

The Welfare Effects of Incentive Schemes *

Adam Copeland[†] Cyril Monnet[‡]

December, 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. We find that the firm's incentive scheme has a large effect on productivity, raising it by 12% 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 a 7% 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%. Under a first-best scheme, we find that the net increase in welfare is 6%.

*We would like to thank Tom Holmes for his advice and encouragement. We also thank Zvi Eckstein, Gautam Gowrisankaran, Matt Mitchell, and Andrea Moro for their helpful comments. In addition, this paper has benefited from participants at numerous seminars. Finally, we thank Jon Stam and Deb Rahn for their assistance in gathering the data and answering our questions about check processing. This paper is based on the first part of Adam Copeland's University of Minnesota Ph.D. dissertation and much of it was written while both authors were Research Associates at the Federal Reserve Bank of Minneapolis. The views expressed herein are those of the authors and not necessarily those of the European Central Bank, the Eurosystem, the Federal Reserve Bank of Minneapolis, or the Federal Reserve Board.

[†]Federal Reserve Board (Adam.M.Copeland@frb.gov)

[‡]European Central Bank (Cyril.Monnet@ecb.int)

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? Within the theoretical framework of the principle-agent problem, we can gauge the success of incentives by answering three questions: Do incentives matter? How significant are they? What are the welfare gains to the worker and the firm from using incentives? This last question is particularly important as it captures both the benefits and costs to using incentives. While a particular bonus scheme may increase worker productivity by a large amount, this gain is costly as it is the result of workers exerting a higher level of effort. Only by measuring workers' disutility from effort and comparing it to the gains in productivity, will we have an accurate account of how well incentives work.

Our paper answers these three questions 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 determine that incentives matter and have significant effects on worker behavior. We then compute the welfare gains from using incentives, by determining how much output and the disutility from effort increase in response to different bonus plans.

This case study is particularly interesting due to special properties of the firm's incentive pay scheme that allow us to identify effort's effect on output and the cost of effort. The structure of the bonus system creates a dynamic effect within the worker's problem. This 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 day if the worker's measure of productivity is low (due, for example, to a bad shock) in the early part of the day. Consequently, 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. Because of the structure of the bonus scheme, we know that after experiencing 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 us with 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 the day 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 day 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 day anew.

With this model, we generate two kinds of results. First, we use reduced form analysis to check that the data supports the basic implications of the model. Second, we employ a simulated maximum likelihood approach to estimate the model's structural parameters.

Using reduced form techniques, we test if a worker's effort level is increasing in the

probability of earning a bonus. To see if this relationship holds in the data, we look at the set of workers' jobs at the end of the day, where the effects of incentives are strongest. We then check whether the unexplained portion of worker productivity is correlated with the worker's probability of earning a bonus. In support of our model, we find that this correlation is significant and has the predicted sign. Using the set of all jobs, we then obtain an upper bound on the effect of effort on productivity. We accomplish this by comparing the set of productivity residuals where workers have an extremely high probability of earning a bonus to the set of residuals where this probability is extremely low. The difference between these two groups of productivity residuals is large and statistically significant.

With these positive results in hand, we turn to estimating the structural parameters of the model. Our estimation method is the simulated 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 the time taken to sort checks, which we use to compute the model's likelihood.

Our parameter estimates show that workers readily respond to incentives, and that effort significantly affects worker output. Over the sample period, the firm's bonus plan increased worker productivity by 12%. 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 7% gain in productivity, which is 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%. We also show that under a first-best scheme, the net increase in welfare is 6%.

Most of the existing empirical work on incentives has focused on answering the first two of three questions listed at the beginning of the introduction: Do incentives matter and, if so, are they significant? The typical approach has been to use reduced form analysis to quantify how much output is affected by a change in incentives. The results from this avenue of research vary widely, attributing an increase in productivity from zero to over twenty percent due to the introduction of incentives.¹ A limitation of the reduced form approach however, is that this methodology cannot answer the third question previ-

¹Lazear (2000) is a recent paper that quantifies the relationship between output and incentives. Blinder (1990) and Prendergast (1999) are good surveys on this literature.

ously posed: What are the welfare gains from using incentives? Answering this question requires estimating the worker's disutility from effort, which can only be done using a structural approach. There are a small number of papers within the personnel literature that use structural estimation. These papers, however, focus on answering other important questions within the literature, rather than attempt to measure the welfare gains from incentives. Paarsch and Shearer (2000), for example, study the optimal contract choice of a firm while Margiotta and Miller (2000) estimate the cost of moral hazard at the senior managerial level. Within this group of papers, our work is closest to Ferrall and Shearer (1999). They estimate the structural parameters of a principal-agent model using a firm's payroll records from the 1920's. 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. Paarsch and Shearer (1999) is another paper close to ours. They use a structural approach to estimate worker effort under piece rates in the tree-planting industry. Our papers differ in our methods of identification and in our modeling of the worker's problem.

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.

Finally, our work is also related to the general literature on testing contract theory. Papers such as Chiappori, Heckman, and Pinquet (2000) and Chevalier and Ellison (1999) use data that captures agents' dynamic behavior to measure the effect of moral hazard in, respectively, automobile insurance and mutual fund markets. We also use this 'dynamic data' approach to identify the moral hazard effect of incentives. Chiappori and Salanié (2000) provides a survey of recent work done in this general research area.

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.

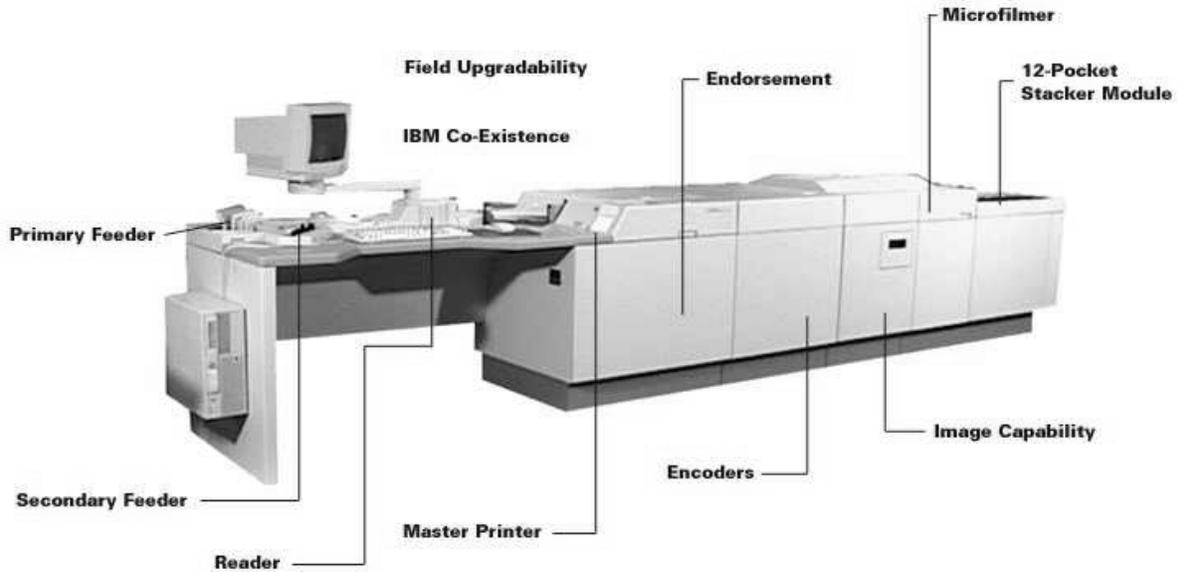
Our data comes from the production and human resource records of the Check Department of the Federal Reserve Bank of Minneapolis. This ‘firm’ provided us with information over a 15 month period (3/01/99 - 5/27/00) on its employees that sort checks in the Low Speed Check Processing Department.² While these workers have several responsibilities, their main task is to sort checks by running them through a sorting machine (see Figure 1). 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.³ 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 in batches. We define a *job* as a batch of checks that needs to be 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 and were employed in the check-sorting department for at least 6 months before the beginning of our 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 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.

²This 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.

³An example of a field is the account number.

Figure 1: Reader/Sorter Machine



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 on average 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:

$$\alpha(\bar{z}) = \rho_0 + \rho_1 z^1 + \dots + \rho_C z^C. \quad (1)$$

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

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

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 7.⁴ 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 fraction 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.

⁴Lower grade employees earn lower wages and are typically newer employees relative to their higher grade counterparts

Table 1: Bonus Payments

Worker	Grade	<i>First IP Regime</i>			<i>Second IP Regime</i>		
		Mean	(Std)	Frac	Mean	(Std)	Frac
All	n/a	\$10.06	(8.54)	0.95	\$6.64	(7.62)	0.81
High grade	6 & 7	\$10.94	(9.19)	0.96	\$8.63	(8.50)	0.85
Low grade	4 & 5	\$8.46	(6.94)	0.92	\$3.32	(4.07)	0.73
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

Frac: Fraction of days with positive bonus pay

In addition, the fraction 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 extreme, Worker 9 received a mean bonus of \$5.45 under the first regime and \$0.22 under the second. Under the 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.⁵ 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 workers.

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. We then prove that a worker's effort level is increasing in the probability of earning a bonus. Finally, we test whether this implication holds in the data.

3.1 The Environment

Each day the firm needs to hire a worker to sort N checks.⁶ As in the standard moral hazard model, the firm cannot determine the effort level exerted by the worker.⁷ We

⁵Hourly wages for employees range from \$8 to \$14 an hour

⁶Each day workers are shown a detailed schedule of their day. This information, along with their experience on the job, enables workers to forecast the amount of checks they will process in their shift.

⁷See Grossman and Hart (1983) and Holmstrom and Milgrom (1987).

accomplish this by modeling the time it takes a worker to process a check as a function 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_ε and variance σ_ε^2 . Using these three variables, we can write the time it takes to complete a job as $\tau(e, \bar{z}, \varepsilon)$. Under this formulation, 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 .

A strong assumption on τ is that both shocks are i.i.d. throughout the day. Focusing first on \bar{z} , it might be the case that the probability of a check jamming increases if the previously sorted check jammed. Using the data however, we were able to reject the hypothesis that check characteristics are correlated within a day.⁸ Turning next to ε , there are two stories why this variable may be correlated within a day. The first story is that shipments of checks may have some common unobserved component. Checks from Bank A, for example, may be packaged and shipped in such a way that they rip more often than checks from Bank B. As we have data on the originating location of checks, we were able to explore and reject this hypothesis. The second story for why ε may be correlated within a day focuses on the worker, rather than the checks. Workers might have good and bad days, which would affect the time it takes them to sort checks. Rather than capture this effect through τ , however, we include a daily cost shock in the worker’s cost of effort function, c . Hence, when workers feel particularly good or bad, the resulting effect on their ability to sort checks will come through their disutility of exerting effort.⁹

Turning to the cost of effort, we specify it as a function of both the level of effort exerted, and a daily cost shock, $\delta \sim \mathcal{N}(\mu_\delta, \sigma_\delta)$. As mentioned previously, this daily shock captures the fact that workers may have good and bad days. We assume that effort is costly, $c(e = 1, \delta) \geq c(e = 0, \delta) \forall \delta$ and that $\frac{dc(0, \delta)}{d\delta} > 0$ and $\frac{dc(1, \delta)}{d\delta} > 0$.

⁸In the data we are also able to identify which machine a worker used to sort checks. As workers mentioned in interviews, we found that all the machines performed the same in sorting checks.

⁹We estimated a version of the model where daily shocks affected the τ function directly. There were no substantial differences in the results.

The timing of events plays an important part in the model. At the beginning of the day, workers draw their daily cost shock. We then assume that workers choose an effort level e before they process each check. While the check is being sorted, the two random shocks, (\bar{z}, ε) , 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}, ε) , 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 \tilde{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, 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 that workers are risk-neutral and that there is no discounting. Utility is then separable in wage and effort, and can be specified as

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

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

¹⁰To 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.

3.2 The Worker's Problem

The worker's problem is to decide, for every check, whether effort should be exerted. To form expectations on the amount of bonus pay a worker will receive at the end of the day, a worker needs to know the history of events in the day as well as the number of checks left in 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 , which is the sum of the difference between $\alpha(\bar{z})$ and $\tau(e, \bar{z}, \varepsilon)$ for all checks a worker has already processed in a day. 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 the latest check, we can define the law of motion of s as $s' = s + \alpha(\bar{z}) - \tau(e, \bar{z}, \varepsilon)$. The number of checks a worker has left in the day to sort is simply denoted as n , where $n \in \{0, 1, 2, \dots, N\}$.

Using the two variables (s, n) and the daily cost shock δ , 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, $c(e, \delta)$, each period. The worker's value function is

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

where the expectation is taken over (\bar{z}, ε) 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 $\tilde{e}(n, s, \delta)$, we use backward induction and determine for which values of s a worker will exert effort, in every period n and for every δ . 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 $\tilde{e}(n, s, \delta)$. We first examine how $\tilde{e}(n, s, \delta)$ changes with respect to n , holding (s, δ) 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 skilled 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}, ε) . Conversely, for an unskilled 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 skill level of a worker, we are unable to predict whether $\tilde{e}(n, s, \delta)$ is increasing or decreasing in n .¹¹

Next, we turn to examining how $\tilde{e}(n, s, \delta)$ changes with respect to s , for a fixed (n, δ) . Due to the structure of the bonus scheme, the probability and size of the worker's bonus is increasing in s . As such, as s increases, so does the worker's effort level. Below, we formally prove this result, assuming that effort is a continuous variable and that $c(e, \delta)$ is increasing and convex. We assume continuity as it simplifies the proof of the theorem and highlights the relevant forces at work.

Theorem 1. *Given effort is a continuous variable, $e \in [0, 1]$, for all (n, δ) , the policy function $\tilde{e}(n, s, \delta)$ is increasing in s .*

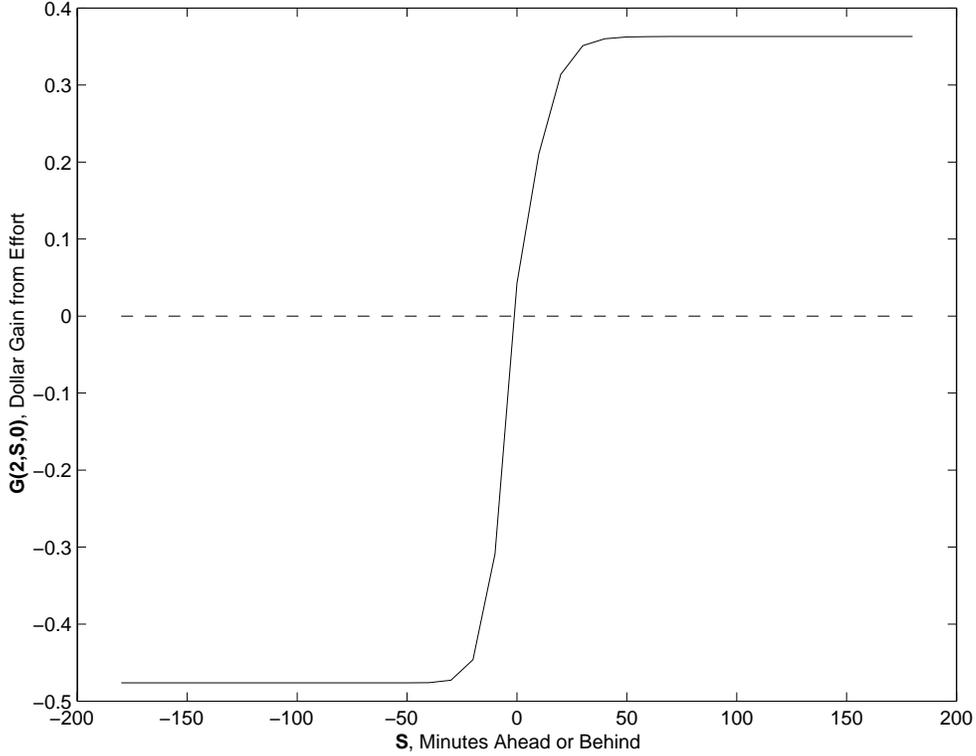
Proof. We first need to show that $V(n, s, \delta)$ is increasing and convex in s for all (n, δ) . This can be shown by induction, using the necessary first order condition, the envelope theorem, and the fact that $V(s, 0, \delta)$ is increasing and convex in s , $\forall \delta$. Then, assuming an interior solution, the first order condition from the worker's effort decision problem is

$$-\frac{dc(e, \delta)}{de} + \int \frac{dV(n-1, s', \delta)}{ds'} \left(-\frac{d\tau(e, \bar{z}, \varepsilon)}{de} \right) = 0,$$

where $s' = s + \alpha(\bar{z}) - \tau(e, \bar{z}, \varepsilon)$ and the integral is over all possible pairs of (\bar{z}, ε) . Given that c , τ , and V are convex functions, we use the implicit function theorem to show that $\frac{d\tilde{e}(n, s, \delta)}{ds} \geq 0$. Whenever the solution is not interior for a given s , the optimal effort level is confined to either 0 or 1, and we have $\frac{d\tilde{e}(n, s, \delta)}{ds} = 0$. \square

¹¹Using a probit model, we infer that the probability of earning a bonus is increasing in n for all workers.

Figure 2: A Plot of $G(n, s, \delta)$



For a given set of parameters, we can calculate the gain from effort, $G(n, s, \delta)$, or the difference in utility from high versus low effort. Let

$$V_e(n, s, \delta) = c(e, \delta) + E[V(n - 1, s, \delta)], \quad (5)$$

then $G(n, s, \delta) = V_1(n, s, \delta) - V_0(n, s, \delta)$. Theorem 1 implies that G is increasing in s for a given (n, δ) . Using our results from the estimation section of our paper, we plot G over s given $n = 2$ and $\delta = 0$. As demonstrated in figure 2, G is non-linear and, as predicted, increasing in s .

3.4 Empirical Tests

Theorem 1 provides a way for us to test our model using reduced form analysis. It implies that for fixed (n, δ) , a worker's effort level is increasing in s . To test this implication

and to estimate an upper bound on the effect of a high level of effort on the time taken to sort checks, we perform two experiments. First, we examine whether worker’s effort levels vary with s as predicted by theory. We accomplish this by looking at the last job workers processed in a day. Looking at this subset of the data allows us to examine worker behavior for a relatively small range of n while still observing a wide range of s . In addition, theory predicts that within a day, workers are more likely to alter their effort levels at the end of the day. As such, analyzing this set of jobs will most clearly highlight the effect of incentives on worker behavior. The second experiment we run estimates the upper bound on the difference between high and low effort on the time it takes to sort checks. We do this by measuring the difference in workers’ productivity residuals for the case when workers have extremely high and extremely low s . In effect, we are measuring the difference in worker productivity when a worker will earn a bonus with probability one versus the case when a worker will earn a bonus with probability zero.

Using the data, we can compute two of the worker’s state variables, (n, s) , for every job (i.e. for every observation). Effort, naturally, is unobserved. However, we can measure the effect of effort on the time taken to process a job by looking at the residuals of a productivity regression. The regression we run has the time taken to complete a job as the dependent variable. The independent variables are the job’s observable characteristics, 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 d_j be the day in which the job was processed, and i_j be the worker who processed the job. Note there are 366 days and 15 workers in the sample. The regression we ran is

$$\ln(\tau_j) = \ln(\beta' \bar{z}_j) + \sum_{l=1}^{366} \xi_l \cdot 1_{l=d_j} + \sum_{k=1}^{15} \eta_k \cdot 1_{k=i_j} + \nu \cdot n_j + \varepsilon_j, \quad (6)$$

where β is a vector of coefficients and $1_{x=y}$ is a dummy variable equal to 1 when $x = y$. As effort is unobserved, its effect on time is captured by the residual term of this regression.

While the above regression captures day and operator effects, it does not control for operator-day shocks. As such, our results from using the residuals of the above regression provide us with an upper-bound on the effect of effort. In the next section, we use a structural approach to separate out the effects of daily shocks and effort on the time

Table 2: Conditional Means of Residuals

	Obs	Mean	Std Error
High Prob	13941	-0.0037	0.0023
Low Prob	319	0.0791	0.0205

Note: ‘High Prob’ is defined as s being greater than half an hour
‘Low Prob’ is defined as s being less than half an hour

taken to sort checks.

After computing these residual terms, we then look at the subset of the data containing the last job of the day that a worker completed, and measure the relationship between these residuals and s . Theorem 1 predicts a negative correlation between s and the residuals, as a higher s induces higher effort, reducing the time taken to sort checks. Confirming the theoretical prediction, we find that the Pearson correlation coefficient is -0.06 and is strongly significant.

To compute an upper bound on the effect of effort on the time taken to sort checks, we look at the subsets of the data where theory predicts workers will exert high effort and where they will exert low effort. As workers spend, on average, 2 to 3 hours a day sorting checks, a worker with an s greater than half an hour has a high probability of earning a bonus and so will exert a high level of effort. In contrast, a worker with an s less than negative thirty (i.e. they are half an hour behind) has a low probability of earning a bonus and will exert a low level of effort. To analyze the gain in time due to a high level of effort, we create two subsets of the data based on s being larger than thirty, and less than negative thirty. We then compute and compare the mean value of the residual terms in both subsets. As mentioned above, the residual term captures the effect of effort on the time taken to process a job. Table 2 contains the results. The difference between the conditional means signifies that a high level of effort reduces the time taken to sort checks by at most 8%, compared to the case where low effort is exerted. Hence, despite the automated nature of the check sorting process, workers do play a large role in determining how fast checks are sorted. Another interesting aspect of Table 2, is that unlike the conditional mean in the low probability case, the conditional mean of the residuals in the high probability case is not significantly different from zero. This is a function of the observations in each category. In the high probability case, there are almost 14,000 observations, while the low probability case only contains 319. This reflects the fact that workers earned a bonus 95% of the time under the first incentive pay

regime and 81% under the second. As the entire data set has 34,076 observations, it is not surprising that the conditional mean in the high probability case is not significantly different from zero. To check the sensitivity of our results to the number of observations, we re-computed our results using different cutoffs for s . There were no significant changes in our results.

These results suggest that effort plays an important role in the time taken to process checks and that workers do respond to incentives as predicted by theory. To better understand how worker's react to incentives and to compute the cost of effort to the worker, we estimate the structural parameters of the model.

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 c , and the distributions of \bar{z} , ε , and δ . In specifying the functional form of τ , there are three main issues that we consider. First, workers are heterogeneous in their productivity, as demonstrated by the wide range in mean bonus payments listed in Table 1. To capture this, we include fixed effects, η_i , in τ .

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. For instance, we checked and rejected the hypothesis that reader-sorter machines processed checks at different speeds.

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. 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 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 of 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.¹²

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 and cubed terms of the observable characteristics to the set of regressors. These

¹²We 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 .

additional non-linear terms did not add any explanatory power, 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

$$\tau^i(e, \bar{z}, \varepsilon) = \left[\beta_0 + (\beta_1 - \beta_3 \cdot e) \cdot z^1 + (\beta_2 - \beta_4 \cdot e) \cdot z^2 \right] \cdot \exp(\eta_i) \cdot \exp(\varepsilon), \quad (7)$$

where $\beta_3, \beta_4 \geq 0$, $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, and $i = 1, \dots, 15$. Taking logs, we get the form of the equation we actually estimate,

$$\ln(\tau^i(e, \bar{z}, \varepsilon)) = \ln \left[\beta_0 + (\beta_1 - \beta_3 \cdot e) \cdot z^1 + (\beta_2 - \beta_4 \cdot e) \cdot z^2 \right] + \eta_i + \varepsilon. \quad (8)$$

In this specification, the heterogeneity in workers is captured by adjusting the coefficients β_0 through β_4 by a fraction η_i . We constrain the coefficients β_3 and β_4 to be non-negative 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.

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 .

Turning to the cost of effort, we specify c as

$$c(e, \delta) = e \cdot \gamma \cdot \exp(\delta), \quad (9)$$

where $\gamma \geq 0$ and $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$. Hence, the cost of low effort is normalized to zero, while the cost of high effort is $\gamma \cdot \exp(\delta)$. Note that as γ is the same for all workers, they are only heterogeneous in η_i .

With these functional forms, we can compute the likelihood of the model. We have a panel data set of I individuals, where we observe each individual for D_i days. A day is composed of $J_{i,d}$ jobs. Individuals are heterogenous in that they have different fixed effects, η_i . In this data set, we have information on the job's observable characteristics, $\hat{z}_{i,d,j}$, the number of checks left to sort in the day, $\hat{n}_{i,d,j}$, the time taken to sort jobs, $\hat{\tau}_{i,d,j}$,

and the benchmark times computed by the firm, $\hat{\alpha}_{i,d,j}$. The likelihood is then

$$L(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \gamma, \{\eta_i\}, \sigma_\varepsilon, \sigma_\delta | \{\hat{z}_{i,d,j}, \hat{n}_{i,d,j}, \hat{\alpha}_{i,d,j}, \hat{\tau}_{i,d,j}\}) = \prod_{i=1}^I \prod_{d=1}^{D_i} \int_{-\infty}^{\infty} \prod_{j=1}^{J_{i,d}} P\left(\ln(\hat{\tau}_{i,d,j}) | \{\hat{z}_{i,d,j}, \hat{n}_{i,d,j}, \hat{\alpha}_{i,d,j}\}, \beta_0, \dots, \beta_4, \gamma, \eta_i, \sigma_\varepsilon, \delta\right) \phi_\delta(d\delta), \quad (10)$$

where $\phi_x = \mathcal{N}(0, \sigma_x^2)$. From equation 8, we know that the likelihood of the observation $\ln(\hat{\tau}_{i,d,j})$, conditional on the data and parameters specified above, is $\phi_\varepsilon\left(\ln(\hat{\tau}_{i,d,j}) - \theta_{i,d,j}\right)$, where

$$\theta_{i,d,j} = \ln\left[\beta_0 + \left(\beta_1 - \beta_3 \cdot e_{i,d,j}\right) \cdot \hat{z}_{i,d,j}^1 + \left(\beta_2 - \beta_4 \cdot e_{i,d,j}\right) \cdot \hat{z}_{i,d,j}^2\right] + \eta_i. \quad (11)$$

The variable $e_{i,d,j}$ is the effort level exerted by worker i , while sorting a job j on day d . 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, δ) . We observe n directly in the data, $\hat{n}_{i,d,j}$, and can compute s from the two sequences $\{\hat{\alpha}_{i,d,j}, \hat{\tau}_{i,d,j}\}$. We integrate over the distribution of the daily shock δ .

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 that a worker's effort level is increasing in the probability of earning a daily bonus. So 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. Theory, then, 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.

Having daily shocks in our model does not affect this avenue of identification. The effect of daily shocks is captured by the variation in worker's productivity across days, while effort's effect on productivity is captured by within day variation in productivity. For instance, when a worker's productivity is low throughout the day, the model ascribes this behavior to a bad daily cost shock, δ , and constant low effort. Similarly, if a worker's productivity is high throughout the day, the model interprets this sequence of events as a good daily shock and constant high effort. In contrast, when a worker begins the day with high productivity but ends the day with low productivity, the model ascribes this behavior to a bad job productivity shock, ε , and the worker's subsequent declining effort level. The identification of effort, then, comes from the dynamics of the data within the day. Estimates of the significance of the daily shock are derived from variation of productivity across days.

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 , and β_4 through the utility-maximizing behavior of workers. These cross-equation restrictions allow us to estimate γ .

In addition to this general source of identification, we can also identify effort through another feature of the firm's bonus scheme. The firm uses an ad hoc formula (equation 1) to determine how well a worker is performing. Hence, it is possible for a worker to be doing well relative to a true measure of productivity, but to be performing badly relative to the firm's incentive scheme. In such a case, despite the fact that the worker is actually performing well, the worker will give up and exert low effort. This difference between true productivity and the firm's formula provides us with another way to separate out the effect of random daily shocks and the effect of bonus-maximizing effort decisions by workers.

A third source of identification comes from the firm's change of parameters in the firm's incentive scheme roughly 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 switch in the bonus scheme is that workers systematically changed when they would exert effort, influencing their productivity. In an environment where effort did not matter, however, a change in the 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 third avenue of identification.

4.3 Estimation Algorithm

As mentioned in the introduction, we use a simulated 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.¹³ 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 $s > \bar{s}_n$. 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. Our technique is *simulated* maximum likelihood, as computing the model’s likelihood entails integrating over the distribution of the daily cost shock.

To find the set of parameters that maximize the likelihood, we use a two step approach. We first search over the parameter space using a simulated annealing program with a large tolerance setting.¹⁴ This algorithm, while slow, performs well at searching over a large parameter space. In addition, the likelihood we are maximizing is a step function along certain dimensions, which the simulated annealing algorithm is adept at handling. We then take the result from the simulated annealing algorithm, and plug it into a standard simplex based algorithm with a small tolerance setting. This algorithm is faster than the simulated annealing one, and searches well within a local area.

¹³See, for example, Rust (1987) and Pakes (1986).

¹⁴A good source on how a simulated annealing algorithm works is Goffe, Ferrier, and Rogers (1994).

Table 3: Parameter Estimates

Parameter		Estimate	Standard Error
		(seconds)	
Intercept	β_0	357.357	1.5e-4
Jams	β_1	31.465	1.4e-5
Fields	β_2	3.831	1.2e-6
Effort on Jams	β_3	6.560	1.5e-5
Effort on Fields	β_4	0.839	1.1e-6
		(dollars)	
Cost of effort	γ	0.455	6.2e-6
		(no units)	
Standard deviation of period shock	σ_ε	0.179	1.3e-6
Standard deviation of daily shock	σ_δ	1.747	3.6e-3

We repeat this second step several times, until the maximum likelihood results from consecutive searches are within 0.1 of each other.

4.4 Parameter Estimates

Using these functional specifications, we obtained the parameter estimates listed in Table 3. All parameter estimates in Table 3 are highly significant as are the fixed effect estimates listed in Table 4.

These estimates imply that when a worker does not exert effort, entering a field takes 3.8 seconds. With effort, a field is entered 0.8 seconds faster, a 21% reduction in time. Clearing a jam when not exerting effort typically takes a worker 31 seconds. With effort, a worker takes 24 seconds, a 23% reduction in time. The gain from effort with regard to clearing jams is easy to see. Motivated workers will simply perform the necessary steps required to clear a jam faster than an unmotivated one. Effort improves the speed with which a worker enters a field (e.g. the account number on a check) in a different manner. Entering numbers on a reader/sorter machine is a task much like touch-typing, where typing speed depends upon the worker's concentration. For this task, then, effort improves the worker's focus on the particular job at hand. 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 17%.

Table 4: Worker Grade and Fixed Effect
High Grade

Worker	1	2	3	4	5	6	7	8	9	10
Grade	7	7	7	6	6	6	6	6	6	6
Fixed Effect	-0.031	0.085	0.096	0.028	0.004	0.014	0.315	0.175	0.429	0.135

Low Grade

Worker	11	12	13	14	15
Grade	5	5	5	4	4
Fixed Effect	0.104	0.240	0.104	0.385	0.301

Turning to the cost of effort, notice that γ is 0.45. 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 \$0.45.

We included fixed effects in order to capture worker heterogeneity. As shown in Table 4, workers widely differ in skill levels. The difference between the best (worker 1) and the worst (worker 9), is considerable. Our results imply that worker 9 takes over 40% longer to sort checks than worker 1. Interestingly, both these workers have high grades, suggesting that skill alone does determine a worker’s grade. It is true, however, that most of the high grade workers sort checks faster than the low grade workers.

4.5 Analysis of Effort Decisions

Using these parameter estimates, we can analyze workers’ effort decisions. The model infers that workers exerted effort 75% and 66% 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 exerted effort by 9%. Recall that the firm’s bonus scheme was changed in two ways. First, the firm made it harder for workers to be eligible for a bonus by altering the parameters in the formula time α . Second, the firm increased the marginal return of effort to workers, conditional on being eligible for a bonus, by increasing the bonus wage B . Note that these two changes have opposing effects on the worker’s effort decision problem. The result that workers exert less effort after the change indicates that overall, the firm’s decision to increase the difficulty of earning a bonus had a much larger effect on workers’ effort decisions than increasing the bonus wage B . Table 5 shows the effects of the change in the incentive scheme at the worker level

Table 5: Percentage of High Effort Decisions

Worker	Grade	First IP Regime	Second IP Regime	Difference
All	n/a	75%	66%	-9%
1	7	78%	75%	-3%
2	7	75%	75%	0%
3	7	77%	74%	-3%
4	6	76%	75%	-1%
5	6	76%	75%	-1%
6	6	76%	75%	-1%
7	6	75%	69%	-6%
8	6	75%	73%	-2%
9	6	69%	16%	-53%
10	6	75%	74%	-1%
11	5	76%	74%	-2%
12	5	73%	73%	0%
13	5	75%	74%	-1%
14	4	73%	27%	-46%
15	4	74%	69%	-5%

While none of the workers increased the percentage of times they exerted effort under the second bonus regime, there is a lot of variation in how much workers decreased how often they exerted effort. Workers 2 and 12 continue to exert the same amount of effort before and after the incentive change. In contrast, workers 9 and 14 drastically lower the amount of effort they exert. This dramatic change, however, is not surprising given the data on their daily bonuses. As shown in Table 1, both these workers went from earning daily bonuses of over \$5, to bonuses of less than a dollar. By looking at the worker's policy rules, we were able to determine why these two workers behaved so differently from the group. Under the first regime, all workers start the day exerting effort, as long as they do not receive a bad daily shock. This is also true under the second regime, except for workers 9 and 14. They perceive that their chances of earning a bonus are so low under the new bonus scheme, that they no longer exert effort in the first period unless they receive a good daily shock. So unlike their co-workers, these two workers only exerted effort after receiving a particularly good daily or period shock.

Knowing how often workers exert effort allows us to determine the increase in productivity due to the firm's bonus plan. Using our parameter estimates, we compute that exerting effort typically decreases the time spent processing checks by 16.5%. This

implies that under the first incentive pay plan, workers' high effort decisions decreased the time spent sorting checks by $(0.75 \cdot 0.16) = 12.0\%$. The decrease in time under the second incentive pay scheme is slightly less, at 10.6%. 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 11.9%.

These results illustrate the large inefficiencies associated with the firm's incentive scheme. The kinked nature of the bonus scheme discourages workers from exerting effort after receiving a bad daily or period shock. Under the first bonus scheme, this results in workers exerting effort only 75% of the time, while under the second workers exert effort 66% of the time. Consequently, the firm's bonus scheme increases worker productivity by only 11.9%, which is significantly less than 16.5% increase in productivity associated with the first best scheme, where workers exert effort all the time.

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 a first-best scheme. We choose these two schemes as we are interested in comparing the case without any incentives to the one where incentives are the most effective. By looking at the worker's problem, it is clear that if workers are only paid an hourly rate, they will never exert a high level of effort. To solve for the first-best scheme, we take the environment we specified earlier and make the additional assumptions that the firm wants to induce the worker to always exert effort and to minimize the cost (i.e. wages).¹⁵ In this environment there are several schemes the firm can use to achieve the first-best. A simple one is for the firm to pay workers a bonus after every effort decision, conditional on whether or not the firm observes the worker exerting effort.

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,

¹⁵By first-best scheme, we mean the optimal scheme for the firm in an environment where the firm can costlessly observe the worker's actions. To solve for this scheme, we use the fact that the firm signs contracts with its customers to sort checks within a short period of time. In this deadline oriented environment, processing checks as fast as possible is valuable to the firm. Hence, we assume that even if the worker receives a large daily cost, it is still worthwhile for the firm to motivate the worker to exert effort.

Table 6: Welfare Analysis per Worker per Day

	Flat Wage ($e=0$)	First Best ($e=1$)	Difference in Welfare
Minutes Spent Sorting	128	107	21
Cost of Effort (γ)	\$0	$5 \cdot \$0.45$	-\$2.25

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 128 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 107 minutes, 21 minutes faster than without effort. The expected disutility the worker experiences from exerting effort is $5 \cdot \$0.45 = \2.25 , 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, though it is clearly positive. As such, we only consider the welfare effects of freeing up the worker's time, and interpret our estimates of the welfare gain as a lower bound. The welfare effects of freeing up the worker's time equals $\$10.55 \cdot \frac{21}{60} = \3.69 . The total expected gain in welfare from using the first-best scheme is then $\$3.69 - \$2.25 = \$1.44$, which is a 6.3% increase in welfare. 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 exerted a high level of effort 75% of time under the first bonus scheme and 66% of the time under the second. Hence, under the first incentive pay scheme, welfare was increased by $0.75 \cdot 6.3\% = 4.7\%$. 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 4.5% due to the firm's bonus plan.

4.7 Goodness of Fit

To determine the goodness-of-fit of our estimated parameters, we perform two tests. First, we perform a likelihood ratio test. Second, we use the firm’s policy shift in the bonus scheme to perform an out-of-sample prediction which we can then check against the data.

The null hypothesis in the likelihood ratio test is that both β_3 and β_4 are equal to zero. The alternative hypothesis, that these parameters are greater than zero, is the model we originally estimated. Finding the maximum likelihood of the model under the null hypothesis is straightforward, as workers will always choose low effort. Taking the ratio of the likelihood for each model yields a test statistic which allows us to strongly reject the null hypothesis. This implies that accounting for workers’ effort decisions is important when analyzing worker productivity in this environment.

The second goodness-of-fit test is an out-of-sample prediction. As discussed in Section 2, two-thirds of the way through our data sample the firm changed the bonus scheme in two ways. First, the formula time α was altered to give less time to workers to complete each job. Second, the bonus wage B was raised, increasing the marginal return to effort for workers, if they are eligible for a bonus. As the first effect creates less incentive for workers to work hard, while the second effect creates more incentive, it is unclear what will happen to worker productivity after the policy change.

This policy shift provides an excellent opportunity to test how well the model is fitting the data. As we have data before and after the policy change, we can estimate the model’s parameters on one part of the data and use it to predict worker productivity on the second part. Then, we can compare the model’s predictions to the actual data and judge how well the model performs.

To measure the change in worker productivity in the data from the policy change, we run the following regression:

$$\ln(\tau_{i,t}) = \ln(\beta' \bar{z}_{i,t}) + \sum_{j=1}^{15} \eta_j \cdot 1_{j=i} + \varphi \cdot 1_{IP} + \varepsilon_{i,t}, \quad (12)$$

where $\tau_{i,t}$ is the time taken to complete job t by worker i . The dummy variable $1_{j=i}$ is equal to 1 when $j = i$, while 1_{IP} is equal to 1 after the switch in the bonus scheme. As shown in Table 7, the estimate of φ from the data is 0.04, which implies that the

Table 7: Actual and Predicted Productivity Change

	Estimate of φ	
Data	0.041	(0.0030)
Simulation	0.019	(0.0005)

Note: Standard errors are in parenthesis

time taken to sort checks increased by 4% after the change in the bonus scheme.¹⁶ We then estimated the model only using the data from after the policy change. We use this subset of the data as there are more instances of workers not earning a daily bonus in this time frame. These observations are crucial for identifying the effect of effort on the time taken to sort checks. Using these parameter estimates, we simulated the model to get predicted times on the jobs completed. Using these simulated times, we re-ran the above regression and estimated φ to be 0.019. Hence, the model correctly predicts the sign of the productivity change, but is off on the magnitude.

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 11.9%. 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 the firm's incentive scheme increases the welfare of the firm and workers by 4.5%.

¹⁶Note that we use the transformed data, where each job has 1000 checks.

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 by a similar amount.

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. 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.

The results from this paper suggest several areas of future research. One extension would be to investigate the effect of incentives on new workers. In particular, do incentives encourage workers to learn faster? More generally, fully modelling the firm’s problem and understanding the constraints it faces is of interest. With such an approach, the effect of adverse selection can be measured. In addition, within this general framework other relationships, such as the connection between worker turnover and incentives, could be explored.

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.
- CHEVALIER, J., AND G. ELLISON (1999): “Risk Taking by Mutual Funds as a Response to Incentives,” *Journal of Political Economy*, 105, 1167–1200.
- CHIAPPORI, P., J. HECKMAN, AND J. PINQUET (2000): “Testing for Moral Hazard on Dynamic Insurance Data,” University of Chicago mimeo.
- CHIAPPORI, P., AND B. SALANIÉ (2000): “Testing Contract Theory: A Survey of Some Recent Work,” Invited Lecture at the World Congress of the Econometric Society.
- 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 (1999): “The Response of Worker Effort to Piece Rates,” *Journal of Human Resources*, pp. 643–667.

- (2000): “Fixed Wages, Piece Rates, and Incentive Effects: Statistical Evidence from Payroll Records,” *International Economic Review*, 41:1, 59–92.
- PAKES, A. (1986): “Patents as Options: Some Estimates of the Value of Holding European Patent Stocks,” *Econometrica*, 54, 755–784.
- PRENDERGAST, C. (1999): “The Provision of Incentives in Firms,” *Journal of Economic Literature*, 37, 7–63.
- RUST, J. (1987): “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55, 999–1033.