Edward P. Lazear
University of Chicago and Hoover Institution

Salaries and Piece Rates*

Some workers receiv
compensation that is
specified in advance
and not directly conti
gent on performance.
Instead, it depends o
an input measure, suc
as hours worked. For
others, compensation
directly related to out
put. This essay is an
attempt to predict a
firm's choice of com-
pensation method.
Piece rates are define
more rigorously.
Among the more im-
portant factors dis-
cussed are worker
heterogeneity, incen-
tives, sorting conside
ations, monitoring
costs, and asymmetri
information. One resu
is that salary workers
tend to be of lower
quality and more ho-
mogeneous than are
their piece-rate
counterparts. Numer-
ous additional results
are provided.

Compensation can take many forms. Remuneration can come as pecuniary payments, as fringes such as health and pension benefits, or as a nonpecuniary reward such as plush office furniture that costs the firm less than it benefits the worker. A significant literature has examined the trade-offs between pecuniary and nonpecuniary compensation, the modern work having been pioneered by Rosen (1974).

More recently, another body of literature has examined the selection of method of total compensation, ignoring the distinction between pecuniary and nonpecuniary payment. This work has focused on risk and incentive factors. It has resulted in comparisons of compensation based on absolute output levels to that based on relative performance.[1] It has also led to explorations of the relation of compensation to experience over the work life.[2]

Little attention has been paid to what may be among the most important and obvious distinction in methods of compensation, namely, the choice between a fixed salary for some period of time, that is, paying on the basis of input and

* Helpful comments by Victoria Lazear and Yoram Weiss are gratefully acknowledged. Support was provided by the Department of Labor and the National Science Foundation.

1. See Lazear and Rosen (1981), Stiglitz (1981), Holmstrom (1982), Green and Stokey (1983).
2. See Lazear (1979, 1981) and Harris and Holmstrom (in press).

(*Journal of Business*, 1986, vol. 59, no. 3)

0021-9398/86/5903-0002$01.50

paying a piece compensation that is specifically geared to output.[3] Two extreme examples are illustrative. Unskilled farm labor often is paid in the classic piece-rate fashion: an amount of payment per pound or piece harvested is specified in advance.

Near the other extreme are middle managers of major corporations whose annual salaries are specified in advance, and who are then paid exactly that amount, independent of output. The qualifier is that, if effort falls below some specified level (e.g., he does not come to work regularly), the manager may be terminated.

Why are some workers paid piece rates based on output while others are paid salaries for their input? There are a number of common explanations, most of which center around monitoring costs. When it is costly to measure output, it is sometimes argued, workers are paid salaries. When monitoring costs are low, piece-rate payment is appropriate. Although there surely is much truth to this, it leaves a number of issues unresolved. Given the lack of clarity, it seems useful to pursue these points in greater depth. This paper focuses on a number of issues that affect the choice between salaries and piece rates. The most important are concerned with sorting workers across jobs, inducing appropriate effort levels, and selecting quantity versus quality of output. Additionally, intertemporal strategic behavior is considered. I begin with an attempt to define more concretely what is meant by "salary" and "piece rate."

There is, of course, a large body of literature on compensation schemes. This essay uses those theories as well as some new analysis and combines it with other work on information and incentives to derive a number of concrete predictions. In particular, the goal is to provide a positive analysis of factors that determine the choice of payment by input over payment by output. The current literature, with a few exceptions, leaves large gaps here.

I conclude with a sketch of an empirical methodology. Results are summarized in the last section.

I. Definitions

The important feature that distinguishes a piece rate from a salary is that, with a piece rate, the worker's payment in a given period is related to output in that period. If the worker is paid a piece rate, then

$$w_t = f(q_t), \tag{1}$$

3. There are two recent exceptions. Pencavel (1977) discusses some of the same issues that are addressed in this paper. He also provides some evidence on punch-press operators in Chicago. Seiler (1984) presents empirical evidence based on 100,000 employees in the footwear industry.

where w_t is compensation in period t, and q_t is worker output in period t.

In its purest sense, salary is defined as compensation that depends on input in the current period. Thus salaried workers receive

$$w_t = g(E_t), \tag{2}$$

where E_t is (some measure of) effort in period t. Payment is contemporaneous with output for piece-rate workers. Salaried workers receive compensation that is not contemporaneous with output but that is contemporaneous with effort. The measure of effort might be hours worked. For the most part, this paper ignores compensation that is based on some relative comparison (see Lazear and Rosen 1981; and Holmstrom 1982) since the focus is on payment by input versus payment by output.

Some examples are useful. Salesmen who are paid on a strict commission basis are piece-rate workers. Magazine, encyclopedia, and cosmetic salesmen often receive no fixed payment but are compensated as a direct and usually linear function of sales. They may choose the number of hours that they work and the effort that they associate with each hour.

Government employees fit the salary classification well. Compensation is independent of output this period and depends strictly on time worked. Certain tasks are required, and dismissal results only when effort falls below some specified standard. Screening through civil-service exams and by monitoring performance during a probationary period is important and allows the government to determine whether workers meet the specified standard.

Most jobs fit somewhere in the middle. For example, many managers in major corporations receive a large proportion of their compensation as a fixed amount specified in advance and independent of that period's output. But at the same time they may often receive a bonus, the size of which is geared directly to this period's output. The bonus component is synchronized to output, is flexible, and is essentially a piece rate. At the top of the corporate hierarchy, senior executives often receive a large proportion of their compensation as bonus and are, in many respects, piece-rate workers.

In what follows, piece rate most often is used to denote the synchronization between output and compensation. Salary implies that workers' pay is independent of this period's output.

II. Sorting

The first issue relates to sorting workers across jobs. The major cost of using a piece rate is that output must be monitored, at least periodically, to determine the worker's salary. The extreme version of a salary requires no monitoring of output. To draw out the differences, let us begin by ignoring all effort considerations. Instead, assume that the worker's lifetime output, q, is given and is not subject to the worker's choice. Two cases are worthy of consideration, namely, symmetric and asymmetric information.

A. Symmetric Information

The first assumes that workers and firms are equally uninformed of q but that both know the distribution of q. Let $q \sim f(q)$ with the distribution function $F(q)$. Assume that the worker can work at an alternative job (or consume leisure) at value $\overline{w}$. One possibility is to pay every worker a salary, S, independent of output level. Another possibility is to pay some piece rate, R, for each unit of q that is produced minus some constant amount to cover monitoring costs. Under the piece-rate scheme, the worker's compensation is

$$w = Rq - \theta, \tag{3}$$

where θ is the per-worker monitoring cost.

The goal is the standard one, namely, to maximize worker's expected wealth subject to a zero profit constraint. Assume risk neutrality so that wealth is the relevant consideration.

If a salary is paid, then no monitoring costs are borne. Zero profits require that

$$S = E(q).$$

As long as the expectation of q exceeds $\overline{w}$ and there are no piece-rate firms, all workers agree to work at this firm.

The problem with this scheme is that it is inefficient to have all workers work at this firm. Those workers for whom $q < \overline{w}$ should take the alternative job, and all can be made better off. Since $F(\overline{w})$ workers have $q < \overline{w}$, separating them from the others causes expected wealth to rise. This is obvious since

$$\begin{aligned}\text{Expected compensation with sorting} &= \overline{w}F(\overline{w}) \\ &+ \int_{\overline{w}}^{\infty} qf(q)dq > \int_{0}^{\infty} qf(q)dq \\ &> E(q) \\ &> S.\end{aligned}$$

This inefficiency can be eliminated if the worker's first day is monitored so that q is revealed. The worker can be induced to leave by paying him a piece rate equal to Rq. Since zero profits are required, in competition $R = 1$ induces both zero profit and efficient separation. But paying a piece rate requires that a monitoring cost, θ, be borne for each worker. The perfect piece rate results in an expected output at the current firm of

$$w = tE(q) + (1 - t)\int_{\overline{w}}^{\infty} qf(q)dq - \theta, \tag{4}$$

where t is the proportion of the work life spent in the initial monitoring period. (With no noise, it is optimal to push t arbitrarily close to zero.) In the first t of the workers' careers, all work at the piece-rate firm. After they learn q, only those with $Rq > \overline{w}$ remain. Setting $R = 1$ and reducing all workers' compensation by θ during the first t of the career results in zero profits and efficient separation. Thus the worker's wage profile is

$$\begin{aligned}
&w_t = tq - \theta && \text{during the first } t \text{ of the career;}\\
&\left.\begin{aligned} w_{1-t} &= (1 - t)q, && \text{for stayers}\\ &= (1 - t)\overline{w}, && \text{for leavers}\end{aligned}\right\} && \text{during the last } 1 - t \text{ of the career.}
\end{aligned} \tag{5}$$

So expected lifetime wealth if the worker starts at the piece-rate job and has the option to move when $w_1 - t < (1 - t)\overline{w}$ is

$$W = tE(q) - \theta + \left[\overline{w}F(\overline{w}) + \int_{\overline{w}}^{\infty} qf(q)dq\right](1 - t). \tag{6}$$

If a straight salary is paid, expected wealth equals $E(q)$.[4]

The condition for selecting a piece rate over a salary is

$$tE(q) - \theta + \left[\overline{w}F(\overline{w}) + \int_{\overline{w}}^{\infty} qf(q)dq\right](1 - t) > E(q)$$

or

$$\overline{w}F(\overline{w}) + \int_{\overline{w}}^{\infty} qf(q)dq - \frac{\theta}{1 - t} > E(q)$$

or

$$\overline{w}F(\overline{w}) - \frac{\theta}{1 - t} > \int_{0}^{\overline{w}} qf(q)dq. \tag{7}$$

4. It is not arbitrary that θ is borne by all workers during the initial period rather than by the stayers alone during the last $1 - t$ of the career. Even if workers are risk neutral, putting the cost θ on the last periods results in inefficient separation, whereas having it borne up front does not. That is, if

$$w_{1-t} = (1 - t)q - \gamma\theta,$$

where γ is chosen to arrive at zero profit, the worker leaves whenever $\overline{w} > q - \left(\frac{\gamma}{1 - t}\right)\theta$. For efficiency, he should leave only if $\overline{w} > q$, so too many leave.

From (7) it is obvious that the piece-rate firm maximizes worker wealth by keeping t as small as possible. Other results are equally intuitive.

First, as θ rises, the likelihood falls that a piece rate will dominate a straight salary. Salaries do not require monitoring; what is lost is the ability to sort workers to their highest valued use. As the cost of sorting rises, it becomes less worthwhile.

For a similar reason, as $\overline{w}$, the alternative use of time, rises, the value of using a piece rate rises. This is true because (7) can be rewritten as

$$\overline{w}F(\overline{w}) - \int_0^{\overline{w}} qf(q)dq > \frac{\theta}{1 - t}.$$

Differentiating the left-hand side with respect to $\overline{w}$ yields

$$F(\overline{w}) + \overline{w}f(\overline{w}) - \overline{w}f(\overline{w}) = F(\overline{w})$$

and

$$F(\overline{w}) > 0.$$

Thus piece rates are more valuable when $\overline{w}$ is large. The intuition is clear in that the better are the alternative opportunities relative to those here, the more is lost by failing to sort workers to their most valued use.

That point can be stated in a slightly different manner. For a given $\overline{w}$, the lower is $E(q|q < \overline{w})$, the more valuable is the piece-rate scheme that sorts workers. If that part of the distribution with $q < \overline{w}$ is very much below $\overline{w}$, then it is valuable to separate them from the firm. A skewed distribution of output with a long left tail is a good candidate for piece-rate pay. If some individuals are extremely bad at performing the task, monitoring and piece rates are more useful. Similarly, the greater the proportion of workers with $q < \overline{w}$, the more valuable is the piece-rate relative to the salary scheme.

The basic idea can be restated. The more heterogeneous workers are, the better it is to use a piece rate with monitoring. But as monitoring costs rise, workers become less willing to foot the bill through reduced salaries. If all workers were of similar abilities, a firm would have a difficult time hiring workers on a piece-rate basis because a firm that paid straight salary could offer a higher average wage. Monitoring costs increases the value of a salary relative to piece rates, and heterogeneity across workers decreases the value of a salary relative to piece rates.[5]

5. Even though θ may only provide private information to the firm, the assumption is that the firm does not renege on the contract, so the promise to pay q if q is observed is kept. Reputation or morale costs associated with violation of the contract may rationalize this assumption.

To add realism, let us recognize that estimates of workers' output do not perfectly reflect ability. In particular, there is some error associated with measurement and some random variation that results because of factors beyond the worker's control. This complicates the problem somewhat. The most important result is that noise reduces the value of the piece-rate scheme relative to a straight salary.

To see this, let

$$\hat{q}_t = tq + \epsilon_t, \tag{8}$$

where t is the period during which monitoring occurs, $\hat{q}_t$ is the observed output during that period, and ϵ_t is random error. A more general formulation would allow t to be endogenous, trading off quicker sorting against more measurement error. We ignore that and assume that t is set at its optimal level.

Given (8), if ϵ_t has the classical properties and if errors are not serially correlated, then the efficient and unbiased estimate of q is

$$\hat{q} \equiv \frac{\hat{q}_t}{t}. \tag{9}$$

Now, from (8) and (9),

$$\hat{q} = q + \frac{\epsilon_t}{t}$$

or

$$\hat{q} = q + \xi, \tag{10}$$

where $\xi \equiv \epsilon_t/t$. Let the density function of ξ be denoted by $g(\xi)$.

The issue is whether piece rates are less likely to be used when the measurement of output is noisy. The worker is given some reading of his output level, $\hat{q}$, and he must decide whether to leave to accept wage $\overline{w}$ or to stay. This amounts to selecting some critical level, q^*, such that, if $\hat{q} < q^*$, the worker leaves and takes the job that pays $\overline{w}$, whereas if $\hat{q} > q^*$, the worker stays.

Piece rates are less likely to be chosen when the estimate of q is noisy. To see this, what needs to be shown is that nondegenerate densities of ξ result in lower expected wages in the piece-rate firm than does the density $g(\xi) = 0$ for all $\xi \neq 0$. If so, then the piece-rate scheme is less advantageous when q is measured with error and is less likely to be chosen over the strict salary (because the strict salary pays $S = E[q]$ to all workers, and this is independent of any measurement error). The proof is tedious and is relegated to the Appendix. The intuition, however, is straightforward. The addition of measurement error causes some workers to remain at the piece-rate firm even though $q < \overline{w}$ because a positive ξ is drawn. The worker does not quit because his measured output is abnormally high during the period, so he is deceived into staying. Additionally, it causes some to leave even when $q > \bar{w}$ because a negative ξ makes $\hat{q} < \bar{w}$. Both types of error produce incorrect sorting and reduce the value of using a piece rate. The salary scheme performs no sorting but saves the monitoring cost, θ, so it is more likely to dominate when the variance in ξ is large.[6]

B. Asymmetric Information

The previous section discussed optimality when firms and workers are symmetrically uninformed about ability. But what if workers have better information than have firms about their output potential? Under these circumstances, as long as monitoring costs, θ, are positive, some workers can always be attracted to a salary firm. The reason is that it costs the worker θ to distinguish himself from his peers. For small differences in ability, it does not pay to bear that cost. There is always some group of least able workers that can be made better off by sorting into firms that do not waste resources on measuring output differences. Those firms assume instead that they have attracted low-ability workers, and they pay accordingly. This is straightforward and shown below.

All workers with $q - \theta < S$ work at the salary firm that pays S since they earn only $q - \theta$ at the piece-rate firm. What is required is that the salary firm can pay some S such that workers with $q < q^*$ select the salary firm and that zero profits can be achieved.

For any given S, workers with $q > S + \theta \equiv q_0$ choose the piece-rate firm. It is necessary to show that there exists an S such that

$$S = \frac{1}{F(S + \theta)} \int_0^{S+\theta} qf(q)dq \tag{11}$$

since the right-hand side is the expected output of a worker at the salary firm. Equation (11) is merely the salary firm's zero profit condition. Alternatively, (11) can be rewritten as

$$q_0 - \frac{1}{F(q_0)} \int_0^{q_0} qf(q)dq - \theta = 0, \tag{12}$$

where $q_0 \equiv S + \theta$. Define the left-hand side of (12) as $H(q_0)$. Then, to show that a salary firm can always compete away some workers from

6. The choice of the optimal sample period is a sequential search problem (see Wald 1947, 1950) and is not addressed here. The idea is that lengthening the number of periods over which the output is sampled before a decision is made provides a more precise estimate of $\hat{q}$ and reduces the amount of incorrect sorting mistakes. The cost is that some workers with very low q work at the piece-rate firm longer than would be necessary. The optimum is likely to take the following form. Examine $\hat{q}$ in the first period. If $q_1 < \hat{q}$, then retain the worker for at least one more period. After the second period, if some function of $\hat{q}$ in period 1 and $\hat{q}$ in period 2 exceeds some q_2, then retain the worker for another period, and so forth. The point is that there will be a tenure-specific standard based on the worker's history of estimated output below which output cannot fall without inducing a termination. (See Harris and Weiss 1981; and Weiss 1984.)

the piece-rate firm, it is necessary only to find a fixed point, $q_0 = q^*$, such that $H(q^*) = 0$ (i.e., to find $q^* - H[q^*] = q^*$). But

$$\lim_{q_0 \to q_{\min}} H(q_0) = q_{\min} - q_{\min} - \theta = -\theta < 0$$

and

$$\lim_{q_0 \to q_{\max}} H(q_0) = q_{\max} - \bar{q} - \theta.$$

If $q_{\max} - \bar{q} > \theta$, then $H(q_{\max}) > 0$. Since $H(q_0)$ is continuous, there exists a q^* such that $H(q^*) = 0$ so that some workers choose the salary firm. If $q_{\max} - \bar{q} < \theta$, then all workers choose the salary firm because even the most able worker who receives only $\bar{q}$ at the salary firm finds $\bar{q} > q - \theta$. The existence of an equilibrium where at least some workers go to the salary firm is proved. (If $\theta = 0$, then a salary firm could exist with only the least able worker at that firm. That worker would be indifferent between employment at piece-rate or salary firms. All others prefer the piece-rate firm.)[7]

The equilibrium value of S is

$$S^* = \frac{1}{F(q^* + \theta)} \int_0^{q^* + \theta} qf(q)dq,$$

and all workers with $q < S^* - \theta$ choose to work at the salary firm, whereas those with $q > S^* - \theta$ work at the piece-rate firm.

The obvious implication is that, for a given occupation, firms that pay workers a straight salary have a lower-quality work force than have firms that pay piece rates. The best workers select firms where performance has a payoff. The worst ones go to firms where ability has no effect on salary. Firms know this, and salaries are adjusted accordingly. Pencavel (1977) presents some evidence that piece-rate workers earn about 7% more than similar time-rate workers. Similarly, Seiler (1984) finds earnings 14% higher for "incentive" workers.

Productivity in piece-rate firms is higher than productivity in salary firms, but this does not imply that, if all salary firms were to pay piece rates, output would rise; the opposite is true. Switching all piece-rate worker to salary by fiat would save measurement costs θ on each worker and would have no effect on output. This is the classic screening result.

Contrast the result when information is asymmetric with that when

7. This result is akin to Riley's (1975) argument in the context of screening that the least able worker never signals ability to employers. Thus, salary firms devote zero resources to screening. It may depend on the discrete nature of θ, however. If monitoring can take place α of the time, and a straight salary paid $(1 - \alpha)$ of the time, then, by selecting different α's, firms may be able to upset the equilibrium. If an equilibrium exists with firms other than straight salary payers, the implication is that as α increases, the average quality in the firm rises.

information is lacking symmetrically. When information is asymmetric, it is always possible to pick off some workers by paying a straight salary. It is certain then that at least some of the firms will be salary firms. Whether there are piece-rate firms as well depends on the costs of monitoring relative to the value to the most able of being sorted from the least able. When information is lacking on both sides, a corner solution is always achieved; that is, either it pays to sort workers, or it does not. Since workers do not know their abilities ex ante, all choose one type of firm or the other.

C. *Capital*

The previous discussion ignores effort considerations and complementarities. It may well pay to place the most able workers in some firms and the least able ones in others for efficiency reasons. Ignoring effort effects for now, consider a production process where capital is important. If capital is important, then it is optimal to separate out low-ability workers. In general, this requires that the piece-rate firm use a two-part wage system,

$$W(q) = a + bq, \tag{13}$$

rather than simply paying $W(q) = Rq$. A possible alternative is to use one-part piece rate, Rq, combined with a standard, $\underline{q}$. If output falls below $\underline{q}$, then the employee is terminated (without compensation). Even when effort is not an issue, this scheme breaks down. However, a fixed salary coupled with a standard is efficient and sustainable but results in workers of only one ability level applying for employment. The scheme in equation (13) looks identical to the piece rate described in the earlier section with $a = -\theta$ and $b = 1$. The difference, and it is a crucial one, is that, there, there is no problem with observability of output or ability. In this section, the focus is on fixed costs with perfect information. It differs from earlier sections, in which both sides were assumed to lack information or in which only one side was assumed to lack information.

First, let us show that a one-part piece rate without a standard is not efficient in general. To focus on this assume that measurement costs are zero. Suppose that the production technology requires that each worker use a machine to produce output and that the rental price of the machine is γ. Then net output at this firm from worker with ability q is $q - \gamma$. For efficiency, it is necessary that only and all workers whose net output levels, $q - \gamma$, exceed the alternative use of time, $\overline{w}$, work at the current firm or one identical to it. In order to induce this to occur it is necessary that $w(q) < \overline{w}$ for $q < \overline{w} + \gamma$ and $w(q) > \overline{w}$ for $q > \overline{w} + \gamma$. If $w(q)$ is continuous, this implies that $w(\overline{w} + \gamma) = \overline{w}$.

A one-part piece rate has the form $w(q) = Rq$. Since $w(\overline{w} + \gamma) = \overline{w}$, this implies that

$$R = \frac{\overline{w}}{\overline{w} + \gamma}.$$

But for this to be an equilibrium it must be true that the firm earns zero profits or that compensation equals output. Required is that

$$\int_{\overline{w}+\gamma}^{\infty} Rqf(q)dq = \int_{\overline{w}+\gamma}^{\infty} (q - \gamma)f(q)dq.$$

After making the substitutions, this implies that

$$\left(1 - \frac{\overline{w}}{\overline{w} + \gamma}\right) \int_{\overline{w}+\gamma}^{\infty} qf(q)dq = \gamma[1 - F(\overline{w} + \gamma)]. \tag{14}$$

Equation (14) cannot hold in general because $\overline{w}$, γ, and $f(q)$ are all exogenous. There is one equation but no unknowns. This proves that a one-part piece rate without a standard is not generally efficient.

(A special case is when there is no capital requirement or when capital is free. In that case, $\gamma = 0$, so eq. [14] holds: $R = 1$, and all workers with $q < \overline{w}$ work elsewhere; the rest are employed here.)

It is a trivial extension, however, to show that a two-part piece rate is efficient. If $w(q) = -\gamma + q$, then profit equals zero since each worker is paid his net output. Further, for those with $q - \gamma < \overline{w}$, $w(q) < \overline{w}$, and for those with $q - \gamma > \overline{w}$, $w(q) > \overline{w}$. So worker sorting is perfect.

An alternative is to pay a one-part piece rate Rq (R not necessarily equal to one) and simply to terminate all workers with $q < \overline{w} + \gamma$. This will not work because of adverse selection. For zero profits, it is necessary that

$$\int_{\overline{w}+\gamma}^{\infty} Rqf(q)dq = \int_{\overline{w}+\gamma}^{\infty} (q - \gamma)f(q)dq$$

or that

$$R = 1 - \frac{\gamma[1 - F(\overline{w} + \gamma)]}{\int_{\overline{w}+\gamma}^{\infty} qf(q)dq}$$

so that $R < 1$.

But if $R < 1$ at this firm, workers at the top of the distribution always prefer a firm that pays $w(q) = q - \gamma$. Figure 1 illustrates this. Only workers with $q < q^*$ work at the firm that pays Rq with $R < 1$, so they are all paid more than their net contribution. The firm cannot break

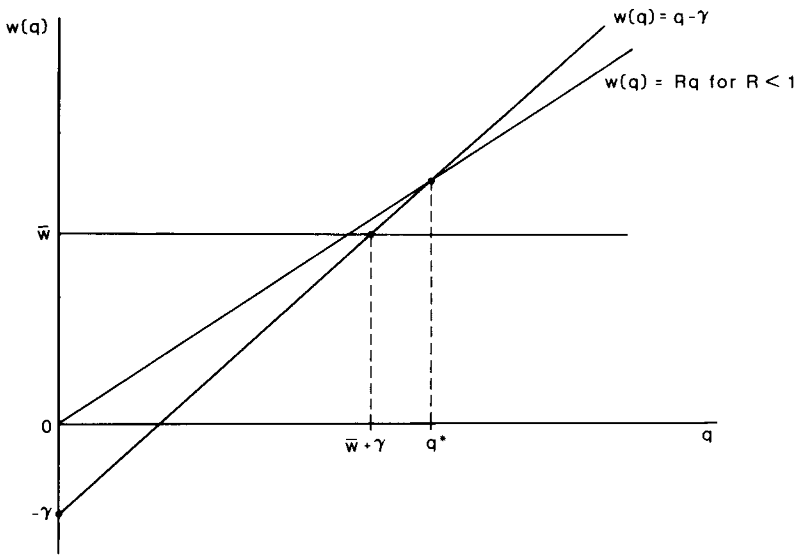

FIG. 1

even, and there is no readjustment of R that will allow zero profit so long as $R < 1$.

In the absence of monitoring cost, a salary firm that pays $S = \overline{w}$ and requires that the worker have $q \geq \overline{w} + \gamma$ can stay in business, but it attracts only those workers with $q = \overline{w} + \gamma$. All others do better at the piece-rate firm. Any higher S with the same standard results in losses, but it is always possible to select a standard consistent with any $S > \overline{w}$ that does not result in losses. All that is necessary is that the required level of performance be equal to $S + \gamma$. The only workers that apply are those with $q = S + \gamma$, which is likely to be a costly way to recruit.

Recapping, a piece rate with a fixed component (in this case $-\gamma$) has salary attributes. The amount $-\gamma$ is paid to input rather than to output. That is, it is a fixed fee that is levied for coming to work and is independent of output. As capital costs go up, γ rises, so the importance of the payment (or fee) by input grows relative to payment by output. In that sense it can be said that the existence of physical capital pushes in the direction of payment by input.

III. Effort

It is commonly argued that piece-rate pay is an incentive device whereas straight salary provides no incentives. To determine the validity of this statement, it is first necessary to distinguish the characteristics that may or may not be attributed to salaries. The first is the invariability of the salary with the current effort level. The second is the invariability of salary with the past effort level. If effort level is observed perfectly and contemporaneously, then there is no reason to expect the salary defined as payment by input to be invariant to the current effort level. An example is a situation in which effort per hour is fixed but hours are variable. Paying the appropriate hourly wage would provide exactly the right incentives.

In this case, and in all cases in which output and input are observed perfectly, the choice of payment by input or output is irrelevant. When everything is observed, a payment by either criterion can be efficient. This is obvious. If E is effort, the cost of effort function is given by $C = C(E)$, and the production function is given by $q = E$,[8] then the efficient level of effort, E^*, is given by

$$E^* = \arg_E \max q - C(E)$$

or

$$= \arg_E \max E - C(E).$$

So E^* solves

$$C'(E^*) = 1.$$

If the worker were paid $S(E) = E$, then he would choose E so as to maximize

$$S(E) - C(E) = E - C(E) \tag{15}$$

and would set $C'(E^*) = 1$, achieving efficiency. Alternatively, if the worker were paid a piece rate on output such that $P(q) = q$, then he would choose E to maximize

$$\begin{aligned} P(q) - C(E) &= q - C(E) \\ &= E - C(E), \end{aligned}$$

so again E^* is chosen, which ensures that $C'(E^*) = 1$.

Still, it is commonly alleged that salaries do not provide the same incentives as do piece rates. What is meant by this statement? The definition of salary implicit in it includes some invariability of payment with effort. For some reason the salary does not respond to effort appropriately. Clearly, at the extreme, if $S(E) = \bar{S}$, then the worker's choice in (15) is to choose E so as to minimize $C(E)$, hardly consistent with efficiency.

The question then becomes, What is it that makes $S(E)$ deviate from $S(E) = E$? The most straightforward answer relates to the inability to

8. Because the form of $C(E)$ is free, any production funtion can always be reduced to $q = E$ with an appropriate redefinition of $C(E)$.

observe E. If E were unobservable, but if q were perfectly observable, then the choice would be clear: payment function $P(q) = q$ results in perfect efficiency whereas $S(E) = \bar{S}$ does not.

But this is extreme. The choice between payment by input and payment by output may be characterized more appropriately by the notion that it may be cheap to observe that minimal effort, $E \geq \bar{E}$ (the worker comes to work and goes through the motions of doing the job), but more expensive to measure his actual level of output (and it may be impossible to measure his exact level of effort). Thus suppose that the firm can determine output exactly if it bears cost θ_1 but can know that effort exceeds some minimum, $\bar{E}$, at cost $\theta_2 < \theta_1$. The issue is when the firm should bear the larger monitoring cost and pay on the basis of output.

As shown earlier, the piece rate that pays $q - \theta_1$ induces efficient effort E^* such that $C'(E^*) = 1$. If the low-monitoring strategy is used, it is clear that the worker supplies exactly $\bar{E}$ of effort. (Note that $\bar{E}$ may be less than or may exceed E^*.) If $\bar{E}$ is the level of effort, then all low-monitoring workers receive a salary, $S = \bar{E} - \theta_2$. The choice of whether to use a salary or a piece rate then implies that a salary is paid if and only if

$$\bar{E} - \theta_2 - C(\bar{E}) > E^* - \theta_1 - C(E^*) \tag{16}$$

since the worker receives $\bar{E} - \theta_2$ if he is on a salary and $E^* - \theta_1$ if he is on a piece rate. Equation (16) can be approximated as

$$\theta_1 - \theta_2 \gtrsim (\bar{E} - E^*)^2 \left[\frac{C''(E^*)}{2}\right], \tag{17}$$

where the right-hand side uses second-order Taylor series expansion around E^* since $C'(E^*) = 1$.

Equation (17) has a number of straightforward implications. First and most obvious is that as $\theta_1 - \theta_2$ increases, the salary tends to be preferred. As the cost of the more precise type of monitoring rises, a salary that depends only on satisfying some minimal level of performance dominates. Note again that this does not imply that effort is lower with a salary than with a piece rate, for $\bar{E}$ is exogenous and may well exceed E^*.

Second, as $|E^* - \bar{E}|$ increases, the value of the piece-rate scheme rises relative to a salary. At one extreme, if $\bar{E} = E^*$, then it is clear that the salary is always preferred because effort is the same and monitoring costs are lower. (Indeed, the firm should set $\bar{E}$ as close to E^* as possible, to the extent that $\bar{E}$ is subject to choice.)

Third, for a given $\bar{E}$, the more elastic is the marginal cost function (i.e., the lower is $C''[\bar{E}]$), the larger is the deviation between $\bar{E}$ and E^*. Then it is necessarily true that the more elastic is the $C'(E)$ function

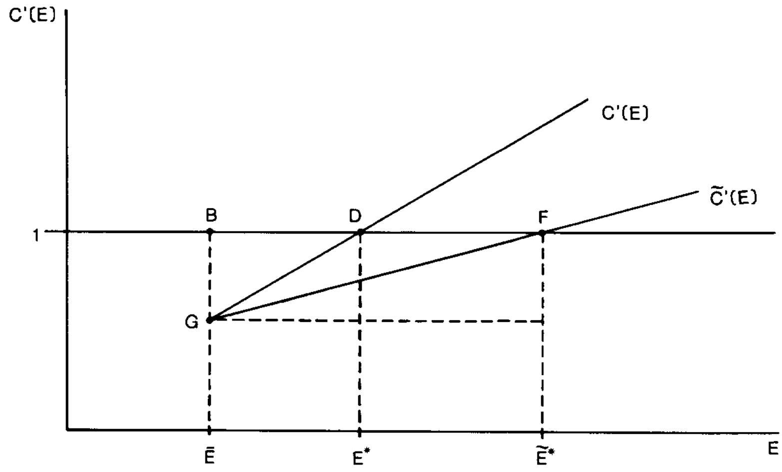

FIG. 2

(i.e., the closer to zero is $C''[E]$), the poorer the salary performs. Figure 2 makes this clear.

Suppose $\bar{E} < E^*$. Then $C'(\bar{E}) < 1$. Consider $C'(E)$ and $\tilde{C}'(E)$ such that $C'(\bar{E}) = \tilde{C}'(\bar{E})$ but $\tilde{C}''(\bar{E}) < C''(\bar{E})$. Then $\tilde{C}'(E)$ is a more elastic marginal cost-of-effort function than is $C'(E)$. If the worker faced $C'(E)$, then a piece rate induces him to supply E^* of effort. If he faced $\tilde{C}'(E)$, he would supply $\tilde{E}^*$ of effort. The salary induces him to supply $\bar{E}$ of effort with either function.

What does the worker lose by stopping at $\bar{E}$ rather than at E^*? The lost output is $E^* - \bar{E}$, which equals the area of rectangle $\bar{E}BDE^*$ (since the height equals one). But in return the worker bears smaller cost. The incremental cost of moving from $\bar{E}$ to E^* is the area of trapezoid $\bar{E}GDE^*$. So the net loss in stopping short of E^* is triangle BDG. Similarly, if the cost function were $\tilde{C}'(E)$, the net loss from stopping at $\bar{E}$ would be triangle BFG. Since $BFG > BDG$, the salary is less likely to be preferred to the piece rate when $C'(E)$ is more elastic.[9]

Summarizing, when effort is endogenous, both piece rates and salaries can act as incentive devices. An alternative to a piece rate is a salary that is contingent on exceeding some performance standard. This may be a cheaper, but less efficient, incentive device. It tends to be preferred to a piece rate when the differences in costs of the two kinds of monitoring are large and when, for a given observable standard level, the marginal cost of effort function is steep, that is, effort is supplied inelastically.

9. A similar argument holds for $\bar{E} > E^*$.

A. *Quantity versus Quality*

It is sometimes alleged that piece rates induce the worker to produce too many low-quality goods and that salaries avoid the problem. Under what conditions is this a correct statement?

The piece rate can be made contingent on the output quality characteristics as well as on the quantity characteristics. More formally, suppose that the firm's revenue function is

$$\text{Revenue} = P(q, Q),$$

where q is the number of units sold, and Q is their quality (arbitrarily defined). Suppose further that the worker can convert effort into output as

$$f(q, Q, E) = 0.$$

Then the efficient competitive firm that maximizes workers' expected wealth subject to zero profits wants to induce q, Q, and E as

$$\max_{q,Q,E} \; P(q, Q) - C(E) \tag{18}$$

subject to

$$f(q, Q, E) = 0.$$

The worker faces some announced piece-rate schedule, $R(q, Q)$, and seeks to maximize

$$R(q, Q) - C(E)$$

subject to

$$f(q, Q, E) = 0.$$

The firm that sets $R(q, Q) \equiv P(q, Q)$ guarantees that the worker's problem becomes identical to (18) and so yields an efficient allocation of resources. This maximizes worker wealth, consistent with zero profits.

When $R(q, Q) \equiv P(q, Q)$, the piece-rate worker does not emphasize quantity to the exclusion of quality. Real-world examples of piece-rate workers who are paid for quality as well as for quantity exist. Executives' bonuses depend on profits, not on the number of units sold, and profit is the measure that weights quantity and quality in exactly the appropriate fashion. Salesmen's commissions depend on sales revenues, not on units sold. Sales revenues weight "big-ticket" high-quality items more heavily. Even agricultural workers may be penalized for turning in bruised fruit.

Of course, it is possible to induce piece-rate workers to favor quantity to the exclusion of quality. All that is required is that $\partial R/\partial Q \leq 0$. But this payment scheme would never be sensible.

Although a salary does not automatically solve the problem, there is an important sense in which salary emphasizes quality relative to the piece rate. Since the standard on which the salary is based is an effort standard rather than an output standard, the worker is indifferent to allocations of q, Q, so long as they satisfy $f(q, Q, \bar{E}) = 0$, where $\bar{E}$ is the required level of effort. A mere request that the worker produce ($\bar{q}$, $\bar{Q}$) should be met with compliance if $f(\bar{q}, \bar{Q}, \bar{E}) = 0$, given that $\bar{E}$ is required anyway.

For salaried workers, one proxy for effort is hours worked. If that were the only dimension of effort, then paying an hourly wage would be equivalent to paying for effort. These workers would be instructed to produce the right combination of q, Q. No monitoring of output would be necessary since there would be no incentive to deviate from the optimum, so long as it required no more than $\bar{E}$ of effort.

When effort is less costly to monitor than both quantity and quality of output, then a salary based on effort may be preferable. An example is an hourly wage, where weekly compensation depends on time worked. Hours are easily measured, and the worker is indifferent between quantity and quality, given hours worked, so that the firm's priorities are likely to be realized.

IV. Risk

There are two sources of income variation associated with a piece rate. The first is variation over time that results because of factors beyond the worker's control, which affect output. The state of the market, exogenous factors of production, and other sources of luck cause the worker's output to vary even if his effort does not. The second kind of variation results from unobserved differences in worker ability. If workers and firms are unaware of each worker's ability, then lifetime output is nondeterministic.[10]

A salary that is not contingent on last period's output tends to smooth out intertemporal variations in income induced by a piece rate. But this kind of risk does not seem to be a major problem. Borrowing and lending in capital markets can smooth much of it. The time period over which output is measured can be lengthened to make compensation per period vary by a smaller amount (see Weiss 1984).

The more important kind of risk that the worker faces is risk in his lifetime wealth. In its purest form, the salary, which is totally independent of output, offers complete insurance against lifetime risk. Since no attempt is made to measure individual's output, there is no difficulty in keeping high-quality workers. Those workers who sign on with the salary firm accept the average salary because they are ignorant about

10. Stiglitz (1975) derives optimal piece rates that trade insurance for efficiency.

their ability levels or are averse to risk. The implication is that salaries are more likely to be paid when workers have a high degree of risk aversion relative to owners.

V. Intertemporal Strategic Behavior

Often, piece rates are negatively related to past output. One problem that arises is that workers artificially depress this period's output because of the effect on next period's compensation. For example, workers pressure peers who show management that the task can be done more rapidly than believed in order to increase their own current salaries. Similarly, salesmen are wary of doing too well today, lest their quotas be raised tomorrow.

Salaries that are contingent on meeting some standard suffer from the same drawback when the standard is adjusted as a function of last period's output. But these effects can be offset if they are anticipated properly.[11]

Consider the piece rate. If workers know that tomorrow's rate is a function of today's output, it would seem that too little effort would be exerted in the current period. But this kind of strategic behavior can be undone by the appropriate piece-rate structure.

In order for the firm to be able to gear next period's rate to this period's output, the firm must enjoy some ex post monopsony power. Otherwise a competitive firm can attract all workers who are poorly treated.[12] The problem arises when specific capital or mobility costs lock the worker into the firm such that there is a wedge between what the worker can receive elsewhere and what he receives here. The firm, knowing this, exploits the worker in an ex post sense. Of course, when workers begin employment with the firm, they are aware of this situation, and total lifetime compensation must be set so as to leave the firm with zero profits. But there still is no way to prevent the firm from behaving opportunistically in the second period.

To formalize this, the worker lives two periods and produces $q_1 = E_1$ and $q_2 = E_2$ in periods 1 and 2, where E_t is effort in period t. Workers have an alternative use of time (say, leisure) at zero effort with value $\overline{w}$ for each period. Further, workers differ in their cost-of-effort functions. Some view effort as less distasteful than others, so

$$\text{Cost}_t = \tilde{C}(E_t),$$

where $\tilde{C}$ is a stochastic function that is known to workers but not to firms.

11. Rogerson (1985) has treated a similar problem in a slightly different context. The problem is analogous to the problem of a division manager spending too much this year because next year's budget depends positively on this year's expenditure.

12. This ignores the possibility of titling the entire age-earnings profile such that all workers receive more than the value of their output in the later years (Lazear 1979).

Firms exploit workers ex post by learning the cost-of-effort function from actions taken in period 1. Low-cost workers find that their piece rates are lower in period 2 than high-cost workers, and they behave strategically as a result, attempting to prevent the firm from learning $\tilde{C}(E)$.[13] Ex post monopsony power implies that the firm need only pay the worker a wage in period 2 that makes him indifferent between work here and work elsewhere:

$$w_2 - \tilde{C}(E_2) = \overline{w}. \tag{19}$$

It is obvious from (19) that the firm sets w_2 lower for workers with low realizations of $\tilde{C}(E_2)$ if the firm can discover that value.

In period 1, the worker receives w_1, and the ex ante zero profit constraint implies that $w_1 + w_2 = q_1 + q_2$ for every worker (because markets are assumed to be competitive). The question is, Does there exist a $w_1(q_1)$ and $w_2(q_1, q_2)$ such that the worker can be induced to behave efficiently in both periods even though he knows that the firm has ex post monopsony power and that the firm will lower his wage if it finds him to be a low $\tilde{C}(E)$ worker? The answer is a definite yes. Even though the worker behaves strategically in period 1, his strategic behavior can be offset by an appropriate piece-rate schedule.

Let us conjecture that the functional form of that compensation scheme is given by

$$w_1 = R_1(q_1) + K_1, \tag{20a}$$

$$w_2 = R_2(q_2) + K_2(q_1). \tag{20b}$$

To derive the efficient scheme that maximizes ex post exploitation but is consistent with zero profits, start by noting that the worker optimizes by selecting $E_1(= q_1)$ and $E_2(= q_2)$ such that

$$\begin{aligned} \tilde{C}'(E_1) &= \frac{\partial w_1}{\partial q_1} \cdot \frac{\partial q_1}{\partial E_1} + \frac{\partial w_2}{\partial q_1} \cdot \frac{\partial q_1}{\partial E_1} \\ &= R_1' + K_2', \end{aligned} \tag{21a}$$

$$\tilde{C}'(E_2) = R_2'. \tag{21b}$$

For efficiency in period 2, it is necessary that $R_2' = 1$. For efficiency in period 1, it is necessary that

$$R_1' + K_2' = 1 \tag{22}$$

or that

$$R_1' = 1 - K_2'.$$

For full ex post exploitation, it is necessary that

$$R_2(q_2) + K_2(q_1) - \tilde{C}(E_2) = \overline{w}. \tag{23}$$

13. Mixed strategies are assumed away.

For any $K_2(q_1)$, selecting $R_1'(q_1) = 1 - K_2'(q_1)$ guarantees efficiency in the first period since substitution into (21a) implies that the worker sets $\tilde{C}'(E_1) = 1$. Letting $R_2(q_2) = q_2$ guarantees efficiency in period 2. The ability of the firm to act as an ex post monopsonist in period 2 depends on its knowledge of the functional form of $\tilde{C}(E)$. Since the firm chooses $R_1' = 1 - K_2'$, it knows that the worker chooses E_1 such that $\tilde{C}'(E_1) = 1$. If the firm knows the functional form of $\tilde{C}(E)$, under some circumstances this allows perfect identification of the worker's actual $C(E)$. But even if the firm cannot perfectly identify $C(E)$, the worker still behaves appropriately. All that happens is that the firm loses some quasi rent because of its inability to act as an ex post monopsonist. (Because of competition, this does not affect the firm's profits, which are zero anyway. Nor can the firm trade off quasi rent in period 2 for inefficiency. Doing so would result in lower lifetime worker wealth and would result in a failure to attract workers in the first place.)

To see this, suppose that the firm forms some estimate of $C(E)$ based on its observation of q_1, and denote this estimate as $\hat{C}(E; q_1)$. (It is not necessary that $\hat{C}$ be unbiased or have any other property.)

Now let the firm define

$$K_2'(q_1) \equiv \frac{\partial \hat{C}}{\partial q_1}(q_1; q_1).$$

Let the firm simply announce that

$$R_1'(q_1) \equiv 1 - \frac{\partial \hat{C}}{\partial q_1}(q_1; q_1)(= 1 - K_2').$$

Equation (22) is satisfied, so efficiency in period 1 is guaranteed. Again, $R_2(q_2) \equiv q_2$ so that period-2 efficiency is guaranteed, and K_1 is set so as to guarantee ex ante zero profits.

The intuition is straightforward. Because the worker takes into account the effect of this period's output on next period's rate, he tends to underproduce this period. But this can be offset by increasing the payment per piece during the first period.

An example is useful. Let $\tilde{C}(E) = \lambda E^2$, where λ is a random variable unknown to the firm. If

$$w_1 = \frac{3}{2} \cdot q_1 - \overline{w}$$

and

$$w_2 = q_2 - \frac{q_1}{2} + \overline{w},$$

then all is fully efficient, zero profits are earned ex ante, and the firm fully exploits the worker ex post. The worker chooses E_1 such that

$$C'(E_1) = \frac{3}{2} - \frac{1}{2} = 1,$$

so E_1 is efficient. He chooses E_2 such that

$$C'(E_2) = 1,$$

so E_2 is efficient. Further, full ex post exploitation is achieved because (23) holds. Required is that

$$R_2(q_2) + K_2(q_1) - \tilde{C}(E_2) = \overline{w}$$

or that

$$q_2 - \frac{q_1}{2} + \overline{w} - \lambda(E_2)^2 = \overline{w}$$

or that

$$q_2 - \frac{q_1}{2} = \lambda(q_2)^2.$$

Since $q_2 = q_1$, this reduces to

$$\frac{q_2}{2} = \lambda q_2^2.$$

But $C'(E_2) = 1$ implies that $q_2 = \dfrac{1}{2\lambda}$, so this becomes

$$\frac{1}{4\lambda} = \lambda\left(\frac{1}{2\lambda}\right)^2 = \frac{1}{4\lambda},$$

so the necessary condition for full exploitation is met. Finally, ex ante profits equal zero since $w_1 + w_2 = q_1 + q_2$.

The conclusion is that intertemporal strategic behavior does not render the piece rate inefficient. Even though workers select effort levels with effects on future rates in mind, all is efficient in equilibrium. This does not negate that workers worry about the effects of today's output on future rates. It merely implies that this worry can be offset with an inflated piece rate in the first period.[14]

14. A similar analysis applies for the setting of standards based on last period's output. Its effects, too, can be offset to yield efficiency. This analysis relates very closely to the literature on setting quotas in a planned economy. The analysis here is most similar to Weitzman's (1980) analysis of the "ratchet effect." More recently, Freixas, Guesnerie, and Tirole (1983) have extended the general theory to consider cases in which the planner does not commit himself to an intertemporal incentive scheme at the start. Earlier papers on the same issue include Yunder (1973), Fan (1975), Bonin (1976), and Weitzman (1976).

VI. Empirical Issues

The preceding analysis presents a number of testable implications about when piece rates and salaries are most likely to be used. Additionally, it offers predictions on the distribution of output within piece-rate jobs as compared to salary jobs.

In order to test these implications, it is necessary to have data, the unit of analysis of which is the job rather than the worker. Even though no data set of this sort exists currently, it is useful to sketch out briefly the empirical issues to be addressed.

First, recall that the definition of a piece rate involved the relation between this period's pay and this period's output. To classify jobs into piece-rate or into salary categories, one would like to examine a time series of output and compensation. A regression equation could be specified of the form

$$y_t = a + bt + c_0 q_t + c_1 q_{t-1} + \ldots + c_i q_{t-i} + dH_t, \qquad (24)$$

where y_t is compensation in t, q_t is output, and H_t is hours worked in t. A pure piece rate implies that $c_0 = 1$ and that $c_1, \ldots, c_1 = 0, d = 0$. A pure salary implies that $c_0, \ldots, c_i = 0$. A salary that is contingent on past performance implies that $c_1, \ldots, c_i \geq 0$, $c_0 = 0$.

The estimated coefficients then become data for successive stages of the empirical analysis. What one would like to predict is the size of $c_0, \ldots, c_i$ across jobs, and the theory presented above provides a structure. Although it is unlikely that reliable data on output could ever be obtained, let us proceed as if that were not a problem to outline the possibilities.

First, heterogeneity in worker abilities implies that piece rates are more likely to be used. This implies that there should be a positive correlation between c_0 and between the variance in q_t across workers. (Seiler [1984] already has found higher variance in the earnings [y_t] of piece-rate workers, but this follows even if the variance in output is higher in salary firms because salaries ignore differences in output.)

Second, costs of monitoring output should be negatively related to c_0. To the extent that a measure of the monitoring costs could be obtained, the implication is testable.[15]

Third, if piece rates act as a sorting device, then there should be a positive correlation between c_0 and turnover rates.

Fourth, if information is asymmetric, then the least able workers work at the salary firms. This implies that, within a given job (whatever that means), there should be a positive correlation between c_0 and the

15. There is a problem in that those jobs for which output measures are not easily obtainable will, simply because of errors in variables, have c_i coefficients that are biased toward zero.

average level of output. For the same reason, c_0 should be positively correlated with the wage rate.

Fifth, the cost of measuring output quality should be negatively correlated with c_0. Additionally, as costs of measuring quality relative to effort rise, $a + b + dH$ should rise as well.

VII. Summary

Some traditional and not so traditional factors that are associated with piece rates are analyzed. Piece-rate workers are distinguished from salary workers in that piece rates depend on current output whereas salaries do not. Salaries are closer to payment by input, broadly defined.

Because salaries are independent of this period's output, it is not necessary to measure output. Although monitoring costs are saved, workers are not sorted as efficiently. Salaries suffer from the drawback that some low-output workers who could do better elsewhere are not induced to leave. The efficient piece rate induces all workers to leave appropriately but carries with it a monitoring cost. These considerations imply that piece rates are likely to be used over salaries when the following conditions hold: (*a*) the cost of measuring output is low; (*b*) the value of the alternative wage is high relative to average output at the current firm; (*c*) workers are heterogeneous in ability levels; and (*d*) output is measured without too much error.

If workers know their ability levels and firms do not, then it is always true that at least some of the firms will pay a straight salary. Piece-rate firms may or may not coexist, depending on the strength of the considerations above. More important is that the least able workers are always in the salary firm. They are the ones who are unwilling to bear the monitoring costs necessary to distinguish abilities. Symmetric ignorance, on the other hand, pushes the solution to a corner; that is, either all workers work in piece-rate firms, or all work in salary firms.

When capital is an important factor of production, firms are not indifferent about which workers are employed by the firm. Fixed costs increase the value of high-ability relative to low-ability workers. Two-part piece rates are efficient and dominate one-part piece rates, which are never efficient when capital is a factor. Salaries, with output requirements, are efficient but tolerate workers of only one ability level. This is yet another reason why piece-rate firms are characterized by a more heterogeneous work force. When effort is less costly to measure than is output quality, a salary based on effort is likely to dominate a piece rate.

Intertemporal strategic behavior is a sometimes noted problem with piece rates. Workers slack off because next period's rate depends on this period's output. This problem is not unique to piece rates. More

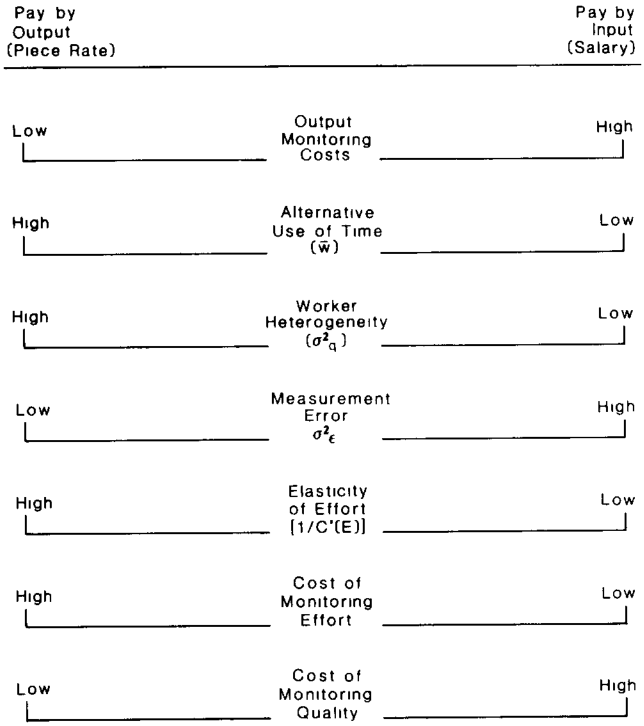

FIG. 3

important, it can always be efficiently offset by the appropriate reward in the early periods. Effects are summarized in figure 3.

Appendix

Proof That Noise Reduces the Value of the Piece Rate

We desire to show that

$$\begin{aligned} t \int_{-\infty}^{\infty}\int_{0}^{q^*-\xi} qf(q)g(\xi)dqd\xi + (1 - t)\overline{w} \int_{-\infty}^{\infty}\int_{0}^{q^*-\xi} f(q)g(\xi)dqd\xi \\ + \int_{-\infty}^{\infty}\int_{q^*-\xi}^{\infty} qf(q)g(\xi)dqd\xi - \theta < t \int_{0}^{w} qf(q)dq \qquad \text{(A1)} \\ + (1 - t)\overline{w}F(\overline{w}) + \int_{\overline{w}}^{\infty} qf(q)dq - \theta. \end{aligned}$$

The θ drops out. Then, taking t of the last term on the left-hand side and adding it to the first term on the left-hand side allows us to rewrite the left-hand side as

$$t\bar{q} + (1 - t)\left[\bar{w}\int_{-\infty}^{\infty}\int_{0}^{q^*-\xi} f(q)g(\xi)dqd\xi + \int_{-\infty}^{\infty}\int_{q^*-\xi}^{\infty} qf(q)g(\xi)dqd\xi\right],$$

where $\bar{q} \equiv E(q)$. Similarly, the right-hand side can be rewritten as

$$t\bar{q} + (1 - t)\left[\bar{w}F(\bar{w}) + \int_{\bar{w}}^{\infty} qf(q)dq\right].$$

After canceling terms, the problem reduces to showing that

$$\bar{w}F(\bar{w}) + \int_{\bar{w}}^{\infty} qf(q)dq > \bar{w}\int_{-\infty}^{\infty}\int_{0}^{q^*-\xi} f(q)g(\xi)dqd\xi + \int_{-\infty}^{\infty}\int_{q^*-\xi}^{\infty} qf(q)g(\xi)dqd\xi. \tag{A2}$$

Define $\mu \equiv q^* - \xi$. Then the right-hand side of (A2) can be written as

$$\int_{-\infty}^{\infty}\left[\bar{w}\int_{0}^{\mu} f(q)dq + \int_{\mu}^{\infty} qf(q)dq\right]g(\xi)d\xi. \tag{A3}$$

Taking the derivative of the inside of (A3) with respect to μ yields:

$$\frac{d}{d\mu} = \bar{w}f(\mu) - \mu f(\mu). \tag{A4}$$

Setting this equal to zero implies that $\mu = \bar{w}$. Since the integral of the maximum is never exceeded by the maximum of the integral, setting $\mu = \bar{w}$ yields an upper bound to the right-hand side. But when $\mu = \bar{w}$, the right-hand side and left-hand side of (A2) are identical. Thus, for any nondegenerate distribution of ξ, the right-hand side exceeds the left-hand side of (A2). Q.E.D.

The worker's choice of a cutoff, q^*, is derived as follows. The problem for the worker is to choose some q^* such that, if $\hat{q} < q^*$, the worker quits. The worker's problem is to use the information given by $\hat{q}$ optimally to infer q and then to make a decision about his work location. The worker can earn $\bar{w}$ elsewhere and quits if

$$\hat{q} < q^*$$

or if

$$q < q^* - \xi.$$

The risk-neutral worker wants to choose q^* to solve

$$\max_{q^*}\left[\bar{w}\int_{-\infty}^{\infty}\int_{-\infty}^{q^*-\xi} f(q)g(\xi)dqd\xi + \int_{-\infty}^{\infty}\int_{q^*-\xi}^{\infty} qf(q)g(\xi)dqd\xi\right](1 - t). \tag{A5}$$

The first term reflects that he earns $\bar{w}$ at the alternative job, and the double integral is the probability of observing $\hat{q} < q^*$. The second term is the expected output at the firm conditional on staying when $\hat{q} > q^*$. Differentiating (A5) with respect to q^* yields

$$\frac{d}{dq^*} = \overline{w}\int_{-\infty}^{\infty} f(q^* - \xi)g(\xi)d\xi - \int_{-\infty}^{\infty} (q^* - \xi)f(q^* - \xi)g(\xi)d\xi$$

or

$$\frac{d}{dq^*} = (\overline{w} - q^*)\int_{-\infty}^{\infty} f(q^* - \xi)g(\xi)d\xi + \int_{-\infty}^{\infty} \xi f(q^* - \xi)g(\xi)d\xi. \quad \text{(A6)}$$

Although (A6) does not appear to be easily interpreted, a simple example proves this false. Suppose that ξ is distributed symmetrically around zero and that q is uniform between q_0 and q_1. When should the criterion level be set so that $q^* = \overline{w}$? Doing so would appear to provide first-best decision making.

For $q^* = \overline{w}$ to be an optimum, equation (A6) must equal zero. Rewrite (A6) as

$$\frac{d}{dq^*} = (\overline{w} - q^*)\int_{-\infty}^{\infty} f(q^* - \xi)g(\xi)d\xi + \int_{-\infty}^{q^*-q_1} \xi f(q^* - \xi)g(\xi)d\xi$$

$$+ \int_{q^*-q_1}^{q^*-q_0} \xi f(q^* - \xi)g(\xi)d\xi + \int_{q^*-q_0}^{\infty} \xi f(q^* - \xi)g(\xi)d\xi.$$

If $f(q)$ is a constant between q_0 and q_1, then this reduces to

$$\frac{d}{dq^*} = \frac{\overline{w} - q^*}{q_1 - q_0}\int_{q^*-q_1}^{q^*-q_0} g(\xi)d\xi + \frac{1}{q_1 - q_0}\int_{q^*-q_1}^{q^*-q_0} \xi g(\xi)d\xi. \quad \text{(A7)}$$

In order for (A7) to equal zero at $q^* = \overline{w}$, it is sufficient that

$$q^* - q_1 = -(q^* - q_0)$$

(since ξ is distributed symmetrically around zero) or that

$$2\overline{w} = q_1 + q_0.$$

When $\overline{w} = 0$ and $q_0 = -q_1$, this holds. But, in general, the condition is not satisfied. In particular, suppose that q_0, $q_1 > 0$ and that $\overline{w} < E(q)$. Then from (A7) it can be shown that $q^* < \overline{w}$. This is intuitive. At the extreme, $\overline{w} < q_0$. Then there is never any situation in which it is efficient to leave for the alternative job: q^* should be set at minus infinity so there is no possibility of $\hat{q} < q^*$. At the other extreme, $\overline{w} > q_1$. Then there is no efficient work at the piece-rate firm: q^* should be set to plus infinity so that no possibility of $\hat{q} > q^*$ exists.

The addition of noise, ξ, to the problem makes the criterion deviate from $\hat{q} \gtreqless \overline{w}$. It also results in inefficient separation and inefficient retention.

References

Bonin, J. 1976. On the design of managerial incentive structures in a decentralized planning environment. *American Economic Review* 66 (September): 682–87.

Fan, L.-S. 1975. On the reward system. *American Economic Review* 65 (March): 226–29

Freixas, X.; Guesnerie, R.; and Tirole, J. 1983. *Planning under Incomplete Information and the Ratchet Effect*. Paris: CERAS, January.

Green, J. R., and Stokey, N. L. 1983. A comparison of tournaments and contracts. *Journal of Political Economy* 91 (June): 349–64.

Harris, M., and Holmstrom, B. In press. Ability, performance, and wage dynamics. *Review of Economic Studies*.

Harris, M., and Weiss, Y. 1981. The tenure phenomenon and job matching. Evanston, Ill.: Northwestern University.

Holmstrom, B. 1982. Moral hazard in teams. *Bell Journal of Economics* 13 (Autumn): 324–40.

Lazear, E. P. 1979. Why is there mandatory retirement? *Journal of Political Economy* 87 (December): 1261–64.

Lazear, E. P. 1981. Agency, earnings profiles, productivity and hours restrictions. *American Economic Review* 71 (September): 606–20.

Lazear, E. P., and Rosen, S. 1981. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* 80 (October): 841–64.

Pencavel, J. 1977. Work effort, on-the-job screening, and alternative methods of remuneration. In R. Ehrenberg (ed.), *Research in Labor Economics*, vol. 1. Greenwich, Conn.: JAI.

Riley, J. G. 1975. Competitive signalling. *Journal of Economic Theory* 10 (April): 174–86.

Rogerson, W. 1985. Repeated moral hazard. *Econometrica* 53 (January): 69–76.

Rosen, S. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82 (January/February): 34–55.

Seiler, E. 1984. Piece rate vs. time rate: The effect of incentives on earnings. *Review of Economics and Statistics* 66, no. 3:363–75.

Stiglitz, J. 1975. Incentives, risk, and information: Notes toward a theory of hierarchy. *Bell Journal of Economics and Management Science* 6 (Autumn): 552–79.

Stiglitz, J. 1981. Contests and cooperation: Towards a general theory of compensation and competition. Unpublished manuscript. Princeton, N.J.: Princeton University, April.

Wald, A. 1947. *Sequential Analysis*. New York: Dover.

Wald, A. 1950. *Statistical Decision Functions*. New York: Chelsea.

Weiss, Y. 1984. Wage contracts when output grows stochastically: The role of mobility costs and capital market imperfections. *Journal of Labor Economics* 2, no. 2 (April): 155–74.

Weitzman, M. 1976. The new Soviet incentive model. *Bell Journal of Economics* 7 (Spring): 251–57.

Weitzman, M. 1980. The "ratchet principle" and performance incentives. *Bell Journal of Economics* 11 (Spring): 302–8.

Yunker, J. A. 1973. A dynamic optimization model of the Soviet enterprise. *Economics of Planning* 13:33–51.