation requirement. To capture district attributes related to district financials and benefits, I include the following variables: an indicator describing the collective bargaining arrangement between teachers and the district, a measure of the district's total revenue per student, and an indicator for whether the district offers tuition reimbursement. Finally, I include variables that account for a district's locale and the demographics of the district population, including: median household income, educational attainment, labor force participation, and poverty status.

The Heckman selection model requires at least one instrumental variable that is related to teacher selection into performance pay districts but is unrelated to a teacher's ability or pay, after accounting for individual and district characteristics. In section 5.1.1, I described a candidate that may satisfy the requirements of a valid instrumental variable: a measure of the concentration of performance pay districts in the state where the teacher attended college. Table 5.5 displays descriptive statistics for this instrument candidate by district type (performance pay or traditional district).

Table 5.5: Descriptive statistics: instrument candidate

2007 Proportion of Teachers in Performance Pay Dsitricts in the State where Earned Bachelor's degree					
	N	**Mean**	**SD**	**Min**	**Max**
full sample	45654	0.13	0.17	0	0.7
if in a performance pay district	4229	0.25	0.23	0	0.7
if in a traditional district	41425	0.12	0.16	0	0.7

The instrument I use in the analysis is the proportion of teachers employed in performance pay districts in 2007 in the state where the teacher received a Bachelor's degree. I argue that this is a proxy measure of a teacher's distance from where she went to college to the nearest performance pay district. If she went to college in a state with a high proportion of teaching jobs in performance pay districts, then at the time of job market entry, it is likely that she is closer to a performance pay district than a teacher from a state with few performance pay districts. The justification for this instrument candidate is that the shorter distance increases the likelihood that the teacher will join a performance pay district, but the distance measure has no relationship to a teacher's pay or ability after controlling for individual and district characteristics.

To construct this instrument candidate, I merged information from the SASS teacher and district surveys. The teacher survey provides a variable indicating the state where the teacher received a Bachelor's degree, and the district survey provides an estimate of the number of teachers employed in each district. Using this information I was able to construct a measure of the proportion of teachers in each state belonging to performance pay districts in 2007-08 using the following procedure. First, I summed (within state) the total number of full-time equivalent (FTE) teachers in each district sampled in the district survey. Second, I summed (within state) the total number of FTE teachers in each performance pay district sampled in the district survey. Finally, for each state, I divided the total number of teachers in performance pay districts by the total number of teachers. This produced a measure of the proportion for teachers belonging to a performance pay district for each state in the year 2007. To finish constructing the instrument candidate, I merged this state level measure with the 03-04 and 07-08 SASS teacher surveys by the state where the teacher received a Bachelor's degree.

In this section I described all of the data restrictions employed in the analysis of the SASS and CCD data, and I described all constructed variables. For the cross-sectional analysis that follows I use the sample and variables described in this section to estimate the Heckman selection models described in section 4.2.

5.3 Results

I begin this section by testing the validity of the instrument candidate, described in previous sections, by reporting the first-stage estimation results. After verifying that the instrument candidate satisfies the first condition of a valid instrument, I proceed with the Heckman selection model estimation. I estimate three specifications of the model for each dependent variable described in the previous section. In the first specification, I estimate the selection model without district level controls, including only individual characteristics and a year effect. In the second specification, I add district controls, and in the third specifications I add region fixed-effects.

5.3.1 first-stage results

The first-stage estimates are a direct test for whether the instrument satisfies the first requirement of a valid instrument: the instrument must be related to the selection decision after controlling for the model covariates. To test this assumption, the first-stage model estimates a teacher's likelihood of selecting a performance pay district as a function of the instrument candidate and individual, district, and region level controls. If the instrument candidate has a statistically significant effect in the first-stage selection equation, then the candidate likely satisfies the first condition of a valid instrument. For the Heckman model, the first stage is specified as a probit binary outcome model. In what follows, I report three Probit model estimates to test whether the instrument candidate is a strong predictor of performance pay district selection after controlling for individual characteristics, district characteristics, and region fixed-effects.

Table 5.6 displays the results of the first stage probit. I report three probit model results including different combinations of individual, district, and regional controls. In each model, I include the instrument candidate, teacher characteristics, and a year effect. Each column of the table corresponds to a separate probit model estimation. In column 1, the probit model estimates the probability that a teacher selects into performance pay district conditional on the instrument candidate, individual characteristics, and a year effect. In column 2, I add district level controls to the model. The district level controls include all of the district characteristics listed in table 5.3, including the 2000 Census variables. In column 3, I add region fixed effects, which control for systematic differences in performance pay participation across the four Census regions (Northeast, Midwest, South, and West). For each column, I report the probit model coefficient estimates for the instrument candidate and teacher characteristics, and I suppress the coefficient estimates on the year effect, district characteristics, and region fixed effects.

The results in table 5.6 suggest that the instrument candidate satisfies the first condition of a valid instrument. In each model, the measure of performance pay concentration is statistically significant at the 1% level. Also, as one would expect, the results suggest that if a teacher attended college in a state with a high concentration of performance pay programs, then the teacher is more likely to teach in a performance pay district.

There is also no reason to believe that there is a systematic relationship between this instrument and a teacher's ability level or pay. In fact, the instrument is statistically insignificant in a regression of teacher total pay on individual characteristics and a year fixed effect. Given that the candidate satisfies the first condition for a valid instrument and there is no evidence suggesting that it is related to a teacher's total pay, I proceed with the Heckman model estimation using the performance pay concentration measure as an instrument.

Table 5.6: First-stage probit estimates of district selection

VARIABLES	(1)	(2)	(3)
Prop. in Perf. Pay in State where Earned BA (2007)	**1.698*** **	**1.402*** **	**1.384*** **
	(0.042)	**(0.046)**	**(0.048)**
Education: Master's Degree	-0.064***	-0.052***	-0.042**
	(0.018)	(0.020)	(0.020)
Education: PhD, Prof. Degree, or Ed. Spec.	-0.078**	-0.118***	-0.110***
	(0.037)	(0.039)	(0.039)
Teaching Experience: 2 or 3 years	-0.017	-0.006	-0.002
	(0.037)	(0.039)	(0.039)
Teaching Experience: 4 or 5 years	0.018	0.036	0.040
	(0.038)	(0.040)	(0.040)
Teaching Experience: 6 or 7 years	-0.020	0.013	0.015
	(0.039)	(0.041)	(0.041)
Teaching Experience: 8 to 10 years	-0.049	-0.020	-0.018
	(0.037)	(0.039)	(0.039)
Teaching Experience: 11 to 13 years	-0.062	-0.028	-0.034
	(0.039)	(0.041)	(0.042)
Teaching Experience: 14 to 16 years	-0.041	-0.017	-0.016
	(0.042)	(0.044)	(0.044)
Teaching Experience: 17 to 20 years	-0.047	0.017	0.012
	(0.040)	(0.042)	(0.042)
Teaching Experience: 21 to 25 years	-0.062	-0.015	-0.018
	(0.039)	(0.041)	(0.041)
Teaching Experience: 26 to 30 years	0.010	0.070*	0.070*
	(0.039)	(0.041)	(0.041)
Teaching Experience: 31 or more years	-0.013	0.038	0.034
	(0.043)	(0.045)	(0.046)
Constant	-1.558***	0.715***	-0.104
	(0.029)	(0.229)	(0.239)
Observations	44561	44561	44561
Year FE	Yes	Yes	Yes
District Controls	No	Yes	Yes
Region FEs	No	No	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.3.2 selection model results

In this section, I present the results of the Heckman selection models proposed in section 4.2. I estimate the Heckman models for each of the four dependent variables described in table 5.2: total pay, performance pay bonus, a bonus indicator, and base pay in a performance pay district. As in table 5.6, I estimate three models for each dependent variable. The first column presents Heckman model estimates after controlling for only individual characteristics and a year effect. In column 2, I add all of the district covariates described in table 5.3. Column 3 presents Heckman model results with the addition of region fixed-effects.

For each model, I report the coefficient estimates for the individual characteristics and suppress the coefficient estimates for the district covariates, region effects, and year effect. In addition to the coefficient estimates, for each model I report the number of censored and uncensored observations, the coefficient on the inverse Mills ratio, the standard error of the coefficient on the inverse Mills ratio, and the implied average difference in the dependent variable.

The total number of observations included in the Heckman model is the sum of two types of observations: those where the dependent variable is observed and those where the dependent variable is not observed. The dependent variables are not observed in censored observations, and they are observed in the uncensored observations. In this analysis the dependent variables are unobserved for teachers who are not in performance pay districts. We do not observe the teacher's total pay or performance bonus if the teacher does not belong to a performance pay district. Therefore, the number of censored observations in each model is equal to the number of teachers in the sample who belong to a traditional district (40,503). Similarly the number of uncensored observations is equal to the number of teachers in the sample who teach in a performance pay district (4,058).

Tables 5.7-5.10 report the implied average difference in the dependent variable.

For example table 5.7, reports the "implied average difference in total pay". This is a measure of the expected difference in total pay between those who self-select into performance pay districts and a representative teacher. This difference is calculated by the following method. First, I estimate the Heckman selection model for the dependent variable and obtain a predicted value for each observation in the data. For the case where the dependent variable is total pay, each predicted value gives the expected total pay for teacher i conditional on the covariates in the model and adjusted to account for nonrandom selection. That is, the predicted values from the Heckman selection model gives the expected total pay for teacher i in district j $\left(Y_{ij}^p\right)$ given his individual characteristics and the characteristics of the district $\left(\mathbf{x}' = \left\{1, educ_i, exp_i, \mathbf{v}_j'\right\}\right)$, where $educ_i$ is the teacher's education level, exp_i is the teacher's experience, and $\mathbf{v}_j$ is a vector of district attributes:

$$E\left[Y_{ij}^p|\mathbf{x}\right]$$

Second, I regress the dependent variable on the same individual and district characteristics as those used in Heckman model, using only the selected sample. I use these model estimates to obtain a predicted value for the entire sample of teachers. In the case where the dependent variable is total pay, the predicted values give the expected total pay conditional on teacher and district characteristics and conditional on selection into a performance pay district. That is the predicted values from the regression using the selected sample gives:

$$E\left[Y_{ij}^p|\mathbf{x}, d_i = 1\right]$$

where, $d_i = 1$ indicates that the teacher selected a performance pay district.

Recall from section 4.1 that by measuring a positive difference in these conditional expectations we are measuring a selection effect. For the case where total pay is the dependent variable, the result is the following: if $E\left[Y_{ij}^{p}|\mathbf{x},d_i=1\right]-E\left[Y_{ij}^{p}|\mathbf{x}\right]>0$ then we are measuring a positive selection effect. The implied average difference value reported in tables 5.7-5.10 is the average (over the entire sample of teachers) of this difference. In the case where total pay is the dependent variable, the "implied average difference in total pay" is the average of the estimated differences in total pay for those who self-select and a randomly drawn teacher $\left(E\left[Y_{ij}^{p}|\mathbf{x},d_i=1\right]-E\left[Y_{ij}^{p}|\mathbf{x}\right]\right)$over the 44,561 observations. A positive value indicates a positive selection effect.

The coefficient on the inverse Mills ratio and its standard error provide the opportunity to test the statistical significance of the difference reported in the "implied average difference measure". A positive implied difference suggests a positive selection effect, and a negative difference implies negative selection. In section4.2 I showed that estimated difference $\left(E\left[Y_{ij}^{p}|\mathbf{x},d_{ij}=1\right]-E\left[Y_{ij}^{p}|\mathbf{x}\right]\right)$ is equal to the following:

$$E\left[Y_{ij}^{p}|\mathbf{x},d_i=1\right]-E\left[Y_{ij}^{p}|\mathbf{x}\right]=\hat{\varphi}\frac{\phi\left(\mathbf{x}'\hat{\boldsymbol{\gamma}}_{\mathbf{1}}+\mathbf{z}'\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)}{\Phi\left(\mathbf{x}'\hat{\boldsymbol{\gamma}}_{\mathbf{1}}+\mathbf{z}'\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)} \tag{5.1}$$

where $\frac{\phi\left(\mathbf{x}'\hat{\boldsymbol{\gamma}}_{\mathbf{1}}+\mathbf{z}'\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)}{\Phi\left(\mathbf{x}'\hat{\boldsymbol{\gamma}}_{\mathbf{1}}+\mathbf{z}'\hat{\boldsymbol{\gamma}}_{\mathbf{2}}\right)}$ is the estimated inverse Mills ratio and $\hat{\varphi}$ is the estimated coefficient on the inverse Mills ratio. The inverse mills ratio is always a positive value, since it it the ratio of a standard normal distribution's probability density function to its cumulative distribution function. Therefore, the difference in equation 5.1 is positive and the selection effect is positive if the underlying parameter φ, estimated by $(\hat{\varphi})$, is positive. In the Heckman two-step model, $\hat{\varphi}$ is estimated as an additional regression coefficient in the second stage. Therefore $\hat{\varphi}$ has the same properties as the other coefficients in the model and it has an associated standard error. I use $\hat{\varphi}$ and its associated standard error to perform a test of the significance of the "implied average difference of the dependent variable" shown in equation 5.1.

Table 5.7 reports the results for the case where teacher total pay is the dependent variable. In column 1, the selection effect is insignificant, but when additional controls are added to the model in columns 2 and 3, the selection effect is positive and statistically significant. The coefficient on the inverse Mills ratio is positive and significant at the 5% level in column 2 and positive and significant at the 10% level in column 3. The implied average difference in total pay in column 2 is $1,761. This suggests that teachers who self-select into performance pay districts earn, on average, $1,761 more in total pay than a teacher with identical observable characteristics who is randomly assigned to an identical performance pay district. A similar result holds for column 3: teachers who self-select earn an extra $1,634 in total pay, on average, relative to an observably identical teacher who is randomly assigned to an identical performance pay district.

There are three possible avenues by which teachers who self-select may earn a higher total pay. One way to interpret this result is from within the context of the theory presented in chapter 3. The theory suggests that teachers who self-select into performance pay districts may be higher ability teachers, on average, and they are therefore more likely to earn a performance bonus. The results presented in table 5.7 are consistent with this theory if teachers who self-select are also more likely to earn a performance pay bonus or earn a higher expected performance bonus. One potential explanation for the results in 5.7 is inconsistent with this theory. Teachers who self-select could potentially earn a higher total pay as a result of earning a higher base salary for unobserved reasons. This explanation seems more likely to result from a data or econometric problem, since base teacher pay is typically awarded based on teacher characteristics that are included in the model.

In the next three tables I explore each of these potential explanations for the estimated total pay differences presented in table 5.7. To reveal which component of total pay is driving this measured difference, I estimate Heckman selection models for each of the following dependent variables: performance pay bonus, a bonus indicator, and base teacher pay in a performance pay district.

Table 5.8 reports the results for the case where teacher bonus pay is the dependent variable. In each column the selection effect is negative and significant. The coefficient on the inverse Mills ratio is negative and significant at the 1% level in each column. The implied average difference in bonus pay in column in each column is about -$1,000. This suggests that on average, teachers who self-select into performance pay districts will earn $1,000 less in bonus pay than a teacher with identical characteristics who is randomly assigned to a similar performance pay district.

Table 5.9 reports the results for the case where an indicator for whether the teacher received a performance bonus is the dependent variable. This model is analogous to a linear probability model adjusted for sample selection. In each column the selection effect is negative and significant at the 1% level. The implied average difference in bonus pay probability is about -0.5. This suggests that on average, teachers who self-select into performance pay districts have a predicted probability of earning a bonus that is 50 percentage points lower, on average, than a teacher with identical characteristics who is randomly assigned to a similar performance pay district. The result suggests that teachers who self-select into performance pay districts are significantly less likely to earn a performance pay bonus than the average teacher with similar observable characteristics.

The results in tables 5.7, 5.8, and 5.9 suggest that teachers who self-select into performance pay districts earn higher total pay than comparable randomly drawn teachers, but these teachers earn a lower expected bonus and are less likely to earn a bonus. Together, these results imply that teachers who self-select into performance pay districts earn a higher than expected base pay in performance pay districts. This result is confirmed in table 5.10.

Table 5.10 reports the results of selection models where teacher base-pay in a performance pay district is the dependent variable. In each column the selection effect is positive and significant. The coefficient on the inverse Mills ratio is positive and significant at the 10% level in column 1 and significant at the 1% level in columns 2 and 3. The implied average difference in base pay in column one is \$1,189 and about \$2,700 in columns 2 and 3. The results in columns 2 and 3 suggest that on average, teachers who self-select into performance pay districts will earn about \$2,700 more in base-pay than a teacher who is randomly assigned to the performance pay district and has identical experience and education.

The estimates in tables 5.7 through 5.10 suggest that, relative to an identical randomly sampled teacher, teachers who self-select into performance pay districts earn a higher expected total pay, earn a lower expected bonus, are less likely to earn a bonus, and earn a higher expected base salary in performance pay districts. One might expect these results to be sensitive to a teacher's age and education level. Younger teachers may be more mobile and less averse to performance pay programs. However, older and more educated teachers may be more aware of their own ability levels, which could also influence the measured selection effects.

Table 5.11 reports results of the selection model estimates by teacher education and age. The table reports the coefficient on the inverse Mills ratio for the Heckman models that include controls for individual characteristics, district characteristics, year effects, and region effects. The first row of the table reports the results for the full sample, which are identical to the results reported in the third column of tables 5.7 through 5.10. The second and third rows of table 5.11 report the estimate for the coefficient on the inverse Mills ratio when the sample is limited to only teachers with a bachelor's degree and master's degree. Rows 4-7 report the results with the sample restricted to teachers of various age levels (under 30, over 30, under 40, and over 40).

The results in table 5.11 suggest that the selection effects estimated in the baseline (full-sample) models, are not overly sensitive to a teacher's age or education. The estimated selection effects all have the same sign for each sample restriction, although in some cases statistical significance is lost. While the estimated direction or sign of the selection effect is consistent throughout table 5.11, the relative magnitudes of the selection effects seem to vary by education and age. For example, table 5.11 suggests that the selection effects are larger among teachers with a master's degree and among teachers between the ages of 30 and 40. This result is consistent with the intuition that teachers with a master's degree and between the ages of 30 to 40 years may be more aware for their own ability levels and may be relatively free to move between district types.

The Heckman model results suggest that teachers who self-select into performance pay districts earn more in total pay than a similar teacher who is randomly assigned to a performance pay district. This result is consistent with the notion that performance pay programs induce a positive selection effect. According to the theoretical framework presented above, this result implies that performance pay districts attract a disproportionate number of high-ability teachers.

However, the results presented in tables 5.8 to 5.11 suggest the opposite. Teachers who self-select into performance pay districts earn a lower expected bonus and are less likely to earn a bonus relative to an identical randomly assigned teacher. Table 5.10 suggests that the positive selection effect measured in table 5.7 is entirely a result of a difference in base teacher pay in performance pay districts. Table 5.10 suggests that teachers who self-select into performance pay districts earn a larger expected base salary than an observably identical randomly assigned teacher. Taken together, the results in tables 5.7 to 5.11 imply that relative to randomly assigned teachers with identical observable characteristics, teachers who self-select into performance pay districts earn a higher expected total pay, earn a lower expected performance bonus, are less likely to earn a bonus, and earn higher expected base salary in performance pay districts.

5.3.3 implications and unanswered questions

This analysis produced results that are inconsistent with the notion that performance pay programs induce a positive selection effect. The finding that teachers who self-select into performance pay programs are less likely to earn a performance bonus is contrary to what one would expect if performance pay programs attract only the highest ability teachers. Furthermore, some of the results are inconsistent with the way in which teachers are paid. In particular, it is difficult to understand why teachers with identical characteristics might earn a different base salary, as the results suggest. This suggests that the methods or data employed in this cross-sectional analysis may be unreliable.

Data misreporting is one possible explanation for these results, and it does not discredit the teacher selection theory nor the empirical methods used in this analysis. What if teachers in the SASS are not reporting bonus and base pay as one might expect? One possibility is that districts are rewarding high performing teachers with additional base salary, rather than bonuses. This might occur if high performing teachers are rewarded with step increases rather than bonuses. Alternatively, teachers in the SASS could be consistently misreporting bonus pay as base pay.

If the SASS measure of base-pay is capturing a significant proportion of teacher pay that is awarded based on measurable teacher quality, then the results presented in this section are at least partially consistent with the notion that performance pay programs induce a positive selection effect. That is, if pay incentives in performance pay programs are consistently awarded as base-pay increases or if teachers in the SASS are misreporting merit-bonus awards as base pay, then the results in tables 5.7 and 5.10 would imply a positive selection effect.

It is unknown whether unexpected base-pay reporting in the SASS explains the results reported in this section. I leave the data reporting question to future research. One might test the data misreporting hypothesis by comparing the SASS self-reported bonus and base salary data to those reported by the district itself. This method may provide insight into how teachers are reporting base-salary and bonus pay in the SASS, and it may provide insight into whether district "rewards for teaching excellence" in the SASS are distributed primarily as base pay increases or bonuses.

Table 5.7: Heckman selection model of teacher total pay

VARIABLES	(1)	(2)	(3)
Education: Master's Degree	5,495.9***	4,302.5***	4,294.0***
	(262.8)	(218.4)	(216.6)
Education: PhD, Prof. Degree, or Ed. Spec.	7,945.6***	6,484.0***	6,507.1***
	(534.7)	(441.2)	(438.0)
Teaching Experience: 2 or 3 years	1,359.5***	1,574.0***	1,484.2***
	(517.7)	(424.7)	(422.0)
Teaching Experience: 4 or 5 years	2,697.4***	2,604.8***	2,564.3***
	(524.8)	(431.0)	(428.2)
Teaching Experience: 6 or 7 years	4,516.4***	4,045.9***	3,996.7***
	(545.3)	(448.2)	(445.3)
Teaching Experience: 8 to 10 years	5,657.3***	5,984.9***	5,850.9***
	(518.7)	(425.5)	(422.9)
Teaching Experience: 11 to 13 years	7,550.2***	7,937.7***	7,927.2***
	(555.1)	(455.4)	(452.3)
Teaching Experience: 14 to 16 years	10,650.8***	10,962.2***	10,833.2***
	(582.7)	(477.7)	(474.7)
Teaching Experience: 17 to 20 years	13,180.9***	13,792.6***	13,766.2***
	(565.8)	(462.7)	(459.7)
Teaching Experience: 21 to 25 years	14,452.4***	15,201.0***	15,180.7***
	(556.8)	(456.2)	(453.2)
Teaching Experience: 26 to 30 years	16,246.0***	16,627.7***	16,517.0***
	(551.6)	(452.4)	(449.6)
Teaching Experience: 31 or more years	16,222.4***	17,078.5***	16,988.0***
	(606.6)	(498.5)	(495.7)
Constant	29,198.6***	30,698.9***	34,808.9***
	(733.8)	(2,815.8)	(2,938.4)
Observations	44561	44561	44561
Censored Observations	40503	40503	40503
Uncensored Observations	4058	4058	4058
Year FE	Yes	Yes	Yes
District Controls	No	Yes	Yes
Region FEs	No	No	Yes
Coefficient on Inverse Mills Ratio	153.9	938.3**	859.2*
S.E. of Coef. on Inverse Mills Ratio	390.6	457.3	472.8
Implied avg. Difference in Total Pay	261.5	1761	1634

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5.8: Heckman selection model of teacher performance pay bonus

VARIABLES	(1)	(2)	(3)
Education: Master's Degree	190.6***	182.3***	176.2***
	(42.7)	(42.5)	(42.5)
Education: PhD, Prof. Degree, or Ed. Spec.	158.9*	167.3*	163.7*
	(86.8)	(85.8)	(86.0)
Teaching Experience: 2 or 3 years	119.8	91.6	93.1
	(84.2)	(82.7)	(83.0)
Teaching Experience: 4 or 5 years	116.4	92.8	91.3
	(85.4)	(84.0)	(84.2)
Teaching Experience: 6 or 7 years	206.9**	174.5**	174.0**
	(88.7)	(87.3)	(87.5)
Teaching Experience: 8 to 10 years	289.6***	248.9***	253.9***
	(84.3)	(82.9)	(83.1)
Teaching Experience: 11 to 13 years	339.2***	314.4***	317.6***
	(90.2)	(88.7)	(88.9)
Teaching Experience: 14 to 16 years	374.0***	336.1***	340.5***
	(94.7)	(93.0)	(93.3)
Teaching Experience: 17 to 20 years	477.0***	455.8***	460.1***
	(91.9)	(90.0)	(90.2)
Teaching Experience: 21 to 25 years	491.8***	459.2***	463.4***
	(90.4)	(88.8)	(89.0)
Teaching Experience: 26 to 30 years	384.3***	369.1***	373.3***
	(89.6)	(88.1)	(88.3)
Teaching Experience: 31 or more years	188.7*	180.5*	186.8*
	(98.5)	(97.0)	(97.3)
Constant	972.7***	294.5	560.7
	(120.0)	(545.8)	(573.8)
Observations	44561	44561	44561
Censored Observations	40503	40503	40503
Uncensored Observations	4058	4058	4058
Year FE	Yes	Yes	Yes
District Controls	No	Yes	Yes
Region FEs	No	No	Yes
Coefficient on Inverse Mills Ratio	-545.9***	-518.2***	-550.4***
S.E. of Coef. on Inverse Mills Ratio	63.73	89.06	92.98
Implied avg. Difference in Bonus Pay	-927.5	-972.8	-1047

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5.9: Heckman selection model of bonus probability

VARIABLES	(1)	(2)	(3)
Education: Master's Degree	0.043***	0.040***	0.038**
	(0.015)	(0.015)	(0.015)
Education: PhD, Prof. Degree, or Ed. Spec.	0.007	0.018	0.015
	(0.031)	(0.031)	(0.031)
Teaching Experience: 2 or 3 years	0.057*	0.043	0.043
	(0.030)	(0.030)	(0.030)
Teaching Experience: 4 or 5 years	0.056*	0.044	0.042
	(0.031)	(0.030)	(0.030)
Teaching Experience: 6 or 7 years	0.108***	0.090***	0.090***
	(0.032)	(0.031)	(0.031)
Teaching Experience: 8 to 10 years	0.114***	0.094***	0.098***
	(0.030)	(0.030)	(0.030)
Teaching Experience: 11 to 13 years	0.101***	0.086***	0.088***
	(0.032)	(0.032)	(0.032)
Teaching Experience: 14 to 16 years	0.116***	0.100***	0.103***
	(0.034)	(0.033)	(0.033)
Teaching Experience: 17 to 20 years	0.109***	0.092***	0.095***
	(0.033)	(0.032)	(0.032)
Teaching Experience: 21 to 25 years	0.186***	0.173***	0.174***
	(0.032)	(0.032)	(0.032)
Teaching Experience: 26 to 30 years	0.109***	0.095***	0.098***
	(0.032)	(0.032)	(0.032)
Teaching Experience: 31 or more years	0.084**	0.071**	0.073**
	(0.035)	(0.035)	(0.035)
Constant	0.560***	-0.096	0.054
	(0.043)	(0.195)	(0.204)
Observations	44561	44561	44561
Censored Observations	40503	40503	40503
Uncensored Observations	4058	4058	4058
Year FE	Yes	Yes	Yes
District Controls	No	Yes	Yes
Region FEs	No	No	Yes
Coefficient on Inverse Mills Ratio	-0.253***	-0.276***	-0.284***
S.E. of Coef. on Inverse Mills Ratio	0.0228	0.0320	0.0334
Implied avg. Diff. in Bonus Probability	-0.430	-0.517	-0.540

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5.10: Heckman selection model of base pay in performance pay districts

VARIABLES	(1)	(2)	(3)
Education: Master's Degree	5,305.3***	4,120.2***	4,117.8***
	(261.3)	(216.4)	(214.5)
Education: PhD, Prof. Degree, or Ed. Spec.	7,786.7***	6,316.8***	6,343.3***
	(531.6)	(437.1)	(433.5)
Teaching Experience: 2 or 3 years	1,239.7**	1,482.4***	1,391.1***
	(514.8)	(420.8)	(417.9)
Teaching Experience: 4 or 5 years	2,581.0***	2,512.0***	2,473.1***
	(521.8)	(427.0)	(424.0)
Teaching Experience: 6 or 7 years	4,309.5***	3,871.4***	3,822.7***
	(542.2)	(444.1)	(440.9)
Teaching Experience: 8 to 10 years	5,367.7***	5,736.0***	5,597.1***
	(515.8)	(421.6)	(418.7)
Teaching Experience: 11 to 13 years	7,211.0***	7,623.3***	7,609.6***
	(551.9)	(451.2)	(447.8)
Teaching Experience: 14 to 16 years	10,276.9***	10,626.1***	10,492.6***
	(579.4)	(473.2)	(470.0)
Teaching Experience: 17 to 20 years	12,703.9***	13,336.8***	13,306.1***
	(562.5)	(458.3)	(455.1)
Teaching Experience: 21 to 25 years	13,960.6***	14,741.8***	14,717.3***
	(553.6)	(451.9)	(448.6)
Teaching Experience: 26 to 30 years	15,861.7***	16,258.6***	16,143.8***
	(548.4)	(448.2)	(445.2)
Teaching Experience: 31 or more years	16,033.7***	16,898.0***	16,801.2***
	(603.2)	(493.8)	(490.7)
Constant	28,225.9***	30,404.4***	34,248.2***
	(729.9)	(2,787.1)	(2,906.2)
Observations	44561	44561	44561
Censored Observations	40503	40503	40503
Uncensored Observations	4058	4058	4058
Year FE	Yes	Yes	Yes
District Controls	No	Yes	Yes
Region FEs	No	No	Yes
Coefficient on Inverse Mills Ratio	699.8*	1456***	1410***
S.E. of Coef. on Inverse Mills Ratio	388.5	453.1	468.2
Implied avg. Difference in Base Pay	1189	2734	2681

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5.11: Selection effects by teacher education and age

	Dependent Variables			
Sample	**Total Pay**	**Bonus Size**	**Bonus Indicator**	**Base Pay ++**
Full Sample	859.191*	-550.390***	-0.284***	1409.581***
Teachers with BA only	343.692	-349.243***	-0.222***	692.934
Teachers with MA only	1671.283*	-1003.495***	-0.389***	2674.778***
age less than or equal to 30	753.938	-317.737***	-0.193***	1071.675*
age greater than 30	946.538*	-624.393***	-0.309***	1570.931***
age less than or equal to 40	1765.030***	-461.531***	-0.285***	2226.561***
age greater than 40	237.644	-674.022***	-0.294***	911.666

*** p<0.01, ** p<0.05, * p<0.1

++ The "Base Pay" dependent variable is the base salary a teacher receives in a performance pay district

Note: Each model includes controls for individual characteristics, district characteristics, year effects, and region effects.

Chapter 6

Longitudinal Methods and Procedures

The cross-sectional method described in chapter 4 relies on several strong assumptions, including: an assumption about the validity of an instrumental variable and the assumption that there is no unobservable difference among teachers who select into at least one performance pay district. These assumptions can be relaxed in an analysis of district level longitudinal data. In this chapter, I describe a method that may be used to identify the selection effect in teacher level longitudinal data that spans the time period during which a district adopted a performance pay program. In section 6.1, I describe the longitudinal data and assumptions that are required to identify the selection effect. In section 6.2 I describe an empirical method that may be used to estimate the selection effect in longitudinal data.

6.1 Identifying the selection effect in longitudinal data

The type of data I employ in this analysis is longitudinal data at the teacher level, which spans the time period during which a district adopted a performance pay program. The data links teachers to their school district, individual characteristics, and performance bonus awards over time, and it provides the opportunity to estimate a selection effect. With it, one can identify a treatment and comparison group of teachers and observe outcomes and characteristics for each group.

Before I describe the empirical design I employ in the longitudinal analysis, it is useful to identify an ideal experiment designed to measure the selection effect of a performance pay program. In a controlled experiment, a scientist might first obtain a random sample of teachers from a population and randomly assign them to a treatment group and a comparison group. The scientist might then give teachers in each group (treatment and comparison) the option to accept or reject a teaching job in performance pay district j, but give the two groups different information about the district. Suppose that teachers in the treatment group are told that the district is indeed a performance pay district and are provided with all information about the performance pay program. Suppose also that this information is withheld from teachers in the comparison group. That is, teachers in the comparison group are given identical information about district j, except they are not told that the district has a performance pay program. In each group (treatment and comparison) a number of teachers will presumably choose to accept the teaching job in district j, albeit with and without prior knowledge of the performance pay program.

The key question is whether teachers who choose the district (selectors) with complete information are different from teachers who choose the district without knowledge of the performance pay program. That is, selectors from the treatment group might differ from the selectors from the comparison group in terms of ability levels and observed characteristics. If a selection effect exists, one might expect that selectors from the treatment group are more likely to possess characteristics that are rewarded in the performance pay program. Treatment group selectors may be more experienced or educated than selectors from the comparison group if those characteristics are more likely to be rewarded in the performance pay program. At a minimum, selectors from the treatment group should be more likely to earn a performance bonus relative to selectors from the control group, if the existence of a performance pay program does indeed influence the teacher's decision to join the district.

In teacher level longitudinal data, I attempt to replicate this experimental design without the benefit of a controlled setting. In this data, one can identify which teachers have joined the district in the years prior to and after the district adopted performance pay. These teachers populate the comparison and treatment groups, respectively. Assuming that teachers who joined the district prior to its adoption of performance pay did so without knowledge of the future performance pay program, these prior joiners are analogous to the control group selectors in the experiment described above. Similarly, teachers who joined the district after the district had adopted performance pay, presumably did so with full knowledge of the district's program. These teachers are analogous to the selectors from the treatment group in the experiment described above. Given this definition of the treatment and comparison groups in the longitudinal data, the final task is to identify key outcome variables that may differ between the groups. Again, if a selection effect exists, one might expect that teachers in the treatment group (post performance pay selectors) may be more experienced or educated or may be more likely to earn a performance bonus relative to teachers in the comparison group (selectors prior to performance pay).

To be more precise about the bonus probability comparison and its implications, define the following indicator of teacher membership in the treatment group:

$T_{ij} = 1$ if teacher i entered district j after the adoption of performance pay.

$T_{ij} = 0$ if teacher i entered district j prior to the adoption of performance pay.

Consider the following comparison of teacher performance pay bonus probabilities among teachers in the treatment and comparison groups:

$$P\left(Y^p_{ijt} = 1 | \mathbf{w}_{it}, \mathbf{v}_{jt}, T_{ij} = 1\right) - P\left(Y^p_{ijt} = 1 | \mathbf{w}_{it}, \mathbf{v}_{jt}, T_{ij} = 0\right) \tag{6.1}$$

where $Y^p_{ijt} = 1$ if teacher i earned a performance bonus in performance pay district j in time period t. The vector $\mathbf{w}_{it}$ contains individual teacher characteristics that may vary over time t, and $\mathbf{v}_{jt}$ is a vector of district j's characteristics that may also vary over time. Expression 6.1 defines the the difference in the probability of earning a performance bonus in performance pay district j for teachers who are in the treatment group and comparison group. Recall from chapter 3 that d_{ij} is an indicator for whether teacher i self-selects into performance pay district j. By definition, $d_{ij} = 1$ for all teachers in the treatment group ($T_{ij} = 1$). That is, all teachers who join the district after the adoption of performance pay are among the teachers who chose performance pay district j over his or her best alternative steps and lanes district option. However, teachers who joined district j prior to its adoption of performance pay may or may not have chosen to enter the district under the performance pay regime. That is, some teachers in the control group are potentially the type that would not have self-selected into district j had she known that it would become a performance pay district. Therefore, the comparison group potentially contains teachers with $d_{ij} = 0$.

Given this result, if the difference in expression 6.1 is positive, then following the logic in proposition 2, district j's performance pay program is inducing a selection effect. That is, if new teachers are more likely to earn a performance bonus relative to teachers who joined the district prior to performance pay, then the performance pay program is attracting higher ability teachers. In the next section, I describe the procedure that I use to estimate the difference in bonus probabilities across the two groups of teachers.

6.2 Empirical model

Here, I describe the method that I employ to estimate the difference in expression 6.1 using teacher level longitudinal data. To estimate expression 6.1, I use a probit model to estimate the marginal effect of membership in the treatment group versus comparison group. Consistent with a probit model, I assume the following functional form for the conditional probability that teacher i earns a performance bonus within district j in time t:

$$P\left(Y_{ijt}^{p} = 1 | \mathbf{w}_{it}, \mathbf{v}_{jt}, T_{ij}\right) = \Phi\left(\alpha T_{ij} + \mathbf{w}_{it}\boldsymbol{\pi}_1 + \mathbf{v}_{jt}\boldsymbol{\pi}_2 + \omega_t\right)$$

where α, $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$, and ω_t are parameters to be estimated: ω_t is a time specific effect, α is a parameter capturing the latent effect of membership in the treatment group, and the vectors $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ are parameters capturing the effects of individual and district characteristics on bonus award probabilities. The scalar Y_{ijt}^{p} is an indicator that denotes whether teacher i earned a performance bonus in district j in time period t. The indicator variable T_{ij} denotes whether the teacher is in district j's treatment or comparison group. That is, whether teacher i joined district j before or after the district adopted performance pay. The vector $\mathbf{w}_{it}$ contains individual teacher characteristics that may vary over time. In previous sections $\mathbf{w}_{it}$ included only teacher education and experience. In some specifications in section 7.3, $\mathbf{w}_{it}$ includes additional individual covariates that may be related to bonus award probabilities or individual behavior, including: years of service in district j, squared terms of teaching experience and tenure, base teacher salary, and an indicator for part-time or full-time employment. The vector $\mathbf{v}_{jt}$ contains district level characteristics that may vary over time. $\Phi()$ is the cumulative distribution function for a standard normal random variable.

For the analysis that follows I estimate this conditional probability within one district over several time periods spanning before and after the district adopted a performance pay program. I assume that district characteristics are constant over the period of analysis so that $\mathbf{v}_{jt}\boldsymbol{\pi}_2$ becomes $\mathbf{v}_j\boldsymbol{\pi}_2$ and is absorbed into a district-specific constant term or fixed-effect in the following specification:

$$P\left(Y_{ijt}^{p} = 1|\mathbf{w}_{it}, \mathbf{v}_{jt}, T_{ij}\right) = \Phi\left(\alpha T_{ij} + \mathbf{w}_{it}\boldsymbol{\pi} + \omega_t\right) \tag{6.2}$$

where the vector $\boldsymbol{\pi}$ contains a district specific constant term and $\mathbf{w}_{it}$ contains a column of ones. The remaining notation in equation 6.2 remains as it was defined above.

The parameters in equation 6.2 can be estimated by maximum likelihood and the results may be used to test whether the difference in expression 6.1 is positive, which would imply a positive selection effect. Using the specification in equation 6.2, the difference in expression 6.1 can be written as:

$$\Phi\left(\alpha + \mathbf{w}_{it}\boldsymbol{\pi} + \omega_t\right) - \Phi\left(\mathbf{w}_{it}\boldsymbol{\pi} + \omega_t\right)$$

Since the cumulative distribution function of a normal random variable is continuous, positive, and increasing, this expression implies that the difference in expression 6.1 is positive if and only if $\alpha > 0$. Using the probit estimate of α ($\hat{\alpha}$), we can test whether $\alpha > 0$. If the results suggest that $\alpha > 0$, then we have evidence that district j's performance pay program is inducing a positive selection effect. In the next chapter I employ this estimation strategy for an individual performance pay district.

Chapter 7

Longitudinal Analysis

In this chapter, I use teacher level longitudinal data from one performance pay district in Minnesota, Osseo Public School District (Osseo PSD), to implement the estimation procedure described in the previous chapter. In section 7.1 I describe the district and its performance pay program. In section 7.2 I describe Osseo PSD's longitudinal data. In section 7.3 I present results of the longitudinal analysis.

7.1 Osseo PSD and its Q-Comp program

Osseo PSD is a district participant in Minnesota's performance pay program, Q-Comp. The Q-Comp program is a state-wide voluntary performance pay program that rewards participating school districts with increased funding up to $260 per student. To participate in Q-Comp, a school district or charter school must design and implement a plan that meets Q-Comp standards in five categories. Q-comp requires that participating districts and charters develop a career ladder or advancement options for teachers, provide job-embedded professional development, administer multiple teacher evaluations, adopt a performance pay regime, and adopt an alternative salary schedule. If a district or charter submits a plan that meets state requirements in each of these criteria, then it is approved by the state to participate in Q-Comp.

Table 7.1: Osseo PSD characteristics vs. state

	MN	Osseo PSD
Percent Minority	24.3	43.6
Percent Free or Reduced Price Lunch	32.7	32.5
Pupil/Teacher Ratio	15.7	18.2
Per Capita Income (Census 2000)	$ 23,198	$ 26,694
Median Family Income (Census 2000)	$ 56,874	$ 71,764
Source: 2008 -2009 Common Core of Data and 2000 Census		

Q-Comp provides flexibility in how districts and charter schools design plans that meet state standards in the five components. Below, I describe Osseo PSD and its approved Q-Comp plan.

Osseo PSD joined the Minnesota Q-Comp program in the 2006-07 school year. Employing over 1,100 full-time equivalent teachers and servicing over 21,000 students, Osseo PSD is the fifth largest school district in Minnesota in both categories and the second largest participant in Q-Comp. Osseo PSD serves cities located primarily in the northwest suburbs of Minneapolis: Brooklyn Center, Brooklyn Park, Corcoran, Dayton, Hassan, Maple Grove, Osseo, and Plymouth. Since it is located in the metro area, Osseo PSD serves a student body that is more affluent than the typical student in Minnesota, and students in Osseo PSD are more likely to come from a racial minority. Table 7.1 presents a comparison of Osseo PSD's student body, student teacher ratio, and income measures to the state of Minnesota. As table 7.1 illustrates, Osseo PSD contains a larger proportion of students from minority groups relative to the state population. The district also maintains a higher student-teacher ratio, which implies Osseo PSD has larger class sizes relative to the state average. Also, table 7.1 suggests that the population served by Osseo PSD is more affluent than the general state population. The per capita income and median household income within the Osseo PSD boundary are both higher than the state values.

In April of 2006, the Minnesota Department of Education notified Osseo PSD that its proposed Q-Comp program had been approved, and it was authorized to implement its Q-Comp program beginning in the 2006-07 school year. Below I describe how Osseo PSD satisfied the five program components.

To meet the state requirement of developing a career ladder or advancement options for teachers, Osseo PSD introduced five teachers positions with varying levels of compensation and responsibility. The five positions are the following: instructional coach, instructional coach facilitator, elementary team leader, secondary site department leader, and district curriculum specialist. Each of these positions carries added responsibility related to teacher training or evaluation and providing group leadership and coordination. Preference in filling these roles is given to experienced teachers or teachers who hold advanced degrees. Each position also provides the teacher with additional income.

A major goal of Osseo's Q-Comp program is to improve annual growth in student achievement on standardized tests in math and reading. To do so, Osseo PSD required each of its 29 schools set school-specific goals related to improving reading and/or math achievement. To meet the state requirement that the Q-Comp plan provide job-embedded professional development, Osseo PSD designed school-specific teams that will meet regularly to help ensure each school meets its annual math and reading growth goal. The teams were designed to provide a wide variety of professional development activities, including: demonstration teaching, team teaching, mentoring, in-school workshops, and professional learning communities.

To meet state requirements for teacher evaluations, Osseo PSD's Q-Comp plan requires that principals and instructional coaches regularly evaluate and meet with classroom teachers. The district plan requires that teachers undergo at least three performance evaluations per year. Osseo PSD uses a district specific performance appraisal system (the Osseo Performance Appraisal), which measures a teacher's proficiency in four domains: planning for student learning, creating an environment of student learning, teaching for student learning, and professionalism. In Osseo's

Q-Comp plan, teachers must meet proficiency in each domain to qualify for a performance bonus or base salary increment in the district's alternative salary schedule, described below.

The Q-Comp program requires that participating districts adopt an alternative teacher salary schedule. Osseo PSD's Q-Comp plan meets this criterion by altering the way in which teachers receive a step increase on the salary schedule. In a traditional steps and lanes schedule, teachers nearly automatically earn a step increase for each year of teaching experience they accumulate. In Osseo's Q-Comp plan, teachers must meet Osseo performance appraisal standards to receive an annual performance increment, or step increase. Teachers must also meet this standard to receive any of the three annual performance bonus awards. The Osseo Q-Comp plan does not alter teacher lane assignments in the salary schedule. All teachers are placed in a lane based on their educational attainment, and lane assignments are not related to teacher performance measures.

The performance pay component of Q-Comp requires that participants base 60% of teacher pay bonuses on teacher performance measures that align with student academic achievement or progress. To meet state requirements for this criterion, Osseo PSD's Q-Comp program awards three types of one-time annual performance bonuses, which in sum total $2,250.

Osseo's largest bonus is determined by a teacher's successful completion of a "professional growth plan". If teachers demonstrate, to their instructional coach, successful implementation of a professional growth plan that is related to student achievement, the teacher earns a bonus of $1,800. This accounts for 80% of the total potential bonus award in Osseo's Q-Comp program. Professional growth plans could include the following: innovative instructional strategies, new curriculum aligned with graduation standards, research journals, and work portfolios.

In Osseo's Q-Comp plan, the remaining ten percent of the potential bonus award is divided equally into two \$225 bonuses linked to the teacher's own student's achievement gains and school-wide student achievement. Teachers can earn a "student learning goals" bonus of \$225 if their students demonstrate achievement growth that exceeds a predetermined learning goal. Teachers may demonstrate student achievement growth through pre and post assessments, student projects, student work portfolios, work completion, and gains in student attendance. Finally, teachers may earn an additional \$225 "site goals" bonus if the teacher's school meets its predetermined growth targets for standardized math and reading scores.

According to figures in Sojourner et al. (2011), Osseo's Q-Comp plan rewards individual teacher performance and inputs more heavily than the average participant's Q-Comp plan and relies less on subjective performance evaluations. For example, according to the definitions in Sojourner et al., the Osseo Q-Comp plan awards \$2,025 for teacher inputs or performance, compared to \$872 for the average Q-Comp plan. Also, the Osseo Q-Comp plan awards \$225 for school-wide incentives and zero dollars for subjective evaluations. These amounts are both lower than the state-wide average Q-Comp plan, which awards \$247 for school-wide performance and \$1,100 for subjective performance evaluations.

7.2 Data, variable definitions, and descriptive statistics

In this analysis, I employ two administrative data sets from two sources: the Minnesota Department of Education (MDE) and Osseo PSD. The data from MDE provides a longitudinal data set for all teachers in the state of Minnesota from the 2000-01 school year through the 2009-10 school year. For each teacher-year observation, the data includes the following variables: years of teaching experience, education level, contract salary, district in which the teacher is employed, and an indicator for whether the teacher works full or part-time. The Osseo PSD data provides a panel of all teachers who have earned both individual Q-Comp bonuses for each year that the district has participated in Q-Comp: from the 2006-07 school year to the 2009-10 school year. For each teacher-year observation, this data set provides an indicator that denotes whether the teacher received both individual Q-Comp bonuses: the "professional growth plan" bonus and the "student learning goals" bonus. This data also provides the amounts of the bonus, but these amounts provide no additional information since each bonus is awarded on an all-or-nothing basis. Osseo's Q-Comp bonus are awarded in partial amounts only if the teacher is less than full-time, and partial bonuses are not awarded for partial completion of bonus requirements.

The data from MDE provides teacher covariates and allows one to track teacher mobility within the state. From the data, one can identify teachers that have taught in more than one district within the state throughout the time period 2001-2010. Also, since this data provides an entire listing of the teacher population in Minnesota, missing teacher-year observations identify years in which a teacher left the Minnesota public school system. These teachers have either retired, decided to work in an alternative field, decided to teach in a private school or other state, or passed away. For this analysis, the crucial information the MDE data provides is teacher control variables (education, experience, full-time indicator, contract salary) and an identifier for the following: which teachers joined Osseo PSD in the years immediately prior to its adoption of Q-Comp (comparison group) and which teachers joined Osseo PSD after its adoption of Q-Comp (treatment group).

The Osseo PSD data provides the primary outcome variable that I employ in the longitudinal analysis: an indicator by teacher and year denoting whether the teacher earned both individual Q-Comp bonuses. The two data sets can be merged by a teacher identifier to create one panel data set that merges the treatment and comparison group identifier and teacher covariates with the Q-Comp bonus indicator.

Table 7.2: Cohort characteristics at time of hiring

	Cohort:	Avg. Exper	%BA	% BA +	% MA	% >MA	N
Pre Q-Comp	joined Osseo PSD in 2002	3.97	45%	24%	22%	9%	96
	joined Osseo PSD in 2003	7.05	17%	32%	27%	24%	41
	joined Osseo PSD in 2004	5.16	28%	30%	34%	7%	82
	joined Osseo PSD in 2005	4.07	27%	41%	23%	9%	75
	joined Osseo PSD in 2006	4.69	41%	27%	22%	10%	117
Post Q-Comp	joined Osseo PSD in 2007	4.68	35%	32%	24%	9%	119
	joined Osseo PSD in 2008	4.73	30%	28%	26%	15%	92
	joined Osseo PSD in 2009	5.02	27%	18%	42%	13%	45
	joined Osseo PSD in 2010	5.66	24%	34%	29%	12%	41
	joined pre Q-Comp	4.74	34%	30%	25%	11%	411
	joined post Q-Comp	4.88	31%	29%	28%	12%	297
	difference (post-pre)	0.14	-0.03	-0.01	0.03	0.01	
	Note: none of these differences are statistically significant (p>0.10)						

In section 7.3, I use this merged panel data set to estimate equation 6.2. In the remainder of this section, I define key variables, provide descriptive statistics, and test for differences in observable characteristics between teachers in the treatment and comparison groups.

Tables 7.2 and 7.3 provide information about the size, education level, and experience levels of teachers in each cohort that has joined Osseo PSD between the years 2002 and 2010. Both tables provide the following information for each cohort of district joiners: the teachers' average experience (Avg. Exper), the percent of teachers with a bachelor's degree (%BA), the percent of teachers with a bachelor's degree plus additional credits (%BA+), the percent of teachers with a master's degree (%MA), the percent of teachers with education level higher than a master's degree (%>MA), and the number of teachers who joined the district in each year (N).

Tables 7.2 and 7.3 differ in that they provide this information at different points in time. Table 7.2 provides cohort statistics for the year in which the cohort joined the district, or year of hire. For example, in the 2001-2002 school year, Osseo PSD hired 96 new teachers. These 96 teachers had an average of 4 years of teaching experience

Table 7.3: Cohort characteristics two years after hiring

	Cohort:	Avg. Exper	%BA	% BA +	% MA	% >MA	N
Pre Q-Comp	joined Osseo PSD in 2002	5.68	38%	27%	21%	14%	78
	joined Osseo PSD in 2003	8.52	18%	27%	33%	21%	33
	joined Osseo PSD in 2004	6.93	25%	29%	35%	12%	69
	joined Osseo PSD in 2005	5.82	20%	26%	30%	25%	61
	joined Osseo PSD in 2006	6.27	28%	30%	26%	16%	109
Post Q-Comp	joined Osseo PSD in 2007	6.99	23%	30%	33%	14%	73
	joined Osseo PSD in 2008	6.50	26%	22%	37%	15%	46
	joined Osseo PSD in 2009	NA	NA	NA	NA	NA	NA
	joined Osseo PSD in 2010	NA	NA	NA	NA	NA	NA
	joined pre Q-Comp	6.40	27%	28%	28%	17%	350
	joined post Q-Comp	6.80	24%	27%	35%	14%	119
	difference (post-pre)	0.40	-0.03	-0.01	0.07	-0.02	
	Note: none of these differences are statistically significant (p>0.10)						

at the time of hire, and almost half of these teachers (45%) possessed only a bachelor's degree at the time of hire. Similarly, in 2009 Osseo PSD hired 45 new teachers and at the time of hire these teachers had an average of 5 years of teaching experience and most held at least a master's degree (55%).

Table 7.3 provides cohort statistics after two years of service in the district. Therefore, table 7.3 contains teachers who were hired two years prior and remained in Osseo PSD for at least 2 years. The table provides a description of those teachers' characteristics in their second year at Osseo PSD. For example, of the 96 teachers that joined Osseo PSD in 2002, two years later, in 2004, 73 teachers remained in the district. These 73 teachers had an average of 5.68 years of teaching experience in 2004. Similarly, of the 92 teachers that joined the district in 2008, 46 of these teachers remained in the district in 2010 and their average experience was 6.5 years, and the majority of those teachers held a master's degree or higher in 2010.

Tables 7.2 and 7.3 also illustrate the definition of the treatment and comparison groups used in this study. Treatment group teachers include all teachers who joined Osseo PSD in the 2006-07 school year through the 2009-10 school year. Teachers in the comparison group are those who joined Osseo PSD in the 5 years prior to its adoption of Q-Comp: teachers who joined Osseo PSD in the 2001-02 school year through the 2005-06 school year.

Using these definitions of treatment and comparison groups, tables 7.2 and 7.3 provide an important test of the selection on observables hypothesis. Above, I suggested that teachers from the treatment group may be more experienced and educated than teachers from the comparison group if those characteristics are related to the probability of earning a performance bonus. That is, teachers who self-select into the performance pay district with knowledge of the program may differ, in terms experience and education, from teachers who joined the district without knowledge of the program. Tables 7.2 and 7.3 provide tests of this hypothesis at the time of hire and two years into the teachers' employment at the district. Both tables show that there is not a statistically significant difference in the experience and education levels of teachers who joined prior to or after the adoption of Q-Comp. That is, neither at the time of hire nor after two years of service am I able to detect a statistically significant difference in education or teaching experience between pre and post Q-Comp joiners.

This result strongly suggests that teacher selection into Osseo PSD after Q-Comp is not related to observable teacher characteristics. However, this does not rule out the possibility that treatment group teachers differ from comparison group teachers in unobservable ways, perhaps in terms of ability, ambition, or teaching quality. To test this hypothesis, I compare bonus award probabilities across treatment and comparison groups. Assuming that Osseo's Q-Comp bonus awards are more likely to be awarded to teachers with high ability, ambition, or teaching quality, differences in award probabilities across groups might suggest that differences in unobserved teacher characteristics exists across groups. For the remainder of this section and in section 7.3 I attempt to detect a difference in Q-Comp bonus award probabilities between the group of teachers who joined Osseo PSD prior to Q-Comp and those who joined

Table 7.4: Cohort bonus probabilities by year

	Earned both Individual Q-Comp Bonuses								
		2007		2008		2009		2010	
	Cohort	%	N	%	N	%	N	%	N
Pre Q-Comp	joined Osseo PSD in 2002	90%	58	94%	54	96%	53	98%	49
	joined Osseo PSD in 2003	80%	30	86%	29	90%	30	93%	29
	joined Osseo PSD in 2004	86%	66	92%	63	98%	58	96%	56
	joined Osseo PSD in 2005	84%	61	88%	58	93%	56	98%	52
	joined Osseo PSD in 2006	70%	117	76%	109	95%	85	98%	84
Post Q-Comp	joined Osseo PSD in 2007	50%	119	59%	107	86%	73	98%	60
	joined Osseo PSD in 2008	.	0	46%	92	82%	51	93%	46
	joined Osseo PSD in 2009	.	0	.	0	69%	45	90%	39
	joined Osseo PSD in 2010	.	0	.	0	.	0	73%	41

after Q-Comp. If the post Q-Comp joiners are more likely to earn bonuses than pre Q-Comp joiners, then we may have evidence that Osseo's Q-Comp plan is inducing a positive selection effect.

In the analysis that follows, I examine two sets of teacher populations. The first population contains all teachers who joined the district in any year between 2001-02 and 2009-10. This population contains teachers who joined the district and either remained in the district through 2009-10 or entered in a year prior to 2009-10 and exited the district before the year 2010. The second population contains only teachers who joined the district in any year between 2002 and 2010 and remained in the district through the 2009-10 school year. I refer to the first population as the unrestricted teacher population and the second as the restricted population, since it is restricted to the population of teachers who join the district and do not exit before 2010.

Tables 7.4 and 7.5 show the proportion of teachers in each cohort that earn both individual Q-Comp bonuses. Table 7.4 shows that percent of teachers earning both bonuses and the number in each cohort each year for the unrestricted teacher population. Table 7.5 shows that percent of teachers earning both bonuses for the restricted teacher population.

Table 7.5: Restricted cohort bonus probabilities by year

	Earned both Individual Q-Comp Bonuses					
	Cohort	**2007**	**2008**	**2009**	**2010**	**N**
Pre Q-Comp	joined Osseo PSD in 2002	98%	96%	98%	98%	49
	joined Osseo PSD in 2003	90%	93%	93%	93%	29
	joined Osseo PSD in 2004	98%	98%	98%	96%	56
	joined Osseo PSD in 2005	98%	96%	98%	98%	52
	joined Osseo PSD in 2006	98%	98%	96%	98%	84
Post Q-Comp	joined Osseo PSD in 2007	98%	97%	98%	98%	60
	joined Osseo PSD in 2008	.	93%	96%	93%	46
	joined Osseo PSD in 2009	.	.	87%	90%	39
	joined Osseo PSD in 2010	.	.	.	73%	41

Table 7.4 illustrates two interesting details regarding Q-Comp bonus awards. First, in each year there appears to be a positive relationship between the number of years a teacher has been employed in the district and the likelihood of earning both bonus awards. For example, in each year, new teachers are the least likely to earn both performance bonuses, and in each year, teachers who joined the district in the earlier years (2002-2004) have a higher probability of earning both performance bonuses. Second, table 7.4 suggests that even among teachers who are relatively new to the district, the probability of earning both bonuses is high. More than half of the teachers with less than three years of service in the district earn both performance bonuses, and in several cohorts over 90% of teachers earn both bonuses in a given year.

Table 7.5 shows the same information as table 7.4, but for the teacher population that joined the district and did not exit before 2010 (the restricted population). I wish to highlight two important differences between this table and table 7.4. First, it appears that teachers in the restricted population are more likely to earn a bonus in every year and cohort. Assuming that the bonus is awarded based on real differences in teacher quality or ambition, this might imply that the teachers who join the district and stay through the 2010 school year are of higher quality or more ambitious, on average. Second, within this population, the relationship between years of service in the district and bonus award probabilities appears to be significantly weakened. For example, for the 2007 and 2008 joiners in the restricted sample, the proportion of teachers earning both bonuses is the same in their first year and in the 2010 school year.

The key insight from table 7.5 is that it appears that the teachers in the restricted sample are different than teachers in the unrestricted sample. For this reason, I test for differences in bonus award probabilities between treatment and comparison groups within both populations.

7.3 Results

In this section, I estimate differences in bonus probabilities, unconditional and conditional on teacher characteristics and year effects. I estimate these differences for both the restricted and unrestricted populations. In tables 7.6 and 7.7, I present a comparison of teacher characteristics and unconditional bonus probabilities for the unrestricted and restricted samples, respectively. In tables 7.8 and 7.9, I present estimates of conditional bonus probabilities (probit estimates of the parameters in equation 6.2) for the unrestricted and restricted populations, respectively. The results suggest that unconditional on teacher characteristics, teachers who joined Osseo PSD after the adoption of Q-Comp are less likely to earn both performance bonuses relative to teachers who joined prior to Q-Comp. However, after accounting for the fact that pre Q-Comp joiners have worked within the district for more years than post Q-Comp joiners, the difference in bonus probabilities between the two groups is statistically insignificant. Overall, I fail to find evidence that Osseo's Q-Comp program is inducing a positive selection effect.

Tables 7.6 and 7.7 compare teacher characteristics and unconditional bonus prob-

Table 7.6: Bonus probabilities and characteristics by group (unrestricted)

Variable	Joined Post Q-Comp	Joined Pre Q-Comp	Difference (post-pre)
Earned both Individual Q-Comp Bonuses	69.4%	88.9%	-0.19***
Total Teaching Experience	5.69	8.48	-2.79***
Education: BA	29%	14%	0.15***
Education: BA (+ credits)	28%	22%	0.06*
Education: MA	30%	28%	0.02
Education: more than MA	13%	35%	-0.22***
Base Contract Salary/1000	45.05	51.85	-6.80***
Years in Osseo PSD	1.91	5.02	-3.10***
Employed Full-time	95%	96%	-0.02
Observations	673	1197	
Note: *** p<0.01, ** p<0.05, * p<0.1			

Table 7.7: Bonus awards and characteristics by group (restricted)

Variable	Joined Post Q-Comp	Joined Pre Q-Comp	Difference (post-pre)
Earned both Individual Q-Comp Bonuses	93.4%	96.9%	-0.04*
Total Teaching Experience	6.16	8.56	-2.40***
Education: BA	29%	13%	0.16***
Education: BA (+ credits)	26%	22%	0.04
Education: MA	31%	29%	0.02
Education: more than MA	14%	36%	-0.22***
Base Contract Salary/1000	44.59	51.77	-7.18***
Years in Osseo PSD	2.08	5.16	-3.08***
Employed Full-time	89%	95%	-0.06***
Observations	497	1080	
Note: *** p<0.01, ** p<0.05, * p<0.1			

abilities across the treatment and comparison groups. Table 7.6 shows the results for the unrestricted, and table 7.7 shows the results for the restricted population. In both tables teacher-year observations are pooled across years, giving a total of 1,870 teacher-year observations in the unrestricted population and 1,577 teacher-year observations in the restricted population.

Within each population, differences between the pre Q-Comp joiners and post Q-Comp joiners are remarkably similar. In both populations, post Q-Comp joiners are less likely to earn both bonuses, but this does not account for differences in characteristics between groups. In both populations, pre Q-Comp joiners differ in their characteristics from post Q-Comp joiners largely because they are separated by the time in which they joined the district or began their teaching careers. For example, in both populations, the average post Q-Comp joiner has fewer years of teaching experience, is less educated, earns a lower base salary, has fewer years of experience within the district, and is less likely to be employed full-time.

These differences in characteristics between the two groups might explain why post Q-Comp joiners are less likely to earn both Q-Comp bonuses. In fact, after controlling for these differences it is possible that post Q-Comp joiners are more likely to earn both bonuses. That is, if we compare a pre Q-Comp joiner to a post Q-Comp with identical characteristics, we might find that the post Q-Comp joiner is more likely to earn the individual Q-Comp bonuses. To explore this possibility, I estimate the probability of earning both Q-Comp bonus awards, conditional on whether the teacher is a post or pre Q-Comp joiner, the teacher's characteristics, and year effects.

Tables 7.8 and 7.9 present maximum likelihood estimates of the parameters in equation 6.2 for each teacher population pooled over the years 2006-07 to 2009-10. Each of the four columns in the table represents results of a separate estimation of the conditional probability of earning both Q-Comp bonuses. Each model includes a different set of individual characteristics as controls, however all of the models include an indicator denoting whether the teacher is a post Q-Comp joiner, to measure the difference in conditional bonus probability across the treatment and comparison groups. All models account for the number of years a teacher has been employed in Osseo PSD. I control for this by including a count measure of the number of years a teacher has been employed in the district at year t . I also include the square of this count variable, which allows for a nonlinear relationship between bonus awards and district specific experience. Also, each model includes a full set of year dummies to account for differences in bonus awards by year.

Table 7.8: Conditional probability of earning Q-Comp bonuses (unrestricted)

VARIABLES	(1)	(2)	(3)	(4)
Joined Post Q-Comp	-0.033	-0.028	-0.029	-0.017
	(0.044)	(0.043)	(0.043)	(0.042)
Total Teaching Experience			0.008	-0.003
			(0.006)	(0.007)
Experience Squared/100			-0.032*	-0.009
			(0.019)	(0.020)
Education: BA (+ credits)			0.026	0.005
			(0.030)	(0.032)
Education: MA			0.011	-0.060
			(0.033)	(0.050)
Education: more than MA			0.043	-0.060
			(0.035)	(0.058)
Employed Full-Time			0.065	-0.018
			(0.064)	(0.057)
Years in Osseo PSD	0.119***	0.116***	0.115***	0.119***
	(0.021)	(0.021)	(0.022)	(0.022)
Years in Osseo PSD Squared	-0.009***	-0.010***	-0.010***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)
Base Contract Salary/1000		0.003**		0.008**
		(0.001)		(0.003)
Observations	1870	1870	1870	1870

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table shows marginal effects evaluated at mean values of the covariates
Standard errors are adjusted for clustering at the teacher level
All models include a full set of year dummies

In both tables the columns represent identical specifications, however table 7.8 presents results for the unrestricted population and table 7.9 presents results from the restricted population. Column 1 includes the fewest number of controls: the year effects and district specific experience. Each additional column adds additional controls to the base specification in column 1. Column 2 adds the teacher's base contract salary as an additional control. This is meant to control for teacher behavior that may vary by income. For example, if the marginal utility of income among teachers is diminishing, teachers who earn a high base salary may have less of an incentive to earn a Q-Comp bonus than low income teachers. However, in a shirking model, teachers who earn high base salary may have more incentive to exert high effort and earn a performance bonus. According to the shirking model, these teachers have more incentive than low income teachers because they have more to lose if they were to be fired for lack of effort. In either case, base salary might influence teacher behavior and bonus probabilities.

In column 3, I drop base contract salary as a control and add measures of teaching experience (a count variable of teaching experience and experience squared), measures of teacher education levels, and an indicator for whether the teacher in employed full-time. In column 4, I include all covariates. In each column, the tables report the marginal effects of a probit model, evaluated at the mean values of the covariates. That is, each coefficient reports the approximate change in the probability of earning both bonuses as a result of a small increase (or a switch from 0 to 1) in the explanatory variable, holding the other covariates constant at their sample means.

The probit results for the restricted and unrestricted samples are similar. The results in tables 7.8 and 7.9 fail to find evidence of a selection effect. After controlling for only the number of years a teacher has been employed in the district, in both populations the difference in the probability of earning a bonus between the pre and post Q-Comp joiners is small and insignificant. In fact, in both tables the only strong

Table 7.9: Conditional probability of earning Q-Comp bonuses (restricted)

VARIABLES	(1)	(2)	(3)	(4)
Joined Post Q-Comp	0.012	0.011	0.012	0.014
	(0.024)	(0.022)	(0.021)	(0.020)
Total Teaching Experience			-0.002	-0.005
			(0.003)	(0.004)
Experience Squared/100			0.010	0.017
			(0.012)	(0.013)
Education: BA (+ credits)			0.022	0.019
			(0.016)	(0.016)
Education: MA			0.007	-0.013
			(0.018)	(0.023)
Education: more than MA			0.031*	0.003
			(0.016)	(0.022)
Employed Full-Time			0.011	-0.019
			(0.026)	(0.017)
Years in Osseo PSD	0.030***	0.025***	0.029***	0.028***
	(0.010)	(0.009)	(0.010)	(0.009)
Years in Osseo PSD Squared	-0.002***	-0.002***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Base Contract Salary/1000		0.002**		0.002***
		(0.001)		(0.001)
Observations	1577	1577	1577	1577

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table shows marginal effects evaluated at mean values of the covariates
Standard errors are adjusted for clustering at the teacher level
All models include a full set of year dummies

predictors of earning a bonus are the number of years a teacher has been employed in the district and the teacher's base contract salary. The relationship between district specific experience and the bonus probabilities is as expected. As a teacher gains in experience within the district one would expect that she would be better prepared to earn the Q-Comp bonuses. She may be better connected to an instructional coach and more aware of the expectations or standards that must be met to earn a bonus. Also, as one would expect, the benefit of district specific experience is nonlinear in that the first few years are more beneficial in earning a bonus than are successive years. Another interesting result is that the marginal effect of base teacher salary is positive and significant in both tables and all specifications. This suggests that the shirking model may be relevant in the case of teachers in Osseo PSD, since the result implies that higher base compensation may be accompanied by greater effort.

The analysis presented here has failed to detect a positive selection effect of a specific performance pay program. There are many reasons why this might be. I could have potentially miss-specified the model or used methods that are not sensitive enough detect such an effect. However, it might be true that Osseo's performance pay program is simply not affecting teacher recruitment or retention. Several characteristics of Osseo's Q-Comp program suggest that this may be the case. As I described in section 3.6, a district stands a better chance of attracting only high ability teachers if it offers a low base compensation relative to comparable competing districts and offers a large performance bonus. In addition, the standards necessary to earn a bonus should be challenging enough to achieve so that low ability teachers will be sufficiently discouraged from applying to positions within the district.

It is unlikely that Osseo's performance pay program sufficiently incorporates these components. First, Osseo's base compensation is among the highest in the state. Controlling for teacher education and experience, teachers in Osseo PSD were paid $5,600 more in base salary than teachers in the average school district in Minnesota during the 2009-10 school year. This ranks Osseo PSD as the state's 14th highest paying district among its 333 regular school districts. Furthermore, Osseo PSD pays its teachers over $1,500 more on average (again controlling for education and experience) than teachers in the comparable neighboring districts of Wayzata and Anoka-Hennepin. Aside from offering high base compensation, Osseo's Q-Comp program offers a nearly automatic pay increase of $2,250 to a full-time teacher. From table 7.4 it appears that over 90% of teachers earn both individual bonus awards. Once a teacher is established within the district, she is almost guaranteed to receive an extra $2,000 in bonuses. This high base compensation and nearly guaranteed bonus is likely to be equally appealing to teachers of low and high ability. For these reasons, it is unlikely that Osseo's performance pay program is inducing differential selection on ability. This may explain why the analysis in this section failed to find evidence of a selection effect.

Chapter 8

Conclusion

Proponents of teacher performance pay often cite selection effects among the potential benefits of the pay reform. They argue that by offering performance pay, districts might attract only high ability or highly motivated teachers. Evidence in other professions supports the claim that individuals respond to monetary incentives, and in some cases it appears that firms offering incentive contracts attract more effective or higher ability workers. However, the evidence regarding the effects of providing teacher incentives through performance pay programs is mixed in the case of effort effects, and up to now, no studies have attempted to estimate whether performance pay programs induce a selection effect.

In this study, I have attempted to fill that void. I showed that when teachers are aware of their own ability levels and performance pay programs offer teachers low base compensation and high bonus awards for true measures of teaching performance, teacher performance pay programs can induce sorting on ability. In addition, I showed that the existence of a selection effect can be measured by estimating the difference in conditional total pay or bonus probabilities between teachers who self-select into performance pay and teachers who are exogenously assigned to the program. If teachers who self-select earn a higher expected total pay or are more likely to earn a bonus than an exogenously assigned teacher, I showed that this implies that self-selectors are higher ability teachers on average.

I used this result to test whether common performance pay programs are inducing a selection effect. I tested for this difference in cross-sectional data from the Schools and Staffing Survey and in longitudinal data from a single performance pay district in Minnesota. In each case I failed to find evidence of a selection effect. In the cross-sectional analysis, I found that while self-selectors earn a higher expected total pay in performance pay districts, they are also less likely to earn a performance bonus. In the longitudinal analysis, I found that teachers who joined the district in the four years after the district had adopted performance pay are not measurably different from teachers who joined the district in the 5 years prior to adoption. Post performance pay joiners are not measurably different in terms of their education, experience, or likelihood of earning a performance bonus.

While I failed to find evidence of a selection effect, that should not be taken as proof that performance pay programs, in general, do not produce their advertised benefits. This study provides a first attempt at measuring any selection effects that performance pay programs might produce. The cross-sectional analysis depends on several strong assumptions about the nature of the teacher's selection decision, the validity of an instrumental variable, the uniformity of performance pay programs, and the unobserved characteristics of teachers who select into at least one performance pay program. These strong assumptions potentially undermine the credibility of the cross-sectional results. On the other hand, the longitudinal analysis relaxes these assumptions, providing a robust empirical design. However, the results of the longitudinal analysis are specific to the program and district under study. The failure to find evidence of a selection effect in Osseo PSD is likely a product of the district's high relative base compensation and nearly guaranteed bonus award.

For future research it would be informative to replicate this longitudinal analysis for several performance pay districts. The theory presented in this study suggests that if performance pay programs do indeed induce a selection effect, we are most likely to find these effects in districts that offer low relative base compensation, offer high bonus awards, and employ an objective and accurate measure of teacher quality. In the future, researchers should focus their energy on estimating selection effects in performance pay districts with these characteristics.

Bibliography

Adams, R. B. and D. Ferreira (2008). Do directors perform for pay? *Journal of Accounting and Economics 46*(1), 154–171.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Anoka-Hennepin School District 11 (2007). Working agreement by and between Anoka-Hennepin Independent School District no. 11 School Board and Anoka-Hennepin Education Minnesota. accessed online 8/19/09 http://www.nctq.org/.

Ballou, D., W. Sanders, and P. Wright (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics 29*(1), 37–65.

Banker, R., S. Lee, and G. Potter (1996). A field study of the impact of a performance-based incentive plan. *Journal of Accounting and Economics 21*(2), 195–226.

Barnett, J., G. Ritter, M. Winters, and J. Greene (2007). Evaluation of Year One of the Achievement Challenge Pilot Project in the Little Rock Public School District. Technical report, University of Arkansas.

Bridges, E. and B. Groves (1984). Managing the incompetent teacher. Number 29, ERIC/CEM School Management Digest Series, University of Oregon.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Working Paper 4483, National Bureau of Economic Research.

Eberts, R., K. Hollenbeck, and J. Stone (2002). Teacher performance incentives and student outcomes. *Journal of Human Resources 37*(4), 913–927.

Fernie, S. and D. Metcalf (1999). It's not what you pay it's the way that you pay it and that's what gets results: jockeys' pay and performance. *Labour 13*(2), 385–411.

Figlio, D. and L. Kenny (2007). Individual teacher incentives and student performance. *Journal of Public Economics 91*(5-6), 901–914.

Florida Department of Education (2007). Florida performance pay guidance: 2007-2008 and beyond. Technical report.

Foster, A. and M. Rosenzweig (1994). A test for moral hazard in the labor market: contractual arrangements, effort, and health. *The Review of Economics and Statistics 76*(2), 213–227.

Glewwe, P., N. Ilias, and M. Kremer (2003). Teacher incentives. Working Paper 9671, National Bureau of Economic Research.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal 113*(485), F64–F98.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization 7*, 24–52.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Working Paper 14607, National Bureau of Economic Research.

Ladd, H. (1999). The dallas school accountability and incentive program: an evaluation of its impacts on student outcomes. *Economics of Education Review 18*(1), 1–16.

Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy 110*(6), 1286–1317.

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review 99*(5), 1979–2011.

Lazear, E. (1986). Salaries and piece rates. *Journal of Business 59*(3), 405–431.

Lazear, E. (2000). Performance pay and productivity. *American Economic Review 90*(5), 1346–1361.

Minnesota Office of the Legislative Auditor (2009). Evaluation report: Q Comp - quality compensation for teachers.

Muralidharan, K. and V. Sundararaman (2009). Teacher Performance Pay: Experimental Evidence from India. Working Paper 15323, National Bureau of Economic Research.

National Center for Education Statistics (2010). Schools and Staffing Survey table library. accessed online 4/8/10, http://nces.ed.gov/surveys/sass/tables.asp.

Neal, D. (1997). The effects of catholic secondary schooling on educational achievement. *Journal of Labor Economics 15*(1), pp. 98–123.

Neal, D. (2011, January). The design of performance pay in education. Working Paper 16710, National Bureau of Economic Research.

Podgursky, M. (2007). Teams versus bureaucracies: personnel policy, wage-setting, and teacher quality in traditional public, charter, and private schools. In M. Berends, M. G. Springer, and H. Walberg (Eds.), *Charter School Outcomes*. Mahwah, NJ: Lawrence Erlbaum Associates.

Podgursky, M. and M. Springer (2007). Teacher performance pay: a review. *Journal of Policy Analysis and Management 26*(4), 909–949.

Sharpes, D. (1987). Incentive pay and the promotion of teaching proficiencies. *The Clearing House 60*, 406–408.

Shearer, B. (2004). Piece rates, fixed wages and incentives: evidence from a field experiment. *Review of Economic Studies 71*(2), 513–534.

Sojourner, A., K. West, and E. Mykerezi (2011, May). When Does Teacher Incentive Pay Raise Student Achievement?: Evidence from Minnesota's Q-Comp Program. Unpublished Manuscript. University of Minnesota.

Springer, M., L. Hamilton, D. McCaffrey, D. Ballou, V. Le, M. Pepper, J. Lockwood, and B. Stecher (2010, A). Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching. Technical report, TN: National Center on Performance Incentives at Vanderbilt University.

Springer, M. G., J. L. Lewis, M. W. Ehlert, M. J. Podgursky, G. D. Crader, L. L. Taylor, T. J. Gronberg, D. W. Jansen, O. S. Lopez, and D. A. Stuit (2010, B). District Awards for Teacher Excellence (D.A.T.E) Program: Final Evaluation Report. Policy evaluation report, National Center on Performance Incentives.

Springer, M. G., J. L. Lewis, M. J. Podgursky, M. W. Ehlert, L. L. Taylor, O. S. Lopez, B. Ghoshdastidar, and A. Peng (2010, C). District Awards for Teacher Excellence (D.A.T.E) Program: Year One Evaluation Report. Policy evaluation report, National Center on Performance Incentives.

Taylor, L. L., M. G. Springer, and M. Ehlert (2009). Teacher-designed performance-pay plans in Texas. In M. G. Springer (Ed.), *Performance Incentives*. Brookings Institution Press, Washington D.C.

Chapter 9

Appendix

9.1 Proofs

9.1.1 definitions and assumptions

Before presenting the proofs, it is useful to define notation and highlight key assumptions and relationships. Define indicator variable s_{ij} to denote whether teacher i has a lower expected utility in performance pay district j under low effort relative to her best available steps and lanes district option. That is, define $s_{ij} = 0$ if $E\left[U(Y_{ij}^{p}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] \geq U(Y_{is}^{s}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$ and $s_{ij} = 1$ if $E\left[U(Y_{ij}^{p}(\underline{e}_i, a_i), \mathbf{v}_j, \underline{e}_i)\right] < U(Y_{is}^{s}(\underline{e}_i, a_i), \mathbf{v}_s, \underline{e}_i)$.

Given this indicator, the conditional probability of each of the five cases in the model is the following:

$$P(1|\mathbf{w}_i, \mathbf{v}_j) = P(a_i \leq a'_{ij}|\mathbf{w}_i, \mathbf{v}_j) P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j, a_i \leq a'_{ij})$$

$$P(2|\mathbf{w}_i, \mathbf{v}_j) = P(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j) P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j, a_i > a'_{ij})$$

$$P(3|\mathbf{w}_i, \mathbf{v}_j) = P(a_i \leq a'_{ij}|\mathbf{w}_i, \mathbf{v}_j) P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j, a_i \leq a'_{ij})$$

$$P(4|\mathbf{w}_i, \mathbf{v}_j) = P(a^*_{ij} > a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j) P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j, a^*_{ij} > a_i > a_{ij})$$

$$P(5|\mathbf{w}_i, \mathbf{v}_j) = P(a_i \geq a^*_{ij}|\mathbf{w}_i, \mathbf{v}_j) P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j, a_i \geq a^*_{ij})$$

Assumption 1: Conditional on individual and district characteristics $(\mathbf{w}_i, \mathbf{v}_j)$, s_{ij} and teacher ability (a_i) are independent. That is, $P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j, a_i) = P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j)$ and $P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j, a_i) = P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)$. This assumption implies that the characteristics of a teacher's best available steps and lanes district are unrelated to the teacher's ability after controlling for the teacher's education and experience and performance pay district $j's$ characteristics. In other words, the teacher's ability level is not related to the likelihood that the her best available steps and lanes district will offer a higher total compensation than is offered by performance pay district j under the low effort choice. Intuitively, this assumption implies that districts are unable to identify or measure teacher ability in the hiring process.

Given assumption 1, the conditional probability of each of the five cases in the model can be expressed as the following:

$$
\begin{aligned}
P(1|\mathbf{w}_i, \mathbf{v}_j) =& P(a_i \leq a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j) \\
P(2|\mathbf{w}_i, \mathbf{v}_j) =& P(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j) \\
P(3|\mathbf{w}_i, \mathbf{v}_j) =& P(a_i \leq a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j) \\
P(4|\mathbf{w}_i, \mathbf{v}_j) =& P(a^*_{ij} > a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j) \\
P(5|\mathbf{w}_i, \mathbf{v}_j) =& P(a_i \geq a^*_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)
\end{aligned}
$$

Given these results, I summarize the relationship among the conditional probabilities in the following equations:

$$
\frac{P\left(1|\mathbf{w}_i, \mathbf{v}_j\right)}{P\left(3|\mathbf{w}_i, \mathbf{v}_j\right)} = \frac{P\left(2|\mathbf{w}_i, \mathbf{v}_j\right)}{P\left(4|\mathbf{w}_i, \mathbf{v}_j\right) + P\left(5|\mathbf{w}_i, \mathbf{v}_j\right)} = \alpha \tag{9.1}
$$

$$
\text{where } \alpha = \frac{P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j)}{P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)} > 0 \tag{9.2}
$$

$$
\Rightarrow P(1|\mathbf{w}_i, \mathbf{v}_j) + P(2|\mathbf{w}_i, \mathbf{v}_j) = \frac{\alpha}{\alpha + 1} \tag{9.3}
$$

$$
\Rightarrow P(3|\mathbf{w}_i, \mathbf{v}_j) + P(4|\mathbf{w}_i, \mathbf{v}_j) + P(5|\mathbf{w}_i, \mathbf{v}_j) = \frac{1}{\alpha + 1} \tag{9.4}
$$

In addition, assumption 1 implies the following results hold regarding probabilities that are conditioned on s_{ij}:

$$
\begin{aligned}
&P(1|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0) = P(3|\mathbf{w}_i, \mathbf{v}_j, s = 1) = P(a_i \leq a'_{ij}|\mathbf{w}_i, \mathbf{v}_j) \\
&P(2|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0) = P(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j) \\
&P(4|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) = P(a^*_{ij} > a_i > a_{ij}|\mathbf{w}_i, \mathbf{v}_j) \\
&P(5|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) = P(a_i \geq a^*_{ij}|\mathbf{w}_i, \mathbf{v}_j)
\end{aligned}
$$

9.1.2 definitions and results related to teacher ability across cases:

Assumption 1 implies the following relations hold regarding the conditional expectation of teacher ability:

$$
\begin{aligned}
E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, s_{ij}=1\right] &= E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, s_{ij}=0\right] = E\left[a_i|\mathbf{w}_i, \mathbf{v}_j\right] \\
E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case1\right] &= E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a_i \leq a'_{ij}, s_{ij}=0\right] = E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a_i \leq a'_{ij}\right] \\
E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case2\right] &= E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a_i > a'_{ij}, s_{ij}=0\right] = E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a_i > a'_{ij}\right] \\
E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case3\right] &= E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a_i \leq a'_{ij}, s_{ij}=1\right] = E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a_i \leq a'_{ij}\right] \\
E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case4\right] &= E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a^*_{ij} > a_i > a'_{ij}, s_{ij}=1\right] \\
&= E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a^*_{ij} > a_i > a'_{ij}\right] \\
E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case5\right] &= E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a_i \geq a^*_{ij}, s_{ij}=1\right] = E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, a_i \geq a^*_{ij}\right]
\end{aligned}
$$

To minimize notation, denote the following conditional expectations:

$$
\begin{aligned}
H &\equiv E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case1\right] = E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case3\right] \\
I &\equiv E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case2\right] \\
J &\equiv E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case4\right] \\
K &\equiv E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, case5\right] \\
K &\geq I \geq J > H \qquad (9.5)
\end{aligned}
$$

Since by assumption 1, s_{ij} and a_i are independent, we know:

$E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, s_{ij}=1\right] = E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, s_{ij}=0\right]$

by the law of iterated expectations, $E\left[a_i|\mathbf{w}_i, \mathbf{v}_j, s_{ij}=1\right]$ is equal to:

$E\left[a_i|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=1\right]=E\left[E\left[a_i|\mathbf{w}_i,\mathbf{v}_j,case\right]\mathbf{w}_i,\mathbf{v}_j,s_{ij}=1\right]$

$=HP(3|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=1)+JP(4|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=1)+KP(5|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=1)$

by the law of iterated expectations, $E\left[a_i|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=0\right]$ is equal to:

$E\left[a_i|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=0\right]=E\left[E\left[a_i|\mathbf{w}_i,\mathbf{v}_j,case\right]\mathbf{w}_i,\mathbf{v}_j,s_{ij}=0\right]$

$=HP(1|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=0)+IP(2|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=0)$

Given this result, we have:

$HP(3|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=1)+JP(4|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=1)+KP(5|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=1)$

$=HP(1|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=0)+IP(2|\mathbf{w}_i,\mathbf{v}_j,s_{ij}=0)$

$\Rightarrow JP(a^*_{ij}>a_i>a_{ij}|\mathbf{w}_i,\mathbf{v}_j)+KP(a_i\geq a^*_{ij}|\mathbf{w}_i,\mathbf{v}_j)=IP(a_i>a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)$

multiplying through by $P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)$ gives:

$IP(a_i>a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)=$

$JP(a^*_{ij}>a_i>a_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)+KP(a_i\geq a^*_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)$

substituting the results from section 9.1.1 gives:

$\Rightarrow IP(a_i>a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)=JP(4|\mathbf{w}_i,\mathbf{v}_j)+KP(5|\mathbf{w}_i,\mathbf{v}_j)$

multiply through by α gives:

$\Rightarrow IP(a_i>a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=0|\mathbf{w}_i,\mathbf{v}_j)=\alpha JP(4|\mathbf{w}_i,\mathbf{v}_j)+\alpha KP(5|\mathbf{w}_i,\mathbf{v}_j)$

$\Rightarrow IP(2|\mathbf{w}_i,\mathbf{v}_j)=\alpha JP(4|\mathbf{w}_i,\mathbf{v}_j)+\alpha KP(5|\mathbf{w}_i,\mathbf{v}_j)$

Given these results, we have the following relationships among the conditional ability levels in cases 2, 4, and 5:

$$IP(2|\mathbf{w}_i,\mathbf{v}_j)=\alpha JP(4|\mathbf{w}_i,\mathbf{v}_j)+\alpha KP(5|\mathbf{w}_i,\mathbf{v}_j) \tag{9.6}$$

9.1.3 definitions and results related to teacher quality across cases:

In this section, I define notation and key relationships regarding expected teacher quality in performance pay district j for teachers in each case. First, recall that in cases 1 and 3, teachers choose low effort such that their teacher quality in performance pay district j is equal to the minimal quality standard $\underline{S}$. In cases 2, 4, and 5 teachers choose high effort in performance pay district j. Therefore, in these cases teacher quality is higher than the minimal standard and varies by case. Define the following notation regarding the expected teacher quality in performance pay district j in cases 2, 4, and 5:

$$T \equiv E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, case2\right]$$
$$V = E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, case4\right]$$
$$W = E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, case5\right]$$
$$T, V, W > \underline{S}$$

Given assumption 1, we can say more about the relative sizes of these conditional expectations. Specifically, we know that since $S_{ij}^* = Q(e_{ij}^*, a_i)$ is strictly increasing in teacher ability, then the following result holds:

$$W \geq T \tag{9.7}$$

Given assumption 1, the following relationship holds:

$E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right] = E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right] = E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j\right]$

By the law of iterated expectations, $E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right]$ is equal to:

$E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right] = E\left[E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, case\right] \mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right]$

$= VP(4|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) + WP(5|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1)$

by the law of iterated expectations, $E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right]$ is equal to:

$E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right] = E\left[E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, case\right] \mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right]$

$= TP(2|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0)$

Setting $E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right] = E\left[S_{ij}^*|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right]$ gives:

$VP(4|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) + WP(5|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) = TP(2|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0)$

substituting the results from section 9.1.1 gives:

$VP(a^*_{ij} > a_i > a_{ij}|\mathbf{w}_i, \mathbf{v}_j) + WP(a_i \geq a^*_{ij}|\mathbf{w}_i, \mathbf{v}_j) = TP(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)$

multiplying through by $P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)$ gives:

$VP(a^*_{ij} > a_i > a_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j) + WP(a_i \geq a^*_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)$

$= TP(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)$

substituting $P(4|\mathbf{w}_i, \mathbf{v}_j) = P(a^*_{ij} > a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)$

and $P(5|\mathbf{w}_i, \mathbf{v}_j) = P(a_i \geq a^*_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)$ gives:

$VP(4|\mathbf{w}_i, \mathbf{v}_j) + WP(5|\mathbf{w}_i, \mathbf{v}_j) = TP(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 1|\mathbf{w}_i, \mathbf{v}_j)$

multiplying through by $\alpha = \frac{P(s_{ij}=0|\mathbf{w}_i,\mathbf{v}_j)}{P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)}$ gives:

$\alpha VP(4|\mathbf{w}_i, \mathbf{v}_j) + \alpha WP(5|\mathbf{w}_i, \mathbf{v}_j) = TP(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j)$

substituting $P(2|\mathbf{w}_i, \mathbf{v}_j) = P(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)P(s_{ij} = 0|\mathbf{w}_i, \mathbf{v}_j)$ gives the following relationship among the expected teacher quality in district j for teachers in cases 2, 4, and 5:

$$TP(2|\mathbf{w}_i, \mathbf{v}_j) = \alpha VP(4|\mathbf{w}_i, \mathbf{v}_j) + \alpha WP(5|\mathbf{w}_i, \mathbf{v}_j) \tag{9.8}$$

9.1.4 definitions and results related to teacher total pay across cases:

Assumption 1, combined with the assumption that the error in teacher quality measurement is unrelated to any other variables in the model, implies that the distribution of teacher total pay is identical for $s_{ij} = 1$ and $s_{ij} = 0$. That implies the following:

$E\left[Y^p_{ij}|\mathbf{w}_i, \mathbf{v}_j\right] = E\left[Y^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right] = E\left[Y^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right]$

To minimize notation, define the following notation regarding the expected total pay for teacher i in performance pay district j in each of the five cases:

$$A \equiv E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, case1\right] = E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, case3\right]$$

$$B \equiv E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, case2\right]$$

$$D \equiv E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, case4\right]$$

$$F \equiv E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, case5\right]$$

Again, since $S_{ij}^{*} = Q(e_{ij}^{*}, a_i)$ is strictly increasing in teacher ability, then the following result holds:

$$F \geq B \geq D > A \tag{9.9}$$

By the law of iterated expectations, $E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right]$ is equal to:

$E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right] = E\left[E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, case\right] \mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right]$

$= AP(3|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) + DP(4|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) + FP(5|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1)$

by the law of iterated expectations, $E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right]$ is equal to:

$E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right] = E\left[E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, case\right] \mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right]$

$= AP(1|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0) + BP(2|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0)$

Setting $E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1\right] = E\left[Y_{ij}^{p}|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0\right]$ gives:

$AP(3|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) + DP(4|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) + FP(5|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1)$

$= AP(1|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0) + BP(2|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0)$

Since $P(1|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0) = P(3|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) = P(a_i \leq a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)$ we have the following:

$DP(4|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) + FP(5|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) = BP(2|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0)$

substituting $P(2|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 0) = P(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)$,

$P(4|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) = P(a_{ij}^{*} > a_i > a_{ij}|\mathbf{w}_i, \mathbf{v}_j)$,

and $P(5|\mathbf{w}_i, \mathbf{v}_j, s_{ij} = 1) = P(a_i \geq a_{ij}^{*}|\mathbf{w}_i, \mathbf{v}_j)$ we have:

$DP(a_{ij}^{*} > a_i > a_{ij}|\mathbf{w}_i, \mathbf{v}_j) + FP(a_i \geq a_{ij}^{*}|\mathbf{w}_i, \mathbf{v}_j) = BP(a_i > a'_{ij}|\mathbf{w}_i, \mathbf{v}_j)$

multiplying through by $P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)$ gives:

$DP(a^*_{ij} > a_i > a_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j) + FP(a_i \geq a^*_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)$

$= BP(a_i > a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)$

substituting $P(4|\mathbf{w}_i,\mathbf{v}_j) = P(a^*_{ij} > a_i > a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)$

and $P(5|\mathbf{w}_i,\mathbf{v}_j) = P(a_i \geq a^*_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)$ gives:

$DP(4|\mathbf{w}_i,\mathbf{v}_j) + FP(5|\mathbf{w}_i,\mathbf{v}_j) = BP(a_i > a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)$

multiplying through by $\alpha = \frac{P(s_{ij}=0|\mathbf{w}_i,\mathbf{v}_j)}{P(s_{ij}=1|\mathbf{w}_i,\mathbf{v}_j)}$ gives:

$\alpha DP(4|\mathbf{w}_i,\mathbf{v}_j) + \alpha FP(5|\mathbf{w}_i,\mathbf{v}_j) = BP(a_i > a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=0|\mathbf{w}_i,\mathbf{v}_j)$

substituting $P(2|\mathbf{w}_i,\mathbf{v}_j) = P(a_i > a'_{ij}|\mathbf{w}_i,\mathbf{v}_j)P(s_{ij}=0|\mathbf{w}_i,\mathbf{v}_j)$ gives the following relationship among the expected total pay for teachers in cases 2, 4, and 5:

$$BP(2|\mathbf{w}_i,\mathbf{v}_j) = \alpha DP(4|\mathbf{w}_i,\mathbf{v}_j) + \alpha FP(5|\mathbf{w}_i,\mathbf{v}_j) \tag{9.10}$$

9.1.5 proof of proposition 1

Throughout each proof I denote $\mathbf{w}_i,\mathbf{v}_j \equiv \mathbf{x}_{ij}$ in order to minimize notation:

Proposition 1 states the following:

$E\left[Q^p_{ij}|\mathbf{w}_i,\mathbf{v}_j,d_{ij}=1\right] > E\left[Q^p_{ij}|\mathbf{w}_i,\mathbf{v}_j,d_{ij}=0\right] \iff$

$E\left[a_i|\mathbf{w}_i,\mathbf{v}_j,d_{ij}=1\right] > E\left[a_i|\mathbf{w}_i,\mathbf{v}_j,d_{ij}=0\right]$

Part I:

$E\left[Q^p_{ij}|\mathbf{x}_{ij},d_{ij}=1\right] > E\left[Q^p_{ij}|\mathbf{x}_{ij},d_{ij}=0\right] \iff$

$E\left[E\left[Q^p_{ij}|\mathbf{x}_{ij},case\right]|\mathbf{x}_{ij},d_{ij}=1\right] > E\left[E\left[Q^p_{ij}|\mathbf{x}_{ij},case\right]|\mathbf{x}_{ij},d_{ij}=0\right]$

$\iff \underline{S}P(1|\mathbf{x}_{ij},d_{ij}=1) + TP(2|\mathbf{x}_{ij},d_{ij}=1) + WP(5|\mathbf{x}_{ij},d_{ij}=1)$

$> \underline{S}P(3|\mathbf{x}_{ij},d_{ij}=0) + VP(4|\mathbf{x}_{ij},d_{ij}=0)$

$\iff \underline{S}\frac{P(1|\mathbf{x}_{ij})}{P(d_{ij}=1|\mathbf{x}_{ij})} + T\frac{P(2|\mathbf{x}_{ij})}{P(d_{ij}=1|\mathbf{x}_{ij})} + W\frac{P(5|\mathbf{x}_{ij})}{P(d_{ij}=1|\mathbf{x}_{ij})} > \underline{S}\frac{P(3|\mathbf{x}_{ij})}{P(d_{ij}=0|\mathbf{x}_{ij})} + V\frac{P(4|\mathbf{x}_{ij})}{P(d_{ij}=0|\mathbf{x}_{ij})}$

$\iff \underline{S}P(1|\mathbf{x}_{ij})P(d_{ij}=0|\mathbf{x}_{ij}) + TP(2|\mathbf{x}_{ij})P(d_{ij}=0|\mathbf{x}_{ij}) +$

$WP(5|\mathbf{x}_{ij})P(d_{ij}=0|\mathbf{x}_{ij}) - \underline{S}P(3|\mathbf{x}_{ij})P(d_{ij}=1|\mathbf{x}_{ij}) - VP(4|\mathbf{x}_{ij})P(d_{ij}=1|\mathbf{x}_{ij}) > 0$

substituting $P(d_{ij}=0|\mathbf{x}_{ij}) = \frac{1}{\alpha+1} - P(5|\mathbf{x}_{ij})$ and $P(d_{ij}=1|\mathbf{x}_{ij}) = \frac{\alpha}{\alpha+1} + P(5|\mathbf{x}_{ij})$ (equation 9.4 and equation 9.3):

$$\Longleftrightarrow \underline{S}P(1|\mathbf{x}_{ij})\left[\tfrac{1}{\alpha+1} - P(5|\mathbf{x}_{ij})\right] + TP(2|\mathbf{x}_{ij})\left[\tfrac{1}{\alpha+1} - P(5|\mathbf{x}_{ij})\right] +$$

$$WP(5|\mathbf{x}_{ij})\left[\tfrac{1}{\alpha+1} - P(5|\mathbf{x}_{ij})\right] - \underline{S}P(3|\mathbf{x}_{ij})\left[\tfrac{\alpha}{\alpha+1} + P(5|\mathbf{x}_{ij})\right]$$

$$-VP(4|\mathbf{x}_{ij})\left[\tfrac{\alpha}{\alpha+1} + P(5|\mathbf{x}_{ij})\right] > 0$$

$$\Longleftrightarrow \tfrac{1}{\alpha+1}\underline{S}P(1|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})\underline{S}P(1|\mathbf{x}_{ij}) + \tfrac{1}{\alpha+1}TP(2|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})TP(2|\mathbf{x}_{ij})$$

$$+\tfrac{1}{\alpha+1}WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})WP(5|\mathbf{x}_{ij}) - \tfrac{\alpha}{\alpha+1}\underline{S}P(3|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})\underline{S}P(3|\mathbf{x}_{ij})$$

$$-\tfrac{\alpha}{\alpha+1}VP(4|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})VP(4|\mathbf{x}_{ij}) > 0$$

substituting $P(1|\mathbf{x}_{ij}) = \alpha P(3|\mathbf{x}_{ij})$ (equation 9.1):

$$\Longleftrightarrow \tfrac{1}{\alpha+1}TP(2|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})TP(2|\mathbf{x}_{ij}) + \tfrac{1}{\alpha+1}WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})WP(5|\mathbf{x}_{ij})$$

$$-\left[1+\alpha\right]P(5|\mathbf{x}_{ij})\underline{S}P(3|\mathbf{x}_{ij}) - \tfrac{\alpha}{\alpha+1}VP(4|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})VP(4|\mathbf{x}_{ij}) > 0$$

substituting $P(3|\mathbf{x}_{ij}) = \frac{1}{\alpha+1} - \frac{P(2|\mathbf{x}_{ij})}{\alpha}$ (equations 9.1 and 9.3):

$$\Longleftrightarrow \tfrac{1}{\alpha+1}TP(2|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})TP(2|\mathbf{x}_{ij}) + \tfrac{1}{\alpha+1}WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})WP(5|\mathbf{x}_{ij})$$

$$-\left[1+\alpha\right]P(5|\mathbf{x}_{ij})\underline{S}\left[\tfrac{1}{\alpha+1} - \tfrac{P(2|\mathbf{x}_{ij})}{\alpha}\right] - \tfrac{\alpha}{\alpha+1}VP(4|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})VP(4|\mathbf{x}_{ij}) > 0$$

$$\Longleftrightarrow \tfrac{1}{\alpha+1}TP(2|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})TP(2|\mathbf{x}_{ij}) + \tfrac{1}{\alpha+1}WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})WP(5|\mathbf{x}_{ij})$$

$$-P(5|\mathbf{x}_{ij})\underline{S} + \left[1+\alpha\right]P(5|\mathbf{x}_{ij})\underline{S}\tfrac{P(2|\mathbf{x}_{ij})}{\alpha} - \tfrac{\alpha}{\alpha+1}VP(4|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})VP(4|\mathbf{x}_{ij}) > 0$$

substituting $TP(2|\mathbf{x}_{ij}) = \alpha VP(4|\mathbf{x}_{ij}) + \alpha WP(5|\mathbf{x}_{ij})$ (equation 9.8) we have:

$$\Longleftrightarrow \tfrac{1}{\alpha+1}\left[V\alpha P(4|\mathbf{x}_{ij}) + W\alpha P(5|\mathbf{x}_{ij})\right] - P(5|\mathbf{x}_{ij})\left[V\alpha P(4|\mathbf{x}_{ij}) + W\alpha P(5|\mathbf{x}_{ij})\right]$$

$$+\tfrac{1}{\alpha+1}WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})\underline{S} + \left[1+\alpha\right]P(5|\mathbf{x}_{ij})\underline{S}\tfrac{P(2|\mathbf{x}_{ij})}{\alpha}$$

$$-\tfrac{\alpha}{\alpha+1}VP(4|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})VP(4|\mathbf{x}_{ij}) > 0$$

$$\Longleftrightarrow \tfrac{\alpha}{\alpha+1}VP(4|\mathbf{x}_{ij}) + \tfrac{\alpha}{\alpha+1}WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})V\alpha P(4|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})W\alpha P(5|\mathbf{x}_{ij})$$

$$+\tfrac{1}{\alpha+1}WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})\underline{S} + \left[1+\alpha\right]P(5|\mathbf{x}_{ij})\underline{S}\tfrac{P(2|\mathbf{x}_{ij})}{\alpha}$$

$$-\tfrac{\alpha}{\alpha+1}VP(4|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})VP(4|\mathbf{x}_{ij}) > 0$$

$$\Longleftrightarrow WP(5|\mathbf{x}_{ij}) - \left[1+\alpha\right]P(5|\mathbf{x}_{ij})VP(4|\mathbf{x}_{ij}) - \left[1+\alpha\right]P(5|\mathbf{x}_{ij})WP(5|\mathbf{x}_{ij}) - P(5|\mathbf{x}_{ij})\underline{S}$$

$$+\left[1+\alpha\right]\tfrac{P(2|\mathbf{x}_{ij})}{\alpha}P(5|\mathbf{x}_{ij})\underline{S} > 0$$

$$\Longleftrightarrow P(5|\mathbf{x}_{ij})\left[W - \left[1+\alpha\right]VP(4|\mathbf{x}_{ij}) - \left[1+\alpha\right]P(5|\mathbf{x}_{ij})W - \underline{S} + \left[1+\alpha\right]\tfrac{P(2|\mathbf{x}_{ij})}{\alpha}\underline{S}\right]$$

$$> 0$$

Substituting $VP(4|\mathbf{x}_{ij}) = T\frac{P(2|\mathbf{x}_{ij})}{\alpha} - WP(5|\mathbf{x}_{ij})$ (equation 9.8):

$\iff P(5|\mathbf{x}_{ij})[W - [1+\alpha]\left[T\frac{P(2|\mathbf{x}_{ij})}{\alpha} - WP(5|\mathbf{x}_{ij})\right]$

$- [1+\alpha] P(5|\mathbf{x}_{ij})W - \underline{S} + [1+\alpha] \frac{P(2|\mathbf{x}_{ij})}{\alpha}\underline{S}] > 0$

$\iff P(5|\mathbf{x}_{ij})[W - [1+\alpha] T\frac{P(2|\mathbf{x}_{ij})}{\alpha} + [1+\alpha] WP(5|\mathbf{x}_{ij}) -$

$[1+\alpha] P(5|\mathbf{x}_{ij})W - \underline{S} + [1+\alpha] \frac{P(2|\mathbf{x}_{ij})}{\alpha}\underline{S}] > 0$

$\iff P(5|\mathbf{x}_{ij})\left[W - [1+\alpha] T\frac{P(2|\mathbf{x}_{ij})}{\alpha} - \underline{S} + [1+\alpha] \frac{P(2|\mathbf{x}_{ij})}{\alpha}\underline{S}\right] > 0$

$\iff P(5|\mathbf{x}_{ij})\left[W - \underline{S} - [1+\alpha] \frac{P(2|\mathbf{x}_{ij})}{\alpha}[T - \underline{S}]\right] > 0$

dividing through by $T - \underline{S}$, which is strictly greater than zero:

$\iff P(5|\mathbf{x}_{ij})\left[\frac{W-\underline{S}}{T-\underline{S}} - [1+\alpha] \frac{P(2|\mathbf{x}_{ij})}{\alpha}\right] > 0$

$\iff P(5|\mathbf{x}_{ij})\left[\frac{W-\underline{S}}{T-\underline{S}} - \left[\frac{P(2|\mathbf{x}_{ij})}{\alpha} + P(2|\mathbf{x}_{ij})\right]\right] > 0$

substituting $\frac{P(2|\mathbf{x}_{ij})}{\alpha} = P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij})$ (equation 9.1) we have:

$\iff P(5|\mathbf{x}_{ij})\left[\frac{W-\underline{S}}{T-\underline{S}} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$

From above, we know that $W \geq T \geq \underline{S}$ (assumption 9.7) which implies that $\frac{W-\underline{S}}{T-\underline{S}} \geq 1$. Also, we know that the sum of probabilities $P(4|\mathbf{x}_{ij}),\ P(5|\mathbf{x}_{ij}),$ and $P(2|\mathbf{x}_{ij})$ is less than one if at least one teacher is in either case 1 or case 3. Given this assumption, we know that $\left[\frac{W-\underline{S}}{T-\underline{S}} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$. Hence, $P(5|\mathbf{x}_{ij})\left[\frac{W-\underline{S}}{T-\underline{S}} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$ if and only if $P(5|\mathbf{x}_{ij}) > 0$.

Therefore, we have shown that $E\left[Q^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1\right] > E\left[Q^p_{ij}|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=0\right]$ if and only if $P(5|\mathbf{w}_i, \mathbf{v}_j) > 0$.

Part II:

$E[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1] > E[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=0] \iff$

$E[E[a_i|\mathbf{w}_i, \mathbf{v}_j, case]|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1] > E[E[a_i|\mathbf{w}_i, \mathbf{v}_j, case]|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=0]$

$\iff HP(1|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1) + IP(2|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1) + KP(5|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1)$

$> HP(3|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=0) + JP(4|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=0)$

Following the same logic used in part one, and substituting the results in equations 9.5 and 9.6 we have the following:

$E[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=1] > E[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij}=0] \iff$

$P(5|\mathbf{x}_{ij})\left[\frac{K-H}{I-H} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$

From above, we know that $K \geq I \geq H$ (assumption 9.5) which implies that $\frac{K-H}{I-H} \geq 1$. Also, we know that the sum of probabilities $P(4|\mathbf{x}_{ij})$, $P(5|\mathbf{x}_{ij})$, and $P(2|\mathbf{x}_{ij})$ is less than one if at least one teacher is in either case 1 or case 3. Given this assumption, we know that $\left[\frac{K-H}{I-H} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$. Hence, $P(5|\mathbf{x}_{ij})\left[\frac{K-H}{I-H} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$ if and only if $P(5|\mathbf{x}_{ij}) > 0$.

Therefore, we have shown that $P(5|\mathbf{w}_i, \mathbf{v}_j) > 0 \iff E[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1] > E[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0]$ and by the proof in part one, we have $E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0\right] \iff P(5|\mathbf{w}_i, \mathbf{v}_j) > 0$. The link is therefore complete:

$E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0\right] \iff$

$E[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1] > E[a_i|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0]$.

9.1.6 proof of proposition 2

Proposition 2 states the following: $E\left[Y_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{w}_i, \mathbf{v}_j\right] \Rightarrow$

$E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0\right]$

$E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{x}_{ij}\right] \iff E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] > E\left[E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij}\right]|\mathbf{x}_{ij}\right]$

$= E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] P(d_{ij} = 1|\mathbf{x}_{ij}) + E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 0\right] P(d_{ij} = 0|\mathbf{x}_{ij})$

so we have: $E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{x}_{ij}\right] \iff E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right]$

$> E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] P(d_{ij} = 1|\mathbf{x}_{ij}) + E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 0\right] P(d_{ij} = 0|\mathbf{x}_{ij})$

rearranging we have:

$E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{x}_{ij}\right] \iff E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 0\right]$

Using the law or iterated expectation we have $E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{x}_{ij}\right]$ iff:

$\iff AP(1|\mathbf{x}_{ij}, d_{ij} = 1) + BP(2|\mathbf{x}_{ij}, d_{ij} = 1) + FP(5|\mathbf{x}_{ij}, d_{ij} = 1)$

$> AP(3|\mathbf{x}_{ij}, d_{ij} = 0) + DP(4|\mathbf{x}_{ij}, d_{ij} = 0)$

Following the same logic used in proposition 1, part one, and substituting the results in equations 9.9 and 9.10 we have the following:

$E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{x}_{ij}\right] \iff$

$P(5|\mathbf{x}_{ij})\left[\frac{F-A}{B-A} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$

From above, we know that $F \geq B \geq A$ (assumption 9.9) which implies that $\frac{F-A}{B-A} \geq 1$. Also, we know that the sum of probabilities $P(4|\mathbf{x}_{ij})$, $P(5|\mathbf{x}_{ij})$, and $P(2|\mathbf{x}_{ij})$ is less than one if at least one teacher is in either case 1 or case 3. Given this assumption, we know that $\left[\frac{F-A}{B-A} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$. Hence, $P(5|\mathbf{x}_{ij})\left[\frac{F-A}{B-A} - [P(4|\mathbf{x}_{ij}) + P(5|\mathbf{x}_{ij}) + P(2|\mathbf{x}_{ij})]\right] > 0$ if and only if $P(5|\mathbf{x}_{ij}) > 0$.

Therefore, we have shown that $E\left[Y_{ij}^p|\mathbf{x}_{ij}, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{x}_{ij}\right] \iff P(5|\mathbf{w}_i, \mathbf{v}_j) > 0$ and by the proof in proposition 1, part one, we have

$P(5|\mathbf{w}_i, \mathbf{v}_j) > 0 \iff E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0\right]$.

The logic is therefore the following:

$E\left[Y_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[Y_{ij}^p|\mathbf{w}_i, \mathbf{v}_j\right] \iff$

$E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 1\right] > E\left[Q_{ij}^p|\mathbf{w}_i, \mathbf{v}_j, d_{ij} = 0\right]$