The Welfare Effects of Incentive Schemes *

Adam Copeland[†]   Cyril Monnet[‡]

May 6, 2002

Abstract

This paper computes the change in welfare associated with the introduction of incentives. Specifically, we calculate by how much the welfare gains of increased output due to incentives outweigh workers' disutility from increased effort. We accomplish this by studying the use of incentives by a firm in the check-clearing industry. Using this firm's production records, we model and estimate the worker's dynamic effort decision problem. This allows us to determine that the firm's incentive scheme has a large effect on productivity, raising it by 16% over the sample period. Using our parameter estimates, we show that the cost of increased effort due to incentives is equal to the dollar value of an 11% rise in productivity. Welfare is measured as the output produced minus the cost of effort, hence the net increase in welfare due to the introduction of the firm's bonus plan is 5%.

1

1 Introduction

Incentives are often used by firms to encourage their employees to work hard. The bonus plans used vary widely, from complicated stock option offerings to simple employee-of-the-month awards. But how well do incentive plans work? Most of the empirical work in this literature addresses this issue by asking how much measured output changes due to the introduction of incentives. Lazear (2000), for example, estimates that productivity jumps 22% due to the incentive effects of a piece-rate compensation scheme at a windshield repair firm. Little attention, however, is paid in quantifying the costs of this extra output. In other words, by how much do the welfare gains of increased output due to incentives outweigh workers' disutility from increased effort? This is an important question, as by not taking the cost of effort into account, we may overstate the importance of incentives.

Our paper addresses this issue by studying the use of incentives by the Check Department of the Federal Reserve Bank of Minneapolis. Given this firm's production records, we develop, solve and estimate a dynamic model of worker behavior. Using these estimates, we measure by how much output and the disutility from effort increase due to the firm's incentive pay scheme. Comparing these welfare changes lets us compute how much the firm's bonus scheme raises overall welfare.

This case study is particularly interesting due to some special properties of the firm's incentive pay scheme that allow us to identify effort's effect on output and the cost of effort. Because of its structure, this bonus system creates a dynamic effect in the worker's problem. The firm's bonus scheme is designed so that employees are only eligible for incentive pay if their *daily* productivity is above a threshold level. Conditional on being eligible, bonus pay is an increasing function of the distance between the worker's productivity and the threshold level. For any level of productivity below the threshold, workers simply earn zero bonus pay. This kink in the bonus profile creates the perverse incentive for a worker to quit working hard in the later part of the shift if the worker's measure of productivity is low (due, for example, to a bad shock) in the early part of the shift. Under this bonus scheme then, the worker's history within the day, and expectation over the probability of the daily bonus pay, affect the worker's effort decision.

This dynamic aspect of the worker's problem is crucial in our model as it is the main source of identification. Theory provides us with the intuitive result that a worker with a small probability of earning incentive pay chooses a lower level of effort compared to the case where the worker has a high probability. Hence, we know that after experienc-

ing a particularly unproductive morning, a worker's chances of beating the firm's daily productivity threshold, and so being eligible for incentive pay, are small. Consequently, the worker will exert a low level of effort. In contrast, if this worker had a productive morning, the chances of earning a bonus are large enough that a high level of effort is exerted. Knowing that a worker's effort level differs under these two circumstances enables us to identify both effort's effect on output and the cost of effort to the worker. We can identify that effort is costly, as workers with a low probability of earning incentive pay will have an unexplained drop in productivity compared to those workers with a high probability. In addition, we can identify effort's effect on output by comparing the difference in productivity between workers with low and high probabilities of earning a bonus.

We are able to exploit this method of identification as the firm has provided with us a detailed data set on worker productivity. We have the firm's production records for 15 full-time, experienced workers over a 15 month time period. These records are at such a fine level of detail that we can track a worker's productivity *within* the day. The data also contain information on a large number of characteristics of the sorted checks. In addition, the firm provided information on its incentive plan, allowing us to measure how well a worker is performing relative to the firm's benchmark productivity level. As such, we can compute a worker's chances of earning a bonus throughout a worker's shift and so take advantage of the avenue of identification described above.

We model the worker's problem as having to make a number of effort decisions within a day, where at the end of the shift the incentive pay formula computes the worker's bonus. We assume effort is a binary variable, and so allow workers to choose a low, costless effort level or a high, costly one. When making an effort decision, the worker knows the past history of events as well as the number of checks left to sort in the day. This information allows the worker to determine how well the worker has performed relative to the incentive pay scheme, and to compute the probability the worker will be eligible for incentive pay at the end of the day. As the bonuses are calculated on a daily basis, a worker starts each shift anew. In this framework, determining the effect of effort on a worker's expected bonus is complicated. A higher level of effort causes both a rise in the level of a worker's expected bonus as well as in the probability of earning that bonus. To simplify this analysis of the worker's policy function, we consider the two special cases where the probability of earning incentive pay is zero and one. We then prove the intuitive result that the level of effort given the probability of earning a bonus

3

is one is greater than that worker's level of effort given the zero probability case.

Using this model, we generate two kinds of results. First, we use non-parametric techniques to test if implications of the model hold in the data. Second, we use a maximum likelihood approach to estimate the model's structural parameters.

We test the prediction of the model that a worker's effort level differs when the probability of earning a bonus is one versus zero. To see if this relationship holds in the data, we test whether the unexplained portion of worker productivity is higher when the worker's probability of earning incentive is close to one, compared to the case when the probability is close to zero. In support of our model, we find the difference between these two groups of productivity residuals is large and statistically significant.

We then turn to estimating the structural parameters of the model. Our estimation method is the maximum likelihood approach, common in the discrete choice estimation literature. This involves solving the worker's problem for a given set of parameters to obtain the worker's policy function. Next, using these policy rules and the data, the model generates a distribution of time taken to sort checks, which we use to compute the model's likelihood.

The parameter estimates from our estimation procedure show that effort significantly affects worker output. Over the sample period, the firm's bonus plan increased worker productivity by 16%. Using our estimate of the cost of effort, we also compute the dollar value of the disutility of workers due to the additional effort they exert under the incentive scheme. We find that this utility cost to workers is equal to a 11% gain in productivity, roughly two-thirds of the gain in output due to incentives. As welfare is measured as the amount of output produced minus the cost of effort, our results show the net increase in welfare from the firm's bonus plan is 5%.

As previously mentioned, most of the existing empirical work on incentives has focused on quantifying how much output is affected by the introduction of incentives. The results from this field of research vary widely, attributing an increase in productivity from zero to over twenty percent due to the use of incentives[1]. In this literature, our work is closest to Ferrall and Shearer (1999) in that we both use the structural estimation approach[2]. They estimate the structural parameters of a principal-agent model,

---

[1]Lazear (2000) and Paarsch and Shearer (2000) are two recent papers that quantify the relationship between output and incentives. Blinder (1990) and Prendergast (1999) are good surveys on this literature.

[2]Margiotta and Miller (2000) also uses the structural estimation approach to measure the cost of moral hazard at the senior managerial level.

using the payroll records from the 1920's of a mining firm. With a static model of effort, they analyze the classic tension between risk averse workers and the firm's use of incentives. We, on the other hand, develop a dynamic model of effort and abstract away from risk aversion. Assuming risk neutrality simplifies the computation of the worker's policy function. In addition, while Ferrall and Shearer study a biweekly bonus scheme, the incentive scheme we analyze is daily and only in effect for part of a worker's shift. As such, given that the gambles workers take are relatively small, using risk neutrality provides us with a plausible approximation of worker behavior.

Two other papers related to our work are Asch (1990) and Oyer (1998). These two papers describe data on the behavior of Navy recruiters and salespeople, respectively, under incentive schemes. They both show that employees significantly alter their behavior during a pay period, conditional on their history of events and future expectations. Hence, they find that dynamic models of worker behavior, such as the one we use, are significant improvements over their static counterparts. Our analysis also points to the importance of dynamic models.

The rest of our paper is organized as follows. Section 2 describes the data. Section 3 lays out the model and derives the worker's problem. Section 4 describes the estimation procedure and reports results. Section 5 concludes.

2   Data Description

In this section, we describe our data set. We first explain the nature of the check-sorting job and in which set of workers we are interested. We then describe the firm's incentive pay scheme and summarize workers' performance under it. Finally, we conclude by describing the firm's goals.

Our data comes from the production and human resource records of the Check Department of the Federal Reserve Bank of Minneapolis. This firm provided us with information over a 15 month period (3/01/99 - 5/27/00) on its employees that sort checks in the Low Speed Check Processing Department[3]. While these workers have several responsibilities, their main task is to sort checks by running them through a sorting machine[4].

---

[3]This Department processes all checks that are 'rejected' from the High Speed Check Processing Department. Checks are first processed by the High Speed Department. But if the High Speed sorting machines have any difficulty processing a check, that check is immediately diverted to the Low Speed Department for further processing.

[4]A picture of this machine is attached at the end of the paper.

Ideally, this sorting machine would process checks without any worker input. However, two events occur that require worker interaction: checks get jammed in the machine, and fields on the check cannot be electronically read by the machine[5]. In the first instance, workers clear the jammed checks and reset the sorting machine. In the second, workers type in the field which the machine failed to read. Checks are processed by workers in batches. We define a *job* as a batch of checks that needs to sorted. The production data we received from the firm is at the job level and includes information on which worker ran a job, what time it was run, how long it took, and its characteristics (e.g. the number of jams that occurred).

The human resource component of the data set provides us with information about the tenure of the worker, as well as the worker's wage-grade level. Using this information, we selected those workers who were full-time, were employed in the check-sorting department for at least 6 months before the beginning of our sample period, and continued working throughout the sample period. Excluding new workers allows us to ignore the effects of learning-by-doing, simplifying our analysis. Of the original 52 workers, only 15 met these three requirements. However, this group of workers completed 34,077 jobs in the data set, which account for roughly half of all checks sorted in the 15 month sample period.

Looking at this subset of the data, we find there is a wide range in the number of jobs a worker completed within a day (1-90) as well as the number of checks sorted by job (1-9000). Typically however, workers run 9 jobs a day, where each one averages 660 checks in length and takes 18 minutes to complete. In addition, workers normally clear 9.3 jams and type in 128 fields per job. As the large number of jams and field corrections indicate, worker input is a large determinant of how fast a job is completed.

The firm uses an incentive pay arrangement that rewards workers based on their daily performance above and beyond their hourly wage. This mechanism works by using a formula that provides a benchmark time for each worker, given the characteristics of the jobs the worker ran that day. Some characteristics that the firm uses are the number of checks sorted, the number of jams cleared, and the number of fields manually typed in. Letting $\bar{z}$ be a vector of C characteristics of a job, we denote this formula as $\alpha(\bar{z})$. This formula is linear in job characteristics, having the following form:

$$\alpha(\bar{z}) = \rho_0 + \rho_1 \cdot z_1 + \ldots + \rho_C \cdot z_C. \tag{1}$$

---

[5]An example of a field is the account number

If a worker completes N jobs in a day, and $\bar{z}_n$ denotes the characteristics of job n, then $\sum_{n=1}^{N} \alpha(\bar{z}_n)$ determines the overall time that a worker needs to beat in order to earn any bonus pay. Conditional on achieving this, the amount of bonus pay a worker receives is an increasing function of the difference between the worker's actual and benchmark time. Let τ_n be the actual time a worker spends on a job and define $s = \sum_{n=1}^{N} [\alpha(\bar{z}_n) - \tau_n]$. The variable s is the amount of time a worker is behind or ahead the benchmark time at the end of the day. Note that s is a function of all the jobs a worker completed in a day. We can then write the firm's bonus payment scheme as

$$\tilde{b}(s) = \begin{cases} 0 & \text{if } s \leq 0 \\ K \cdot s & \text{otherwise} \end{cases} \tag{2}$$

where $K > 0$ is some constant. So given $s > 0$, the bonus amount that a worker earns is a function of the total time s, in hours, by which the worker beat the benchmark formula, multiplied by a wage K. The firm has provided us with the constant K as well as its incentive pay formula, α, which enables us to reproduce the daily cutoff times each worker faced.

An important aspect of this data set is a change in the incentive pay formula in January of 2000, roughly two-thirds of the way through the sample period. Before the switch, $K = \$7.17$ for all employees. After the switch, $K = \$9.75$ for workers with a grade of 4 or 5 and $K = \$12.76$ for those with a grade of 6 or 7.[6] In addition, the parameters of the formula α and the set of job characteristics $\bar{z}$, were changed in order to raise the threshold level of productivity. So the switch in the incentive scheme involved two opposing effects: on one hand, it became harder to earn incentive pay, while on the other hand, bonuses were potentially bigger as K was increased.

Table 1 contains the average bonus payments for workers under the two regimes, as well as the percentage of total jobs where a positive incentive amount was earned. This table demonstrates that after the switch, workers earned smaller bonus payments less often. Notice that the average daily bonus of all workers decreased from $10.06 to $6.64. In addition, the percent of days where a worker earned a bonus fell from 95% to 81%.

Table 1 also shows the large amount of heterogeneity in worker productivity. Worker 5 is one of the most productive workers in the sample, earning an average bonus of over $16 under both bonus regimes. Also, this worker *always* earned a bonus. At the other

---

[6]Lower grade employees earn lower wages and are typically newer employees relative to their higher grade counterparts

Table 1: Bonus Payments

Worker	Grade	First IP Regime			Second IP Regime		
		Mean	(Std)	Prct	Mean	(Std)	Prct
All	n/a	$10.06	(8.54)	0.95	$6.64	(7.62)	0.81
High grade	6 & 7	$10.85	(9.13)	0.96	$8.94	(8.75)	0.86
Low grade	4 & 5	$8.50	(6.94)	0.92	$3.47	(4.26)	0.74
1	7	$14.60	(12.43)	1.00	$11.19	(11.30)	0.96
2	7	$8.52	(7.96)	0.95	$7.28	(6.33)	0.93
3	7	$9.74	(11.33)	0.97	$10.50	(8.80)	0.92
4	6	$14.29	(10.03)	0.98	$14.04	(9.32)	0.99
5	6	$18.22	(6.48)	1.00	$16.13	(5.33)	1.00
6	6	$15.23	(8.94)	0.98	$12.59	(8.88)	1.00
7	6	$9.14	(4.78)	1.00	$2.75	(2.80)	0.72
8	6	$6.10	(6.33)	0.93	$5.44	(5.78)	0.82
9	6	$5.45	(3.98)	0.84	$0.22	(0.74)	0.12
10	6	$6.95	(6.38)	0.95	$4.59	(4.58)	0.85
11	5	$11.05	(8.14)	0.98	$4.67	(4.63)	0.92
12	5	$5.35	(6.02)	0.80	$4.64	(3.31)	0.89
13	5	$14.71	(6.24)	1.00	$5.30	(5.37)	0.85
14	4	$5.27	(3.39)	0.90	$0.54	(1.07)	0.33
15	4	$5.40	(3.42)	0.95	$1.86	(2.32)	0.76

Prct: Percentage of days with positive bonus pay

8

extreme, Worker 9 received a mean bonus of $5.45 under the first regime and $0.22 under the second. Under second regime, this worker earned a bonus only 12% of the time. Workers also differ in how they reacted to the change in incentives. Workers 3 and 7 both earned an average bonus payment of roughly $9 a day under the first incentive pay regime. Under the second regime however, Worker 3's average bonus increased to $10.50 while Worker 7's dropped to $2.75.

How important, though, are these bonus payments to workers? To check this, we computed the mean of the ratio of bonus pay to total wages. As these workers performed tasks throughout the day not under the purview of the incentive scheme, the focus of this ratio is restricted to time spent on sorting machines. Total wages are thus computed as the time spent on a sorter multiplied by the worker's hourly wage plus the bonus amount.[7] The mean value over all workers for this ratio was 0.25 under the first incentive scheme and 0.18 for the second. The mean value by worker ranged from 0.10 to 0.34 and 0.003 to 0.33 for the first and second incentive schemes respectively. These results suggest that when looking at the time spent sorting checks, incentives are indeed significant to employees.

While incentives are beneficial in that they increase worker productivity, they also have the disadvantage of adversely affecting the quality of output. In this check-sorting environment, workers, in an attempt to increase productivity, could decrease quality by entering incorrect numbers for the fields that cannot be electronically scanned. However, there is another department within the firm that is able to detect when incorrect field numbers have been entered. When a mistake is found, an analyst goes through the records to fix the error and find the responsible worker. Due to the time this error-checking process takes, the firm views such mistakes seriously. A worker who makes such an error is docked pay, and repeat offenders are fired. By reviewing the payroll records for workers over the sample period, we found that workers rarely made these quality errors, and no worker continually made them over time. As such, in this paper we do not model a quality trade-off.

3   The Model

In this section we layout the model and describe some theoretical predictions. We first define the environment of the worker and derive the worker's effort decision problem.

---

[7]Hourly wages for employees range from $8 to $14 an hour

Then we prove a theorem describing the worker's policy function. Finally, we test whether this implication holds in the data.

3.1 The Environment

Each day the firm needs to hire enough workers to sort N checks[8]. As in the standard moral hazard model, we assume that the firm cannot determine the effort level exerted by the worker[9]. We accomplish this by modelling the time it takes a worker to process a check as a function τ of three variables: the worker's effort level and two random shocks. We model effort as a binary choice, $e \in \{0, 1\}$. The first random shock is a vector $\bar{z}$ of the characteristics of the check that are observed both by the firm and the econometrician. An example of this characteristic would be whether or not a check jams the machine. We assume that $\bar{z}$ is independent over checks and let F denote its cdf. The second shock ε is a characteristic of a check that is unobservable to the firm and the econometrician. An example of this shock would be a particularly tricky jam or ripped check. This shock is also independent over checks, and we assume it is distributed normally, with mean μ and variance σ^2. Using these three variables, we can write the time it takes to complete a job as as function $\tau(e, \bar{z}, \varepsilon)$. Note that even though the firm knows τ and observes $\bar{z}$, it cannot determine the worker's effort level because of the unobserved effects of ε.

The timing of events plays an important part in the model. We assume that workers choose an effort level e before they process each check. Then, while the check is being sorted, the two random shocks, $(\bar{z}, \varepsilon)$, are realized. This implies that when the worker makes an effort decision before sorting a check, all checks look identical. It is only after the check has been sorted, after the realization of $(\bar{z}, \varepsilon)$, that checks are distinguishable.

As this paper focuses on worker behavior, we take the contracts offered to workers as exogenous. Hence, we do not model the firm's problem. As discussed in section 2, this contract includes a fixed wage $\bar{w}$ and a variable incentive component, the function $\bar{b}$ (see equation 2). Note that the firm's incentive scheme is how this model differs from the standard moral hazard problem. Unlike in the standard model, workers in this firm do not receive compensation after every effort decision, but rather at the end of the day after sorting N checks and making N effort decisions. As shown in the following section,

---

[8]The assumption that workers know how many checks they will sort in a day is not unreasonable. Each day, the workers' supervisor makes out a schedule detailing the tasks workers are to perform during their shift.

[9]See Grossman and Hart (1983) and Holmstrom and Milgrom (1987).

10

this change makes the worker's problem dynamic, which is a departure from the standard model.

Workers' utility depends upon their wage and how much effort they exert. We assume there is a cost to effort to the worker, $c(e)$, where $c(1) > c(0)$. We further assume that workers are risk-neutral and that there is no discounting. Utility is then separable in wage and effort, and can be specified as

$$\bar{w} + \tilde{b}\left(\sum_{n=1}^{N} \alpha(z_n) - \tau(e_n, z_n, \varepsilon_n)\right) - \sum_{n=1}^{N} c(e_n). \tag{3}$$

where the summation is over the N checks a worker processed in a day. The first term in the utility function is the workers' base wage, the second term is their daily bonus, and the last term the cost of effort over the entire day. The assumption that a worker is risk neutral, as opposed to risk averse, simplifies the worker's policy function and allows us to directly compare the dollar gains from output with the cost of effort in our welfare analysis. This assumption however, is not crucial for our results. As the variation in a worker's income due to bonuses is small and as the bonuses are paid out at a high frequency, worker's behavior under risk neutrality is a good first order approximation of the worker's behavior under risk aversion[10].

3.2   The Worker's Problem

The worker's problem is to decide, for every check, whether effort should be exerted. Naturally, a worker's decision is affected by the history of events in the day as well as the number of checks left in the day. Both sets of information are needed by the worker to form expectations of the bonus pay the worker will receive at the end of the day.

The history of events a worker observes is the triplet $(e, \bar{z}, \varepsilon)$ for all checks already sorted. This information allows the worker to determine how well the worker is doing with respect to the firm's formula α. A sufficient statistic for this history is the variable s, where s is the sum of the difference between $\alpha(\bar{z})$ and $\tau(e, \bar{z}, \varepsilon)$ for all checks a worker has already processed. Hence, $s = 0$ for the first job in the day. Letting $(e, \bar{z}, \varepsilon)$ be the choice of effort and realizations of the two random shocks that occurred when sorting

---

[10]To provide a stronger basis for this claim, we computed the certainty equivalent of an agent with a standard level of risk aversion when faced with the varying income stream of a typical worker in the data. We found that the percentage difference between the certainty equivalence and the mean level of income to be less than one-half of a percent for standard levels of risk aversion.

the latest check, we can define the law of motion for s as $s' = \alpha(z) - \tau(e, \bar{z}, \varepsilon) + s$. The number of checks a worker has left in the day to sort is simply denoted as n, where $n \in \{0, 1, 2 \ldots, N\}$.

Using the two variables (s, n), we can write the worker's problem recursively as an N period stochastic dynamic problem. In this environment, the worker only gets paid at the end of the day, $n = 0$, but incurs the cost of effort, $c(e)$, each period. The worker's value function is

$$V(n, s) = \begin{cases} \max_{e \in \{0,1\}} \left\{ -c(e) + E[V(n - 1, \tilde{s}(e, \bar{z}, \varepsilon, s)] \right\} & \text{if } n = \{1, 2, \ldots, N\}, \\ \bar{w} + \tilde{b}(s) & \text{if } n = 0, \end{cases} \tag{4}$$

where the expectation is taken over $(\bar{z}, \varepsilon)$ and

$$\tilde{s}(e, \bar{z}, \varepsilon, s) = s + \alpha(\bar{z}) - \tau(e, \bar{z}, \varepsilon).$$

To solve for the worker's policy function $\tilde{e}(n, s)$, we use backward induction and determine for which values of s a worker will exert effort, in every period n . In deciding whether or not to choose $e = 1$, the worker computes whether the expected value of the bonus at the end of the day is larger than the cost of effort this period and the expected cost of effort in future periods.

3.3 Theoretical Results

To better understand the worker's problem, we analyze the comparative statics of the worker's policy function $\tilde{e}(n, s)$. We first examine how $\tilde{e}(n, s)$ changes with respect to n, holding s fixed. Increasing n, or increasing the number of checks a worker has left to sort, provides the worker with more opportunities to affect s. For a highly productive worker, this increase in opportunities is beneficial as the worker has more chances to increase s and so offset any bad draws of $(\bar{z}, \varepsilon)$. Conversely, for an unproductive worker who struggles to sort checks faster than the firm's estimated time, increasing n is not beneficial. More opportunities mean that this worker is less likely to have a positive s at the end of the day, and so be eligible for incentive pay. Hence, without knowing the productivity of a worker, we are unable to say whether $\tilde{e}(n, s)$ is increasing or decreasing in n.

Next, we turn to examining how $\tilde{e}(n, s)$ changes with respect to s, for a fixed n.

12

At first glance, one might expect that $\tilde{e}(n, s)$ is increasing in s. A higher s, after all, weakly increases the worker's expected bonus. This monotonicity result, however, only holds under specific assumptions on the joint distribution of the random shocks, $(\bar{z}, \varepsilon)$. The difficulty in showing this result comes from the fact that effort has two effects on a worker's bonus pay. A worker's effort level has both a level effect on the bonus and an effect on the probability of earning that bonus. This second effect on the probability of earning incentive pay is not monotone in s. As such, unless we impose a particular structure on $(\bar{z}, \varepsilon)$, we are unable to determine if the worker's policy function is monotone in s for a fixed n.

To gain a better understanding of how the worker's policy function varies in s, we analyze two special cases. First, consider the case where the worker has probability one of earning a bonus. The worker's decision of whether to choose low or high effort depends on the comparison of the difference in the costs of effort against the difference in the expected bonus. In this case, the effort decision has no effect on the probability of earning a bonus. So the difference in the expected bonus pay is simply the extra expected time saved because of high effort, multiplied by K, the incentive pay wage parameter. Formally, we write this as

$$D \equiv K \cdot \int_{-\infty}^{\infty} \int_Z \Big(\tau(0, \bar{z}, \varepsilon) - \tau(1, \bar{z}, \varepsilon) \Big) F(d\bar{z}) \Phi(d\varepsilon). \tag{5}$$

We assume that

Assumption 1. $D > c(1) - c(0)$.

Naturally, a worker will only exert effort if this assumption holds. As shown in a separate appendix, in our model this special case occurs when s goes to infinity. The intuition is that as s gets increasingly large, the probability of receiving bad enough shocks of $(\bar{z}, \varepsilon)$ such that a worker is no longer eligible for a bonus goes to zero.

The second special case we consider is when the probability of earning incentive pay is zero. Obviously, if the worker has no chance of earning incentive pay, there is no point in exerting a high level of effort. We show in a separate appendix, that in our model this special case holds when s goes to negative infinity.

To formalize the above intuition, let $G(n, s)$ be the difference in the utility gain from

13

high versus low effort. Hence defining the expected gain from effort e as

$$V_e(n, s) = c(e) + \int_{-\infty}^{\infty} \int_Z V(n - 1, \tilde{s}(e, \bar{z}, \varepsilon, s)) F(d\bar{z}) \Phi(d\varepsilon), \qquad (6)$$

we know that $G(n, s) = V_1(n, s) - V_0(n, s)$. We then have the following theorem.

Theorem 1. *Under assumptions on τ and α, given in a separate appendix,*

$$\lim_{s \to \infty} G(n, s) = c(0) - c(1) + K \cdot D, \qquad (7)$$

where D is defined in equation 5, and

$$\lim_{s \to -\infty} G(n, s) = c(0) - c(1). \qquad (8)$$

Proof. The proof is available upon request. $\square$

The assumptions on τ and α are simply technical assumptions that insure that the difference $V_1(n, s) - V_0(n, s)$ does not go to infinity as s goes to infinity. Equation 7 formalizes the intuition discussed above for the case when the probability of earning incentive pay is one. In the limit, the return from high versus low effort is the difference in the cost of effort plus the term D. Equation 8 shows that the return from high versus low effort as s goes to negative infinity is simply the difference in the cost of effort.

These limit results show that the $\lim_{s \to \infty} G(n, s)$ is higher then $\lim_{s \to -\infty} G(n, s)$. Clearly, low effort is exerted when the probability of earning incentive pay is zero. In addition, as we assume that $D > c(0) - c(1)$, we know that effort will be exerted when the probability of earning incentive is one. Determining how $G(n, s)$ varies for intermediate values of s is complicated, as this variation is a function of the joint distribution of $(\bar{z}, \varepsilon)$. The difficulty arises in that the difference in the increase in the probability of earning incentive pay when high effort is exerted versus low effort is not monotone in s. In the two limits cases, neither low nor high effort change the probability. But for intermediate values of s, high effort does increase the probability of earning the bonus by more than low effort. Without more specific assumptions, we are unable to determine the magnitude of this non-monotonicity, and so show if $G(n, s)$ is monotone in s. When estimating our model, however, we calculate $G(n, s)$ for all relevant values of s and n. Figure 1 illustrates how $G(n, s)$ varies with s, given our functional form assumptions and the estimated parameters of our model. It plots $G(n = 2, s)$ over a range of s, showing that this 'gain

14

Figure 1: A Plot of $G(n, s)$
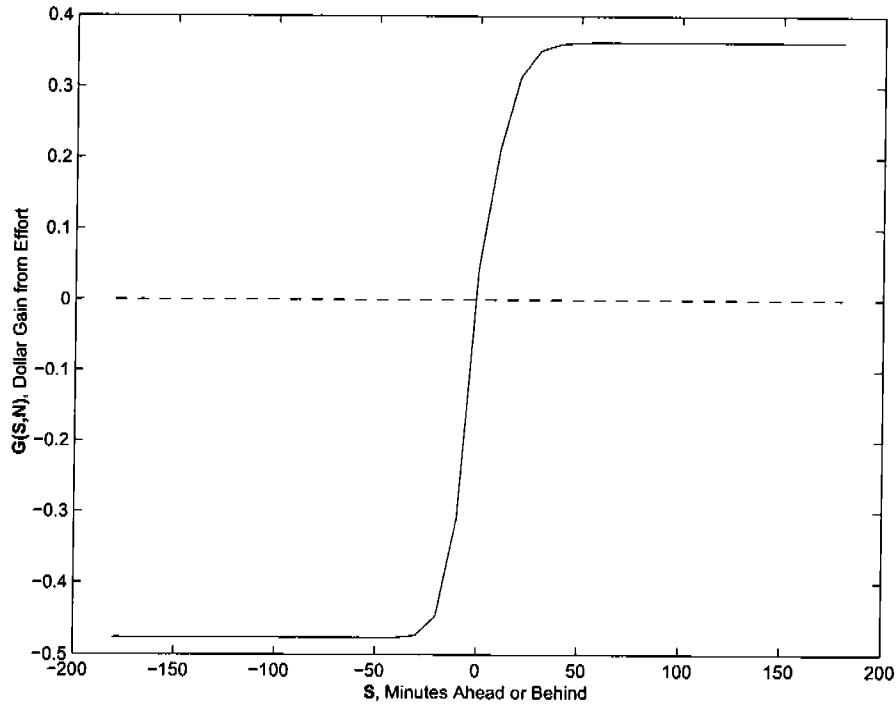
from effort' function, for the estimated parameter values, is monotone and non-linear. We also examine how $G(n, s)$ varies with n. We find that for all workers in the sample, $G(n, s)$ is increasing in n.

3.4 Empirical Tests

The theorem provides a way for us to perform a non-parametric test of our model. It implies that for a fixed n, a worker's effort level given a high probability of earning incentive pay is higher than the worker's effort level given a low probability of earning a bonus. To test this implication and to quantify the effect of a high level of effort on the time taken to sort checks, we conduct the following experiment. In the data, we can compute a worker's state variables, (n, s), for every job (i.e. for every observation). We then run a productivity regression, where the dependent variable is the time taken to complete a job, and the independent variables are the job's observable characteristics,

15

day effects, worker dummy variables, and the worker's state variable n. Recall that τ_j denotes the time taken to sort a job j, $\bar{z}_j$ are a job's characteristics, and n_j is the number of checks in the day that the worker processing job j has left to sort, including those checks in job j. Let $\bar{d}_j$ be the day in which the job was processed, and $\bar{i}_j$ be the worker who processed the job. Note there are 320 days and 15 workers in the sample. The regression we ran is

$$ln(\tau_j) = ln\left(\beta'\bar{z}_j\right) + \sum_{l=1}^{320} \xi_l \cdot 1_{l=\bar{d}_j} + \sum_{k=1}^{15} \eta_k \cdot 1_{k=\bar{i}_j} + \nu \cdot n_j + \varepsilon_j, \qquad (9)$$

where β is a vector of coefficients and $1_{x=y}$ is a dummy variable equal to 1 when $x = y$[11]. As effort is unobserved, its effect on time is captured by the residual term of this regression. We then turn to calculating the probability of earning incentive pay for every observation. The theory states that this probability is a non-linear function of the worker's state variables and characteristics. To approximate this relationship, we run a probit model where the dependent variable is 1 if the worker earned a bonus that day and 0 otherwise. The independent variables are worker dummy variables, s, n, and higher ordered terms of s and n. Combining the probit and regression results, we have, for every job, the worker's probability of earning incentive pay and a residual term capturing the effect of the worker's effort. The theory implies that when the probability of earning incentive pay is high, the residual should be low, as effort is high. Conversely, when the probability is low, the residual should be high due to low effort. We test this implication by computing the mean of the residuals conditional on the probability of earning incentive pay being over 95% and then being under 5%. Table 2 contains the results.

In this table, all the means of the residuals are highly significant, except for the mean of -0.0003 in the second incentive pay regime. More importantly, the difference between the means within an incentive pay regime are highly significant and have the anticipated sign. The size of difference in the regimes is about 0.09 under the first regime and 0.05 under the second. As the productivity regression used the natural log of time, these differences in residuals are in percents. As such, these results indicate that a high level of effort reduces the time taken to sort a job by 5 to 10 percent.

An interesting aspect of Table 2 is the number of observations in each category. Under

---

[11]We re-ran our computations using different specifications of this regression, and found there were no significant changes in the final results

Table 2: Conditional Means of Residuals

		Obs	Mean	Std Error
1st IP	High Prob	21,569	-0.0059	0.0018
	Low Prob	91	0.0892	0.0375
2nd IP	High Prob	5,644	-0.0003	0.0038
	Low Prob	464	0.0487	0.0158

Note: 'High Prob' is defined as the probability being over 95%
'Low Prob' is defined as the probability being under 5%

the first incentive pay regime, it is hard to find observations where a worker has a low probability of earning a bonus. There were only 91 out of 23,098 jobs in the first incentive pay regime where the probability of earning incentive pay was under 5%. Recall from Table 1, that under this regime workers earned a bonus 95% of the time. To check the sensitivity of our results to the number of observations, we re-computed our results using a 10% cutoff for the probability. There were no significant changes in our results. In first incentive pay regime, for the 'Low Probability' category, the number of observations increased to 126 and the new mean of the residuals was 0.0859.

4 Estimation

In this section, we describe how we estimate the structural parameters of this model. We begin by stating and justifying our functional form specifications. Then we show how the model is identified and summarize our estimation technique. Finally, we report our parameter estimates and discuss their implications.

4.1 Specification

In order to estimate this model, we need to know the functional form of τ and the distributions of $(\bar{z}, \varepsilon)$. In specifying the functional form of τ, there are three main issues that we consider. First, workers are heterogenous in their productivity, as demonstrated by the wide range in mean bonus payments listed in Table 1. To capture this, we assume there are I types of workers, who differ by a productivity parameter η_i. This parameter proportionately affects τ. A worker's type is not assigned a priori, but rather the model assigns types to worker so as to maximize the model's likelihood. To determine the total

17

number of types of workers, we re-estimated the model for an increasing number of types until there are no significant changes in the parameter estimates.

The second issue we consider is which characteristics of a job are important in determining how long a job takes to finish. We have previously mentioned three observed characteristics of jobs, the number of jams cleared, fields corrected, and checks sorted. There are, however, a number of other characteristics in the data, such as the area of the country from which the checks originated and which sorting machine was used, that may help determine τ. To find the set of characteristics that are important predictors of time spent processing a job, we regressed all observable characteristics in the data on time. The end result is that three characteristics - the number of jams, the number of fields corrected, and the number of checks sorted - explain 90% of the variation in time. Other characteristics either do not add any explanatory power or only have a marginal effect.

In our specification of τ, we only include the number of jams and fields corrected as a job's observable characteristics. We do this because of a problem computing the worker's policy functions. We found that solving the worker's problem for a large number of checks, N, took a prohibitively long time. As such, we decided to approximate the model by assuming that workers made an effort decision every 1000 checks[12]. This, however, raised a problem with the structure of our data. As discussed in the Data Description (Section 2), an observation in our data set is a job, where a job ranges from 1 to over 8000 checks. To bring the model specification and data into line, we re-arranged the data to construct observations of 1000 checks. We accomplish this by first chronologically lining up the jobs for every worker in a day. Then, starting at the end of the day, we cut and spliced jobs together to make new jobs of uniform length. Depending on their number, the residual checks left at the beginning of the day were either discarded or expanded into a 1000 check job. Under this modification, the number of checks per job in the data is constant and so can be captured by τ's intercept term. Consequently, we construct the random variable $\bar{z}$ of a job's characteristics, as a 2×1 vector, where z_1 is the number of jams that occurred and z_2 is the number of fields corrected.

Reducing the vector of characteristics $\bar{z}$ to two dimensions is advantageous as it decreases the computation burden of solving the worker's problem. However, it also

---

[12]We are planning to estimate the model where workers make an effort decision every 500 checks to check our model's sensitivity in this respect.

complicates the problem of computing a worker's expectations over $\alpha(\bar{z})$. The firm's actual incentive pay formula uses a number of characteristics other than jams and fields corrected. As such, when we compute the worker's expectations over the state variable s in the next period, we use an approximation the firm's actual incentive pay formula. Like the firm's actual formula, the approximation we use is a linear function of jams and fields corrected[13].

Finally, the last issue we consider with respect to τ's specification is how $e, \bar{z}$, and ε interact with one another. The result mentioned above, that a linear regression of the job's observable characteristics on time has an R-squared of 90%, is strong evidence for a linear specification of τ in $\bar{z}$. To check the magnitude of a non-linear relationship between the observable characteristics and time, we re-ran the above regression adding squared terms of the observable characteristics to the set of regressors. The estimated coefficients on these squared terms are insignificant, adding to the credibility of τ being linear in $\bar{z}$. Lastly, with regard to effort, our observations in the workplace and conversations with workers lead us to believe that effort has a direct effect on how quickly jams are cleared and fields are entered. Consequently, the effort term in τ needs to interact with both elements of $\bar{z}$. With these issue in mind, we specify τ as

$$\tau^j\left(e, \bar{z}, \varepsilon\right) = \left[\beta_0 + \left(\beta_1 - \beta_3 \cdot e\right) \cdot z_1 + \left(\beta_2 - \beta_4 \cdot e\right) \cdot z_2\right] \cdot exp(\eta_j) \cdot exp(\varepsilon), \qquad (10)$$

where $\beta_3, \beta_4 \geq 0$, $\varepsilon \sim N(0, \sigma^2)$ and $j = 1, \ldots, J$. Taking logs, we get the form of the equation we actually estimate,

$$ln(\tau^j(e, \bar{z}, \varepsilon)) = ln\left[\beta_0 + \left(\beta_1 - \beta_3 \cdot e\right) \cdot z_1 + \left(\beta_2 - \beta_4 \cdot e\right) \cdot z_2\right] + \eta_j + \varepsilon. \qquad (11)$$

In this specification, the heterogeneity in workers is captured by adjusting the coefficients β_0 thro ugh β_4 by a fraction η_j. Rather than specify exogenously which worker belonged to which type j, we assigned worker's types so as to maximize the model's likelihood given the parameter values.[14] We constrain the coefficients β_3 and β_4 to be greater than zero as our prior is that effort lowers the time it takes to complete a job. This also reduces the parameter space and so speeds up the estimation algorithm.

---

[13]We estimated the coefficients and intercept term of this function using ordinary least squares. The R-squared for this regression under the first regime is 0.70 while under the second regime it is 0.98. The lower R-squared under the first regime is concerning. We plan on improving upon this approximation.

[14]Another way we considered introducing worker heterogeneity is by having differing costs of effort, $c_j(e)$. We are planning to re-estimate the model with heterogenous costs.

Having specified τ and the distribution of ε, we are left with F, the cdf of $\bar{z}$. As we have data on realizations of $\bar{z}$, we use these observations to construct an empirical distribution of F. Finally, we normalize the cost of low effort, $c(0)$, to zero, and let $c(1) = \gamma$. Notice that γ is the same for all workers, and so workers are only heterogeneous in $_i\eta$.

With these functional forms, we can compute the likelihood of the model. We have a panel data set of I individuals, where for each individual we have T_i observations. Individuals are heterogenous in that they can be of different types $\{\eta_j\}$. In this data set, we have information on the job's observable characteristics, $\hat{z}_{i,t}$, the number of checks left to sort in the day, $\hat{n}_{i,t}$, the time taken to sort jobs, $\hat{\tau}_{i,t}$, and the benchmark times computed by the firm, $\hat{\alpha}_{i,t}$. The likelihood is then

$$L(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \gamma, \{\eta_j\}, \sigma \,|\, \{\hat{z}_{i,t}, \hat{n}_{i,t}, \hat{\alpha}_{i,t}, \hat{\tau}_{i,t}\}) =$$

$$\prod_{i=1}^{I} \max_{\{\eta_j\}} \left[\prod_{t=1}^{T_i} P\Big(ln(\hat{\tau}_{i,t}) | \{\hat{z}_{i,t}, \hat{n}_{i,t}, \hat{\alpha}_{i,t}\}, \beta_0, \ldots, \beta_4, \gamma, \eta_j, \sigma \Big) \right]. \quad (12)$$

From equation 11, we know that the likelihood of the observation $ln(\hat{\tau}_{i,t})$, conditional on the data and parameters specified above, is $\phi\Big(\frac{ln(\hat{\tau}_{i,t}) - \delta_{i,t,j}}{\sigma} \Big)$, where

$$\delta_{i,t,j} = ln\left[\beta_0 + \Big(\beta_1 - \beta_3 \cdot e_{i,t,j} \Big) \cdot \hat{z}_{i,t}^1 + \Big(\beta_2 - \beta_4 \cdot e_{i,t,j} \Big) \cdot \hat{z}_{i,t}^2 \right] + \eta_j. \quad (13)$$

The variable $e_{i,t,j}$ is the effort level exerted by worker i, while sorting a job at time t when the worker is type j. We compute which effort level a worker chooses by solving the worker's dynamic problem given the parameter values and obtaining the worker's policy function. This policy function depends upon the worker's state variables (n, s). We observe n directly in the data, $\hat{n}_{i,t}$, and can compute s from the two sequences $\{\hat{\alpha}_{i,t}, \hat{\tau}_{i,t}\}$.

We assign workers' types so as to maximize the likelihood. We do this by calculating the likelihoods associated with a worker being each possible type $\{\eta_j\}$. We then compare likelihoods across types, and assign each worker the type with the largest corresponding likelihood.

4.2 Identification

The general identification issue in this check-sorting environment is determining if effort significantly affects the time taken to sort checks. The main source of identification

comes from the incentive effects of the firm's kinked bonus system. Looking back at the incentive pay program defined by equation 2, note that this formula generates a bonus wage profile that is flat at $0 dollars for all negative values of the state variable s. Then, for positive values of s, the wage profile is linearly increasing in s. This kink at 0 creates a perverse incentive for workers to quit working hard once they have fallen too far behind in terms of s. In Theorem 1, we prove this intuitive result, showing for large, positive values of s, a worker will choose high effort while for values of s that are negative and large in absolute value, workers will choose low effort. So theory predicts that workers with low, negative values of s will have low productivity (high τ), while those with large positive values of s will have high productivity (low τ). Conversely, in an environment where incentives did not matter, there is no correlation between s and τ. In the data, we observe the values of s and τ for a worker for every job. As the distribution of s in the data ranges from under a low -180 minutes to over a high 200 minutes, we have a straightforward way of determining if effort matters in this environment.

For our particular model and functional form specifications, this general identification problem reduces to showing that we can measure effort's effect on time, captured by (β_3, β_4), and the cost of effort, γ. As described above, theory predicts when effort will and will not be exerted in certain cases. Using this information, we are able to precisely measure β_3 and β_4. Further, the model imposes a specific structure on the relationship among γ, β_3, β_4 and τ through the utility-maximizing behavior of workers. These cross-equation restrictions allow us to precisely estimate γ.

A second source of identification comes from the firm's change of parameters in the incentive scheme two-thirds of the way through the sample period. As described in the data section of the paper, the firm altered the productivity thresholds workers need to beat to earn bonus pay, as well as the formula used to compute bonuses conditional on eligibility. A consequence of this shift is that workers systematically changed when they exerted effort, influencing their productivity. In an environment without effort, however, the change in incentive schemes would have no effect on worker productivity. These different predictions on how productivity should change with the switch in incentives provides us with a second avenue of identification.

One difficulty with our current model is that we are unable to separate out the effect of daily shocks to the worker from the effect of effort. We plan to address this problem by incorporating a daily shock into the model and re-estimating the model's parameters.

4.3 Estimation Algorithm

As mentioned in the introduction, we use a maximum likelihood approach to estimate the structural parameters of the worker's problem. To compute the model's likelihood, we use a simple three step algorithm that is common to the discrete-choice structural estimation literature. The first step is to specify the functional forms of the cost of time function, τ, and the cost of effort function, c. Note that we have already assumed that workers' utility is additive in wages and effort. In addition, we need to choose values for all the parameters in the model. The second step involves using the parameter values and newly specified functions to solve the worker's problem. For the functional forms we consider in this paper, the worker's policy rules are cutoff rules. Hence, for every period n we find a threshold value $\bar{s}_n$ where a worker will only choose $e = 1$ if $s > \bar{s}_n$. As described in the appendix, these policy rules are computed using backward induction. Finally, the last step is to use the policy rules to infer the worker's effort decisions. Using this information along with the data, we then calculate the likelihood.

To find the set of parameters that maximize the likelihood, we use a simulated annealing program to search over the parameter space. We tried to use faster search methods, such as simplex or derivative based algorithms, but found they continually got stuck on 'flat spots'. As effort is discrete and the worker's policy function is a cutoff rule, this likelihood is a step function along certain dimensions. We found that only the simulated annealing algorithm was able to perform a wide enough search of the parameter space to find the global optimum.[15]

4.4 Parameter Estimates

Using these functional specifications, we obtained the parameter estimates listed in Table 3. Note that all parameter estimates are highly significant[16].

These estimates imply that when a worker does not exert effort, entering a field takes 4.5 seconds. With effort, a field is entered 1.3 seconds faster, a 28% reduction in time. In contrast, how fast a worker clears jams is only marginally affected by the worker's effort decision. This is not unexpected however, as clearing a jam entails a series of steps (e.g. opening a panel component, pulling out the stuck checks, etc) that are not conducive to being hurried. Entering a field (e.g. a check's bank account number), on the other hand,

---

[15] A good source on how a simulated annealing algorithm works is Goffe, Ferrier, and Rogers (1994).

[16] The standard errors reported here were computed given each worker's type

Table 3: Parameter Estimates

Parameter		Estimate	Standard Error
		(seconds)	
Intercept	β_0	380.106	1.4616
Jams	β_1	27.558	0.0144
Fields	β_2	4.465	0.1362
Effort on Jams	β_3	0.083	0.0252
Effort on Fields	β_4	1.254	4.3e-5
		(no units)	
Type 2	η_2	0.210	4.0e-6
Type 3	η_3	0.342	1.4e-5
Standard deviation	σ	0.224	4.3e-7
		(dollars)	
Cost of effort	γ	0.476	1.0e-7

Note: η_1 is normalized to 0

is a task much like touch-typing, whose speed depends upon the worker's concentration. To gain a better understanding of how important effort is in reducing time, consider that a typical batch of 1000 checks requires a worker to clear 14 jams and type in 193 fields. Our parameter estimates imply that with effort, the time a worker spends processing checks decreases by 15%. For Type 1 workers, this is a decrease of 4 minutes, while for Type 2 and 3 workers this is closer to 5 minutes.

Turning to the cost of effort, notice that γ is 0.48. As we assumed that utility is additively separable in effort and wage, this value can be interpreted as the dollar cost to a worker for choosing $e = 1$. Thus, the disutility from working hard while sorting 1000 checks is $0.48.

We included three types of workers in the model as we found that having more than three did not significantly change our results. Type 1 workers are the best and benchmark workers (i.e $\eta_1 = 0$). Type 2 workers are worse than Type 1 workers, taking 21% more time to process checks than Type 1. Finally, Type 3 workers are the least efficient workers, sorting checks 34% slower than Type 1 workers. Table 4 provides a comparison of these assigned types to the workers' actual grades. Note that most high grade workers (those with grades of 6 or 7) are Type 1 and most low grade workers are not Type 1. There are, however, some low skilled, high grade workers and high skilled, low grade workers.

Table 4: Worker Grade and Type

Worker	High Grade										Low Grade				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Grade (Data)	7	7	7	6	6	6	6	6	6	6	5	5	5	4	4
Type (Model)	1	1	1	1	1	1	2	2	3	1	1	2	3	2	1

4.5 Analysis of Effort Decisions

Using these parameter estimates, we can analyze workers' effort decisions. The model infers that workers exerted effort 98% and 84% of the time respectively, under the firm's two incentive schemes. Hence, the changes the firm made to the bonus scheme decreased the number of times workers choose $e = 1$ by 14%. Knowing how often workers exert effort allows us to determine by how much the firm's bonus plan increases productivity. Using our parameter estimates, we compute that exerting effort decreases the time spent processing checks by 15%. This implies that under the first incentive pay plan, workers' high effort decreased the time spent sorting checks by $(0.98 \cdot 0.15) = 14.7\%$. The decrease in time under the second incentive pay scheme is slightly less, at 12.6%. Taking the average of these two numbers, weighted by the number of observations under each regime, we find that over the sample period, the firm's bonus plan decreased the time spent sorting checks by 14%. This translates into a 16% increase in worker productivity.

When exactly though, do workers choose low effort? Assuming that workers start out the day exerting effort, theory predicts that low effort decisions will most likely occur in the second half of a shift, after a worker experiences an unproductive morning. As discussed in the theoretical implications part of this paper, Section 3, this is due to the firm's kinked incentive pay scheme. Given the estimated parameters, we can test this prediction as the model computes workers' policy rules and effort decisions over the sample period.

From the model, we know that workers choose low effort 1,038 times out of a possible 18,120 times during the sample period. To determine if workers are choosing effort over an entire day or mixing high and low effort decisions within a day, we constructed Table 5. This table lists the number of days when a high level of effort was never exerted (Zero Effort), days when both high and low effort were exerted (Mixed Effort), and the days when a high level of effort was always exerted (Full Effort). It is interesting to note that under the first incentive pay regime there are no 'zero effort' days, while under the second regime there are 138. This increase in 'zero effort' days is the result of the two

Table 5: Effort Decisions within the Day

	Zero Effort	Mixed Effort	Full Effort
First IP Regime	0	119	2493
Second IP Regime	138	74	939

type 3 workers in the sample. Under the second incentive pay regime, these workers do not start the day expecting to receive a large enough bonus to compensate them for the cost of effort. Hence, they will only choose $e = 1$ if they receive high productivity shocks during the day. In contrast, all workers under the first incentive pay regime and type 1 and 2 workers under the second regime, always start the day with a high effort level.

By focusing on workers who start the day exerting a high level of effort, we can test the theoretical prediction of when in the day low effort decisions typically occur. Out of the 1,038 low effort decisions, 350 of them were made by workers who started the day choosing $e = 1$. Of these decisions, all occurred in the second half of a worker's shift, supporting our theoretical prediction. In addition, this result illustrates that there is a large inefficiency associated with the dynamic nature of the firm's incentive scheme. The use of daily productivity numbers and the resulting effects on worker behaviour account for roughly a third of all low effort decisions made. By re-structuring its incentive pay scheme to compensate workers more frequently, the firm could significantly reduce the number of times workers choose low effort.

4.6   Welfare Analysis

Now that we understand how often workers choose to exert a high level of effort, we analyze by how much the firm's bonus plan increases welfare. We address this issue by first considering the welfare in the economy under two extreme cases: a flat-wage scheme and the first-best scheme. We choose these two schemes as we are interested in comparing the case where workers never exert effort to the one where they always exert effort. By looking at the worker's problem, it is clear that if workers are only paid an hourly rate, they will never exert effort. Under the first-best scheme, we know that it is profit-maximizing for the firm to always motivate its workers to choose $e = 1$. This is because the expected gain to the firm when the worker exerts effort is larger than the disutility from effort. Our parameters imply that effort causes, at the least, a 4 minute reduction in time for a typical batch of 1000 checks. The mean wage of workers is $10.55,

25

Table 6: Welfare Analysis per Worker per Day

	Flat Wage (e=0)	First Best (e=1)	Difference in Welfare
Minutes Spent Sorting	136	115	21
Cost of Effort (γ)	$0	5 · $0.48	-$2.38

which means the firm expects to save $\frac{4}{60} \cdot \$10.55$, or $0.70. As the disutility from effort is only $0.48, there is a gain to always have the worker choose $e = 1$.

With these compensation schemes in mind, we turn to measuring welfare. Our strategy is to compare the welfare associated with processing a typical day's worth of checks, about 5000 checks, under the two schemes. From the data, we know that while sorting 5000 checks, a worker will typically clear 70 jams and type in 966 fields. Using these numbers, we find that in the flat-wage case (where no effort is exerted), a worker will typically take 136 minutes to sort 5000 checks. Naturally, as a worker never exerts effort under this scheme, the disutility from effort is $0. Moving to the first-best, we find that a worker will typically sort 5000 checks in 115 minutes, 21 minutes faster than without effort. The disutility the worker experiences from exerting effort is $5 \cdot \$0.48 = \$2.38$, as $e = 1$ is chosen 5 times (once for every 1000 checks). To compute the net change in welfare when moving from a flat-wage to the first best scheme, we need to compare the gain of 21 minutes to the disutility from exerting effort. The firm gains in two ways from the decrease in time taken to sort checks. First, the worker now has an extra 21 minutes to perform other tasks for the firm. To compute the value of this extra time to the firm, we use the mean wage of workers, $10.55. Second, the sorting machine is free for 21 minutes. The value of this extra time to the firm is harder to quantify[17]. As such, we only consider the welfare effects of freeing up the worker's time, which equals $\$10.55 \cdot \frac{21}{60} = \$3.69$. The total expected gain in welfare from using the first-best scheme is then $3.69 - $2.38 = 1.31, which is a 5% increase in welfare.[18] Table 6 summarizes these results.

Using this method of analysis, we can easily compute the welfare gain from using the firm's incentive schemes. As previously mentioned, our estimates imply that workers

---

[17]We are currently working on a good approximation of this value

[18]These results were computed using a Type 1 worker. Using a Type 2 or 3 worker increases the firm's gains from using incentives and increases the percent change in welfare.

exerted a high level of effort 98% of time under the first bonus scheme and 84% of the time under the second. Hence, under the first incentive pay scheme, welfare was increased by $0.98 \cdot 5\% = 4.9\%$. We similarly calculate the welfare increase under the second incentive pay regime and then take the average of the two percentages, weighted by number of observations under each regime. We find that over the sample period, welfare increased by 4.7% due to the firm's bonus plan.

5 Summary and Conclusion

Most empirical work on incentive pay has only focused on measuring by how much incentives increase output. As such, it is unclear at what cost this extra output is obtained. Our paper adds to this literature by examining both the increase in output and the corresponding rise in disutility from higher effort due to incentives. This allows us to measure by how much the welfare of the firm and workers rises due to the introduction of incentives.

We accomplish this by studying a check-clearing firm's use of incentives. Using the firm's production records, we develop and estimate a dynamic model of worker behavior. This allows us to determine by how much the welfare gains of increased output due to incentives outweigh the disutility from increased effort. We find that compared to an environment without incentives, the firm's bonus scheme lowers the time taken to sort checks by 14%. Roughly two-thirds of this gain, however, is needed to compensate workers for their higher effort levels. By comparing these two welfare changes, we compute that the introduction of incentives increases the welfare of the firm and workers by 4.7%.

Although this paper only looks at one firm, we believe our results have broad applicability. The 'continuous flow' production technology used by the firm has general characteristics common to a large portion of the manufacturing sector of the economy. Specifically, the automation of the check-sorting process and the worker's role in maintaining the operation of a machine, are production characteristics found throughout a variety of manufacturing industries. In these industries, then, we believe that the introduction of incentives would increase the welfare of firms and workers.

Another lesson we draw from our results concerns the dynamic effects of contracts. In the case of the check-clearing firm, the combination of compensating workers after they have made multiple effort decisions, with a minimum productivity requirement significantly reduces the bonus system's effectiveness. Roughly a third of the low effort

decisions made by employees over the sample period were caused by this dynamic effect. Simply by paying workers more often, the firm could get rid of this inefficiency. The general lesson, then, is that when designing incentive schemes, close attention should be paid to how the bonus scheme affects the worker's effort decision over time.

References

ASCH, B. (1990): "Do Incentives Matter?," *Industry Labor Relations Review*, 43, 89–107.

BLINDER, A. (ed.) (1990): *Paying for Productivity: A Look at the Evidence*. Brookings Institution, Washington, DC.

FERRALL, C., AND B. SHEARER (1999): "Incentives and Transactions Costs Within the Firm: Estimating an Agency Model Using Payroll Records," *Review of Economic Studies*, 66, 309–338.

GOFFE, W. L., G. D. FERRIER, AND J. ROGERS (1994): "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65–99.

GROSSMAN, S., AND O. HART (1983): "An Analysis of the Principle-Agent Problem," *Econometrica*, 51, 7–46.

HOLMSTROM, B., AND P. MILGROM (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303–328.

LAZEAR, E. P. (2000): "Performance Pay and Productivity," *American Economic Review*, 90, 1346–1361.

MARGIOTTA, M., AND R. MILLER (2000): "Managerial Compensation and the Cost of Moral Hazard," *International Economic Review*, 41, 669–719.

OYER, P. (1998): "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality," *Quarterly Journal of Economics*, 113, 149–185.

PAARSCH, H., AND B. SHEARER (2000): "Fixed Wages, Piece Rates, and Incentive Effects: Statistical Evidence from Payroll Records," *International Economic Review*, 41:1, 59–92.

PRENDERGAST, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37, 7–63.

MicroFilmer

Field Upgradability

Endorsement

12-Pocket Stacker Module

IBM Co-Existence

Primary Feeder

Image Capability

Encoders

Secondary Feeder

Master Printer

Reader