Incentives Work: Getting Teachers to Come to School*

Esther Duflo, Rema Hanna, and Stephen P. Ryan†

May 30, 2010

Abstract

We use a randomized experiment and a structural model to test whether monitoring and financial incentives can reduce teacher absence and increase learning in rural India. In treatment schools, teachers' attendance was monitored daily using cameras, and their salaries were made a nonlinear function of attendance. Absenteeism by teachers fell by 21 percentage points relative to the control group, and children's test scores increased by 0.17 standard deviations. We estimate a structural dynamic labor supply model and find that teachers responded strongly to the financial incentives, and that this alone can explain the difference between the two groups. Our model is used to compute cost-minimizing compensation policies.

*This project is a collaborative exercise involving many people. Foremost, we are deeply indebted to Seva Mandir, and especially to Neelima Khetan and Priyanka Singh, who made this evaluation possible. We thank Ritwik Sakar and Ashwin Vasan for their excellent work coordinating the fieldwork. Greg Fischer, Shehla Imran, Callie Scott, Konrad Menzel, and Kudzaishe Takavarasha provided superb research assistance. For their helpful comments, we thank referees, Abhijit Banerjee, Rachel Glennerster, Michael Kremer and Sendhil Mullainathan. We owe a special thank to the referees, who made substantial suggestions that considerably improved the paper. For financial support, we thank the John D. and Catherine T. MacArthur Foundation.

†The authors are from MIT (Department of Economics and J-PAL), the Kennedy School of Government of Harvard University and J-PAL, and MIT (Department of economics)

1 Introduction

Over the past decade, many developing countries have expanded primary school access. However, these improvements have not been accompanied by improvements in school quality. For example, in India, a nationwide survey found that 65 percent of children enrolled in grades 2 through 5 in government primary schools could not read a simple paragraph (Pratham, 2006). These poor learning outcomes may be due, in part, to teacher absenteeism. Using unannounced visits to measure attendance, a nationally representative survey found that 24 percent of teachers in India were absent during normal school hours (Chaudhury, et al., 2005).[1] Improving attendance rates may be the first step needed to make "universal primary education" a meaningful term.

Solving the absenteeism problem poses a significant challenge (see Banerjee and Duflo (2006) for a review). In many countries, teachers are a powerful political force, able to resist attempt to impose stricter attendance on them. As such, many governments have begun to shift from hiring government teachers to instead hiring "para-teachers." Para-teachers are teachers that are hired on short, flexible contracts to work in primary schools and in non-formal education centers (NFEs) run by non governmental organizations (NGOs) and local governments. Unlike government teachers, it may be feasible to implement greater oversight and incentives for para-teachers since they do not form a entrenched constituency, they are already subject to yearly renewal of their contract, and there is a long queue of qualified job applicants. Thus, providing para-teachers with incentives may be an effective way to improve the quality of education, provided that para-teachers can teach effectively.

In this paper, we use both experimental and structural methods to empirically test whether the direct monitoring of the attendance of para-teachers (referred to simply as teachers in the rest of the paper), coupled with high-powered financial incentives based on their attendance, improves teacher attendance and school quality.

The effect of incentives based on presence is theoretically ambiguous. While simple labor supply models would predict that incentives should increase effort, the incentives could fail for a variety of reasons. First, the incentives may not be high enough. Second, the incentive schemes may crowd out the teacher's intrinsic motivation to attend school (Benabou and Tirole, 2006). Finally, some teachers, who previously believed that they were required to work every day, may decide to stop working once they have reached their target income for the month (Fehr and Goette, 2002).

[1] Although teachers do have some official non-teaching duties, this absence rate is too high to be fully explained by this particular story.

Even if incentives increase teacher attendance, it is unclear whether child learning levels will actually increase. Teachers may multitask (Holmstrom and Milgrom, 1991), reducing their efforts along other dimensions.[2] Such schemes may also demoralize teachers, resulting in less effort (Fehr and Schmidt, 2004), or may harm teachers' intrinsic motivation to teach (Kreps, 1997). On the other hand, incentives can improve learning levels if the main cost of working is the opportunity cost of attending school and, once in school, the marginal cost of teaching is low. In this case, an incentive system that directly rewards presence would stand a good chance of increasing child learning. Thus, whether or not the incentives can improve school quality is ultimately an empirical question.

To address these questions, we study a teacher incentive program run by the NGO Seva Mandir. Seva Mandir runs single-teacher NFEs in the rural villages of Rajasthan, India. Teacher absenteeism is high, despite the fact that Seva Mandir tries to combat it, berating frequently-absent teachers, and threatening dismissal for repeated absences. In our baseline study, evaluated in August 2003, the absence rate was 44 percent. In September of 2003, Seva Mandir gave teachers in 57 randomly selected program schools a camera, along with instructions to have one of the students take a picture of the teacher and the other students at the start and close of each school day. The cameras had tamper-proof date and time functions, allowing for the collection of precise data on teacher attendance that could be used to calculate teachers' salaries. Each teacher was then paid according to a nonlinear function of the number of valid school days for which they were actually present, where a "valid" day was defined as one for which the opening and closing photographs were separated by at least five hours and both photographs showed at least eight children. Specifically, they received Rs 500 if they attended fewer than 10 days in a given month, and Rs 50 for any additional day (up to a maximum of 25 or 26 days depending on the month). In the 56 comparison schools, teachers were paid a fixed rate for the month (Rs 1000) and were reminded (as usual) that regular presence was a requirement of their job, and they could in principle be dismissed for repeated, unexcused absences.

The program resulted in an immediate and long-lasting improvement in teacher attendance rates in treatment schools, as measured through monthly unannounced visits in both treatment and comparison schools. Over the 30 months in which attendance was tracked, teachers at program schools had an absence rate of 21 percent, compared to 44 percent

[2] This is a legitimate concern as other incentive programs (based on test scores) have been subject to multitasking (Glewwe, Ilias and Kremer, 2003), manipulation (e.g., Figlio and Winicki, 2002; Figlio and Getzler, 2002) or outright cheating (Jacob and Levitt, 2003). On the other hand, Lavy (2003) and Mulharidharan and Sundaraman (2009) find very positive effects of similar programs.

baseline and the 42 percent in the comparison schools.

While the reduced form results inform us that this program was effective in reducing absenteeism, it does not tell us what the effect of another scheme with a different payment structure would be. Moreover, it does not allow us to identify the response to the financial incentive separately from a possible independent effect of collecting daily data on absence.[3] To answer these questions, we estimate a structural dynamic model of teacher labor supply using the daily attendance data in the treatment schools. Our estimation strategy leverages the fact that the financial incentive for a teacher to attend school on a given day changes with the number of days previously worked in the month and the number of days left in the month. This is because teachers have to attend at least 10 days in a month before they begin to receive the incentive, and the implied shadow value of working changes as the teacher builds up the option to work for Rs 50 per day at the end of the month. Indeed, regression discontinuity design estimates illustrate that teachers who were "out of the money" at the end of the previous month work significantly more at the beginning of the next month, while the opposite is true for those who were "in the money."

In order to understand the effect of the financial and monitoring incentives on teacher attendance, we estimate two complementary structural models of the teachers' labor supply functions. The two models are conceptually similar in that they both model the dynamic decision process facing teachers as they accumulate days worked towards the bonus at the end of the month. Both approaches allow for unobserved heterogeneity at the teacher level, and differ in their treatment of serial dependence in opportunity cost of working. In the first set of models, the opportunity cost is allowed to depend on whether the teacher attended work on the previous day. The second set of models posits that the opportunity cost to working is subject to an autocorrelated shock that follows an AR(1) process. The two models deliver very similar results. A nice institutional feature of the experiment is that the incentives shift discontinuously with the change in month, which is the source of the identification of the responsiveness to the bonus. As a robustness check on our results, we combine the spirit of the regression-discontinuity approach with the structural model by estimating a model with a three-day sample window around the change in month. Our results are very close to those found using the whole sample.

To our knowledge, this is one of the few papers to estimate dynamic labor supply decisions with unobserved heterogeneity and a serially correlated error structure.[4] Stinebrickner (2000)

[3] Another possibility is that the existence of this scheme discourages teachers in the *control* group. This would lead us to overestimate the impact of the program.

[4] See Aguirregabiria and Mira (2010), Todd, Keane and Wolpin (2010), and Todd and Wolpin (2010) for

discusses some of the econometric issues associated with this problem. Three related papers are Bound, Stinebrickner, and Waidman (2005), Sullivan (2009), and Stinebrickner (2001).[5]

We find that teachers are responsive to the financial incentives: our estimates suggest that the elasticity of labor supply with respect to the level of the financial bonus is between 0.20 and 0.30 in our preferred specifications. In most specifications, we do not use the data from the control group in our structural model. This allows us to use the data to test whether the model accurately predicts teacher presence in the control group. Models that include both serial correlation and teacher heterogeneity do well in these out of sample tests: when we set the incentive to zero, it closely predicts the difference in attendance in the treatment and the control group, as well as the number of days worked under a new incentive system initiated by Seva Mandir after the experiment.

The idea of hold out samples for validation has been used in several papers, staring with at least McFadden (1977) (see Keane, Todd and Wolpin (2010) for a recent review of discrete choice dynamic model in labor economics). A smaller number of papers use randomized control experiments to validate a structural model. Wise (1985) estimates a model of housing demand on a control group data, and validates the model using the forecast of the effect of a housing subsidy. More recently, Todd and Wolpin (2006) used data from the PROGRESA program, a conditional cash transfer program where transfers are in part conditioned on school attendance. Using only the control villages, they estimated a structural model of fertility, school participation and child labor. The model was validated to compare the predicted effect of PROGRESA to the experimental estimates of program effects. Lise, Seitz and Smith (2004) use data from the Self Sufficiency Program in Cana to validate a search model of the labor market. Our paper differs from all of these in that we use the data from the *treatment* group to evaluate the model, and forecast what would happen to the control group under the model in order to validate it, as well what would happen with a change in rules. This is the approach in Keane and Moffitt (1998), which estimates a model of labor supply and welfare program participation after a change in program rules, and then uses the model to predict behavior prior to that policy change. Other papers which combine structural methods and experimental data (without using the control group for out of sample validation) include Attanasio, Meghir, and Santiago (2006), Card and Hyslop

recent surveys of the estimation of dynamic choice structural models).

[5]Interestingly, Stinebrickner (2001) also estimate a dynamic model of teacher labor supply, although the dynamic problem they solve in this case is whether or not teacher work in a particular year as a function of the wage that year. Other papers who study dynamic labor supply decision of teachers in this context include Wilbert van der Klaauw (2005) and Stinebrickner (2001).

(2008) and Ferrall (2008).

An advantage of the structural model is the parameters can be used to estimate the effects of other possible rules (see Todd and Wolpin (2010) for different applications of this method to development policy). We use the parameters of the model to compute the optimal incentive scheme for a given number of days worked on average in a month. We calculate that Seva Mandir could achieve the same number of days worked (17) by increasing both the bonus cutoff to 21 days and the bonus to 75 Rupees per day while saving 193 Rupees per teacher per month, an average cost savings of 22 percent.

Although we find that teachers are sensitive to the financial incentives, we see no evidence of multitasking. When the school was open, teachers were as likely to be teaching in treatment as in comparison schools, suggesting that the marginal costs of teaching are low conditional on attendance. Student attendance when the school was open was similar in both groups, so the students in the treatment group received more days of instruction. A year into the program, test scores in the treatment schools were 0.17 standard deviations higher than in the comparison schools. Two and a half years into the program, children from the treatment schools were also 10 percentage points (or 62 percent) more likely to transfer to formal primary schools, which requires passing a competency test. The program's impact and cost are similar to other successful education programs.

The paper is organized as follows. Section 2 describes the program and evaluation strategy. The results on teacher attendance are presented in Section 3, while the estimates from the dynamic labor supply model are presented in Section 4. Section 5 presents the results on other dimensions of teacher effort, as well as student outcomes. Section 6 concludes.

2 Experimental Design and Data Collection

2.1 Non-formal Education Centers

Since the enactment of the National Policy on Education in 1986, non-formal education centers (NFEs) have played an important role in India's drive toward universal primary education. They have been the main instrument for expanding school access to children in remote and rural areas. They have also been used to transition children who may otherwise not attend school into a government school. As of 1997, 21 million children were enrolled in NFEs across India (Education for All Forum, 2000). Similar informal schools operate throughout most of the developing world (Bangladesh, Kenya, etc.).

Children of all ages may attend the NFE, though, in our sample, most are between 7 and

10 years of age. Nearly all of the children are illiterate when they enroll. In the setting of our study, the NFEs are open six hours a day and have about 20 students each. All students are taught in one classroom by one teacher, who is recruited from the local community and has, on average, a 10th grade education. Instruction focuses on basic Hindi and math skills. The schools only have one teacher; thus, when the teacher is absent, the school is closed.

2.2 The Incentive Program

Seva Mandir runs about 150 NFEs in the tribal villages of Udaipur, Rajasthan. Udaipur is a sparsely populated, hard to access region. Thus, it is difficult to regularly monitor the NFEs, and absenteeism is high. A 1995 study (Banerjee et al., 2005) found that the absence rate was 40 percent, while our first observation in the schools included in our study (in August 2003, before the program was announced) found that the rate was 44 percent.

Seva Mandir relies on occasional visits to the schools, as well as reports by the local village workers, and uses bi-monthly teacher meetings to berate teachers who are known to be frequently absent. However, they realize that this level of oversight is insufficient. To reduce teacher absenteeism, Seva Mandir implemented an external monitoring and incentive program starting in September 2003. They chose 120 schools to participate, with 60 randomly selected schools serving as the treatment group and the remaining 60 as the comparison group.[6] In the treatment schools, Seva Mandir gave each teacher a camera, along with instructions for one of the students to take a photograph of the teacher and the other students at the start and end of each school day. The cameras had a tamper-proof date and time function that made it possible to precisely track each school's openings and closings.[7] Camera upkeep (replacing batteries, and changing and collecting the film) was conducted monthly at regularly scheduled teacher meetings. If a camera malfunctioned, teachers were instructed to call the program hotline within 48 hours. Someone was then dispatched to replace the camera, and teachers were credited for the missing day. Rolls were collected every two months, and payments were distributed every two months.

At the start of the program, Seva Mandir's monthly base salary for teachers was Rs 1000 ($23 at the real exchange rate, or about $160 at PPP) for at least 20 days of work per

[6]After randomization but prior to the announcement of the program, 7 of these schools closed. The closures were equally distributed among the treatment and controls schools, and were not due to the program. We thus have 57 treatment schools and 56 comparison schools.

[7]The time and date buttons on the cameras were covered with heavy tape, and each had a seal that would indicate if it had been tampered with. Fines would have been imposed if cameras had been tampered with (this did not happen) or if they had been used for another purpose (this happened in one case).

month. In the treatment schools, teachers received a Rs 50 ($1.15) bonus for each additional day they attended in excess of the 20 days (where holidays and training days, or about 3 days per month on average, are automatically credited as working days), and they received a Rs 50 fine for each day of the 20 days they skipped work. Seva Mandir defined a "valid" day as one in which the opening and closing photographs were separated by at least five hours and at least eight children were present in both photos to indicate that the school was actually functioning. Due to ethical and political concerns, Seva Mandir capped the fine at Rs 500. Thus, salaries ranged from Rs 500 to Rs 1,300 (or $11.50 to $29.50). In the 56 comparison schools, teachers were paid the flat rate of Rs 1,000, and were reminded that regular attendance was a compulsory part of their job, and they could be in principle dismissed for poor attendance. However, this happens very rarely, and did not happen during the span of the evaluation.[8]

2.3 Data Collection

An independent evaluation team led by Vidhya Bhawan (a consortium of schools and teacher training institutes) and J-PAL collected the data. We have two sources of attendance data. First, we collected data on teacher attendance through one random unannounced visit per month in all schools. By comparing the absence rates obtained from the random checks across the two types of schools, we can determine the program's effect on absenteeism.[9] Second, Seva Mandir provided us with access to the camera and payment data for the treatment schools.

We collected data on teacher and student activity during the random check. For schools that were open during the visit, the enumerator noted the school activities: how many children were sitting in the classroom, whether anything was written on the blackboard, and whether the teacher was talking to the children. While these are crude measures of teacher performance, they were chosen because each could be easily observed before the teachers could adjust their behavior. In addition, the enumerator also conducted a roll call and noted whether any of the absent children had left school or had enrolled in a government school, and then updated the evaluation roster to include new children.

To determine whether child learning increased as a result of the program, the evaluation

[8]Teachers in the control schools knew that the camera program was occurring, and that some teachers were randomly selected to be part of the pilot program.

[9]Teachers understood that the random checks were not linked with an incentive. We cannot rule out the fact that the random check could have increased attendance in comparison schools. However, we have no reason to believe that they would differentially affect the attendance of comparison and treatment teachers.

team, in collaboration with Seva Mandir, administered three basic competency exams to all children enrolled in the NFEs in August 2003: a pre-test in August 2003, a mid-test in April 2004, and a post-test in September 2004. The pre-test followed Seva Mandir's usual testing protocol. Children were given either a written exam (for those who could write) or an oral exam (for those who could not). For the mid-test and post-test, all children were given both the oral exam and the written exam; those unable to write, of course, earned a zero on the written section. The oral exam tested simple math skills (counting, one-digit addition, simple division) and basic Hindi vocabulary skills, while the written exam tested for these competencies plus more complex math skills (two-digit addition and subtraction, multiplication and division), the ability to construct sentences, and reading comprehension. Thus, the written exam tested both a child's ability to write and his ability to handle material requiring higher levels of competency relative to the oral exam.

2.4 Baseline and Experiment Integrity

Given that schools were randomly allocated to the treatment and comparison groups, we expected school quality to be similar across groups prior to the program onset. Before the program was announced in August 2003, the evaluators were able to randomly visit 41 schools in the treatment group and 39 in the comparison.[10] Panel A of Table 1 shows that the attendance rates were 66 percent and 64 percent, respectively. This difference is not statistically significant. Other measures of school quality were also similar prior to the program: in all dimensions shown in Table 1, the treatment schools appear to be slightly better than comparison schools, but the differences are always small and never significant. Finally, to determine the joint significance of the treatment variable on all of the outcomes listed in Panels B through D, we estimated a Seemingly Unrelated Regression (SUR) model. The F-statistic is 1.21, with a p-value of 0.27, implying that the comparison and treatment schools were similar to one another at the program's inception.

Baseline academic achievement, as measured by the pre-test, was the same for students across the two types of schools (Table 1, Panel E). On average, students in both groups appeared to be at the same level of preparedness before the program. There is no significant difference in either probability to take the written test or scores on the written tests.

[10] Due to time constraints, only 80 randomly selected schools of the 113 were visited prior to the program. There was no significant (or perceivable) difference in the characteristics of the schools that were not observed before the program. Moreover, the conclusion of the paper remains unchanged when we restrict all the subsequent analysis to the 80 schools that could be observed before the program was started.

3 Results: Teacher attendance

3.1 Reduced form results: Teacher Behavior

The effect on teacher absence was both immediate and long lasting. Figure 1 shows the fraction of schools found open on the day of the random visit, by month. Between August and September 2003, teacher attendance increased in treatment schools relative to the comparison schools. Over the next two and a half years, the attendance rates in both types of schools followed similar seasonal fluctuations, with treatment school attendance systematically higher than comparison school attendance.

As Figure 1 shows, the treatment effect remained strong even after the post-test, which marked the end of the formal evaluation. Since the program had been very effective, Seva Mandir maintained it. However, at the end of the study, they only had enough resources to keep the program operating in the treatment schools. The random checks conducted after the post-test showed that the higher attendance rates persisted at treatment schools even after the teachers knew that the program was permanent, suggesting that teachers did not alter their behavior simply for the duration of the evaluation.

Table 2 presents a detailed breakdown of the program effect on absence rates for different time periods. On average, the teacher absence rate was 21 percentage points lower (or about half) in the treatment than in the comparison schools (Panel A).[11] The effects on teacher attendance were pervasive—teacher attendance increased for both low and high quality teachers. Panel B reports the impact for teachers with above median test scores on the teacher skills exam conducted prior to the program, while Panel C shows the impact for teachers with below median scores.[12] The program impact on attendance was larger for below median teachers (a 24 percentage point increase versus a 15 percentage point increase). However, this was due to the fact that the program brought below median teachers to the same level of attendance as above median teachers (78 percent).

The program reduced absence everywhere in the distribution. Figure 2 plots the observed density of absence rates in the treatment and comparison schools for 25 random checks. The figure clearly shows that the program shifted the entire distribution of absence for treatment

[11] This reduction in school closures was comparable to that of a previous Seva Mandir program, which tried to reduce school closures by hiring a second teacher for the NFEs. In that program, school closure only fell by 15 percentage points (Banerjee, Jacob and Kremer, 2005), both because individual teacher absenteeism remained high and because teachers did not coordinate to come on different days.

[12] Teacher test scores and teacher attendance are correlated: in the control group, below median teachers came to school 53 percent of the time, while above median teachers came to school 63 percent of the time.

teachers.[13] Not one of the teachers in the comparison schools was present during all 25 observations. Almost 25 percent of teachers were absent more than half the time. In contrast, 5 of the treatment teachers were present on all days, 47 percent of teachers were present on 21 days or more, and all teachers were present at least half the time. Thus, the program was effective on two margins: it eliminated very delinquent behavior (less than 50 percent attendance) and increased the number of teachers with high attendance records.

A comparison of the random check data and the camera data suggests that, for the most part, teachers did not "game" the system. Out of the 1337 cases where we have both camera data and a random check for a day, 80 percent matched. In 13 percent of the cases, the school was found open during the random check, but the photos indicated that the day was not considered "valid", often because the photos were not separated by five hours. There are 88 cases (7 percent) in which the school was closed and the photos were valid, but only 54 (4 percent of the total) of these were due to teachers being absent in the middle of the day during the random check and shown as present both before and after. In the other cases, the data did not match because the random check was completed after the school had closed for the day, or there were missing data on the time of the random check or photo.

One interesting question is whether the effect of the program would be very different in the long run, because the program would induce different teachers to join schools with cameras. As of October 2009, the program was still in place in the same schools (Seva Mandir has recently extended it to all schools). We monitored the schools for a year, from September 2006 to September 2007. After four years, teacher attendance was still significantly higher in the camera schools (72 percent versus 61 percent). Thus, this program seems to have a very long lasting effect on teacher attendance.

3.2 The Impact of Financial Incentives: Preliminary Evidence

The program had two components: the daily monitoring of teacher attendance (which complements Seva Mandir's usual random checks and reports from village or zonal workers) and an incentive that was linked to attendance. In addition, a monthly random check was performed in both treatment and control schools by the research team. However, we believe that this last check was not perceived by the teacher as being part of Seva Mandir's program: it was performed by a separate team, associated with a different organization, and

[13] We also graphed the estimated cumulative density function of the frequency of attendance, assuming that the distribution of absence follows a beta-binomial distribution (not shown for brevity). The results are similar to that of Figure 2.

the informed consent signed by all teachers clearly disclosed that this data was not going to be shared with Seva Mandir.

Aside from the financial incentive, teachers may respond to the fact that Seva Mandir now obtains daily data on them (what we refer to as "the monitoring effect"), either because of a fear of being fired if the data reveals that they are absent most of the time, or because Seva Mandir may, without firing them, punish them for absence. At the bi-monthly in-service teacher training, teacher absence is discussed, and Seva Mandir workers berate teachers whom they know to be absent a lot. On the other hand, since Seva Mandir continues to inspect treatment and control schools on a random basis, it is also possible that teachers believe that daily data on attendance does not increase the chance that they are punished for absence in expectation: they may believe that one absence found out during a surprise visit would be as costly as several absences. Whether or not there is a direct effect of obtaining daily attendance data (an " increased monitoring effect") or not is thus an open question.

Ideally, to disentangle the effect of the financial incentive from a direct increased monitoring effect, we would have provided different monitoring and incentive schemes in different, randomly selected schools. Some teachers could have been monitored daily, but without receiving incentives. Some could have received a small incentive, while others could have received a larger one. This was not feasible for Seva Mandir. However, the nonlinear nature of the incentive scheme provides us with an opportunity to try to isolate the effect of the financial incentives, if we are willing to assume that the effect of the threat of monitoring does not follow exactly the same time pattern. Consider a teacher who, because he was ill, was unable to attend school on most of the first 20 days of the 26 days of the month. By day 21, assuming he has attended only 5 days so far, he knows that, if he works every single day remaining in the month, he will have worked only 10 days. Thus, he will earn Rs 500, the same amount he would earn if he did not work any other days that month. Although he is still monitored (and may worry that if he does not attend at all in a month he may be punished), his monetary incentive to work in these last few days is zero. At the start of the next month, the clock is re-set. He now has incentive to start attending school again, since by attending at the beginning of the month he can hope to be "in the money" by the end of the month, thereby benefiting from the incentive. Consider another teacher who has worked 10 days by the 21st day of the month. For every day he works in the five remaining days, he earns Rs 50. By the beginning of the next month, his incentive to work is no higher. In fact, it could even be somewhat lower since he may not benefit from the work done the first day of the month if he does not work at least 10 days in that month.

This leads to a simple regression discontinuity design test for whether financial incentives matter, under the assumption that the teacher's outside option if he does not go to school does not also jump discontinuously when the month changes. Figure 3 gives a graphical representation of the approach. It shows a regression of the probability that a teacher works if she is in the money by day 21 of the month (with 4 days left), in the last 10 days of that month and the first 10 days of the next month. We fit a third order polynomial on the left and the right of the change in month. The figure shows a jump up for teachers who were not in the money, and no jump for those who were in the money. This is exactly what we would expect: the change in incentive at the beginning of a month is important for teachers who were not in the money since, in the data, we see that 65 percent of the teachers who are out of the money in a month will be in the money the following month. The teachers who were in the money, however, have a 95 percent chance to be in the money again. In addition, these teachers value the fact that the first days worked help them work toward the 10 day threshold. Correspondingly, we do not see a sharp drop in presence for the teachers who had been in the money the previous month.

Table 3 presents these results in regression form. Specifically, for teachers in the treatment group, we created a dataset that contains their attendance records for the first and the last day of each month. The last day of each month and the next day of the following month form a pair, indexed by m. We run the following equation, where $Work_{itm}$ is a dummy variable equal to 1 if teacher i works in day t in the pair of days m (t is either 1 or 2):

$$Work_{itm} = \alpha + \beta 1_{im}(d > 10) + \gamma Firstday_t + \lambda 1_{im}(d > 10) * Firstday_t + \upsilon_i + \mu_m + \epsilon_{itm}, \quad (1)$$

where $1_{im}(d > 10)$ is a dummy equal to 1 for both days in the pair m if the teacher had worked more than 10 days in the month of the first day of the pair, and 0 otherwise. $Firstday_t$ is a dummy that indicates that this is the first day of month (i.e. the second day of the pair). We estimate this equation treating υ_i and μ_m as either fixed effects or random effects. If the teachers are sensitive to financial incentives, we expect β to be positive (teachers should work more when they are in the money than out of the money), γ to be positive (a teacher who is out of the money in a given month should work more in the first day of the following month) and λ to be negative and as large as γ (there is no increase in incentive for teachers who had worked at least 10 days before).[14]

[14]Note that even with teacher fixed effects, β does not have a causal interpretation, because the shocks may be auto-correlated. For example, a teacher who has been sick the entire month, and thus has worked less than 10 days, may also be less likely to work the first day of the next month. However, because when

The results clearly show that teachers are more likely (46 percentage points in the random effect specification in Column 3) to attend school at the beginning of a month if they were not in the money in the previous month, which we do not see for teachers who were in the money. This holds after controlling for teacher fixed effects (Column 4), and even if we restrict the sample to the first and last day of the month (Columns 1 and 2).

These results imply that teachers are responsive to the financial incentives, unless there are other factors affecting teachers that happen to have exactly the same structure. Shocks to teachers' outside options are unlikely to change discontinuously when the calendar month changes, as no other teacher activity is linked to the calendar month. Could the effect of daily attendance monitoring also be month specific, independent of the financial incentives? It could be if teachers were afraid that Seva Mandir would use the sum of monthly absence to punish or berate the teacher (and possibly to fire them): in that case they would also worry about total absence in a month, and each month would be a new beginning. If teachers thought that Seva Mandir's probability to fire them changed discontinuously every month that their attendance was less than 10 days it would be impossible to separate that from the incentives, since it would have exactly the same time-structure.

Since we do not observe any dismissal in our data (let alone teachers' belief about the probability of dismissal), and we do not have data on non-penuniary punishment either, we are not able to estimate a function that relate firing or non-pecuniary punishment to past presence, so this cannot be tested directly. However, the very fact that no teacher was fired, even though some teachers were absent almost the entire month, suggests that Seva Mandir takes a much longer perspective when they consider teacher performance. Indeed, according to Seva Mandir's head of the education unit, it is a teacher's record over an entire year or more that determines their assessment, not how it distributes across a month. In fact, in the control group, we find no relationship between the calendar day in the month and the chance that we see a teacher at work.[15]

Furthermore, even if Seva Mandir paid attention to monthly totals, we may expect them to matter exactly in the opposite way. Seva Mandir's official policy is that teachers should attend at least 20 days per month. Therefore, a teacher who has attended 20 days would have had no reason to attend any more days from an "official policy" perspective. We expect

a month starts and finishes is arbitrary and should not be related to the underlying structure of shocks, a positive γ indicates that teachers are sensitive to financial incentives, unless there is a common "first day of the month" effect unrelated to the incentives. A negative λ will be robust even to this effect, since it would suggest that only teachers who are "out of the money" experience a "first day of the month" effect.

[15]This result is available from the authors upon request.

the opposite from a financial perspective, and this is also what we observe. Conversely, it is reasonable to think that Seva Mandir would be particularly likely to punish teachers who have attended very few days, so that the incentive to attend to avoid displeasing Seva Mandir should be strong for teachers who have attended less than ten days in a month, precisely when the financial incentive is the weakest.

Thus, these findings suggest that teachers respond to the incentives. However, without more structure, it is not possible to conclude whether the effect of the program was due to financial incentives per se. To analyze this problem, we set up a dynamic labor supply model and we use the additional restrictions that the model provides to estimate its parameters.

4 A Dynamic Model of Labor Supply

We propose and estimate a simple partial equilibrium model of dynamic labor supply, which incorporates the teacher response to the varying incentives over a month. The model is highly stylized and does not necessarily fully reflect teachers' optimization problem, but illustrates how within-month optimization can help us identify the responsiveness to financial incentive.

Let m signify the month and t the day within the month, where $t = \{1, \ldots, T_m\}$.[16] The teacher's utility function over consumption, C_{tm}, and leisure, L_{tm}, each day in the month is as follows:

$$U_{tm} = U(C_{tm}, L_{tm}) = \beta C_{tm}(\pi_m) + (\mu_{tm} - P)L_{tm}, \tag{2}$$

where P is the non-pecuniary cost of missing work.[17] We have assumed that utility is linear in consumption and that consumption and leisure are additively separable. This formulation implies that there will not be a dependency in behavior between months. For example, a teacher would not decide to work more in one month because she worked little in previous months. A more general utility function might generate such behavior.[18]

Consumption is a function of earned income, π_m. Since we assume that there is no

[16] T is 25 in most months. Note that out of the 25 days, there are also several days of training and holidays, which are automatically credited as days worked for the teachers by the payment algorithm. Our estimation procedure follows the same rule, but when we report the number of days worked, we report it out of the days where teachers actually had to make a decision, which is on average 22 days per month.

[17] Although it is clear that μ and P are not separately identified, we have written the costs and benefits of missing work in this way to make explicit the difference between these two countervailing forces on the teacher's labor decision. Moreover, this makes explicit what we mean by "monitoring effect:"the experiment may affect P, but we assume it does not affect μ.

[18] For example, if utility was logarithmic, and teachers could borrow and save between month, they could decide to work little in a particular month, where the opportunity cost of working is high, and borrow against work in future months. This would make teacher behavior dependent on the entire history of work so far.

discounting within months and utility is linear in consumption, we can assume that the teacher consumes all his income on the last day of the month, when he is paid.[19] The parameter β converts consumption, measured in Rupees, into utility terms. We let L_{tm} equal one if the teacher does not attend work on that day and zero otherwise.

The coefficient on the value of leisure, μ_{tm}, has a deterministic and stochastic component:

$$\mu_{tm} = \mu + \epsilon_{tm}. \tag{3}$$

The deterministic component, μ, is the difference between the value of leisure and the intrinsic value of being in school, including any innate motivation. To the extent that teachers value teaching, or do not want to disappoint students and parents, μ will be more negative. The stochastic shock, ϵ_{tm}, captures variation in the opportunity cost of attending work on a given day; we assume that it has a normal distribution.

Teachers who do not go to work face two types of penalties. First, an agent who does not attend school on a particular day is assumed to pay a non-pecuniary cost, P. This term captures the idea that teachers are verbally rebuked by their supervisors at bi-monthly meetings if they have been found out to be absent during the month: each day of absence makes it more likely that the teacher is found out shirking.[20] Second, we introduce the possibility that an agent is fired for poor attendance; we denote the probability of being fired in a given period by $p_m(t, d)$, where this probability depends on the number of days previously worked, d, by time t in month m.[21] Teachers may be fired at the beginning of each day before attending work; teachers who are fired receive a one-time payment of F, the outside option to being a teacher. F may potentially be related to μ, the opportunity cost of working a day. This would be the case, for example, if the next best option to working as a teacher is working as a day laborer.[22]

Letting d_{m-1} denote the number of days worked in a month, the agent's income earned

[19] A more realistic alternative, which would lead to no difference in any of our estimation, would be to assume that consumption of π_m is spread throughout the next month.

[20] Note that this way of introducing the non-pecuniary cost is a restrictive functional form assumption: if it was allowed to vary with the total number of absences in the month in a non-restrictive way, this would preclude identification of the incentive effects. As we explain above, we think that this is a reasonable functional form.

[21] In principle, the probability of being fired can be a function of the teacher's complete past work history, but for expositional clarity we consider a specification that only depends on days worked in the current month.

[22] To slightly anticipate our results below, we are not able to identify F in the data, since we do not observe any firing in our study period; it will therefore be impossible to estimate it in more detail.

in the last period in the treatment group is given by the following function:

$$\pi_m = 500 + 50 \max\{0, d_{m-1} - 10\}. \tag{4}$$

In the control group, the agent's income in the last period is Rs 1000, irrespective of attendance. The payoff function is such that teachers who work every day in a month will receive more in the treatment group than in the control group.

The teacher is assumed to maximize the present value of lifetime utility.[23] For teachers in the control group, this implies that they face a simple repeated binary choice problem. The Bellman equation for agents in the control group every day of the month except the last day is:

$$\begin{aligned} V_m(t, d; \epsilon_{tm}) = p_m(t, d) \cdot F + (1 - p_m(t, d)) \max\{&\mu - P + \epsilon_{tm} + EV_m(t + 1, d; \epsilon_{t,m+1}), \\ &EV_m(t + 1, d + 1; \epsilon_{t,m+1})\}, \end{aligned} \tag{5}$$

where, without loss of generality, we have set the current-period utility of attending work to zero. The expectation over future value functions is taken with respect to the distribution of next period's shock, $\epsilon_{t,m+1}$. Agents weigh the marginal change in the possibility of being fired in future periods against the immediate benefits of skipping work. From Equation 5, it is clear that μ and P are not separately identified. Therefore, without loss of generality, we redefine the outside option of not working for the control group as $\tilde{\mu} = \mu - P$. $\tilde{\mu}$ could easily be negative if P is large enough. At the end of each day, for $t < T_m$, t increases by one and d increases by one if the teacher worked that day. After time T_m, the state variables of time and days worked reset to zero. On the last day of the month, the value function is almost identical, with $\beta * 1000$ added to the utility of not being fired.

Teachers in the treatment group face a very different decision problem. First, the structure of the financial incentives induces an additional dynamic concern, as teachers trade off immediate gratification against the possibility of increased wages at the end of the month. Second, the cameras provide Seva Mandir with better information on absences, which can lead to changes in both P, the non-pecuniary cost paid for each absence, and the probability of being fired $p_m(t, d)$. Whether P should increase or decrease with the financial incentive is an open question: on the one hand, Seva Mandir now has perfect information on pres-

[23] We assume that there is no discounting within or across months. With idiosyncratic shocks to the outside option our model is equivalent to one where there is discounting across months but not within months, as idiosyncratic errors imply that the relevant decision horizon is only the current month.

ence, whereas in the control group, they visit the school infrequently, so most absences go undetected. On the other hand, one can imagine that Seva Mandir puts more weight to an absence they find out during one of their inspection visits, and thus that the expected cost of a missed day is similar in the treatment and the control groups. Moreover, if Seva Mandir feels that teachers are sufficiently punished for not attending school by the financial penalty, they may lower the non-pecuniary punishment relative to a situation without incentive, which would lower P. For expositional clarity, and to emphasize that it may differ from the control group, we denote the punishment in the treatment group by $\overline{P}$.

Given this payoff structure, for $t < T_m$, we can write the value function for each teacher as follows:

$$\begin{aligned} V_m(t, d; \epsilon_{tm}) = {} & p_m(t, d) \cdot F + (1 - p_m(t, d)) \\ & \times \max\{\mu - \overline{P} + \epsilon_{tm} + EV_m(t + 1, d; \epsilon_{t,m+1}), EV_m(t + 1, d + 1; \epsilon_{t,m+1})\}. \end{aligned} \tag{6}$$

At time T_m:

$$\begin{aligned} & V_m(T_m, d; \epsilon_{T_m,m}) = p_m(T_m, d) \cdot F + (1 - p_m(T_m, d)) \\ & \times \max\{\mu - \overline{P} + \epsilon_{T_m,m} + \beta\pi(d) + EV_{m+1}(1, 0; \epsilon_{t,m+1}), \beta\pi(d + 1) + EV_{m+1}(1, 0; \epsilon_{t,m+1})\}. \end{aligned} \tag{7}$$

Note that the term $EV_{m+1}(1, 0; \epsilon_{t,m+1})$ enters into both arguments of the maximum operator in Equation 7. Since the expectation of this term is independent of any action taken today, in the context of the present model we can ignore any dynamic considerations that arise in the next month when making decisions in the current month. This is useful since we can think about solving the value function by starting at time T_m and working backward, which breaks an infinite-horizon dynamic program into a repeated series of independent finite-time horizon dynamic programs.

Equation 7 also motivates several of our normalizing assumptions. First, the mean of the shock and the mean level of utility of not working are not separately identified; as a result we set the mean of the shock to be equal to zero. Second, Equation 7 is only identified up to scale, as multiplying both sides by a positive constant does not change the work decision. Therefore, we follow a common standard in the discrete choice literature and normalize the variance of the error term to one. Third, as in Equation 5, μ and $\overline{P}$ are not separately identified, and without loss of generality we let $\overline{P} = 0$. We note, however, that we can calculate the difference between $\tilde{\mu}$ and $\mu - \overline{P}$ by comparing the predicted attendance rates across the treatment and control groups when we set the financial incentives to zero. This difference identifies the effect of the cameras on teacher attendance absent the financial incentives.

4.1 Estimators

We estimate several specifications of the general dynamic program described by Equations 6 and 7.[24] The models vary according to what we assume about μ and the distribution of ϵ, and whether we use information from the control schools directly in the estimation. We start with the simplest i.i.d. model, and progressively add observed and unobserved heterogeneity and allow for auto-correlation in the shock to the outside option, μ. We first estimate these models using only the treatment group data and, following Todd and Wolpin (2006), use the means from the control group as an out-of-sample check under the null hypothesis that the non-pecuniary cost of not working (P) is the same in the treatment and the control group. We also estimate a second set of models, where we allow the outside option to vary with teacher test scores and the average attendance of the control group in the local block area. The inclusion of block-level control group attendance scores controls for spatially-correlated shocks to working at the month level, for example, particularly hot or cold weather that made going to school unattractive to both the treatment and control groups. We also allow the outside option to vary with teacher scores to control for the fact that teachers with higher scores may be more diligent and may tend to work more often.

Before describing the empirical specifications in detail, we note that we never observe any teachers being fired in the data. This is despite the fact that some teachers have months where they never worked. Therefore, a consistent estimator for $p_m(t, d)$ in the model above is $\hat{p}_m(t, d) = 0$. We therefore proceed as if the teachers perceive the probability of being fired as being identically equal to zero. This may not be completely correct, in that teachers may believe that if they do not come at all in the year they will be fired, but we never observe anyone being fired in either the treatment or control groups. This is despite the fact that 2.3 percent of the teacher-months in the treatment data have recorded zero attendance, with the worst teacher missing 50 percent of the days in 2005. In the control group, of teachers with 20 or more random checks, 34 percent of teachers were present 50 percent of the time or less, and 8 percent present less than 35 percent of the time. For these reasons, we are fairly comfortable positing that the probability of being fired is equal to zero regardless of

[24]We briefly discuss the identification of these models below.

work history, at least in the range of work history we observe.[25]

Models with i.i.d. Errors The simplest model that we estimate is one where all agents share the same marginal utility of income, β, and average outside option of not working, μ, and the shocks to the utility of not working are i.i.d. We use all of the days in the month in our estimation by utilizing the empirical counterpart of Equation 6 for $t < T$:

$$\begin{aligned} Pr(\text{work}; t, d, \theta) &= Pr(\mu + \epsilon_{tm} + EV(t+1, d) < EV(t+1, d+1)) \\ &= Pr(\epsilon_{tm} < EV(t+1, d+1) - EV(t+1, d) - \mu) \\ &= \Phi(EV(t+1, d+1) - EV(t+1, d) - \mu), \end{aligned} \tag{8}$$

where $\Phi(\cdot)$ is the standard normal distribution. Each of the value functions in Equation 8 is computed using backward recursion from period T_m. Let w_{imt} be an indicator function equal to one if teacher i worked on day t in month m, and zero otherwise. The log-likelihood function for the model without serial correlation in the error terms is then:

$$LLH(\theta) = \sum_{i=1}^{N} \sum_{m=1}^{M_i} \sum_{t=1}^{T_m} [w_{imt} Pr(\text{work}; t, d, \theta) + (1 - w_{imt})(1 - Pr(\text{work}; t, d, \theta)], \tag{9}$$

where each agent is indexed by i, the months they work are indexed by $m = \{1, \ldots, M_i\}$, and the days within each of those months are indexed by $t = \{1, \ldots, T_m\}$. This likelihood is well-behaved and can be evaluated quickly since numerical integration is not necessary.

Models with Serial Correlation It is reasonable to think that the shock to teacher's outside option may be correlated over periods. For example, when a teacher is sick, she may be sick for a few days. Indeed, Appendix Table I suggests that serial correlation is prevalent in the data. The table shows empirical sequences of days worked in the last five days of a month by teachers who were already "in" and "out" of the money. These teachers did not face any dynamic incentives within the month, and we can see evidence of autocorrelation by just looking at these empirical probabilities. Intuitively, a lack of autocorrelated shocks should imply that the distribution of sequences is uniform. While the table with teachers who are definitely out of the money is inconclusive, due to a small sample size, the table for teachers in the money is more clear. The two most frequent sequences are the two most

[25] Our discussions with Seva Mandir suggest that the discussions with teachers and verbal rebuking are the binding constraints for teachers, which explain why they never have to fire anyone.

positively autocorrelated, 00111 and 11100. At the other end of the spectrum, the most negatively autocorrelated sequence, 10101, is the second least frequent, appearing ten times less frequently than 00111. This suggestive evidence motivates several specifications that can handle serial correlation in shocks to the outside option.

Our first approach allows for a simple form of serial correlation in the preference shock by allowing the value of leisure to depend on the observed lagged absence:[26]

$$\mu_{mt} = \mu + w_{m,t-1} \cdot \gamma, \tag{10}$$

where, as above, $w_{m,t-1}$ is an indicator function for whether the agent worked in the previous period. If $\gamma > 0$, then we expect that working today increases the probability of working tomorrow, and therefore days worked or missed will be clustered together in a month. The likelihood of this model is given by:

$$\begin{aligned} LLH(\theta) = \sum_{i=1}^{N} \sum_{m=1}^{M_i} \sum_{t=1}^{T_m} [& w_{imt} Pr(\text{work}; t, d, \theta, w_{m,t-1}) \\ & + (1 - w_{imt})(1 - Pr(\text{work}; t, d, \theta, w_{m,t-1})], \end{aligned} \tag{11}$$

Note that the probability of working today expressly depends on whether the agent worked yesterday through Equation 10.[27] This specification has the advantage of using all of the data within a month and is simple to implement. An important advantage is that we can estimate this specification using only a narrow window around the change in month. The main source of identifying variation in this case is that, when the month changes, the financial incentives to work increase for teachers who were "out of the money" at the end of the previous month, while they decrease for those who were previously "in the money." This is the specification of the structural model that is closest to the regression discontinuity estimate presented above. If the results are similar to those using all the data, this will increase our confidence in the other models.

The second approach is to model the shock process as following an AR(1) process:

$$\epsilon_{mt} = \rho \epsilon_{m,t-1} + \nu_{mt}, \tag{12}$$

where ρ is the persistence parameter and ν_{mt} is a draw from the standard normal distri-

[26] We thank an anonymous referee for suggesting this specification.

[27] We drop the first month of observations for each teacher since we lack data on whether the teacher worked the last day in the previous month.

bution. Autocorrelation could be either positive (illness) or negative (teacher has a task to accomplish). Irrespective of whether ρ is positive or negative, we can no longer directly apply the estimator used in the i.i.d. case. This is because the ϵ_{mt} will be correlated with d, as teachers with very high draws on ϵ_{mt} are more likely to be in the region where $d < 10$ if ρ is positive (the converse will be true if ρ is negative). In this case, the expectation is that $\epsilon = 0$ is invalid, and will bias our estimates of the other parameters.

Our solution is to consider only the sequence of days worked at the beginning of the month.[28] Heuristically, we match the empirical frequencies of sequences of N days worked at the beginning of each month as closely as possible to the frequencies predicted by our model. This results in $2^N - 1$ linearly independent moments, where we have subtracted one to correct for the fact that the probabilities must sum to one. In our estimation, we match sequences of length $N = 5$, which generates 31 moments. In this approach, we treat the draw of the error term at the beginning of each month as coming from the unconditional distribution of ϵ. This is justified by the observation that true distribution of ϵ, conditioning on the work history in the previous month's first five days, is essentially identical to the unconditional distribution after 25 days, even for high ρ values.[29] The MSM estimator does not directly exploit the variation coming from the discontinuous change in incentives at the end of the month, as this would require an enormous number of moments. For example, modeling the sequences across a month with 25 days would require at least sequences of 26 days, generating $2^{26} - 1 = 67,108,863$ moments. However, the discontinuity and the nonlinear payment rule is still the source of identification in our model, as expressed through changes in the probability of working throughout the month as a function of days previously worked.

Observed and Unobserved Heterogeneity We consider two types of extensions to the specifications above. The first extension incorporates observed characteristics into the specification for μ. We introduce the attendance in the control group in the same geographic block as a shifter for the outside option, to exploit the informational content of absence in the control group: the parallel monthly fluctuations in the attendance of control and treatment teachers in Figure 2 suggest that there the behavior of treatment and control group teachers across months may indeed be correlated. Secondly, we allow the teacher's score on Seva

[28]This model is estimated using the method of simulated moments, which is described in detail in the Appendix.

[29]Simulation results are available upon request from the authors.

Mandir's admission exam to shift μ.[30]

The second extension is to relax the assumption that the outside option is equal across all agents after conditioning for observed heterogeneity. We estimate specifications with either fixed effects or random coefficients. In the fixed effects model, teacher types are fixed across time, and we allow μ_i to be estimated separately for each teacher.[31] In the random coefficient models, we estimate two specifications which differ through the distribution of outside options. In the first specification, μ_{im} is drawn anew from a normal distribution each month. In the second specification, we allow for a mixture of two types, where each type is distributed normally with proportion p and $(1-p)$ in the population.

4.2 Parameter Estimates

We present the results of these various specifications in Table 4. We present the main parameters of the model, as well as the implied labor supply elasticity, the percentage increase in the average number of days worked caused by a one percent increase in the value of the bonus and the semi-elasticity with respect to the bonus cut-off, and the percentage increase in the average number of days worked in response to an increase in one day in the minimum number of days necessary for a bonus.

The first two columns present the results from specifications without any controls for autocorrelated shocks. Model I is the simplest model, estimating a common β and μ for all teachers. The estimate for β indicates that teachers respond positively to the financial incentives, and will work more often the closer they are to being in the money. The predicted number of days worked in the treatment group, 17.23, tracks very closely to the empirical number, which was 17.16. However, because the estimated opportunity cost of working, $\mu = 1.564$, is greater than zero, this model vastly under-predicts the number of days that teachers work in the control group. Teachers in the control group attended 12.9 days of work per month on average; model I predicts these teachers would work 1.31 days per month. A potential explanation for this result is that teachers vary in their outside options. Therefore, model II relaxes the assumption of a common μ and allows for teacher fixed effects. The estimated β is lowered but still positive, and the model still under-predicts attendance in the control group.

[30]In our current model, this will reflect teacher heterogeneity. In a richer model, it will also capture dependency in behavior between months.

[31]The model with fixed effects has the usual panel model bias, although we expect it will be attenuated here given the relatively long length of the panel.

An additional explanation for these findings is that the models are picking up the confounding effects of serial correlation in the errors and the financial incentives. In order to control for this possibility, we estimate two sets of models, one set (III, IV, and V) controlling for serial correlation using an AR(1) specification for the per-period error term, and one set (VI, VII, and VIII) using a shifter for the outside option that depends on whether the teacher worked the previous day.

Model III adds the AR(1) error process to model I. The model estimates imply that agents still respond strongly the financial incentives and finds strong evidence of positive serial correlation ($\rho = 0.422$), but the control group prediction is still far too low. Model IV adds in one degree of unobserved heterogeneity by allowing μ to be drawn anew from a normal distribution at the beginning of each month. These estimates are largely the same as in model III, with similar poor results for the control group prediction. Finally, model V adds in a second type of unobserved heterogeneity. In this model, the outside option is drawn from one of two normal distributions with a probability p. The estimates from model V suggest that there are two types of workers in the data: a majority with a μ less than zero, and a small proportion ($p = 0.024$) who have a μ drawn from a much higher distribution. In contrast to previous models, this model predicts that most teachers have a negative μ, which implies that teachers will work most days in the control group. The model predicts control group attendance of 12.9 days per month, which is the same as the empirical attendance rate.[32]

While model V finds that response to incentives is positive and does an excellent job of predicting attendance in the control group, one drawback is that it only uses the first five days of data from each month. In order to test the robustness of our findings while using the entire data set, we also estimate several models which incorporate serial correlation through a shifter on the outside option, which depends on whether the teacher worked in the previous period. This is a simple method of introducing serial correlation into the model while retaining the ability to use maximum likelihood approach, which, in principle, can be estimated on the entire data set.

In model VI, we estimate a random coefficients specification with μ drawn from one

[32]A natural question arises: why stop at two types of heterogeneity? Bajari, Fox, Kim, and Ryan (2009) use the same data as an application of an estimator which is non parametric with respect to the distribution of μ. Their estimator allows for up to 80 discrete types and 40 continuous types for μ, holding the other parameters at their values from Model V. The results suggest that the mixture of two normals captures the unobserved heterogeneity extremely well. The Bajari, et. al. result suggest that most of the weight is put on a few points centered around -0.5 and a small diffuse mass of weight around points at higher levels of μ near 2.0, just as the two-type Model V does.

normal distribution and a simple shifter, *yesterday*, which is added to μ if the teacher did not work in the previous period. The results of this model are similar to model V, as β and the mean of the outside option are estimated to be similar. This model estimates a higher level of variance in the outside option, and finds that the yesterday shifter is positive. A positive shifter implies that if a teacher did not work in the previous day, μ is shifted up by 0.094. Holding the financial incentives fixed, for the average teacher with $\mu = -0.304$, this implies a decrease in the probability of working in the current period by 4.5 percent. Compared to model IV, this model predicts attendance in the control quite better, although it under-predicts attendance slightly. Model VII is the same specification as model VI, but uses a restricted sample of three days on either side of the change in the month to approximate a structural version of a regression discontinuity model. The counterfactual predictions are similar to model VI. This is very reassuring, as this identifying variation in this specification is the sharp change in incentives around the change in month.

Model VIII is the same specification of model VI, estimated on the full sample, with the inclusion of μ shifters for control group attendance in the same geographic block and teacher test scores. Both variables enter μ positively, so the estimates of -0.132 and -0.005 for attendance and test scores, respectively, imply that teachers work more when teachers in the control group geographically proximate to them work more and when they have higher scores, as the negative coefficients on these coefficients imply that the μ decreases with these two variables. The coefficient on the behavior of the control group teachers is significant, which is consistent with the parallel seasonal pattern we observed in the reduced form. The addition of these two controls improves the efficiency of the estimator, but does not result in a large change to the other parameters or our predictions for the number of days worked in both groups.

The similarity of results from different estimation methods for the model is encouraging. Particularly reassuring is the fact that the estimate of β, or all the implied elasticities for the model, is very similar (ranging between 0.2 and 0.3) when using all the data (model VI and VIII), the first five days (model V), or the three days window (model VII). This suggests that the identification based on the shift in incentive at the end of the month drives the results in all the models. In these three models, the mean outside option (which includes the punishment for not working) is negative for the majority of the teachers.[33] This suggest that, taking into account the non-pecuniary cost of absence, teachers are willing to work more than half the days, even without financial incentives. This is consistent with the fact

[33] In the model with two types, the average outside option is estimated to be -0.375.

that the teachers work a little over half the time in the control group.

Note that all these models predict about the same rate of absence in the control group as we predict when setting the incentive to zero for our treatment group teacher. It suggests that $\overline{P}$ is close to P, or that there is no direct impact of the daily monitoring.

4.3 Goodness-of-Fit and Out-of-Sample Tests

To provide a sense of the fit of each model, we report the predicted number of days worked under each specification. This is not a good test for the models estimated using maximum likelihood (Models I, II, III, and VI-VIII), which use all the days worked to compute the parameters of the model, and should therefore do a good job of matching the average number of days worked. However, note that this is not a parameter that our method of simulated moments estimation tried to match (since we matched only the first five days of teacher behavior), so it provides a partial goodness of fit metric for these models. Moreover, the model using the three days window does not match this moment mechanically either.

Figure 4A plots the density of days worked predicted by Model V, and its 95 percent confidence interval, and compares it to the actual density observed in the data. Since the estimation is not calibrated to match this shape, as we are only using the history of the first five days in our estimation, the fit is surprisingly good. The model reproduces the general shape of the distribution, although the mode of the distribution in our predicted fit is to the left of the mode in the data by one day. The model tends to slightly over-predict the frequency of 17 to 21 days worked and under-predict the frequency of days worked between 3 and 10. With the exception of a small proportion of teachers who work a small number of days in a month, the true distribution lies comfortably within the 95 percent confidence interval of the prediction.

As an extra test of goodness of fit, Table 2 in the Appendix shows the empirical moments and the predicted moments from Model V. Our model generally does well in predicting the patterns of days worked, under-predicting some of the extreme not work/work sequences (00011, 00111, and 01111) and over-predicting others (11100 and 11110).

A change in the incentive system at Seva Mandir, after the first version of this paper was written and our model was estimated, provides us with a very nice counterfactual experiment. In December 2006, Seva Mandir increased the minimum monthly payment to Rs 700, which teachers receive if they work 12 days or less (rather than 10 days). For each additional day they work, teachers earn an additional Rs 70 per day. Seva Mandir provided us with the camera data in the summer of 2007, a few months after the change in policy. The average number of days worked since January 2007 increased very slightly, from 17.16 to 17.39 days. The predicted number of days worked for each model is reported in the last row in Table 4. Here again, our preferred specifications (Models V through VIII) performs very well: they predict between 17.8 and 20.2 days worked under the new incentive scheme, an increase from the predicted number of days under the main scheme (as in the actual data). Figure 4B shows the actual distribution of days worked and the predicted one for model V. The model does a good job of predicting the distribution of days worked in the out-of-sample test, although the empirical distribution has more variance than in the original experiment.

4.4 Counterfactual Optimal Policies

A primary benefit of estimating a structural model of behavior is the ability to calculate outcomes under economic environments not observed in the data. In our case, we are interested in finding the cost-minimizing combination of the two policy instruments, the size of the bonus and the threshold to get into the bonus, that lead to a minimum number of days worked in a month. Using Model V as our foundation, we calculated the expected number of days worked and expected size of the financial payout for a wide range of potential policies under our preferred model with autocorrelation and two types of heterogeneity.[34] We let the minimum number of days to obtain a bonus range from zero to 23, which is the upper limit of days that a teacher could work in any month. At the same time, we varied the bonus paid for each day over the cutoff from zero to 300 Rp/day in increments of 25 Rp/day. Table 5 shows the lowest-cost combinations of those two policy variables that achieved a minimum expected number of days worked under model V.[35] The table also shows the gain in test scores for each of these combinations (calculated using the estimate of the effect of each extra day on presence and test score, which we estimate below).

The results of this simulation show two general trends: the cost-minimizing cutoff generally decreases and the bonus increases in the expected number of days worked that the policymaker wants to achieve. Both of these trends lead to drastically increasing costs as the target increases. This result directly follows from the model: as we continue to increase the target, the marginal teacher has increasingly higher opportunity costs of working. This becomes quite expensive, as soon it is necessary to incentivize the "slacker" teacher types

[34]We also calculated the optimal policies under Model VI to test the robustness of our results. Those policies are reported in Appendix Table 4. The results are roughly comparable to those under Model V, with an increasing per-day bonus and an inverted-U shaped cutoff function.

[35]The expected outcomes were subject to a very small amount of variance as we drew model primitives from their estimated distributions 50 times for each combination of policy instruments.

in our sample. It is interesting to note that for about the same amount of money spent on both the treatment and control groups in the actual experiment (roughly 1000 Rp/month), teachers under the optimal counterfactual policy would have worked approximately 20 days, an improvement of roughly 16 percent and 56 percent over the treatment group and control group, respectively. Our counterfactual calculations show that while the actual intervention was successful in increasing teacher attendance, the NGO could have induced higher work effort with approximately the same amount of expenditures by doubling the bonus threshold and nearly tripling the per-day bonus. This is due to the fact that teachers in our sample appear to be on average more likely than not to attend school even without incentives and be forward looking. A higher threshold avoids rewarding infra-marginal days, and provides incentives to teachers to work to accumulate the number of days necessary to get the larger prize.

5 Was Learning Affected?

5.1 Teacher Behavior

Though the program increased teacher attendance and the length of the school day, it could still be considered ineffective if the teachers compensated for increased attendance by teaching less. We used the activity data that was collected at the time of the random check to determine what the teachers were doing once they were in the classroom. Since we can only measure the impact of the program on teacher performance for schools that were open, the fact that treatment schools were open more may introduce selection bias. That is, if teachers with high outside options (who are thus more likely to be absent) also tended to teach less when present, the treatment effect may be biased downward since more observations would be drawn from among low-effort teachers in the treatment group than in the comparison group. Nevertheless, Table 6 shows that there was no significant difference in teacher activities: across both types of schools, teachers were as likely to be in the classroom, to have used the blackboard, and to be addressing students when the enumerator arrived. This does not appear to have changed during the duration of the program.

The fact that teachers did not reduce their effort in school suggests that the fears of multitasking and loss of intrinsic motivation were perhaps unfounded. Instead, our findings suggest that once teachers were forced to attend (and therefore to forgo the additional earnings from working elsewhere or their leisure time), the marginal cost of teaching must have been small. This belief was supported during in-depth conversations with 15 randomly selected NFE teachers regarding their teaching habits in November and December of 2005. We found that teachers spent little time preparing for class. Teaching in the NFE follows an established routine, with the teacher conducting the same type of literacy and numeracy activities every day. One teacher stated that he decides on the activities of the day as he is walking to school in the morning. Other teachers stated that, once they left the NFE, they were occupied with household and field duties, and thus had little time to prepare for class outside of mandatory training meetings. Furthermore, despite the poor attendance rates, many teachers displayed a motivation to teach. They stated that they felt good when the students learned, and liked the fact they were helping to educate disadvantaged students.

The teachers' general acceptance of the incentive system may be an additional reason why multitasking was not a problem. Several months into the program, teachers filled out feedback forms. Seva Mandir also conducted a feedback session at their bi-annual sessions, which were attended by members of the research team. Overall, teachers did not complain about the principle of the program, although many teachers had some specific complaints about the inflexibility of the rules. For example, many did not like the fact that a day was not valid even if a teacher was present 4 hours and 55 minutes (the normal school day is six hours, but an hour's slack was given). On the other hand, many felt empowered by the fact that the onus of performing better was actually in their hands: "Our payments have increased, so my interest in running the center has gone up." Others described how the payment system had made other community members less likely to burden the teacher with other responsibilities once they knew that a teacher would be penalized if he did not attend school. This suggests that the program may actually have stronger effects in the long run, as it signals a change in the norms of what teachers are expected to do.

5.2 Child Presence

On the feedback forms, many teachers claimed that the program increased child attendance: "This program has instilled a sense of discipline among us as well as the students. Since we come on time, the students have to come on time as well." Unfortunately, conditional on whether a school was open, the effect of the program on child attendance cannot be directly estimated without bias, because we can only measure child attendance when the school is open. For example, if schools that were typically open also attracted more children, and the program induced the "worst" school (with fewer children attending regularly) to be open more often in the treatment schools than in the comparison schools, then this selection bias will tend to bias the effect of the program on child attendance downwards. The selection bias could also be positive, for example if the good schools generally attract students with better earning opportunities, who are more likely to be absent, and the "marginal" day is due to weak schools catering to students with little outside opportunities. Selection bias is a realistic concern (and likely to be negative) since, for the comparison schools, there is a positive correlation between the number of times a school is found open and the number of children found in school. Moreover, we found that the effect of the program was higher for schools with originally weak teachers, which may attract fewer children.

Keeping this caveat in mind, child attendance was not significantly different in treatment and comparison schools. In Table 7, we present the child attendance rates in an open school, by treatment status (Panel A). An average child's attendance rate was the same in treatment and comparison schools (46 percent). Excluding children who left the NFE, child attendance is higher overall (62 percent for treatment and 58 percent for comparison schools), but the difference is not significant.

However, treatment schools had more teaching days. Even if the program did not increase child attendance on a particular day, the increase in the number of days that the school was open should result in more days of instruction per child. The program's impact on child instruction time is reported in Panel B of Table 7. Taking into account days in which the schools were closed, a child in a treatment school received 9 percentage points (or 30 percent) more days of instruction than a child in a comparison school. This corresponds to 2.7 more days of instruction time a month at treatment schools. Since there are roughly 20 children per classroom, this figure translates into 54 more child-days of instruction per month in the treatment schools than in comparison schools. This effect is larger than that of successful interventions that have been shown to increase child attendance (Glewwe and Kremer, 2005; Banerjee, Jacob and Kremer, 2005). The effect on presence does not appear to be affected by student ability (proxied by the whether or not the child could take a written test in the pre-test). While presence increased slightly more for those who could not write prior to the program (14 versus 10 percentage points), this difference is not significant.

In summary, since children were as likely to attend class on a given day in the treatment schools as in the comparison schools, and because the school was open more often, children received significantly more days of instruction in the treatment schools. This finding suggests that the high teacher absence rate we observed is not likely to be the efficient response to a lack of interest by the children: if it were the case that children came to school 55 percent of the time because they could not afford to attend more than a certain number of days, then we would see a sharp reduction in child attendance in treatment schools on days when the school was open. On the other hand, we do not see a sharp increase in the attendance of children in the treatment schools. This suggests that either the teacher absence rate is not the main cause of the children's irregular attendance, or that the children have not yet had time to adjust. The latter explanation is not entirely plausible, however, since the program has now been in place for over two years, and we do not see a larger increase in the attendance of children in the later periods than in the earlier period.

5.3 Child Learning

Children in the treatment schools, on average, received about 30 percent more instruction time than children in the comparison schools, with no apparent decline in teacher effort. Some, however, argue that because para-teachers are less qualified than other teachers, it is not clear that they are effective despite the support and in-service training they get from NGOs like Seva Mandir. If para-teachers are indeed ineffective, the fact that it is possible to induce them to attend school more often is not particularly relevant for policy. Understanding the effect of the program on learning is therefore critical.

5.3.1 Attrition and Means of Mid- and Post-Test

Before comparing test scores in the treatment and comparison schools, we must first ensure that selective attrition does not invalidate the comparison. There are two possible sources of attrition.[36] First, some children leave the NFEs, either because they drop out of school altogether or because they start attending regular primary schools. Second, some children were absent on testing days. To minimize the impact of attrition on the study, we made considerable attempts to track down the children (even if they had left the NFE to attend a formal school or had been absent on the testing day) and administered the post-test to them. Consequently, attrition was fairly limited. Of the 2,230 students who took the pre-test, 1,893 also took the mid-test, and 1,760 also took the post-test. Table 8 shows the attrition rate in both types of schools, as well as the characteristics of the attriters. At the time of the mid-test, attrition was higher in the comparison group than in the treatment group. At the time of the post-test, attrition was similar across both groups, and children who dropped out of the treatment schools were similar in their test scores to children who dropped out of the comparison schools.

[36] As mentioned earlier, seven centers closed down prior to the start of the program. We made no attempt to test the children from these centers in the pre-test.

Table 8 also provides some simple descriptive statistics, comparing the test scores of treatment and comparison children. The first row presents the percentage of children who were able to take the written exam, while subsequent rows provide the mean exam score (normalized by the mid-test comparison group). Relative to the pre-test and mid-test, many more children, in both the treatment and comparison schools, were able to write by the post-test. On the post-test, students did slightly worse in math relative to the mid-test comparison, but they performed much better in language.

Finally, Table 8 also shows the simple differences in the mid- and the post-test scores for students in the treatment and comparison schools. On both tests, in both language and math, the treatment students did better than the comparison students (a 0.16 standard deviation increase and 0.11 standard deviations in language at the post-test score), even though the differences are not significant. Since child test scores are strongly auto-correlated, we obtain greater precision by controlling for the child's pre-test score level.

5.3.2 Test Results

In Table 9, we report the program's impact on test scores. We compare the average test scores of students in the treatment and comparison schools, conditional on a child's pre-program competency level. In a regression framework, we model the effect of being in a school j that is being treated (Treat_j) on child i's score (Score_{ijk}) on test k (where k denotes either the mid- or post-test exam):

$$\text{Score}_{ijk} = \beta_1 + \beta_2 \text{Treat}_j + \beta_3 \text{ Pre\_Writ}_{ij} + \beta_4 \text{Oral\_Score}_{ij} + \beta_5 \text{ Written\_Score}_{ij} + \varepsilon_{ijk}. \quad (13)$$

Because test scores are highly autocorrelated, controlling for a child's test scores before the program increases the precision of our estimate. However, the specific structure of the pre-test (i.e., there is not one "score" on a comparable scale for each child because the children either took the written or the oral test in the pre-test) does not allow for a traditional difference-in-difference (DD) or "value added" (child fixed effect) strategy. Instead, we include a variable containing the child's pre-test score for the oral test if he took the oral pre-test and 0 otherwise (Oral_Score_{ij}), the child's pre-test score on the written test if he took the written test and 0 otherwise ($\text{Written\_Score}_{ij}$), and an indicator variable for whether he took the written test at the pre-test (Pre_Writ_{ij}).[37] This fully controls for the

[37] At the pre-test, children were given either the oral or the written score. At the mid- and post-test, every child took the oral part, and every child who could write took the written exam (all children were given a chance to try the written exam; if they could not read, they were given a zero for the written test).

child pre-test achievement, and is thus similar in spirit to a DD strategy. Standard errors are clustered by school. Each cell in Table 9 represents the treatment effect (β_2) obtained in a separate regression. For ease of interpretation, the mid-test results (Columns 1 to 4) and post-test results (Columns 5 to 8) are expressed in the standard deviation of the distribution of the mid-test score in the comparison schools.[38]

The tables reveal that the program had a significant impact on learning, even as early as the mid-test. Children in treatment schools gained 0.16 standard deviations of the test score distribution in language, 0.15 standard deviations in math, and 0.17 overall (Panel A). Including controls for school characteristics—location, teacher test scores, and the infrastructure index of school—does not significantly change our findings (Panel B). Children who could write at the time of the pre-test gained the most from the program. For example, they had mid-line test scores 0.25 standard deviations higher in treatment schools than in comparison schools (Panel D). Interestingly, the children who could write at the time of the pre-test do not increase their attendance rate in response to the greater teacher attendance rate relatively more than those who could not write at the time of the pre-test. Therefore, it is not that they have relatively more days of schooling than the students who could not write as a result of the program, but rather that they seem better equipped to make the most out of the additional days of schooling that they receive.

We compare the program's impact on girls versus boys in Panels E and F. Girls gained as much, if not more, from the program as boys. On the mid-test, 7 percentage-points more of girls in the treatment schools were able to write relative to the comparison schools, compared to only 2 percentage-points of boys (this five percentage point difference is significant).

The differences between students in the treatment and comparison schools persisted in the post-test (Columns 5 to 8). Children in treatment schools gained 0.21 standard deviations in language, 0.16 in math, and 0.17 overall (Panel A). Similar to the mid-test, much of the gains came from children who could write at the time of the pre-test. The post-test also suggests that girls gained slightly more from the program than the boys, but these differences are not significant. The treatment effect of 0.17 standard deviations is similar to other successful educational interventions, such as the Balsakhi Remedial Education Program in India during its first year (Banerjee, et al., 2005).

[38] Scores are normalized such that the mean and standard deviation of the comparison group at the time of the mid-test exam is zero and one, respectively. (Specifically, we subtract the mean of the comparison group in the pre-test, and divide by the standard deviation.) This allows for comparison across samples, as well as with the results from other studies. We could not normalize with respect to the pre-test score distribution since not every child took the same test at the pre-test.

Finally, we examined other sources of heterogeneity in program effect (school infrastructure, teacher level of education, teacher proficiency measured by the test scores). The results (omitted for brevity, but available from the authors), suggest little or no heterogeneity in program effect along these dimensions.

5.3.3 Leaving the NFE

NFEs prepare children, who might not otherwise attend school, to enter government schools at the age-appropriate grade level. To do so, children must demonstrate proficiency for a grade, either by passing an exam or through vetting by a government teacher. The ability of his students to join government schools is, therefore, a strong signal of success for a NFE teacher. The program increased the graduation rate to the government schools. As shown in Table 10, 26 percent of students in the treatment schools graduated to the government schools, compared to only 16 percent in the comparison schools (by February 2006). This 10 percentage point difference implies a 62 percent increase in the graduation rate and is significant.

In the final row of Table 10, we present the dropout rates for children who left school entirely (i.e. left the NFE and did not join a government school). The dropout rate is slightly lower for the treatment schools, but this difference is insignificant.

5.3.4 Estimating the Effect of Teacher Presence on Learning

The previous sections presented the reduced form analysis of the effect of the incentive program on child learning. Table 11 interprets what these estimates can tell us about the impact of teacher attendance.[39] Columns 1 to 3 report simple correlations between the teacher attendance rate and the child test scores. Specifically, we report the coefficient estimate of the number of times a school was found open (Open_j) on a regression of either the mid-test or post-test scores:

$$\text{Score}_{ijk} = \beta_1 + \beta_2 \text{Open}_j + \beta_3 \text{ Pre\_Writ}_{ij} + \beta_4 \text{Oral\_Score}_{ij} + \beta_5 \text{ Written\_Score}_{ij} + \varepsilon_{ijk}. \quad (14)$$

We continue to control for the child's pre-test score and to cluster standard errors by school.

Column 1 reports OLS estimation of Equation 2 for the comparison schools. In this case, the random check data are used to estimate the number of times a school is found open.

[39]This estimate are the effect of being present at a random check, which combines the effect of having come at all, and having come for a longer time.

The coefficient is 0.20, indicating that the test scores of children in centers open 100 percent of the time would be 0.10 standard deviations higher than those of children in a center open 50 percent of the time. Note that this coefficient is insignificant.

This point estimate is similar to those reported in other studies (Chaudhury, et al., 2005) and, taken at face value, would imply that the effect of teacher attendance on learning is not that large. Chaudhury et al. (2005) conjectures that the measurement of absence rates based on a few random visits per school have considerable error, and may thus bias the results downwards. Consistent with this theory, the effect on the post-test scores, where having more months of random check data allows us to better estimate the absence rate per school, becomes larger (0.58 standard deviations). Our study provides a much more direct test of this hypothesis, since, for treatment teachers, the photograph data gives us the actual attendance. We present the OLS estimate of the effect of attendance for treatment teachers using the random check data (Column 2) and camera data (Column 3). Overall, the effect of teacher attendance is larger in the treatment schools than the comparison schools (0.39 in Column 2 to 0.20 in Column 1, both obtained with random check data). More interestingly, consistent with the measurement error hypothesis, the effect of teacher attendance is larger and much more significant when using the more accurate measure of attendance from the camera data, especially for the mid-test scores (the estimate is 0.87 standard deviations in the Column 3 as compared to 0.39 in Column 2). For the post-test, where we have a much more accurate measure of attendance from the random check data, the results from the two methods are similar (0.98 in Column 3 versus 1.17 in Column 2).

Finally, in Column 4, we pool both samples and instrument Open_j (as measured by the random check) with the treatment status of the school to obtain exogenous variation in the percentage of time the school was found open. Since we have shown that the program had a direct effect on the length of the school day, as well as whether or not the school opened at all, the 2SLS estimate captures the joint effect of outright absence and of a longer school day. The 2SLS estimates are higher than the OLS results found in Column 1, and they are indistinguishable from the OLS results in Column 3, obtained with the precisely measured absence rate. This suggests that the relatively low correlation between teacher absence and test scores that was observed in previous studies is indeed likely to be due to measurement error in the teacher absence data. The more precise IV estimates suggest that even a 10 percentage point reduction in the absence rate would result in a 0.10 standard deviation increase in child test scores.

Extrapolating these estimates (which must be done with caution, since the local effect may be different from the overall effect), we can conclude that the effect of being enrolled in an NFE for a year with a teacher present every day is about one standard deviation. This point estimate is similar to the effect of attending remedial education classes with a para-teacher for one year in urban India for children who are enrolled in regular primary school, but have not yet achieved basic numeracy or literacy (Banerjee et al., 2005). In that study, the point estimate was 1.12 standard deviations. Both of these studies therefore suggest that para-teachers can be extremely effective teachers, at least when an NGO provides them with proper training.

6 Conclusion

In this paper, we show that direct monitoring, combined with simple and credible financial incentives based on teacher attendance, leads to large increases in attendance among para-teachers in informal schools. Absenteeism fell from an average of 42 percent in the comparison schools to 21 percent in the treatment schools, without affecting the teachers' effort while in school. As a result, the students in treatment schools benefited from about 30 percent more instruction time. The program had an economically significant impact on test scores: after one year, child test scores in program schools were 0.17 standard deviations higher than in comparison schools. Children were also much more likely to be admitted to government schools.

This paper contributes to a small but growing literature that exploits structural modeling and carefully controlled randomized experiments to answer an economic question. On the substantive front, our results suggest that providing incentives for attendance in non-formal schools can increase learning levels. However, the question arises as to whether incentive programs can be instituted for government teachers. Teachers in government schools are often more politically powerful than para-teachers. Thus, it may prove difficult to institute a system in which they would be monitored daily using a camera or similar device. However, our findings suggest that the barriers currently preventing teachers from attending school regularly (e.g., distance, other activities, lack of interest by children) are not insurmountable. Given political will, it is possible that solutions to the absence problem could be found in government schools as well.[40]

A recent experiment demonstrates the external validity of these results outside the NGO

[40] A sign that these results did indeed generate interest in the government is that Seva Mandir was awarded the annual Government of India "Digital Learning Award" of 2007 for this project.

context, as well as the difficulty of extending such programs in government settings (Banerjee, Duflo and Glennerster, 2007). Following the results of the cameras program, the government of Rajasthan created a similar system for government nurses, whose absence rate was about 44 percent. The nurses were monitored using time and date stamps. The announced incentive system was severe: it called for a 50 percent reduction in the pay of nurses who were absent 50 percent of the time, and termination of persistently absent nurses. In the first few months, when these punishments were carried out, the program lead to about a 50 percent reduction in absenteeism. However, after a few months, the government started granting a large number of ex-post "exemptions" (although the monitoring did continue). The absence rate in the treatment group quickly converged to that of the control group. This further confirms that monitoring is effective, but only when coupled with real incentives, as is suggested by the results of our structural model.

The program for nurses suggests that barriers exist to the implementation of incentive systems for government employees. However, our findings also imply that para-teachers can be effective. If implementing monitoring within the government system turns out to be impossible, our results provide support for the policy of increasing teaching staff through the hiring of para-teachers.

References

Aguirregabiria, Victor and Pedro Mira (2010). "Dynamic Discrete Choice Structural Models: A Survey." *Journal of Econometrics*, 156(1): 38-67.

Attanasio, Orazio, Costas Meghir and Ana Santiago (2006). "Education Choices in Mexico. Using a Structural Model and a Randomized Experiment to Evaluate Progresa," MIMEO, UCL.

Bajari, Patrick, Jeremy Fox, Kyoo-il Kim, and Stephen P. Ryan (2009). "A Simple Nonparametric Estimator for the Distribution of Random Coefficients," NBER Working Paper 15210.

Banerjee, Abhijit, Esther Duflo and Rachel Glennerster (2007). "Putting Band-Aid on a Corpse. Incentives for nurses in the Indian Public Health Care System" *Journal of the European Economics Association* Vol. 6(2-3): 487-500.

Banerjee, Abhijit, and Esther Duflo (2006). “Addressing Absence,” *Journal of Economic Perspectives* 20(1): 117-132.

Banerjee, Abhijit, Suraj Jacob and Michael Kremer, with Jenny Lanjouw and Peter Lanjouw (2005). “Moving to Universal Education: Costs and Tradeoffs” MIMEO, MIT.

Benabou, Roland, and Jean Tirole (2006). “Incentives and pro-social Behavior” *American Economic Review*, 95(6): 1652-1678

Bound, John, Todd Stinebrickner, and Timothy A. Waidman (2005). “Using a Structural Retirement Model to Simulate the Effect of Changes to the OASDI and Medicare Programs,” ERIU Working Paper, University of Michigan.

Card, David and Dean Hyslop (2005.) “Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare Leavers” *Econometrica*, 73(6): 1723-1770.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, F. Halsey Rogers (2005). “Teacher Absence in India: A Snapshot,” *Journal of the European Economic Association.*3(2): 658-667.

Education for All Forum (2000). *EFA Country Assessment Country Reports.*

Fehr, Ernst, and Schmidt (2004). “Fairness and Incentives in a Multi-task Principal–Agent Model” *Scandinavian Journal of Economics*, 106(3), 453–474.

Fehr, Ernst and Lorenz Goette (2007). ”Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment,” American Economic Review 97(1): 298-317.

Ferral, Christopher (2008). “Explaining and Forecasting Results of the Self-Sufficiency Project” Queen’s Economics Department Working Paper No. 1165.

Figlio, David and Lawrence S. Getzler (2002). “Accountability, Ability and Disability: Gaming The System,” NBER Working Paper 9307.

Figlio, David and Josh Winicki (2002). “Food for Thought? The Effects of School Accountability Planson School Nutrition, “ NBER Working Paper 9319.

Glewwe, Paul, Nauman Ilias and Michael Kremer (2003). “Teacher Incentives,” NBER working paper no. 9671.

Glewwe, Paul and Michael Kremer (2005). "Schools, Teachers, and Education Outcomes in Developing Countries,", forthcoming in *Handbook on the Economics of Education.*

Holmstrom Bengt and P. Milgrom (1991). "Multi-Task Principal-Agent Problems: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization*, VII: 24–52.

Jacob, Brian A. and Steven D. Levitt (2003). "Rotten Apples: An Investigation Of The Prevalence And Predictors Of Teacher Cheating," Quarterly Journal of Economics, 118(3): 843-877.

Keane, Michael and Ken Wolpin (1994). "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76: 648-672.

Keane, M. and R. Moffitt (1998). "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review*, 39(3): 553-589.

Kreps, David (1997). "Intrinsic Motivation and Extrinsic Incentives," *American Economic Review*, 87(2): 359-364.

Lavy, Victor (2003)."Paying for Performance: The Effect of Teachers' Financial Incentives on Students' Scholastic Outcomes,"CEPR Discussion Papers 3862.

Lise , Jeremy, Shannon Seitz and Jeffrey Smith (2004). "Equilibrium Policy Experiments nad the evaluation of social Programs" NBER working paper no. 10283.

McFadden, Daniel and A. P. Talvitie and Associates (1977)."Validation of Disaggregate Travel Demand Models: Some Tests" in Urban Demand Forecasting Project, Final Report, Volume V, Institute of Transportation Studies, University of California, Berkeley.

McFadden, D. (1989). "A Method of Simulated Moments for Estimation of Multinomial Discrete Response Models Without Numerical Integration," *Econometrica*, 57:995-1026.

Muralidharan, Karthik and Venkatesh Sundaraman (2009). "Teacher Performance Pay: Experimental Evidence from India," NBER Working Paper 15323.

Pakes, A. and D. Pollard, (1989). "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57: 1027-1057.

Pratham (2006). *Annual Status of Education Report*.

Stinebrickner, Todd (2000). "Serially Correlated State Variables in Dynamic, Discrete Choice Models," *Journal of Applied Econometrics*, 15, 595-624.

Stinebrickner, Todd R (2001a). "A Dynamic Model of Teacher Labor Supply." *Journal of Labor Economics*, 19(1): 196-230.

Stinebrickner, Todd R (2001b). "Compensation Policies and Teacher Decisions." *International Economic Review*, 42(3): 751-779.

Sullivan, Paul (2009). "A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search," Forthcoming, *International Economic Review*.

Todd, Petra and Kenneth Wolpin (2006). "Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico:," *American Economic Review*, 2006, 96(5): 1384–1417.

Todd, Petra, and Kenneth I. Wolpin (2009). "Structural Estimation and Policy Evaluation in Developing Countries." PIER Working Paper No. 09-028.

Todd, Petra, Michael P. Keane, and Kenneth I. Wolpin (2010). "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications." Forthcoming in the Handbook of Labor Economics, Volume 4.

Van der Klaauw, Wilbert. 2005. "The Supply and Early Careers of Teachers." MIMEO, UNC Chappel Hill.

Wise, David (1985). "A Behavioral Model vs.Experimentation: The Effects of Housing Subsidies on Rent," in P.Brucker and R.Pauly (eds), Methods of Operations Research50, Verlag Anton Hain, 441-89.

Appendix: Estimation of model with AR(1) errors

To estimate a model with this stochastic process on the outside option, we discretize the error term into 200 states and solve the dynamic programming problem defined by Equations

6 and 7 for an initial guess of our parameters.[41] We then simulate many work histories from this model, forming an unbiased estimate of the distribution of *sequences* of days worked at the beginning of each month. We simulate the model by drawing sequences $\epsilon = \{\epsilon_0, \ldots, \epsilon_t\}$ and following the optimal policy prescribed the dynamic program. For different sequences of ϵ's we obtain different work histories. Repeating this process many times results in unbiased estimates of the probabilities of all possible sequences. We then match the model's predicted set of probabilities over these sequences against their empirical counterparts.

To address the "incidental parameters" problem of not knowing the first ϵ, we treat the the draw on ϵ at $t = 1$ as coming from the limiting distribution of Equation 12, given by $\bar{F} = N(0, 1/(1-\rho^2))$. The key assumption here is that the dependence of the error term at the beginning of next month on the error term at the fifth day of the previous month is nearly zero, even for high values of ρ.[42] As a result, we can treat the error term next month as drawn from the unconditional distribution of ϵ.

Denoting a sequence of days worked as A, for each teacher-month sequence, we form a vector of moment conditions:

$$E[Pr(A_{im}; X_m) - \widehat{Pr}(A_{im}; X_m, \widehat{\theta})] = 0, \tag{15}$$

where $Pr(A_{im,}; X_m)$ is the empirical probability of observing a sequence of days worked conditioning on X_m, a vector containing the number of holidays and the maximum number of days in that month an agent could potentially work in that month.[43] The moments used in estimation sum across all months and all teachers to form the unconditional expectation of observing a sequence of days worked, $Pr(A)$. We form $\widehat{Pr}(A)$, the model's predicted counterpart, through Monte Carlo simulation:

$$\widehat{Pr}(A; X, \widehat{\theta}) = \frac{1}{N \cdot M \cdot NS} \sum_{i=1}^{N} \sum_{m=1}^{M} \sum_{j=1}^{NS} 1(A_{ijm} = A; X_m, \widehat{\theta}), \tag{16}$$

where A_{ijm} is the simulated work history associated with teacher i and simulation j, as

[41] The number of states for ϵ was determined by increasing the number of points in the discretization of the error term until there was no change in the expected distribution of outcomes. For alternative approaches to estimating dynamic discrete choice models with serially-correlated errors, see Keane and Wolpin (1994) and Stinebrickner (2000).

[42] For example, the dependence of an error term for the first day of next month, 22 periods in the future, on the error term from the fifth day of the current month is ρ^{22}, or 0.00039 for $\rho = 0.7$.

[43] This is necessary since the maximum payoff a teacher could obtain varies across months with the length of the month and the number of holidays in that month, which count as a day worked in the bonus payoff function if they fall on a workday.

derived from a dynamic program constructed in accordance with the parameters $\widehat{\theta}$ and the characteristics of the month m. The number of simulations used to form the expected probability of observing a sequence of days worked is denoted as NS. In all of our estimations we use $NS = 200,000$. Note that we are also drawing ϵ_1 anew from the distribution $\bar{F}$ for every simulated path, where we keep track of the seeding values in the random number generator, as to ensure that the function value is always the same for a given $\widehat{\theta}$. The unconditional moments are:

$$E[Pr(A;X_m) - \widehat{Pr}(A;X_m,\widehat{\theta})] = 0, \tag{17}$$

The objective function under the method of simulated moments is:

$$\min_{\theta} g(X_i,\theta)'\Omega^{-1}g(X_i,\theta), \tag{18}$$

where $g(X_i,\theta)$ is the vector of moments formed by stacking the $2^N - 1$ moments defined by Equation 17, and Ω^{-1} is the standard two-step optimal weighting matrix. For more details concerning the implementation and asymptotic theory of simulation estimators, see McFadden (1989) and Pakes and Pollard (1989).

Matching sequences of days worked from the first N days in each month produces $2^N - 1$ linearly independent moments, where we subtract one to correct for the fact that the probabilities have to sum to one. In our estimation, we match sequences of length $N = 5$, which generates 31 moments. Experimentation with shorter and longer sequences of days worked did not result in significant changes to the coefficients.[44]

We also relax the assumption that the outside option is equal across all agents by allowing the outside option, μ_{im}, to be drawn anew for all teachers every month from a known parametric distribution $G(\mu)$. When forming moments in the MSM estimator in Equation 18 we need to integrate out this unobserved heterogeneity. The modification to the expected

[44] There is also a related econometric problem: the more moments one has to match, the lower the number of observations corresponding to each sequence. As the number of moments gets large, the number of teachers who actually followed any specific sequence diminishes towards zero. The number of days we match reflects a tradeoff in the additional information embodied in a longer sequence of choice behavior against the empirical imprecision of measuring those moments. This is a conceptually separate problem from the computational burden of simulating the model probabilities precisely, which also contributes to noisy estimates. For example, using the first 16 days of the months, where we will start seing some teachers "in the money," generates 65,535 moments, which is more than double the number of observations in the data.

probability of observing a sequence of days worked for teacher i is then:

$$\widehat{Pr}(A_i; X, \widehat{\theta}) = \frac{1}{M \cdot NS \cdot U} \sum_{m=1}^{M} \sum_{i=1}^{NS} \sum_{u=1}^{U} 1(A_{im} = A; X_m, \widehat{\theta}_1, u_{im}), \tag{19}$$

where u_{im} is a draw of the mean level of the outside option from $G(\widehat{\theta}_2)$, the unknown distribution of heterogeneity in the population. In practice we set $U = 200$. For clarity, we have partitioned the set of unknown parameters into $\widehat{\theta}_1 = \{\beta, \rho\}$ and $\widehat{\theta}_2$, the set of parameters governing the distribution of unobserved heterogeneity. Note that this model is slightly different than the fixed-effects model considered in the i.i.d. case above, as it allows the draw of the outside option to vary across both months and agents.

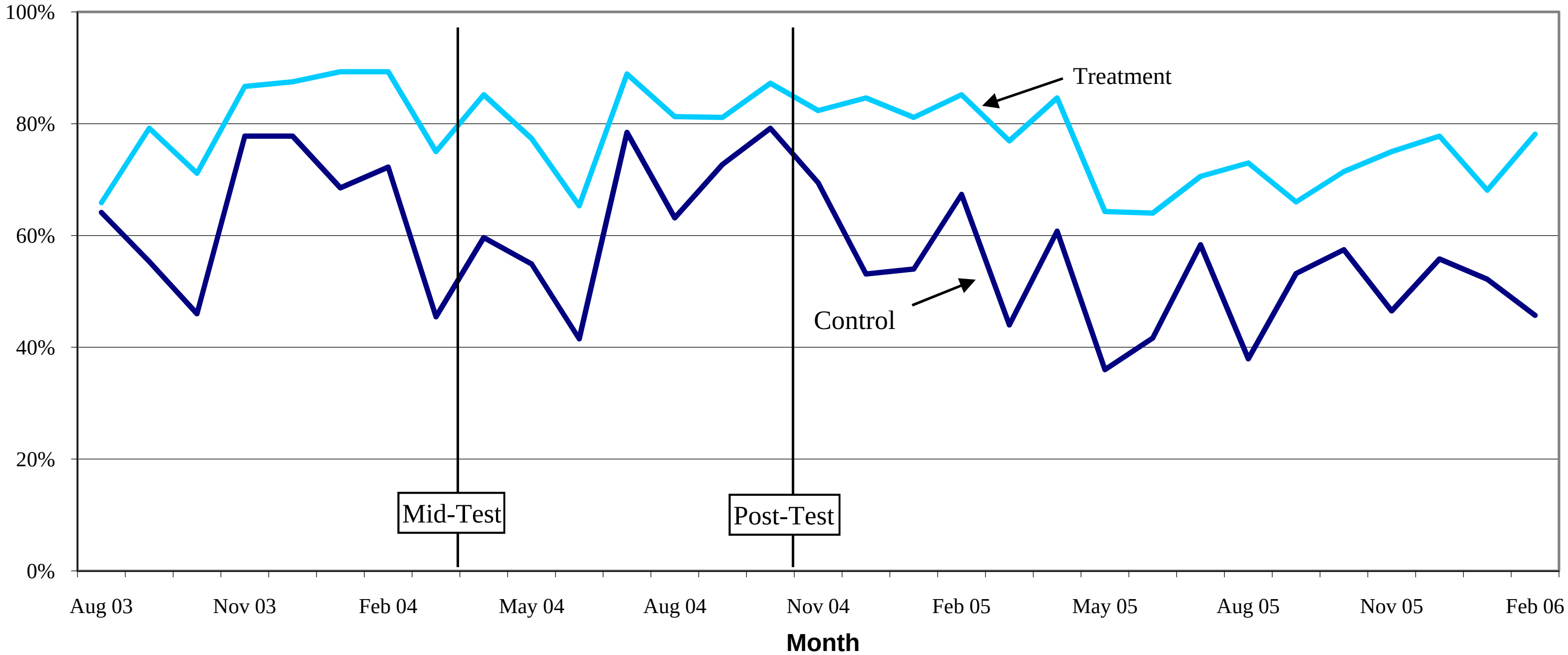
We estimate models with two types of unobserved heterogeneity which differ through the specification of $G(\widehat{\theta}_2)$. In the first model $G(\widehat{\theta}_2)$ is distributed normally with mean and variance $\widehat{\theta} = \{\mu_1, \sigma_1^2\}$. In the second model, our preferred specification, we allow for a mixture of two types, where each type is distributed normally with proportion p and $(1 - p)$ in the population.

Table 1: Baseline Data

	Treatment	Control	Difference
	(1)	(2)	(3)
A. Teacher Attendance			
School Open	0.66	0.64	0.02
			(0.11)
	41	39	80
B. Student Participation (Random Check)			
Number of Students Present	17.71	15.92	1.78
			(2.31)
	27	25	52
C. Teacher Qualifications			
Teacher Test Scores	34.99	33.62	1.37
			(2.01)
	53	56	109
D. Teacher Performance Measures (Random Check)			
Percentage of Children Sitting Within Classroom	0.83	0.84	0.00
			(0.09)
	27	25	52
Percent of Teachers Interacting with Students	0.78	0.72	0.06
			(0.12)
	27	25	52
Blackboards Utilized	0.85	0.89	-0.04
			(0.11)
	20	19	39
Fstat(1,110)			1.21
p-value			(0.27)
E. Baseline Test Scores			
Took Written Exam	0.17	0.19	-0.02
			(0.04)
	1136	1094	2230
Total Score on Oral Exam	-0.08	0.00	-0.08
			(0.07)
	940	888	1828
Total Score on Written Exam	0.16	0.00	0.16
			(0.19)
	196	206	402

Notes: (1) Teacher Performance Measures from Random Checks only includes schools that were open during the random check. (2) Children who could write were given a written exam. Children who could not write were given an oral exam. (3) Standard errors are clustered by school.

Figure 1: Percentage of Schools Open during Random Checks



Note: (1) The program began in September 2003. August only includes the 80 schools checked before announcement of program. September includes all random checks between August 25 through the end of September. (2) Child learning levels were assessed in a mid-test (April 2004) and a post-test (November 2004). After the post-test, the "official" evaluation period ended. Random checks continued in both the treatment and control schools.

Table 2: Teacher Attendance

Sept 2003-Feb 2006			Difference Between Treatment and Control Schools		
Treatment	Control	Diff	Until Mid-Test	Mid to Post Test	After Post Test
(1)	(2)	(3)	(4)	(5)	(6)
A. All Teachers					
0.79	0.58	0.21	0.20	0.20	0.23
		(0.03)	(0.04)	(0.04)	(0.04)
1575	1496	3071	882	660	1529
B. Teachers with Above Median Test Scores					
0.78	0.63	0.15	0.15	0.15	0.14
		(0.04)	(0.05)	(0.05)	(0.06)
843	702	1545	423	327	795
C. Teachers with Below Median Test Scores					
0.78	0.53	0.24	0.21	0.14	0.32
		(0.04)	(0.05)	(0.06)	(0.06)
625	757	1382	412	300	670

Notes: (1) Child learning levels were assessed in a mid-test (April 2004) and a post-test (November 2004). After the post-test, the "official" evaluation period was ended. Random checks continued in both the treatment and control schools. (2) Standard errors are clustered by school. (3) Panels B and C only include the 109 schools where teacher tests were available.

Figure 2: Impact of the Cameras
(out of at least 25 visits)

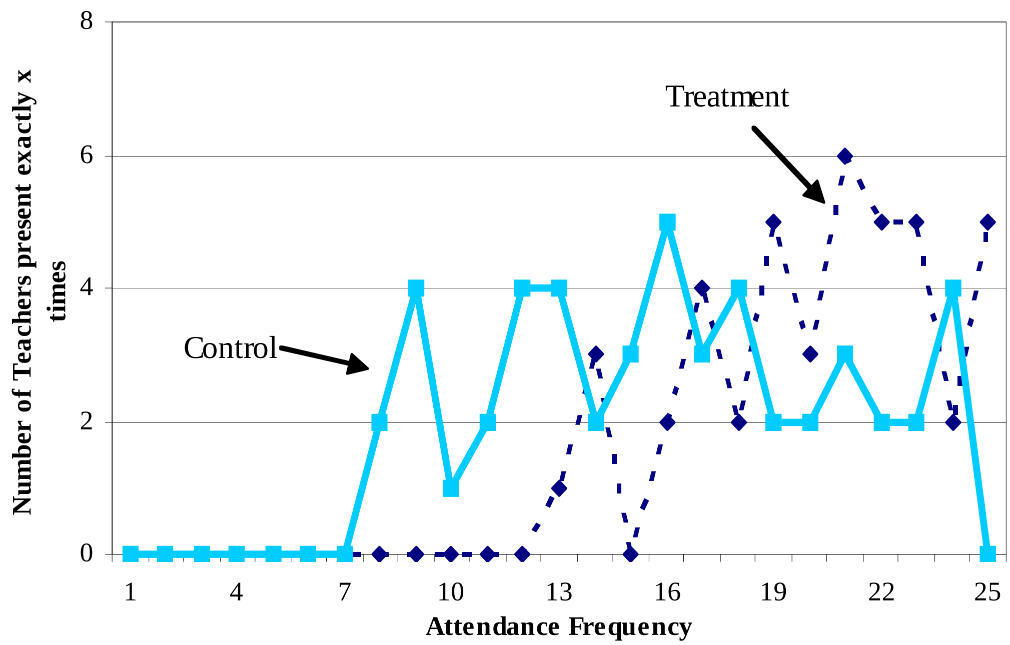
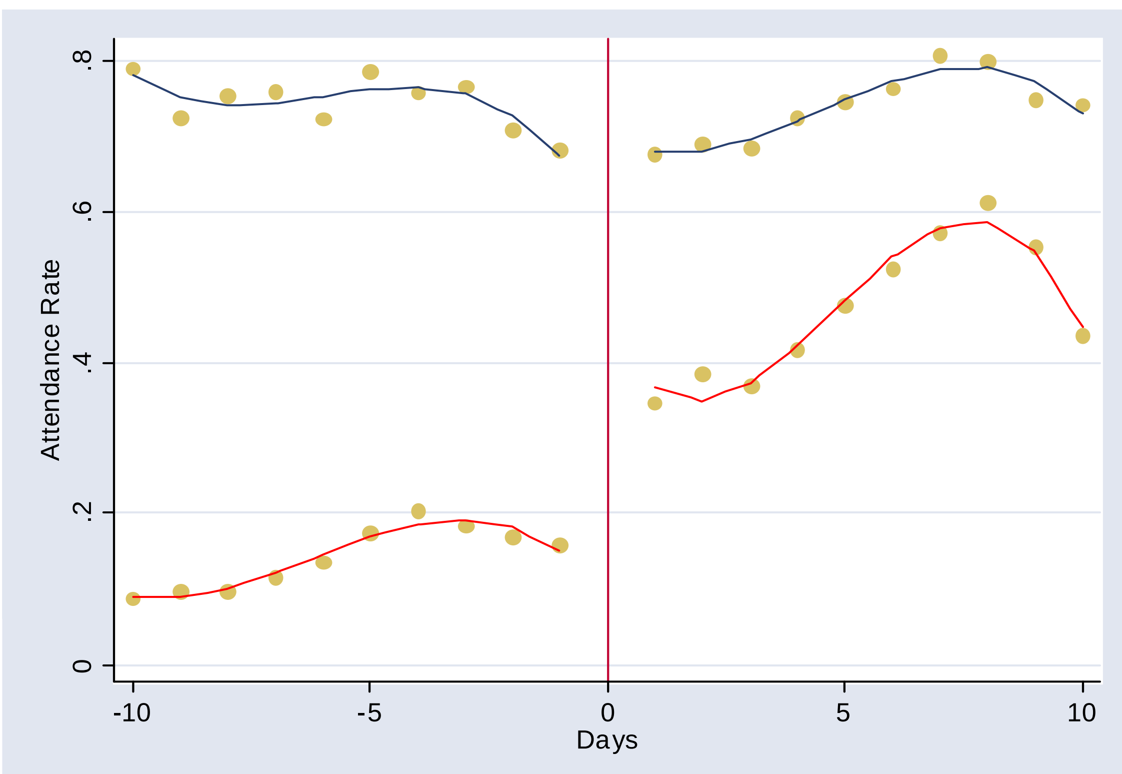


Table 3 : Do Teachers Work More When They are "In the Money"?

	(1)	(2)	(3)	(4)
Beginning of Month	0.19	0.12	0.46	0.39
	(0.05)	(0.06)	(0.04)	(0.03)
In the Money	0.52	0.37	0.6	0.48
	(0.04)	(0.05)	(0.03)	(0.01)
Beginning of the Month * In the Money	-0.19	-0.12	-0.34	-0.3
	(0.06)	(0.06)	(0.04)	(0.02)
Observations	2813	2813	27501	27501
R-squared	0.06	0.22	0.08	0.16
Sample	1st and last day of month	1st and last day of month	1st 10 and last 10 days of month	1st 10 and last 10 days of month
Third Order Polynomial on Days on each side			X	X
Teacher Fixed Effects		X		X
Month Fixed Effects		X		X
Clustered Standard Errors	X		X	

Note: (1) The dependent variable in all models is an indicator variable for whether the teacher worked on a particular day, as measured by the photographs for the treatment schools.

Figure 3: RDD Representation of Teacher Attendance at the Start and End of the Month



Note: (1) The top lines represent the months in which the teacher is in the money, while the bottom lines represent the months in which the teacher not in the money. (2) The estimation includes a third order polynomial of days on the left and right side of the change of month.

Table 4: Results from the Structural Model

Parameter	Model I (1)	Model II (2)	Model III (3)	Model IV (4)	Model V (5)	Model VI (6)	Model VII (7)	Model VIII (8)
β	0.049	0.027	0.055	0.057	0.013	0.017	0.017	0.016
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	(0.001)
μ_1	1.564		1.777	1.778	-0.428	-0.304	-0.160	-0.108
	(0.013)		(0.013)	(0.021)	(0.045)	(0.042)	(0.092)	(0.057)
ρ			0.422	0.412	0.449			
			(0.030)	(0.021)	(0.043)			
σ_1^2				0.043	0.007	0.252	0.418	0.235
				(0.012)	(0.019)	(0.015)	(0.052)	(0.028)
μ_2					1.781			
					(0.345)			
σ_2^2					0.050			
					(0.545)			
p					0.024			
					(0.007)			
Yesterday Shifter						0.094	0.024	0.095
						(0.010)	(0.009)	(0.014)
Attendance								-0.132
								(0.095)
Test Score								-0.005
								(0.002)
Heterogeneity	None	FE	None	RC	RC	RC	RC	RC
Three day Window	No	No	No	No	No	No	Yes	No
LLH	10269.13	9932.71				9286.03	3320.70	9287.33
ϵ_{Bonus}	1.09	0.592	1.299	1.82	0.196	0.298	0.279	0.283
	(0.147)	(0.062)	(0.123)	(0.136)	(0.053)	(0.026)	(0.038)	(0.064)
ϵ_{bonus_cutoff}	-18.26	-1.90	-16.94	-14.07	-0.14	-0.074	-0.454	-0.100
	(2.023)	(0.564)	(0.889)	(1.609)	(0.144)	(0.050)	(0.252)	(0.137)
Predicted Days Worked	17.23	17.30	16.87	16.28	16.75	18.381	17.596	18.213
	(0.361)	(0.153)	(0.260)	(0.566)	(0.391)	(0.391)	(0.809)	(0.974)
Days Worked BONUS=0	1.31	6.96	1.35	1.174	12.90	9.774	11.314	10.605
	(0.041)	(0.101)	(0.049)	(0.072)	(0.281)	(0.605)	(0.916)	(1.454)
Out of Sample Prediction	21.47	19.975	21.48	21.550	17.77	20.157	19.281	19.948
	(0.046)	(0.164)	(0.030)	(0.060)	(0.479)	(0.287)	(0.753)	(0.678)

Note: Models I, II, VI, VII, and VIII are estimated using maximum likelihood. Models III, IV, and V are estimated using the method of simulated moments with an optimal weighting matrix. We report the elasticity of days worked with respect to the bonus, ϵBonus, and the semi-elasticity with respect to a bonus cutoff, ϵBonus_cutoff. The last three rows report the expected number of days worked under the original incentives, a counterfactual where BONUS=0, and the second set of financial incentives

Figure 4A: Predicted Fit From Model V

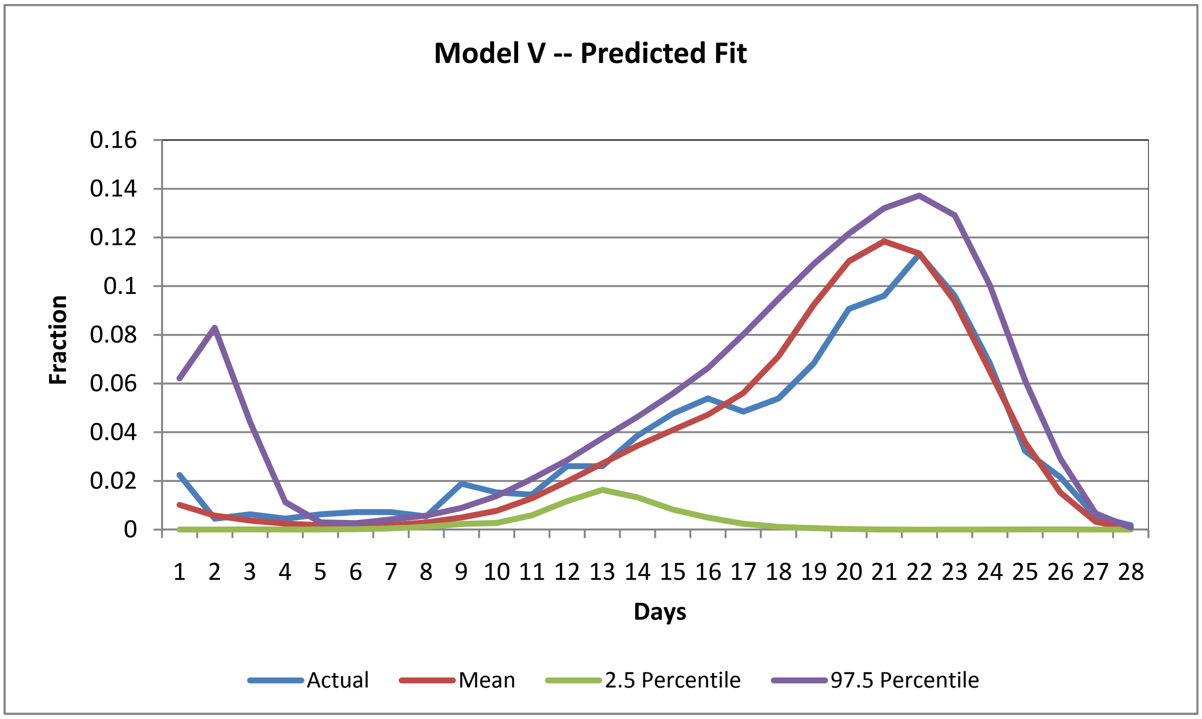


Figure 4B: CounterFactual Fit From Model V

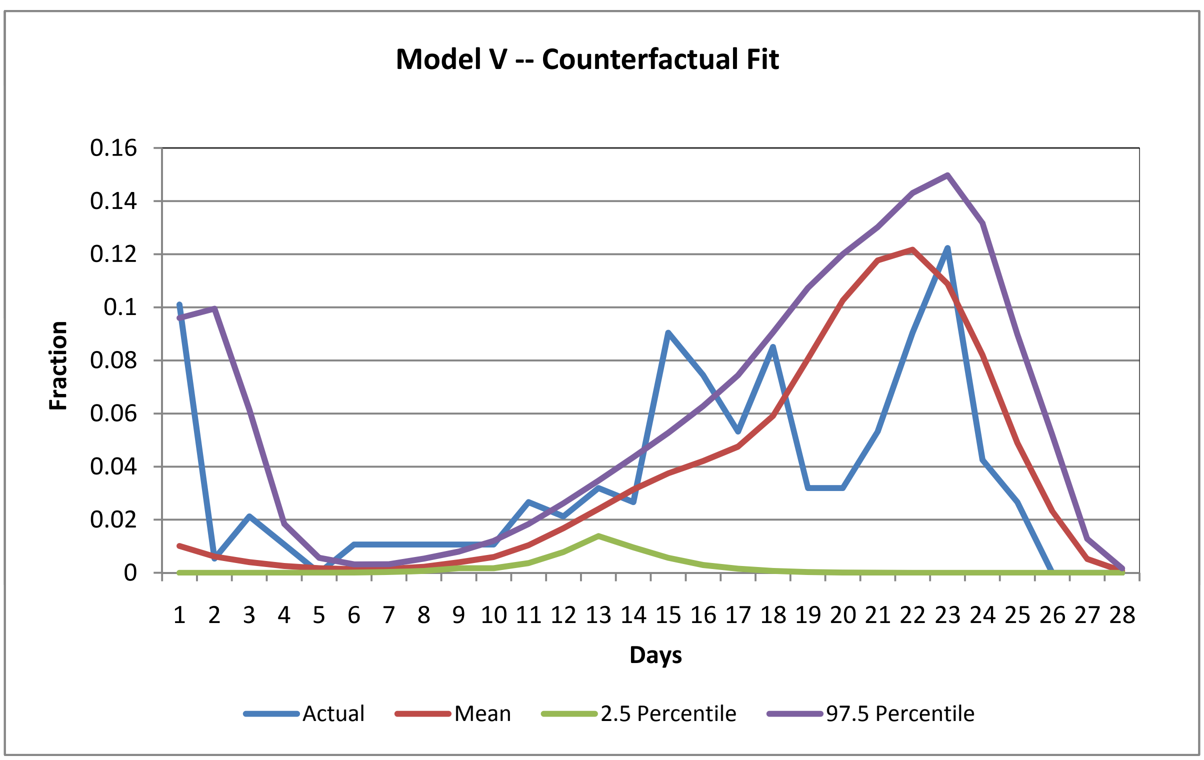


Table 5: Counterfactual Cost-Minimizing Policies

Expected Days Worked (1)	Bonus Cutoff (2)	Bonus (3)	Expected Cost (4)	Test Score Gain over Control Group (13 days) (5)
14	0	0	500	0.04
15	21	25	521	0.07
16	22	75	664	0.11
17	21	75	672	0.15
18	20	75	755	0.18
19	20	100	921	0.22
20	20	125	1112	0.26
21	16	225	2642	0.29
22	11	275	4604	0.33

Table 6: Teacher Performance

	Sept 2003-Feb 2006			Difference Between Treatment and Control Schools		
	Treatment (1)	Control (2)	Diff (3)	Until Mid-Test (4)	Mid to Post Test (5)	After Post Test (6)
Percent of Children Sitting Within Classroom	0.72	0.73	-0.01	0.01	0.04	-0.01
			(0.01)	(0.89)	(0.03)	(0.02)
	1239	867	2106	643	480	983
Percent of Teachers Interacting with Students	0.55	0.57	-0.02	-0.02	0.05	-0.04
			(0.02)	(0.04)	(0.05)	(0.03)
	1239	867	2106	643	480	983
Blackboards Utilized	0.92	0.93	-0.01	-256766.00	0.01	-0.01
			(0.01)	(0.02)	(0.02)	(0.02)
	990	708	1698	613	472	613

Notes: (1) Teacher Performance Measures from Random Checks only includes schools that were open during the random check. (2) Standard errors are clustered by school.

Table 7: Child Attendance

	Sept 03-Feb 06			Difference Between Treatment and Control Schools		
	Treatment	Control	Diff	Until Mid-Test	Mid to Post Test	After Post Test
	(1)	(2)	(3)	(4)	(5)	(6)
A. Attendance Conditional on School Open						
Attendance of Students Present at Pre-Test Exam	0.46	0.46	0.01	0.02	0.03	0.00
			(0.03)	(0.03)	(0.04)	(0.03)
	23495	16280	39775			
Attendance for Children who did not leave NFE	0.62	0.58	0.04	0.02	0.04	0.05
			(0.03)	(0.03)	(0.04)	(0.03)
	12956	10737	23693			
B. Total Instruction Time (Presence)						
Presence for Students Present at Pre-Test Exam	0.37	0.28	0.09	0.10	0.10	0.08
			(0.03)	(0.03)	(0.04)	(0.03)
	29489	26695	56184			
Presence for Student who did not leave NFE	0.50	0.36	0.13	0.10	0.13	0.15
			(0.03)	(0.04)	(0.05)	(0.04)
	16274	17247	33521			
C. Presence, by Student Learning Level at Program Start (for those who did not leave)						
Took Oral Pre-Test	0.50	0.36	0.14	0.11	0.14	0.15
			(0.03)	(0.03)	(0.05)	(0.04)
	14778	14335	29113			
Took Written Pre-Test	0.48	0.39	0.10	0.07	0.07	0.11
			(0.06)	(0.07)	(0.06)	(0.07)
	1496	2912	4408			

Notes: (1) Standard errors are clustered at the level of the school. (2) Child attendance data were collected during random checks. (3) The attendance at the pre-test exam determined the child enrollment at the start of the program.

Table 8: Descriptive Statistics for Mid Test and Post Test

	Mid Test			Post Test		
	Treatment	Control	Difference	Treatment	Control	Difference
	A. Attrition Process					
Percent Attrition	0.11	0.22	-0.10	0.24	0.21	0.03
			(0.05)			(0.04)
Difference in Percent Written of Pre-Test attriters-stayers	0.01	0.03	0.02	0.06	-0.03	0.10
			(0.06)			(0.06)
Difference in Verbal Test of Pre-Test attriters-stayers	0.05	0.08	-0.03	0.02	0.12	-0.10
			(0.14)			(0.14)
Difference in Written Test of Pre-Test attriters-stayers	-0.41	-0.23	-0.18	-0.19	-0.13	-0.06
			(0.34)			(0.29)
	B. Exam Score Means					
Took Written	0.36	0.33	0.03	0.61	0.57	0.04
			(0.04)			(0.05)
Math	0.14	0.00	0.14	-0.08	-0.24	0.16
			(0.10)			(0.15)
Language	0.14	0.00	0.14	1.71	1.60	0.11
			(0.10)			(0.11)
Total	0.14	0.00	0.14	0.35	0.24	0.12
			(0.10)			(0.11)

Notes: (1) Test Scores in Panel B are normalized by the mean of the mid-test control. (2) Standard Errors are clustered by school.

Table 9: Estimation of Treatment Effects for the Mid- and Post-Test

Mid-Test				Post-Test			
Took Written	Math	Lang	Total	Took Written	Math	Lang	Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			A. All Children				
0.04	0.15	0.16	0.17	0.06	0.21	0.16	0.17
(0.03)	(0.07)	(0.06)	(0.06)	(0.04)	(0.12)	(0.08)	(0.09)
1893	1893	1893	1893	1760	1760	1760	1760
			B. With Controls				
0.02	0.13	0.13	0.14	0.05	0.17	0.13	0.15
(0.03)	(0.07)	(0.05)	(0.06)	(0.04)	(0.10)	(0.07)	(0.07)
1893	1893	1893	1893	1760	1760	1760	1760
			C. Took Pre-Test Oral				
	0.14	0.13	0.15		0.2	0.13	0.16
	(0.08)	(0.06)	(0.07)		(0.14)	(0.09)	(0.10)
	1550	1550	1550		1454	1454	1454
			D. Took Pre-Test Written				
	0.19	0.28	0.25		0.28	0.28	0.25
	(0.12)	(0.11)	(0.11)		(0.18)	(0.11)	(0.12)
	343	343	343		306	306	306
			E. Girls				
0.07	0.18	0.18	0.2	0.07	0.22	0.17	0.18
(0.03)	(0.07)	(0.07)	(0.07)	(0.05)	(0.12)	(0.09)	(0.09)
891	891	891	891	821	821	821	821
			F. Boys				
0.02	0.12	0.14	0.14	0.05	0.19	0.16	0.16
(0.04)	(0.09)	(0.07)	(0.07)	(0.04)	(0.15)	(0.10)	(0.10)
988	988	988	988	929	929	929	929

Notes: (1) The table presents the coefficient estimate of being in a treated school on the sum of a child's score on the oral and written exams. All regressions include controls for the child's learning levels prior to the program. (2) The mid and post test scores are normalized by mid test control group. (3) Controls in Row B include Block, Teacher Test Scores, and Infrastructure Index. (4) Standard errors are clustered by school.

Table 10: Dropouts and Movement into Government Schools

	Treatment (1)	Control (2)	Diff (3)
Child Left NFE	0.44	0.36	0.08
			(0.04)
Child Enrolled in Government School	0.26	0.16	0.10
			(0.03)
Child Dropped Out of School	0.18	0.20	-0.02
			(0.03)
N	1136	1061	2197

Notes: (1) Standard errors are clustered at the level of the school. (2) Dropouts are defined as those who were absent for the last five random checks in which a school was found open.

Table 11: Does the Random Check Predict Test Scores?

Method:	OLS	OLS	OLS	2SLS
Sample:	Control Schools	Treatment Schools	Treatment Schools	All Schools
Data:	Random Check	Random Check	Photographs	Random Check
	(1)	(2)	(3)	(4)
		A. Mid-test (Sept 03-April 04)		
Took Written	0.02	0.28	0.36	0.26
	(0.10)	(0.08)	(0.11)	(0.19)
Total Score	0.20	0.39	0.87	1.07
	(0.19)	(0.21)	(0.22)	(0.43)
N	878	1015	1015	1893
		B. Post-test (Sept 03 -Oct 04)		
Took Written	0.24	0.51	0.59	0.33
	(0.16)	(0.15)	(0.20)	(0.22)
Total Score	0.58	1.17	0.98	0.97
	(0.35)	(0.36)	(0.53)	(0.47)
N	883	877	877	1760

Notes: (1) The table presents the coefficient estimate of the teacher's attendance on the sum of a child's score on the oral and written exams. All regressions include controls for the child's learning levels prior to the program. (2) The mid and post test scores are normalized by the mid test control group. (3) Standard errors are clustered by school.

Appendix Table 1: Empirical Sequences of Days Worked in the Last Five Days of a Month

In the Money	Sequence of Days	Frequency	Percent	Number of Days Worked
(1)	(2)	(3)	(4)	(5)
No	00000	38	.4367816	0
No	01000	1	.0114943	1
No	10000	1	.0114943	1
No	00010	2	.0229885	1
No	00101	1	.0114943	2
No	00110	1	.0114943	2
No	01001	1	.0114943	2
No	00011	3	.0344828	2
No	11000	4	.045977	2
No	10011	1	.0114943	3
No	11010	1	.0114943	3
No	01101	1	.0114943	3
No	01011	1	.0114943	3
No	10101	1	.0114943	3
No	11100	2	.0229885	3
No	10110	2	.0229885	3
No	00111	3	.0344828	3
No	11001	3	.0344828	3
No	10111	2	.0229885	4
No	01111	3	.0344828	4
No	11110	4	.045977	4
No	11111	9	.1034483	5
Yes	00000	20	.019305	0
Yes	10000	2	.0019305	1
Yes	01000	2	.0019305	1
Yes	00010	3	.0028958	1
Yes	00100	7	.0067568	1
Yes	00001	9	.0086873	1
Yes	10010	2	.0019305	2
Yes	10001	2	.0019305	2
Yes	10100	2	.0019305	2
Yes	01010	3	.0028958	2
Yes	01001	4	.003861	2
Yes	00101	5	.0048263	2
Yes	00110	5	.0048263	2
Yes	01100	6	.0057915	2
Yes	11000	9	.0086873	2
Yes	00011	26	.0250965	2
Yes	10110	3	.0028958	3
Yes	10101	4	.003861	3
Yes	01101	5	.0048263	3
Yes	11010	5	.0048263	3
Yes	01110	6	.0057915	3
Yes	10011	8	.007722	3
Yes	01011	14	.0135135	3
Yes	11001	19	.0183398	3
Yes	11100	31	.0299228	3
Yes	00111	48	.046332	3
Yes	10111	23	.0222008	4
Yes	11110	48	.046332	4
Yes	11011	49	.0472973	4
Yes	11101	55	.0530888	4
Yes	01111	74	.0714286	4
Yes	11111	537	.5183398	5

Appendix Table 2: Fitted Moments

Sequence	Empirical	Fitted
(1)	(2)	(3)
0 0 0 0 0	0.062	0.066
0 0 0 0 1	0.009	0.01
0 0 0 1 0	0.004	0.006
0 0 0 1 1	0.021	0.016
0 0 1 0 0	0.004	0.004
0 0 1 0 1	0.001	0.003
0 0 1 1 0	0.006	0.003
0 0 1 1 1	0.045	0.024
0 1 0 0 0	0.005	0.006
0 1 0 0 1	0.005	0.004
0 1 0 1 0	0.001	0.002
0 1 0 1 1	0.012	0.008
0 1 1 0 0	0.007	0.005
0 1 1 0 1	0.012	0.008
0 1 1 1 0	0.008	0.008
0 1 1 1 1	0.105	0.058
1 0 0 0 0	0.015	0.008
1 0 0 0 1	0.005	0.01
1 0 0 1 0	0.002	0.004
1 0 0 1 1	0.018	0.02
1 0 1 0 0	0.002	0.004
1 0 1 0 1	0.002	0.006
1 0 1 1 0	0.006	0.006
1 0 1 1 1	0.042	0.041
1 1 0 0 0	0.015	0.013
1 1 0 0 1	0.016	0.016
1 1 0 1 0	0.009	0.008
1 1 0 1 1	0.048	0.038
1 1 1 0 0	0.011	0.023
1 1 1 0 1	0.036	0.039
1 1 1 1 0	0.055	0.064

Appendix Table 3: Alternative Specifications

Parameter	Appendix Model I	Appendix Model II
	(1)	(2)
β	0.043	0.016
	(0.002)	(0.000)
μ_1	1.236	-1.166
	(0.094)	(0.011)
σ_1^2	0.143	
	(0.039)	
μ_2	1.870	
	(0.672)	
σ_2^2	2.051	
	(0.305)	
p	0.003	
	(0.002)	
Yesterday Shifter		1.145
		(0.017)
Heterogeneity	RC	None
LLH		9654.887
ϵ_{Bonus}	1.371	0.722
	(0.436)	(0.075)
ϵ_{bonus_cutoff}	-6.06	-0.058
	(3.635)	(0.028)
Predicted Days Worked	14.98	17.264
	(1.541)	(0.360)
Days Worked BONUS=0	2.63	1.992
	(0.362)	(0.086)
Out of Sample Prediction	19.83	20.566
	(1.725)	(0.117)

Note: Both models are estimated using maximum likelihood.

Appendix Table 4: Model VI Cost-Minimization Policies

Days Worked (1)	Bonus Cutoff (2)	Bonus (3)	Cost (4)
14	13	25	642
15	21	75	686
16	18	50	698
17	20	75	782
18	19	75	877
19	23	175	967
20	17	75	1051
21	19	100	1134
22	21	200	1515