The Welfare Effects of Incentive Schemes

Adam Copeland† Cyril Monnet ‡

May 8, 2008

Abstract

This paper computes the change in welfare associated with the introduction of incentives. 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 for the average worker. Using our parameter estimates, we show that the cost of increased effort due to incentives is equal to the dollar value of a 5% rise in productivity. Welfare is measured as the output produced minus the cost of effort, hence the net increase in the average worker’s welfare due to the introduction of the firm’s bonus plan is 7%. Under a first-best scheme, we find that the net increase in welfare is 9%.

*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 suggestions. In addition, this paper has benefited from participants at numerous seminars and comments from an editor and anonymous referees. 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 Federal Reserve Bank of Philadelphia, the Federal Reserve System, the U.S. Bureau of Economic Analysis, or the U.S. Department of Commerce.

†Bureau of Economic Analysis (adam.copeland@bea.gov)
‡Federal Reserve Bank of Philadelphia (Cyril.Monnet@phil.frb.org)
1 Introduction

Incentives are often used by firms to encourage their employees to work hard. The bonus plans used vary widely, from complicated profit sharing plans to simple employee-of-the-month awards, and, in large part, work well.\(^1\) Introducing a simple piece-rate scheme can increase productivity over 20\% (Lazear 2000), while using group bonus schemes to take advantage of team-based work structures can also significantly boost worker productivity.\(^2\)

But boosting productivity through incentives is costly to workers, because higher effort is physically or mentally draining. 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. Little empirical work has focused on measuring the cost of effort and determining the welfare effects of incentive schemes. Our paper helps fill this void in the literature 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 compute the welfare gains from using incentives, by determining how much output and the disutility from effort vary 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 estimate effort’s effect on output. 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 earn zero bonus pay. This kink in the bonus profile creates the perverse incentive for a worker to quit working hard in the latter 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. 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. This result allows us to measure effort’s effect on output by comparing the difference in productivity between workers with low and high probabilities of earning a daily bonus.

We are able to make these productivity comparisons, because the firm has provided

---

\(^1\)Prendergast (1999), Lazear and Oyer (2008) for recent surveys of the personnel literature.

\(^2\)For example, see Ichniowski, Shaw, and Premnushi (1997), Hamilton, Nickerson, and Owan (2003), and Boning, Ichniowski, and Shaw (forthcoming).
us with a detailed data set on worker output. 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 us with the details of its incentive plan.

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. 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. With this information, the worker is able to determine how well he has performed relative to the incentive pay scheme, and to compute the probability of receiving incentive pay at the end of the day. Because the bonuses are calculated on a daily basis, a worker starts each day anew.

We estimate the model using a maximum likelihood approach. 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 the average worker’s 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 the average worker is equal to a 5% gain in productivity, which is almost half of the gain in output due to incentives. Because welfare is measured as the amount of output produced minus the cost of effort, our results show the net increase in the average worker’s welfare from the firm’s bonus plan is 7%. We also show that under a first-best scheme, the net increase in welfare would be 9%. As we discuss in more detail later, this measure of welfare ignores issues of selection, focusing only on the moral hazard effects of incentives.

The paper adds to the empirical personnel literature in two ways. First, in addition to measuring the effect of incentives of productivity, we compute the welfare gains from using incentives. Second, our paper is one of the first in this literature to rigorously study dynamic incentive issues. This paper adapts the “dynamic data” approach used in the general literature on testing contract theory to measure the effects of moral hazard on worker productivity.\(^3\)

Our paper builds upon the body of work studying the interaction of incentives and the production process. Our paper is closest to Ferrall and Shearer (1999), who estimate the structural parameters of a principal-agent model using a mining firm’s payroll

\(^3\)Chiappori and Salanić (2000) provides a survey of recent work using this “dynamic data” approach.
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. Introducing dynamics is important for two reasons. We show dynamic forces play an important role in understanding worker’s effort decisions when firms employ non-linear incentive schemes. Further, the dynamics of the worker problem, in conjunction with the firm’s non-linear incentive scheme, provide the opportunity for us to measure the effect of effort on productivity. In contrast, the vast majority of the literature uses static models of worker behavior and relies on observing changes to workers’ incentives to measure effort’s effect on productivity (e.g. Lazear (2000), Hubbard (2003), and Hamilton, Nickerson, and Owan (2003)), or convincing firms to run controlled experiments (e.g. Shearer (2004), Bandiera, Barankay, and Rasul (2005), and Bandiera, Barankay, and Rasul (2007)). In this respect, our paper is similar to the work of Chevalier and Ellison (1999), who use data that capture agents’ dynamic behavior to measure the effect of moral hazard in mutual fund markets.

From a broader perspective, our paper is related to the empirical literature of managerial incentive pay (e.g. Jensen and Murphy (1990) and Margiotta and Miller (2000)). This literature measures the moral hazard effects of bonuses on managerial performance, though mostly within static settings. Our results emphasize the importance of considering dynamic behavioral models, an issue especially relevant for executive compensation given the long-term decisions managers make and the equity bonuses they often receive. Indeed, Oyer (1998) presents empirical results consistent with employees adjusting effort over the fiscal year in order to maximize their year-end compensation.

Finally, since we have little information on the firm’s objective function, we abstract from most issues related to the firm’s payoff structure in our paper and focus instead on how workers react to the bonus scheme in place. This scheme is likely to be sub-optimal if we consider a standard objective for the firm. In the case where the principal and agent have linear or constant relative risk aversion (exponential) utility, Holmstrom and Milgrom (1987) show that the optimal dynamic incentive scheme is linear in accounts, whereby the principal counts the number of realizations of each possible output, and remunerates a realization according to a transfer which is optimal in the non-repeated set-up. The final payment to the agent, given its optimal strategy, is then the sum of the one-period payments appropriately weighted by the number of accounts. In the current paper, the firm’s choice of a kinked bonus profile can be reconciled with the prediction
of Holmstrom and Milgrom (1987), if the firm’s payoff is not a simple linear function of output. This may indeed be the case, because sorting checks is a deadline-oriented business, where missing a deadline by an hour is disproportionately more costly than beating a deadline by an hour.

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. Workers’ main task is to sort checks by running them through a sorting machine (see figure 1). Ideally, the sorting machine processes checks without any worker input. If checks jam in the machine, however, or fields on the check cannot be electronically read, then workers are required to intervene by clearing jammed checks or typing 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. Our production data is at the job level and includes information on which worker ran a job, when it was run, how long it took, and its characteristics. The time taken to run a job is measured as the time elapsed between the first and last check sorted, and so excludes the time necessary to setup a sorting machine. This measure of time also excludes workers’ breaks, because they can stop the clock by logging off of the sorting machine. Hence, this is a precise measure of the time taken to sort checks.

The human resource component of the data set provides us with information about the tenure of workers, as well as their wage-grade level. There is a notable amount of turnover, and three-fourths of exiting workers are relatively new. Management interprets

---

4 Checks are first processed by the High Speed Check Processing 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.

5 An example of a field is the account number.
this turnover as new workers learning about the difficulties of the job and laments the
cost of training workers who leave after only a few weeks on the job. Because learning
confounds effort’s effect on worker productivity, we decided to only analyze full-time
workers who were employed in the check-sorting department for at least 6 months before
the beginning of our sample period. Of the original 45 workers, only 15 met these
requirements. This group of workers completed 34,056 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, there is a wide range in the number of jobs a
worker completed within a day (1-50) as well as the number of checks sorted by job
(1-8,500). Typically however, workers run 6 to 7 jobs a day, where each one averages
660 checks in length and takes 18 minutes to complete. In addition, workers on average
clear 9 jams and type in 128 fields per job. As evidenced by the large number of jams
and field corrections, worker input is a large determinant of how fast a job is completed.
As the statistics indicate, workers typically spend roughly two hours a day on the sorting machine. Workers usually perform jobs back-to-back, and often a shift’s work on the sorter is continuous. The remainder of a worker’s 8 hour shift is spent on other tasks, such as integrating checks sorted by this department with those sorted by the High Speed Checking Department.

The firm uses an incentive pay arrangement that rewards workers based on their daily performance on sorting machines above and beyond their hourly wage. This mechanism uses 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 $z$ be a vector of $C$ characteristics of a job, the firm’s formula, $\alpha(z)$, has the following form:

$$
\alpha(z) = \rho_0 + \rho_1 z^1 + \ldots + \rho_C z^C,
$$

(1)

where $(\rho_0, \rho_1, \ldots, \rho_C)$ are coefficients set by the firm. When a worker completes $N$ jobs in a day, and $z_n$ denotes the characteristics of job $n$, then $\sum_{n=1}^{N} \alpha(z_n)$ 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 $\tau_n$ be the actual time a worker spends on a job and define $s = \sum_{n=1}^{N} [\alpha(z_n) - \tau_n]$. The variable $s$ is the amount of time a worker is behind or ahead the benchmark time at the end of the day. We can write the firm’s bonus payment scheme as

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

(2)

where $B > 0$ is some constant. The firm has provided us with $B$ and the coefficients of the bonus scheme, which enables us to reproduce the daily cutoff times each worker faced.

Importantly, the firm changed the parameters of the bonus scheme in January of 2000, roughly two-thirds of the way through the sample period. Before the switch, $B = \$7.17$ for all employees. After the switch, $B = \$9.75$ for workers with a grade of 4 or 5 and $B = \$12.76$ for those with a grade of 6 or 7. In addition, the coefficients of $\alpha$

---

6Lower grade employees earn lower wages and are typically newer employees relative to their higher grade counterparts.
were changed 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 $B$ was increased.

The incentive scheme was changed because the original incentive structure had been calibrated on older, slower sorting machines. One of management’s goals with the bonus scheme was to only reward workers who performed above average. With the new sorting technology, however, workers were earning a bonus most of the time. Consequently, management decided to re-calibrate the incentive scheme to make it more difficult for workers to be eligible for a bonus. This change in the bonus structure was not part of a larger re-organization of the check-sorting operation. As such, the shift in the compensation schedule is similar to a natural experiment, because other aspects of the workers’ environment (e.g. the degree of monitoring by supervisors) were held constant over the time period.

Table 1 contains the average bonus payments by worker under the two regimes, as well as the fraction of total jobs where a positive bonus was earned. After the switch, workers earned smaller bonuses less often. The average daily bonus of all workers decreased from $10.05 to $6.64, and 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 regimes. At the other extreme, worker 9 received a mean bonus of $5.45 under the first regime and $0.22 under the second. Workers also reacted differently to the change in incentives. Workers 3 and 7 both earned an average bonus of roughly $9 a day under the first incentive pay regime. Under the second regime however, worker 3’s average bonus increased to $10.49 while worker 7’s dropped to $2.75.

To check the importance of these bonus payments to workers, we computed the mean of the ratio of bonus pay to total wages. Because these workers performed other tasks not under the purview of the incentive scheme, we restricted this ratio 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. 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 0.004 to 0.33 for the first and second

---

7 The introduction of these newer sorting machines occurred over a year before the start of our sample.

8 Hourly wages for employees range from $8 to $14 an hour.
Table 1: Bonus Payments

<table>
<thead>
<tr>
<th>Worker</th>
<th>Grade</th>
<th>First IP Regime</th>
<th>Second IP Regime</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (Std)</td>
<td>Frac</td>
</tr>
<tr>
<td>All n.a.</td>
<td>$10.05 (8.54)</td>
<td>0.95</td>
<td>$6.64 (7.62)</td>
</tr>
<tr>
<td>High grade 6 &amp; 7</td>
<td>$10.93 (9.19)</td>
<td>0.96</td>
<td>$8.63 (8.50)</td>
</tr>
<tr>
<td>Low grade 4 &amp; 5</td>
<td>$8.46 (6.94)</td>
<td>0.92</td>
<td>$3.32 (4.07)</td>
</tr>
<tr>
<td>1 7</td>
<td>$14.56 (12.46)</td>
<td>1.00</td>
<td>$11.19 (11.30)</td>
</tr>
<tr>
<td>2 7</td>
<td>$8.52 (7.96)</td>
<td>0.95</td>
<td>$7.28 (6.33)</td>
</tr>
<tr>
<td>3 7</td>
<td>$9.74 (11.33)</td>
<td>0.97</td>
<td>$10.49 (8.80)</td>
</tr>
<tr>
<td>4 6</td>
<td>$14.29 (10.03)</td>
<td>0.98</td>
<td>$14.04 (9.32)</td>
</tr>
<tr>
<td>5 6</td>
<td>$18.22 (6.48)</td>
<td>1.00</td>
<td>$16.13 (5.33)</td>
</tr>
<tr>
<td>6 6</td>
<td>$15.23 (8.94)</td>
<td>0.98</td>
<td>$12.59 (8.88)</td>
</tr>
<tr>
<td>7 6</td>
<td>$9.14 (4.78)</td>
<td>1.00</td>
<td>$7.25 (2.80)</td>
</tr>
<tr>
<td>8 6</td>
<td>$6.10 (6.33)</td>
<td>0.93</td>
<td>$5.44 (5.78)</td>
</tr>
<tr>
<td>9 6</td>
<td>$5.45 (3.98)</td>
<td>0.84</td>
<td>$0.22 (0.74)</td>
</tr>
<tr>
<td>10 6</td>
<td>$6.95 (6.38)</td>
<td>0.95</td>
<td>$4.59 (4.58)</td>
</tr>
<tr>
<td>11 5</td>
<td>$11.05 (8.14)</td>
<td>0.98</td>
<td>$4.67 (4.63)</td>
</tr>
<tr>
<td>12 5</td>
<td>$5.35 (6.02)</td>
<td>0.80</td>
<td>$4.64 (3.31)</td>
</tr>
<tr>
<td>13 5</td>
<td>$14.71 (6.24)</td>
<td>1.00</td>
<td>$5.30 (5.37)</td>
</tr>
<tr>
<td>14 4</td>
<td>$5.27 (3.39)</td>
<td>0.90</td>
<td>$0.54 (1.07)</td>
</tr>
<tr>
<td>15 4</td>
<td>$5.40 (3.42)</td>
<td>0.95</td>
<td>$1.86 (2.32)</td>
</tr>
</tbody>
</table>

Frac: Fraction of days with positive bonus pay

While incentives are beneficial in that they increase worker productivity, they also adversely affect 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. The firm is able to detect when incorrect field numbers have been entered and, because of the time that correcting mistakes takes, 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.

Paarsch and Shearer (2000) consider a tradeoff between quantity and quality with regard to the use of incentives.
3 The Model

In this section 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. Each check $n = 1, ..., N$ is characterized by a pair $(z_n, \varepsilon_n)$, where $z_n$ is a random vector of characteristics that is observed by the firm and the worker, and $\varepsilon_n$ is a random shock that is only observed by the worker. Both $z_n$ and $\varepsilon_n$ are i.i.d. throughout the day. All workers also face a common daily random shock to productivity $\nu$ that is i.i.d. across days. Workers can exert effort to sort each check $n$, where effort is a binary choice, $e_n = \{0, 1\}$. A low level of effort corresponds to a baseline level of effort that workers need to exert in order to keep their jobs at the firm. Given the daily shock $\nu$, check characteristics $(z_n, \varepsilon_n)$ and effort $e_n$, the time it takes to sort check $n$ is given by $\tau(e_n, z_n, \varepsilon_n, \nu)$, which is a decreasing function of $e_n$. We assume effort is costly and that workers receive an individual shock $\delta$ to their cost of effort, where $\delta$ is i.i.d. across workers and days. The cost of effort is denoted by

$$c(e, \delta) = e\gamma \exp(\delta)$$

where $\gamma > 0$. Workers have a linear utility function that is separable in wage and effort and we assume there is no discounting within the day.

Because the firm does not observe $\varepsilon$, the firm cannot determine the effort level exerted by the worker—as in the standard moral hazard model. Unlike the standard moral hazard model, workers do not receive compensation after each job, but rather at the end of the day after sorting $N$ checks and making $N$ effort decisions. Because this paper focuses on worker behavior, we take the contracts offered to workers as exogenous and

---

10 Each 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.

11 We are implicitly assuming that the determinants of a worker’s baseline level of effort, for example the threat of being fired and a worker’s outside opportunities, remain constant over the sample period. Supervisors ensure that workers exert a baseline level of effort by periodically checking on workers’ progress throughout the day.

so 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 $\hat{b}$ (see equation 2). A worker’s utility after having processed $N$ checks is

$$\bar{w} + \hat{b} \left( \sum_{n=1}^{N} \alpha(z_n) - \tau(e_n, z_n, \epsilon_n, \nu) \right) - \sum_{n=1}^{N} c(e_n, \delta).$$

The first term is the worker’s base wage, the second term is the daily bonus, and the last term is the cost of effort over the entire day.

The timing of events plays an important part in the model. At the beginning of the day, workers observe $(\nu, \delta)$, and then start sorting checks. Before processing each check, workers choose their effort level. While the check is being sorted, the two random shocks, $(z, \epsilon)$, 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 that they are distinguishable.\(^{13}\)

Several assumptions deserve further discussion. First, a strong assumption on both $z_n$ and $\epsilon_n$ is that they are i.i.d. throughout the day. However, it may 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 $z_n$ are correlated within a day. Turning to $\epsilon_n$, it may be correlated within the day because of a common unobserved characteristic across checks (e.g. one day’s shipment of checks may be wet, and so harder to sort) or a common cost shock to the worker (e.g. a worker feels sick). To capture the first effect, we introduced the daily shock to productivity $\nu$ that is common to all workers. To model the second effect, we introduced the daily shock $\delta$ to the worker’s cost of effort. Even with the addition of daily productivity shocks, we acknowledge that the additivity and independence of the per-period unobserved productivity shocks are very specific modeling assumptions. As detailed in Magnac and Thesmar (2002) and Aguirregabiria and Mira (2002), however, these assumptions are inline with the vast majority of empirical work on dynamic decision problems, following, among others, Rust (1988).\(^{14}\) Nevertheless, the independence assumption could lead to biased estimates if the true model involves serial correlation in $\epsilon$ over the day.

Second, our assumption that a worker’s utility is linear in wages and effort implies risk

\(^{13}\)We interpret effort as mental exertion, where a high level of effort enables workers to quickly respond when the sorting machine requires their input.

\(^{14}\)See Eckstein and Wolpin (1989) for an older, but still relevant, survey of the different empirical strategies to estimate dynamic discrete choice models.
neutrality. This allows us to directly compare the dollar gains from output with the cost of effort in our welfare analysis, and simplifies the worker’s decision problem. Further, because the variation in a worker’s total income due to bonuses is small and 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. In addition, from an empirical perspective, risk aversion is likely not identified in this environment where a worker’s actions are not observed.

Finally, we implicitly assume that workers cannot make payments to the firm ($\bar{w} \geq 0$) and so are subject to limited liability. Otherwise, it would be possible to achieve the first best level of effort by proposing a two-parts wage. By requiring the worker to make a fixed payment to the firm in case of a bad outcome, the firm provides enough of an incentive for the worker to exert the first best effort level. However Innes (1990) showed in a static environment with limited liability, the payment to the worker will be an increasing function of the outcome, when it is above a threshold, and a constant below that threshold. Hence, the optimal static contract with limited liability shares some features of the dynamic incentive scheme we study here.

### 3.2 The Worker’s Problem

The worker decides, for every check, whether effort should be exerted, given the bonus scheme. When $\bar{n}$ checks have been sorted, the history of events a worker observes is the tuple $(\delta, \nu, \{(e_n, z_n, \varepsilon_n)\}_{n=1,\ldots,\bar{n}})$. This information allows the worker to determine how well the worker is doing with respect to the firm’s formula $\alpha$. A sufficient statistic for this history is the variable $s_n$, where

\[
\begin{align*}
    s_{n+1} &= s_n + \alpha(z_n) - \tau(e_n, z_n, \varepsilon_n, \nu), \quad \text{and} \\
    s_1 &= 0.
\end{align*}
\]

We can then write the worker’s problem recursively as an $N$ period stochastic dynamic problem. Given the worker only gets paid at the end of the day, but incurs the cost of

---

15To provide a stronger basis for this claim, we computed the certainty equivalence 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.
effort each period, the worker’s value function is

\[ V(n, s_n, \nu, \delta) = \begin{cases} 
\max_{e \in \{0,1\}} \{-c(e, \delta) + E[V(n + 1, s_{n+1}(e, z, \varepsilon, s_n, \nu), \nu, \delta)]\} & \text{if } n = 1, 2, \ldots, N \\
\bar{w} + \bar{b}(s_n) & \text{if } n = N + 1 
\end{cases} \tag{6} \]

where the expectation is taken over \((z, \varepsilon)\).

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

### 3.3 Theoretical Results

To better understand the worker’s problem, we analyze two comparative statics of the worker’s policy function. First, we examine how \(e(n, s, \nu, \delta)\) changes with respect to \(n\), holding \((s, \nu, \delta)\) fixed. Decreasing \(n\) provides the worker with more opportunities to affect \(s\). For a highly skilled worker, this increase in opportunities is beneficial because the worker has more chances to increase \(s\) and so offset any bad draws of \((z, \varepsilon)\). Conversely, for a less skilled worker who struggles to sort checks faster than the benchmark time, decreasing \(n\) is not beneficial. Hence, without knowing the skill level of a worker, we are unable to predict whether \(e(n, s, \nu, \delta)\) is increasing or decreasing in \(n\).\(^{16}\)

Second, we turn to examining how \(e(n, s, \nu, \delta)\) changes with respect to \(s\), for a fixed \((n, \nu, \delta)\). 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.

**Theorem 1.** For all \((n, \nu, \delta)\), the policy function \(e(n, s, \nu, \delta)\) is increasing in \(s\).

**Proof.** The proof is by induction. Suppose first \(V(n, s, \nu, \delta)\) is increasing and convex in \(s\). We first show that if \(e(n - 1, s_0, \nu, \delta) = 1\) then for all \(s_1 > s_0\), \(e(n - 1, s_1, \nu, \delta) = 1\). Suppose \(e(n - 1, s_0, \nu, \delta) = 1\), then it must be the case that,

\[
E[V(n, s_0 + \alpha - \tau(1, z, \varepsilon, \nu), \nu, \delta) - V(n, s_0 + \alpha - \tau(0, z, \varepsilon, \nu), \nu, \delta)] > c(1, \delta) - c(0, \delta).
\]

\(^{16}\)Using a probit model, we infer that the probability of earning a bonus is decreasing in \(n\) for all workers.
because otherwise no effort would have been exerted. Given $V(n, s, \nu, \delta)$ is increasing and convex in $s$, we know that for $s_1 > s_0$,

$$E[V(n, s_1 + \alpha - \tau(1, z, \varepsilon, \nu), \nu, \delta) - V(n, s_1 + \alpha - \tau(0, z, \varepsilon, \nu), \nu, \delta)] >$$

$$E[V(n, s_0 + \alpha - \tau(1, z, \varepsilon, \nu), \nu, \delta) - V(n, s_0 + \alpha - \tau(0, z, \varepsilon, \nu), \nu, \delta)],$$

which implies that $e(n - 1, s_1, \nu, \delta) = 1$.

We can then show that $V(n - 1, s, \nu, \delta)$ is increasing and convex in $s$, using the fact that $e(n - 1, s, \nu, \delta)$ is increasing in $s$ as shown above. The proof then follows as $V(N + 1, s, \nu, \delta)$ is increasing and convex by construction.

\[\square\]

### 3.4 Empirical Tests

Theorem 1 shows how to measure effort’s effect on productivity. Based on a worker’s probability of earning a daily bonus, we are able to distinguish moments within a shift when a worker exerts a low level of effort from instances when a high level of effort is exerted. By comparing a worker’s productivity given the inference of low and high levels of effort, we are able to measure effort’s effect on productivity. Figure 2 illustrates such an example in the data. The horizontal axis lists, chronologically, the jobs a worker completed within the day, while the worker’s state variable is on the vertical axis. Over the first three jobs the worker performs well relative to the benchmark time. Consequently, we infer that the worker exerts a high level of effort over these jobs. The fourth job, however, is a disaster, and leaves the worker 15 minutes behind the benchmark time, with half the work day over. At the start of the fifth job, the worker recognizes the low probability of earning a bonus, and so we infer the worker exerts a low level of effort. Indeed, throughout the remainder of the day the worker’s state variable remains negative.

In contrast, figure 3 depicts a day when the worker experiences a high probability of earning a daily bonus throughout the day, and so we infer the worker always exerts a high level of effort. If we only observed instances where we infer that a single effort level is exerted, we would not be able to measure effort’s effect on productivity.

To test whether the implications of Theorem 1 hold in the data, we use a reduced form approach to estimate how productivity in the first and second half of a worker’s shift changes with variation in the probability of earning a bonus. To accomplish this, we select 2 subsets of the data: (a) workers who, at the midpoint of their shift, are more
Figure 2: Changing Effort During a Shift

Figure 3: Constant Effort During a Shift
than 20 minutes ahead of the firm’s benchmark time, \( s > 20 \), and (b) workers who are 20 minutes behind, \( s < -20 \). At the midpoint of their shift, the first subset of workers have a higher probability of earning a bonus, relative to when they started their shift and \( s = 0 \). Consequently, we expect these workers to exert a higher level of effort and have a higher level of productivity in the second half of their shift relative to the first. The same reasoning applies to the second subset of workers; these workers should have a lower level of productivity in the second half of the shift relative to the first half.

To measure the change in productivity from the first to second half of a shift, for each subset of workers we ran a regression of the time taken to complete a job on that job’s characteristics, \( \mathbf{X}_t \), allowing the estimated parameters to differ between the first and second halves of workers’ shifts. Let \( L_t \) equal 1 if a job is part of a worker’s second shift, and equal 0 otherwise. Let \( \text{oper}_t = \{1, 2, \ldots, 15\} \) denote the operator number of each worker. The regression we estimate is

\[
\ln(\tau_t) = \ln\left( (1 - L_t) \left[ \mathbf{X}_t \beta^1 \right] + L_t \left[ \mathbf{X}_t \beta^2 \right] \right) + \sum_{i=2}^{15} I_{\text{oper}_t = i} \eta_i.
\]

The job characteristics are the number of checks sorted, jams cleared and operator-corrected fields. Both \( \beta^1 \) and \( \beta^2 \) are \( 3 \times 1 \) parameter vectors to be estimated, \( I_{x=y} \) is an indicator variable equal to 1 when \( x = y \), and \( \eta \) are worker fixed effects. The afternoon dummy variable allows us to compare the average time taken to sort checks in the first and second half of the same day, controlling for daily productivity shocks. Because it is difficult to determine the magnitude of the changes in the time taken to complete jobs by directly comparing the estimated parameters, \( \hat{\beta}^1 \) and \( \hat{\beta}^2 \), we compute an expected time taken to complete jobs using the distribution of job characteristics in the data. Precisely, we let

\[
T_1 = \frac{1}{N} \sum_{t=1}^{N} \mathbf{X}_t \hat{\beta}^1, \quad T_2 = \frac{1}{N} \sum_{t=1}^{N} \mathbf{X}_t \hat{\beta}^2,
\]

where \( N \) is the total number of jobs in the data.

For the first subset of workers we find that worker’s productivity increases from the first to second shift. On average, these workers complete jobs 20 minutes faster in the afternoon after they had a good morning (see table 2). For the second subset of workers, we find that worker’s productivity dramatically declines. On average, workers take 275 minutes longer to complete jobs in the afternoon given they had a bad morning, a steep falloff in productivity.
Workers with a good first half

<table>
<thead>
<tr>
<th>Median</th>
<th>Mean</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>400.5</td>
<td>920.2</td>
</tr>
<tr>
<td>$T_2$</td>
<td>394.9</td>
<td>900.4</td>
</tr>
</tbody>
</table>

Workers with a bad first half

<table>
<thead>
<tr>
<th>Median</th>
<th>Mean</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>531.5</td>
<td>1102.4</td>
</tr>
<tr>
<td>$T_2$</td>
<td>624.5</td>
<td>1377.3</td>
</tr>
</tbody>
</table>

Table 2: Within Day Comparisons of Worker Productivity

These results suggest that effort plays an important role in the time taken to process checks and that workers 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 the model. We begin by stating and justifying our functional form specifications. Then we summarize our estimation technique, report our parameter estimates, and discuss their implications.

4.1 Specification

In order to estimate the model, we need to specify the functional form of $\tau$, and the distributions of $z, \varepsilon, \nu$, and $\delta$. There are three main issues that we consider. First, because workers are heterogeneous in their productivity (see table 1), we include worker fixed effects, $\eta$, in $\tau$.

Second, we consider which of a number of check characteristics are important predictors of $\tau$. To this end, we regressed all observable characteristics in the data on time. We find 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 spent sorting checks. Other characteristics have at most marginal explanatory power. For instance, we rejected the hypothesis that reader-sorter machines processed checks at different speeds.
Our specification of \( \tau \) only includes the number of jams and fields corrected as a job's observable characteristics because of computational constraints. Solving the worker's problem for a large number of checks took a prohibitively long time. As such, we decided to approximate the model by assuming that workers made an effort decision every 1,000 checks. This, however, raised a problem with the structure of our data. As discussed in section 2, an observation in our data set is a job, which ranges from 1 to over 8,000 checks. To bring the model specification and data into line, we re-arranged the data to construct observations of 1,000 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 1,000 check job. Under this modification, the number of checks per job in the data is constant and so can be captured by \( \tau \)’s intercept term. Consequently, we construct the random variable \( z \) of a job’s characteristics as a \( 2 \times 1 \) vector, where \( z^1 \) is the number of jams that occurred and \( z^2 \) is the number of fields corrected.

While reducing \( z \) to two dimensions is computationally advantageous, it complicates the problem of computing a worker’s expectations over \( \alpha(z) \). 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 a linear approximation of the firm’s actual incentive pay formula based on the number of jams and fields corrected.\(^{17}\)

Third, we consider how \( e, z, \varepsilon, \) and \( \nu \) interact with one another. We found that a log-log specification of the regression of time-to-sort on check characteristics fit the data the best.\(^{18}\) With regard to effort, our observations in the workplace and conversations with workers lead us to believe that effort has a direct effect on both how quickly jams are cleared and fields are entered. Consequently, we interact the effort term in \( \tau \) with both elements of \( z \). With these issues in mind, we specify \( \ln(\tau) \) as

\[
\ln(\tau_i(e, z, \varepsilon, \nu)) = \ln\left[ \beta_0 + \left( \beta_1 - \beta_3 \cdot e \right) \cdot z^1 + \left( \beta_2 - \beta_4 \cdot e \right) \cdot z^2 \right] + \eta_i + \varepsilon + \nu. \tag{7}
\]

where \( \beta_3, \beta_4 \geq 0, \varepsilon \sim N(0, \sigma_\varepsilon^2), \nu \sim N(0, \sigma_\nu^2), \) and \( i = 1, \ldots, 15 \). Heterogeneity across

\(^{17}\)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 .

\(^{18}\)We also ran the above regression adding squared and cubed terms of the observable characteristics to the set of regressors. These additional non-linear terms did not add any explanatory power.
workers is captured by adjusting the coefficients $\beta_0$ through $\beta_4$ by a fixed effect $\eta_i$. We constrain the coefficients $\beta_3$ and $\beta_4$ to be non-negative because our prior is that effort lowers the time it takes to complete a job.

To fully specify the cost of effort (see equation 3), we assume that $\delta \sim N(0, \sigma_\delta^2)$. Finally, we use the observations of characteristics in the data to construct an empirical distribution for $z$.

With these functional form assumptions, we can compute the likelihood of the model. Because workers receive common day shocks, an observation is a day. We have a panel data set of $I$ individuals over $D$ days. A day is composed of $J_{di}$ jobs. We have data on the job’s observable characteristics, $\hat{z}_{dij}$, the number of checks sorted in the day, $\hat{n}_{dij}$, the time taken to sort jobs, $\hat{\tau}_{dij}$, and the benchmark times computed by the firm, $\hat{\alpha}_{dij}$. Letting $\phi_x$ be the pdf of a normal distribution with mean zero and variance equal to $\sigma_x^2$, we write the likelihood as

$$L(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \gamma, \{\eta_i\}, \sigma_\varepsilon, \sigma_\nu, \sigma_\delta | \{\hat{z}_{dij}, \hat{n}_{dij}, \hat{\alpha}_{dij}, \hat{\tau}_{dij}\}) =$$

$$\prod_{d=1}^{D} \int_{-\infty}^{\infty} \prod_{i=1}^{I} \int_{-\infty}^{\infty} \prod_{j=1}^{J_{di}} \frac{1}{\sqrt{2\pi\sigma_\varepsilon}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} [\ln(\hat{\tau}_{dij}) - \ln(\tau_{dij})]^2\right) \phi_\delta(\delta) d\delta \phi_\nu(\nu) d\nu,$$

where, from equation 7,

$$\ln(\tau_{dij}) = \ln\left[\beta_0 + \left(\beta_1 - \beta_3 \cdot e_{dij}\right) \cdot \hat{z}_{dij}^1 + \left(\beta_2 - \beta_4 \cdot e_{dij}\right) \cdot \hat{z}_{dij}^2\right] + \eta_i + \nu.$$

The variable $e_{dij}$ 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 data, through backwards induction.

Identification in this paper rests upon the threshold in the firm’s daily bonus plan, which is determined by the coefficients of the firm’s formula time (see equation 1). If firm chooses to set the threshold to negative infinity, then $\beta_3$ and $\beta_4$ in equation 7 would not be separately identified from $\beta_1$ and $\beta_2$, respectively. Under these circumstances, each worker would always exert the same level of effort throughout the day. Because we incorporate daily productivity shocks, the lack of variation in effort within the day results in the likelihood being flat in $\{\beta_3, \beta_4\}$, the parameters which measure the gains from effort.\(^{19}\)

\(^{19}\)A similar argument can be made if the kink is set to positive infinity, which results in workers always earning a flat hourly wage.
4.2 Estimation Algorithm

We use a maximum likelihood approach to estimate the structural parameters of the worker’s problem. We choose maximum likelihood because it has the minimum variance achievable by a consistent and asymptotically normally distributed estimator, and computing the likelihood of the model is not computational burdensome. We did not use the conditional choice probability approach (see Hotz and Miller (1993) and Aguirregabiria and Mira (2002)), because the advantages of this alternative approach depend upon observing an agent’s choices, which is not the case here.

To find the set of parameters that maximize the likelihood function, we use a two step approach. We first search over the parameter space using a simulated annealing program with a large tolerance setting.\(^20\) 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. We repeat the second step several times, until the maximum likelihood results from consecutive searches are within 0.0001 of each other.

4.3 Goodness-of-Fit

Before discussing our estimated parameters, we assess how well the model fits the data. To do this, we examine the model’s ability to predict the time-to-sort values in the data, and we analyze how well the model performs in an out-of-sample test.

To determine how well the model’s predicted time-to-sort values match the data, we plot these values against a 45 degree line (see figure 4). Two features stand out. First, there is a band of observations where the model predicts much longer time-to-sort values relative to the data. These observations occurred in the same week, across all workers and all types of jobs. Clearly something unusual and unobserved happened this week that is not in the model. As detailed in table 3, while the model misses on these 653 observations, it does well in matching the remaining 21,795 observations.

Second, the model tends to under-predict the time taken on longer runs, which not surprisingly have more jams and more operator corrections. Most likely, these longer runs had difficult jams or operator corrections for which the model has a hard time

\(^{20}\)A good source on how a simulated annealing algorithm works is Goffe, Ferrier, and Rogers (1994).
Table 3: Mean Values of Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Obs.</th>
<th>ln(actual time)</th>
<th>ln(predicted time)</th>
<th>Jams</th>
<th>Opcorr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odd week</td>
<td>653</td>
<td>3.09 (0.02)</td>
<td>2.52 (0.02)</td>
<td>15.24 (0.42)</td>
<td>196.49 (5.29)</td>
</tr>
<tr>
<td>Rest of data</td>
<td>21,795</td>
<td>3.23 (0.00)</td>
<td>3.10 (0.00)</td>
<td>13.44 (0.05)</td>
<td>192.86 (0.76)</td>
</tr>
</tbody>
</table>

Note: Obs. stands for number of observations and mean standard errors are in parenthesis.
accounting. The curvature of the log-log specification of the time-to-sort production function dampens the influence of these observations.

Our second approach to assessing how well the model fits the data is to perform an out-of-sample test. As discussed in Section 2, two-thirds of the way through our data sample the firm changed the bonus scheme. Taking advantage of this policy change, we test how well the model, when estimated only using data from one regime, predicts the change in worker productivity associated with the switch in bonus schemes.

To measure the change in worker productivity in the data due to the change in regimes, we run a regression of the log of the time taken to sort checks on a constant term, the number of jams and operator corrections that occurred, worker fixed-effects, a trend variable, and a dummy for the change in regime.\(^{21}\) The estimated coefficient on the dummy variable is 0.025 and has a standard error of 0.0053, implying that on average the time taken to sort checks fell 2.5% with the change in regime.

To obtain the model’s out-of-sample prediction, we first estimate the model only using data from the second regime. Using these estimated parameters (see the estimates labeled “2nd Regime” in table 4), we simulate the model under both incentive regimes. We then run the same regression described above on the simulated times. The estimated coefficient on the dummy variable for the incentive regime is 0.021 with a standard error of 0.0044. Hence, the model predicts that the time taken to sort checks will fall 2.1% under the second bonus scheme, correctly predicting the sign of the productivity change, and explaining 84% of the observed change in productivity.\(^{22}\)

### 4.4 Parameter Estimates

Having confidence that the model fits the data well, we report our estimated parameters in table 4. We report two sets of parameters and their standard errors.\(^{23}\) The “Full Sample” estimates are based on all the data in our sample, while the “2nd Regime” estimates are based on data after the change in the incentive regime.\(^{24}\) The two sets of

---

\(^{21}\) We used the transformed data, where each job has 1000 checks.

\(^{22}\) Using the non-linear, time-to-sort-check specification of our model, we estimate a 2.7% fall in worker productivity in the data with the introduction of the new incentive scheme. Using the simulated data, productivity falls 2.2%; thus the model accounts for 81% of the observed change in productivity.

\(^{23}\) We computed the standard errors of the parameters by inverting the information matrix of the sample.

\(^{24}\) We did not estimate the model on the sample of data only from the first regime because we infer that there is little variation in workers’ effort levels. As reported on table 1, workers earned a bonus 95% of the time in the first regime. Hence, there are few opportunities to observe worker productivity
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Full Sample Estimates</th>
<th>2nd Regime Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>388.894 (1.1259)</td>
</tr>
<tr>
<td>Jams</td>
<td>$\beta_1$</td>
<td>37.208 (0.0873)</td>
</tr>
<tr>
<td>Fields</td>
<td>$\beta_2$</td>
<td>3.644 (0.0102)</td>
</tr>
<tr>
<td>Effort on Jams</td>
<td>$\beta_3$</td>
<td>10.690 (0.0877)</td>
</tr>
<tr>
<td>Effort on Fields</td>
<td>$\beta_4$</td>
<td>0.571 (0.0055)</td>
</tr>
<tr>
<td>Cost of effort</td>
<td>$\gamma$</td>
<td>0.423 (0.0047)</td>
</tr>
<tr>
<td>Std. dev. of period shock</td>
<td>$\sigma_\varepsilon$</td>
<td>0.139 (0.0002)</td>
</tr>
<tr>
<td>Std. dev. of daily cost shock</td>
<td>$\sigma_\delta$</td>
<td>5.453 (1.1167)</td>
</tr>
<tr>
<td>Std. dev. of daily productivity shock</td>
<td>$\sigma_\nu$</td>
<td>0.031 (0.0006)</td>
</tr>
<tr>
<td>Number of observations</td>
<td></td>
<td>22,448</td>
</tr>
<tr>
<td>Percentage of High Effort in 2nd regime</td>
<td></td>
<td>68</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parenthesis.

estimates are quite close to one another, reassuring us of their robustness. Further, both models produce similar predictions about workers’ effort decisions, as evidenced by the same predicted percentage of overall high effort decisions in the second regime.

The “Full Sample” estimates imply that when a worker exerts a low level of effort, entering a field takes 3.6 seconds. With a high level of effort, a field is entered 0.6 seconds faster, a 17% reduction in time. Clearing a jam when not exerting high effort typically takes 37 seconds. With a high level of effort, it takes 26.5 seconds, a 28% reduction in time. To better understand the impact of exerting a high level of effort, consider that a typical batch of 1,000 checks requires a worker to clear 14 jams and type in 193 fields. Our parameter estimates imply that with a high level of effort, the time a worker spends processing checks decreases by 16.1%.

We estimate that the average cost of a high level of effort, $\gamma$, is 0.42. Because we assumed that utility is additively separable in effort and wage, this value can be

when the probability of earning a bonus is low. In contrast, during the second regime workers earned a bonus 81% of the time.
Table 5: Worker Grade and Fixed Effect

<table>
<thead>
<tr>
<th>Worker</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Fixed Effect</td>
<td>0.08</td>
<td>0.07</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.27</td>
<td>0.17</td>
<td>0.41</td>
<td>0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Worker</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Fixed Effect</td>
<td>0.13</td>
<td>0.22</td>
<td>0.11</td>
<td>0.40</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Note: Worker 1’s fixed effect is set to 0.

interpreted as the average dollar cost to a worker for choosing \( e = 1 \). Thus, the expected disutility from exerting a high level of effort while sorting 1,000 checks is $0.42.

Finally, our estimates of worker fixed effects show that worker heterogeneity is important. As shown in Table 5, workers widely differ in skill levels, where the fixed effect of worker 1 is normalized to 0. 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 not determine a worker’s grade. In general, however, high grade workers sort checks faster than low grade workers.

4.5 Analysis of Effort Decisions

Using the parameter estimates, we can analyze workers’ effort decisions. The model infers that workers exerted effort 76% and 68% of the time respectively under the firm’s two incentive regimes. Hence, the changes the firm made to the bonus scheme decreased the number of times workers exerted effort by 8%. Our results show 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 6 shows the effects of the change in the incentive scheme at the worker level. None of the workers increased the percentage of times they exerted effort under the second regime, although there is a lot of variation in how much workers decreased the frequency with which they exerted effort. For example, worker 2 continues to exert the same amount of effort before and after the incentive change. In contrast, workers 9 and 14 drastically lowered the amount of effort they exerted, a result inline with the data on their daily bonuses. As shown in Table 1, both
Table 6: Percentage of High Effort Decisions

<table>
<thead>
<tr>
<th>Worker</th>
<th>Grade</th>
<th>First IP Regime</th>
<th>Second IP Regime</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>n.a.</td>
<td>76%</td>
<td>68%</td>
<td>-8%</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>80%</td>
<td>77%</td>
<td>-3%</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>77%</td>
<td>77%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>79%</td>
<td>76%</td>
<td>-3%</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>78%</td>
<td>77%</td>
<td>-1%</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>78%</td>
<td>77%</td>
<td>-1%</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>78%</td>
<td>77%</td>
<td>-1%</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>76%</td>
<td>71%</td>
<td>-5%</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>76%</td>
<td>75%</td>
<td>-1%</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>70%</td>
<td>20%</td>
<td>-50%</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>77%</td>
<td>76%</td>
<td>-1%</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>78%</td>
<td>76%</td>
<td>-2%</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>75%</td>
<td>74%</td>
<td>-1%</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>77%</td>
<td>76%</td>
<td>-1%</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>74%</td>
<td>27%</td>
<td>-47%</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>76%</td>
<td>69%</td>
<td>-7%</td>
</tr>
</tbody>
</table>

these workers went from earning daily bonuses of over $5, to bonuses of less than a dollar with the introduction of the new incentive scheme.

Knowing how often workers exert effort allows us to determine the increase in productivity due to the firm’s bonus plan. As mentioned earlier, our parameter estimates imply that exerting effort typically decreases the time spent processing checks by 16.1%. Hence, under the first regime, workers’ high effort decisions decreased the time spent sorting checks by \((0.76 \cdot 16.1) = 12.2\%\). The decrease in time under the second regime is slightly less, at 10.9%. 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. Consequently, the firm’s bonus scheme increases worker productivity by only 11.9%, significantly less than the 16.1% increase in productivity when workers exert effort all the time.
4.6 Welfare Analysis

We now analyze the welfare effects of the firm’s bonus plan. We consider welfare in the economy under two extreme cases: a flat-wage scheme and a first-best scheme. It is clear workers will never exert a high level of effort under a flat-wage scheme. To solve for the first-best scheme, we take the environment we specified earlier and make the additional assumptions that the firm observes effort, wants to induce the worker to always exert effort, and seeks to minimize wages.\(^{25}\) 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 observing the worker exerting a high level of effort.

With these two 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 5,000 checks, under the two schemes for each worker. From the data, we know that while sorting 5,000 checks, a worker will typically clear 70 jams and type in 966 fields. Using these numbers, we compute the welfare gains for each worker as well as the average welfare gain across all workers (see table 7). Because workers all have the same cost of effort, they only differ in their productivity. Worker 1, one of the fastest sorters, typically takes 156 minutes to sort 5,000 checks in the flat-wage case where no effort is exerted, assuming the daily productivity shock equals its mean of zero. Naturally, the disutility from effort is equal to 0. Under the first-best scheme, we find worker 1 will typically sort 5,000 checks in 131 minutes, 25 minutes faster than without effort. The expected disutility worker 1 experiences from exerting effort is \(5 \cdot 0.42 = 2.10\), as \(e = 1\) is chosen 5 times (once for every 1,000 checks), and assuming the daily cost shock equals its mean of zero. 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 25 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 25 minutes to perform other tasks. 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 25 minutes. The value of this

\(^{25}\)By 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 7: Welfare Analysis per Worker per Day

<table>
<thead>
<tr>
<th>Worker</th>
<th>Payment Schemes</th>
<th>Difference in Welfare</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Flat Wage (e=0)</td>
<td>First Best (e=1)</td>
</tr>
<tr>
<td></td>
<td>(minutes)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>156</td>
<td>131</td>
</tr>
<tr>
<td>2</td>
<td>169</td>
<td>141</td>
</tr>
<tr>
<td>3</td>
<td>167</td>
<td>140</td>
</tr>
<tr>
<td>4</td>
<td>156</td>
<td>130</td>
</tr>
<tr>
<td>5</td>
<td>157</td>
<td>131</td>
</tr>
<tr>
<td>6</td>
<td>157</td>
<td>132</td>
</tr>
<tr>
<td>7</td>
<td>204</td>
<td>171</td>
</tr>
<tr>
<td>8</td>
<td>185</td>
<td>155</td>
</tr>
<tr>
<td>9</td>
<td>236</td>
<td>198</td>
</tr>
<tr>
<td>10</td>
<td>175</td>
<td>146</td>
</tr>
<tr>
<td>11</td>
<td>177</td>
<td>149</td>
</tr>
<tr>
<td>12</td>
<td>194</td>
<td>162</td>
</tr>
<tr>
<td>13</td>
<td>232</td>
<td>195</td>
</tr>
<tr>
<td>14</td>
<td>201</td>
<td>169</td>
</tr>
<tr>
<td>15</td>
<td>174</td>
<td>146</td>
</tr>
<tr>
<td>Average</td>
<td>183</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>5 · $0.42</td>
<td></td>
</tr>
</tbody>
</table>

Cost of Eff
extra time to the firm is positive, though harder to quantify. 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 worker 1’s time equals $10.55 \cdot \frac{25}{60} = $4.40. The expected gain in welfare from using the first-best scheme for worker 1 is then $4.40 - $2.10 = $2.30, which is a 8.3% increase in welfare (see columns 4 and 5 of table 7). Averaging across all workers in the sample, we find that workers typically take 183 minutes to sort 5,000 checks when exerting a low level of effort. With a high level of effort, the average worker takes 153 minutes, resulting in a 9.3% increase in welfare. Across workers, the welfare gains range from a low of 8.3% to a high to 11.0%.

Focusing on the firm, we compare how different compensation schemes impact aggregate productivity. Using the estimated parameters, we simulate the model and compute the average time spent sorting checks and the average percentage of jobs when workers exerted effort, under a flat-wage scheme, the firm’s actual bonus scheme, and the first-best incentive scheme (see table 8). Not surprisingly, as workers are coaxed to exert effort, the time taken to sort checks over our sample falls. Under a flat wage compensation scheme, workers take 11,866 hours to sort all the checks in our sample. Driven by worker effort, the time taken to sort checks falls 10.7% to 10,595 hours under the firms actual incentive scheme. Moving from the actual scheme to the first-best scheme further drops the time taken to sort checks by another 6.0%.

Using this approach, we can easily compute the welfare gain from using the firm’s incentive schemes. Recall our estimates imply that workers exerted a high level of effort 76% and 68% of time under the first and second regime respectively. Hence, under the first regime, the typical worker’s welfare increased by $0.76 \cdot 9.3\% = 7.1\%$. We similarly calculate the welfare increase in the second 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 6.9% due to the firm’s bonus plan.

Because the workers we analyze worked throughout the sample period, this paper
does not address the selection effects associated with the change in the incentive scheme. Presumably, bonus schemes would encourage higher skilled workers to the firm, increasing productivity and so raising the welfare gains from implementing a bonus scheme. On the other hand, if workers learn their skill level while on the job, bonus schemes may permanently increase turnover, resulting in a higher level of hiring and training costs. As discussed in section 2, there is a notable amount of turnover in the check-sorting department that we study. Preliminary analysis showed that exiting workers included both unproductive and highly productive people, suggesting that exit decisions are based on more than just the check-sorting activities of this job or there is significant heterogeneity in outside options. Further, only one worker started at the firm after the switch in the incentive scheme. As such, we found the data had little to say about the selection effects of the firm’s bonus scheme. Exploring the relationship between the incentive pay scheme and the supply of labor for check-sorters is an interesting question and particularly pertinent to this firm. Unfortunately, because our sample of the data contains little information on selection, we leave these important issues for further research.

A surprising conclusion from our analysis is that welfare dropped when the firm implemented its new incentive scheme. This is driven by the 8% decline in the average level of effort that workers’ exerted under the new bonus regime. This result questions the firm’s decision to alter the bonus system in such a way as to lower the probability of worker’s earning a bonus. This conclusion, however, is based on a simplified firm’s problem that ignores the selection effect of incentives mentioned above. To fully evaluate the switch in incentive schemes from the firm’s perspective requires more detailed modeling of the firm and labor market. Analyzing this more complicated environment, however, is beyond the scope of this paper.

5 Summary and Conclusion

Most empirical work on incentive pay has considered its effects on productivity within static environments. This paper builds upon the literature by fully accounting for the effects of incentives—measuring both the gain in worker productivity and the cost of effort exerted. The paper further extends the literature by considering the dynamic effects of incentive contracts when measuring how incentives can change worker productivity.

We accomplish this by studying a check-clearing firm’s use of incentives. Using the

---

26See Russell (2005) for recent work in this area.
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%. A little less than half 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 6.9%.

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.

This paper, however, has focused on only part of the relationship between the firm and worker—how full-time, experienced workers react to changes in their compensation scheme. There are broader questions with regard to incentives that would be interesting to analyze within a complete principal-agent model. There is, for example, significant turnover within the firm which may have been influenced by the bonus structure. Further, there may be selection effects as different types of workers choose jobs to sort checks under the new bonus scheme. Lastly, the use of incentives may influence how fast new worker’s learn their job. These are important issues that require modeling the firm’s problem and understanding how labor is supplied in this industry.
References


