

IZA DP No. 1229

Incentive Effects of Bonus Payments: Evidence from an International Company

Axel Engellandt
Regina T. Riphahn

July 2004

Incentive Effects of Bonus Payments: Evidence from an International Company

Axel Engellandt

University of Basel

Regina T. Riphahn

*University of Basel,
DIW Berlin and IZA Bonn*

Discussion Paper No. 1229
July 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Incentive Effects of Bonus Payments: Evidence from an International Company*

This study uses panel data describing about 6,500 employees in a large international company to study the incentive effects of performance related pay. The company uses two performance related remuneration mechanisms. One is an individual "surprise" bonus payment. The other is a more structured system, where part of the salary is determined by individual performance evaluations. We hypothesize that effort is higher in departments where (i) performance evaluation results are more spread out, (ii) person-specific performance evaluations are more flexible over time, (iii) surprise bonuses are used more frequently. These hypotheses are tested using days of absence and overtime work as effort indicators. The tests yield that hypotheses (ii) and (iii) are supported, and that (i) cannot be tested reliably due to possible simultaneity bias in our data. We investigate and confirm the robustness of these findings. They suggest that surprise bonus payments and flexibility in the evaluation of individual performances over time provide effective incentives for employee effort.

JEL Classification: J33, M12, J24, J41, M50, C25

Keywords: bonus, merit pay, pay for performance, worker effort, social status, mobility, absenteeism, overtime

Corresponding author:

Regina T. Riphahn
WWZ
University of Basel
Post Box 517
4003 Basel
Switzerland
Email: regina.riphahn@unibas.ch

* We thank the company which provided the data and Michael Gerfin, Jennifer Hunt, the participants of the meeting of the Ausschuss für Bevölkerungsökonomie in Vienna, of the third Transatlantic Meeting of Labor Economists in Buch, and of the Annual Meeting of the European Society of Population Economists (ESPE) in Bergen for helpful comments.

1. Introduction

Does performance-related pay help to increase worker effort? Most economists and managers believe so. Yet the empirical evidence remains amazingly sparse even years after Prendergast (1999) pointed out that there has been little empirical assessment of incentive provisions for workers. This is surprising given the increasing reliance of employers on merit pay.

To address this gap in the literature we apply panel data from a large international company and investigate whether workers provide more effort when it pays. Most of the available evidence in the labor economics literature looks at aggregate outcomes only, comparing e.g. the performance of companies and entire industries with different human resource practices.¹ On the other hand, psychological studies investigate work attitudes and job satisfaction as opposed to objective measures of effort and performance.²

These two literatures reach different conclusions. While the studies applying aggregate data to estimate company and industry production functions (e.g. Groves et al. 1994, or Jones and Kato 1995) yield that bonus payments are correlated with higher output and productivity, analyses which rely on employee surveys are not as optimistic. Taylor and Pierce (1999) conclude "The appraisal/merit component (...) clearly had deleterious effects on staff members' organizational commitment and attitudes towards supervision, particularly among high

¹ Examples are Groves et al. (1994) investigating the productivity consequences of managerial autonomy in Chinese industries, Jones and Kato (1995) measuring productivity effects of employee stock-ownership and bonuses for Japanese firms, Lee and Rhee (1996) estimating similar models with South Korean time series on 8 industries, and Morton (1998) who applied quarterly industry data from Taiwan. Cable and Wilson (1989, 1990) provide evidence on productivity enhancing effects of profit-sharing in the United Kingdom and Germany. For further sources see Kahn and Sherer (1990) and Prendergast (1999).

² Taylor and Pierce (1999) evaluate the consequences of introducing a performance management system in a small government organization in New Zealand on employee attitudes. Similarly, Marsden and Richardson (1994) and Marsden et al. (2001) evaluate the effect of merit pay on employee motivation in the United Kingdom. Armstrong-Stassen et al. (1993) use survey data on 121 individuals to investigate the determinants of employee attitudes to profit sharing. They find that prior company commitment and the perception of a connection between performance and payout are important.

performing employees." They report that the bonus system lead to a perception of unfairness and cite studies with similar outcomes. Marsden and Richardson (1994) describe a performance-pay system with demoralizing effects deriving from perceived unfairness and favoritism of supervisors.

Prendergast (1999, p.56) concludes his survey on the provision of incentives in firms: "All in all, the available empirical evidence on contracts does not yet provide a ringing endorsement of the theory." He points to several shortcomings of available studies, among them the weakness of tests, the lack of reliable personnel data, empirical identification and selection problems, and the focus on workers with easily observable outputs. We attempt to overcome these limitations.

Our empirical tests are based on panel data describing about 6,500 employees of a large international company. Among these employees are production workers, researchers, administrators, and managers. We analyze objective outcomes using individual-level data. To identify the effect of performance pay on behavior, we compare workers across company departments and over time. Departments differ in the intensity with which the performance pay system is applied: While in some departments performance evaluations fully use the available spectrum of evaluation outcomes, others differentiate to a lesser degree. The validity of this identification strategy hinges on the exogeneity of worker assignment to departments, which we discuss and examine below.

This identification strategy, which is based on a comparison within a given firm, is most similar to the one used by Kahn and Sherer (1990). They applied panel data describing a firm's 92 managers and find that managers with a high bonus impact of performance work harder. However, since this result does not hold with respect to merit pay, the authors conclude that only bonus differences affect subsequent performance.

Using absenteeism and overtime work as indicators of worker effort we investigate the role of performance pay for richer and more recent data. The longitudinal nature of the data

allows us to handle carefully issues of causality, and to control for unobserved effects. Our analysis does not just look at managers, who have received much attention in the pay-performance literature, but at all workers from a Swiss unit of our company.³ By using objective outcomes as opposed to subjective performance evaluations we avoid the measurement problems that arise when worker evaluations are compared across supervisors (see e.g. Prendergast 1999 or Schwab and Olson 1990 for a discussion) or which are inherent in workers' statements on their attitudes, efforts, and performance. Clearly, objective indicators of individual behavior provide a more reliable reflection of effort than the firm or industry productivity measures applied in the literature.⁴

We confirm Kahn and Sherer's (1990) finding of a productivity enhancing effect of bonus payments. Also, worker effort appears to respond positively to the variability of individual ratings over time. Workers in departments where performance ratings vary over time and so ensure a higher degree of competition in the internal labor market provide more effort than workers in departments where individual ratings hardly change.

2. Institutional Background

We received personnel data on white and blue collar employees from an international company. The data cover one organizational unit in Switzerland between 1999 and 2002. In this time the unit employed 6,425 individuals for most of whom we have repeated annual observations.⁵

³ The literature on the pay-performance relation for managers is surveyed by Murphy (1999). For further evidence see e.g. Aggarwal and Samwick (2003).

⁴ Booth and Frank (1999) use earnings increases as reflections of productivity changes. Since we do not observe workers' earnings we cannot apply this approach.

⁵ The data excludes a very small fraction of top managers as well some "social responsibility employees" such as disabled individuals, who formally are on the payrolls even though they do not engage in productive activity in the firm anymore. In addition, we disregard even in this description those workers for whom department number and performance rating are unavailable.

The firm is organized in departments, of which we observe a total of 78 different ones in the full sample over the years. Due to organizational changes only about half of these departments exist over all four years of our data: 11 exist over three consecutive years, and 21 are observed over two consecutive periods. The remaining 7 units cannot be followed over time. On average there are 106 employees in a department, where the department size varies between 1 and 858.

In this company two performance-related remuneration mechanisms are in place: One is an individual "surprise" bonus at the order of about 1,400-3,400 EURO, which is granted for special achievements.⁶ This bonus payment is at the discretion of supervisors who can spend a given annual budget for this purpose. On average, they apply this instrument about five times a year, with significant heterogeneity across departments. About one quarter of the departments did not pay out bonuses at all. The others paid on average 7 bonuses per 100 employees per year.

The other is a more complex performance pay system, where depending on the hierarchical level of the employee between 10 and 85 percent of the annual salary is determined by the outcome of an annual individual performance evaluation.⁷ After a goal-setting session in the beginning of the year, supervisors rate the performance of employees typically on a scale between 0 and 150 percent of the originally envisioned objectives at the end of the year. The ratings are cross-checked by the supervisors' managers. The absolute amount of the payout depends on individual performance, the individual base salary, and the performance of the entire division, adhering to the following formula:

⁶ Depending on employees' salary the bonus can amount to between 10 and 100 percent of a gross monthly salary.

⁷ The salary rate of merit pay for regular employees are at 10 to 13 percent, they reach 20 percent for middle management, and increase to up to 70 and 85 percent for top management.

$$\text{Bonus pay / salary} = \text{individual performance rating} \cdot \text{division performance} \cdot \text{salary rate}^8$$

In 2002, the previously required distribution of performance ratings per department was abandoned but management clearly communicated the expectation that future rating distributions should be in the same range.⁹ Also, in 2002 mid-year reviews were introduced to complement the set of goal-setting and evaluation meetings. As the general character of the evaluation system did not change by these reforms they are disregarded in our analysis. Performance oriented pay was extended to all employees already in the mid 1990s, several years before our data was gathered.¹⁰

As a share of the firm's base salary cost the expenditures for surprise bonus payments amount to 0.3 percent and those for evaluation-based performance pay to about 15 percent. The company prides itself on paying performance-related remunerations and salaries above industry averages. It also offers an employee stock ownership plan and a stock option plan for executives.

3. Theory and Hypotheses

Within this institutional framework we test whether performance-related pay enhances worker effort. The literature shows that beneficial effects on worker behavior are not at all

⁸ If e.g. the individual performance rating is 120 percent, the division performance is 105 percent and the salary rate 10 percent, this worker receives performance pay of $1.2 * 1.05 * 0.1 = 12.6$ percent of base salary. For managers with identical personal and divisional ratings but a higher fixed salary rate of 20 percent, the payout would amount to $1.2 * 1.05 * 0.2 = 25.2$ percent of the base salary. The workers in our data belong to divisions with at least 1,000 employees such that divisional results should be exogenous for the individual employee.

⁹ About ten percent of the workers used to be grouped as low achievers, two thirds were in the medium range and about one quarter were in the group of top performers, leaving supervisors substantial leeway in their rating decisions.

¹⁰ Taylor and Pierce (1999) and Kahn and Sherer (1999) point out that a system's effects may differ depending on whether it was just introduced or has been in place for several years already. In this respect, we consider our study as an evaluation of an ongoing system where initial employee responses to its introduction already faded.

granted and that they may vary depending on the mode of performance-related remuneration.¹¹ The following determinants of the effectiveness of performance-related pay are discussed in the literature (see e.g. Baron and Kreps 1999): (i) the saliency of a payment matters, where payments that attain more employee attention such as surprise remunerations are more effective than expected payments. (ii) Workers may reduce effort when they perceive inequities, e.g. if current salary differences result from past performances that are no longer known to the individuals. (iii) Risk-averse workers with uncertain tenure value bonuses more highly than long term pay raises. (iv) Payment systems differ in the degree to which payment is determined by individual effort; effort may be lower if the payout formula allows for substitution between individual and divisional performance. (v) The transparency of an evaluation system and its reliability affect its motivational effectiveness. (vi) Tournament theory suggests that higher wage dispersion can motivate higher worker effort.

While we cannot test each of these predictions separately, we propose three hypotheses on the effectiveness of the firms' performance-based pay based on these insights. As departments differ systematically in their utilization of the worker evaluation system we observe heterogeneity in the extent to which employees' performance affects their remuneration. If performance pay enhances motivation and effort we would expect to see higher effort levels among those individuals whose performance is more strongly reflected in their pay. To measure the degree to which pay reflects performance we propose two measures:

One describes the cross-sectional variance of performance ratings in each department. We expect that individuals who work in a department with little heterogeneity in evaluation outcomes perceive less of an incentive to improve their performance than individuals for whom a change in performance is rewarded with large jumps in ratings and subsequent pay. This is our first

¹¹ Kahn and Sherer (1990) found managers to be more responsive to bonus payments than to merit pay. Schwab and Olson (1990) conclude their Monte Carlo study with the opposite result of a higher pay-performance correlation under merit-pay than under a bonus system.

hypothesis, which agrees with the predictions of tournament theory, i.e. that effort should be largest when wage dispersion is highest (H1).

The second measure describes the variability of person-specific evaluations in a given department over time: In a low variability scenario, workers receive about the same performance rating every year even if their performance varies. In a high variability scenario, individual ratings in one year have little predictive power for next year's rating as they closely match actual performance. Based on the above arguments (ii) and (iv) which favor a high correlation of performance pay to current and person-specific performance we hypothesize that performance incentives are stronger in the second, high variability scenario. Here the performance of every employee is evaluated anew every year as opposed to a situation where established judgements predominate over time (H2).

Our third hypothesis builds on numerous of the above listed arguments: we propose that worker effort is higher in departments which over the course of the year provide a larger number of surprise bonus payments per worker. In these departments performance pay is highly salient, it clearly relates to current as opposed to past performance, it depends (directly) on the effort of the individual, and the final payout in these departments is more dispersed than in departments that do not pay surprise bonuses (H3).

4. Empirical Approach and Data Description

Empirical Approach: After a descriptive analysis of the correlation between worker effort and performance pay we investigate whether the observed correlations are statistically significant in the framework of multivariate regressions which control for potential composition effects. As we observe individual workers (i) who are associated with departments (j) over several periods (t), we estimate the following baseline specification:

$$Y_{it} = \alpha + \beta I_{jt-1} + \gamma X_{ijt} + \delta_j + v_t + \mu_i + \epsilon_{it}$$

Y represents an indicator of worker i 's effort in period t , I measures the pre-period value of the performance pay incentives implied by our three hypotheses, which we consider both, jointly and individually, and X provides background indicators describing worker and department. δ controls for department type fixed effects,¹² ν reflects year fixed effects, μ represents individual-specific unobserved heterogeneity which we assume to be uncorrelated with the explanatory variables, and ϵ is the random error term. We use a least squares estimator for hours of overtime work, which in the case of our firm can be negative,¹³ and a tobit model for the always positive number of yearly days of absence. We then simulate the estimated effects to gauge their magnitude.

The analysis concludes with a number of robustness tests which address several potential limitations to the reliability of our results such as endogenous worker sorting between departments and simultaneity bias if supervisors apply the incentive instruments in response to observed worker effort.

Dependent Variables: In order to test whether the cross-sectional and the intertemporal variability of departmental performance evaluations or bonus payments affect worker behavior we need indicators of worker effort. We apply two measures which are used frequently in the literature: The first describes workers' days of absence due to health problems,¹⁴ the second indicates how many hours of overtime an individual accumulated at the end of the year.¹⁵ Swiss

¹² In our company we find three types of departments: production, administration, and research.

¹³ Workers can end a month with less than the contracted number of hours. While on average a share of about 13 percent of all workers end a month with negative overtime its average amount is much less than half of the average amount of positive overtime (-14.4 hours vs. 37.1 hours).

¹⁴ For other studies using absenteeism as an indicator of effort see e.g. Treble (2001), or Ichino and Riphahn (2004), or Engellandt and Riphahn (2004).

¹⁵ Landers et al. (1996) discuss scenarios where firms use hours of work as an indicator of employee characteristics. Drago (1991) uses the willingness to work an extra unpaid 20 minutes and the propensity to go to work even if not feeling well as performance indicators. Sousa-Poza and Ziegler (2003) consider overtime work as an indicator of worker productivity.

labor law demands that employees are compensated for overtime work (Rehbinder 2002). In our firm overtime work is not remunerated financially but is used to substitute for working hours at a later time. However, workers cannot carry balances of more than 120 hours from one month to the next.¹⁶ The data documents that numerous workers accumulate more than 120 hours of overtime which ends up being a gift to the company. Overtime within the 120 hours limit can be interpreted as a time credit that the employee grants the company. The lower an individual's motivation the less likely the person may be to work beyond contractual requirements and the lower the balance on the overtime account will be. Therefore a worker's number of accumulated overtime hours is a plausible indicator of motivation and effort.

Explanatory Variables: As our three key indicators of performance pay incentives we consider the standard deviation of the department's performance ratings (H1) and an indicator for the average change in individual ratings over time in this department to measure the flexibility in the assignment of individual ratings (H2). The latter measure is calculated in two steps: First, we obtain for every person the difference in performance ratings for the years $t-2$ and $t-1$. In a given department the average of these differences should approximate zero. Therefore, we calculate the standard deviation of these person-specific changes by department: The higher the departmental rating changes the larger the standard deviation of this measure. Finally, we consider the department-specific number of bonus payments per year and employee (H3).

Our control variables (X) describe the individual worker and her department. Among the individual characteristics we consider age and its square, sex, and marital status. To proxy human capital we use indicators of requirements for the person's job, which are available in twenty discrete categories (levels). The department is characterized by the number of employees, their

¹⁶ Workers who e.g. accumulate 140 hours of overtime at the end of one month will start out the next month with a surplus of only 120 hours. 20 hours are deducted from their overtime account.

average age, job level, and the share of male employees. We can distinguish production, administration, and research departments and control for these group fixed effects. Descriptive statistics are presented in Table 1.

Sample: As our effort measures may vary by type of contract we restrict our sample to include only full time employees who were employed with the firm year-round. We drop a few observations with missing or extreme performance ratings (values 0 and 1 or beyond 200), which reflect individual circumstances that are not necessarily correlated to their actual performance.

Since our identification strategy relies on assigning department characteristics to individual outcomes we have to drop observations for whom the department indicators are missing. Also, we can measure incentive indicators only if individuals are employed in departments that exist over the three periods required to generate the indicators.¹⁷ Therefore we lose workers employed in departments that did not exist for at least three periods. As it is unlikely that reorganizations of the firm respond to the heterogeneity of departmental ratings or the frequency of bonus payments we consider the selection based on department stability as an exogenous criterion. After these selections our analysis sample consists of 8,872 worker-year observations. These employees are employed in 44 different departments of which 32 are observed over the two periods of 2001 and 2002.¹⁸

5. Results

Table 2 describes the average effort outcomes for individuals in departments where performance incentives are above or below department average. We would expect less effort, i.e.

¹⁷ To generate the lagged standard deviation in individual rating changes by department we need the difference between ratings in periods t-1 and t-2.

¹⁸ We have only two observation years in the final sample because the definition of our main independent variables requires information on two prior periods and our panel covers four years overall.

more days of absence and fewer overtime hours in departments with small standard deviations of annual performance ratings and with small intertemporal rating changes (H2), and with few bonus payments (H3). However, not all of the table's entries confirm our expectations.

The first row yields that contrary to expectations individuals in departments with a small dispersion in annual performance ratings provide significantly more hours of overtime and accumulate significantly fewer days of absence. The next row provides partial support for hypothesis 2: We find more overtime hours - but not less absenteeism - in departments where individual performance ratings are more flexible over time. The last row corroborates hypothesis 3: Individuals who work in departments which use bonuses more frequently put in more overtime and are significantly less absent. Table 2 therefore yields only limited support for our general hypothesis that greater incentives are correlated with more effort. Next, we investigate in multivariate regression analyses whether this outcome is robust to controls for potential composition effects.

The results of the estimations are presented in Tables 3a and 3b. We apply random effects models to account for individual-specific unobserved heterogeneity. The last rows present the estimated variances of the individual-specific error component μ_i . Using a LM test we reject the null hypothesis that the variance of μ_i is equal to zero at high levels of statistical significance in all specifications.¹⁹

The first specification in Tables 3a and 3b considers only three explanatory variables describing a worker's department: the standard deviation of individual ratings in period t-1, the standard deviation of individual-specific rating changes between periods t-1 and t-2, and the number of bonuses paid per 100 employees in year t-1. The significant coefficients on the

¹⁹ As an alternative to the random effects estimator we applied regressions with standard errors clustered by individual. The results remain unchanged.

standard deviation of ratings in both tables indicate a correlation between incentive and effort outcome that is contrary to hypothesis 1: The higher the rating dispersion and the more it pays to provide effort, the lower the number of overtime hours and the more days of absence are accumulated by workers.

The coefficients on the other two incentive measures are of the expected sign: the more variable performance ratings are over time the higher overtime and the lower absenteeism. Workers in departments with more frequent bonus payments provide more effort. The coefficient estimates in the overtime model are highly statistically significant, those in the absences model are not.²⁰

In order to account for potential composition effects the subsequent columns in Tables 3a and 3b add covariates to the model. In specification 2 a set of 14 individual-specific covariates is considered, in specification 3 we add four measures describing department characteristics, and in specification 4 a year dummy and department type effects are considered, as well. However, the nature of the correlation between the incentive and effort measures does not vary substantially when different specifications are considered. With one exception none of the coefficients in the first three rows changes sign, and only in a few cases is the precision of the estimates affected.

Therefore, the conclusions based on the first column appear to be robust across specifications: The data support the second and third hypotheses regarding the intertemporal variability in individual ratings over time and the effort enhancing effects of bonus payments. In contrast, the hypothesis that effort is higher in departments that evaluate with greater rating dispersion seems to be rejected.²¹

What may explain these surprisingly different incentive effects of rating dispersion and

²⁰ The same holds when a linear estimator replaces the Tobit estimator.

²¹ These results also are robust to considering only one of the three incentive measures at a time. Only the effect of the rating change in the absence equation turns insignificantly positive when the other incentive indicators are not controlled. This corresponds to the bivariate correlation pattern observed in Table 2.

rating flexibility over time? The coefficient estimates could suffer from simultaneity bias if supervisors chose departmental ratings and bonus generosity *in response to* observed effort e.g. as a discipline device, instead of - as considered so far - as incentives for future behavior. In that case our results were misleading. Since our data lacks instruments we cannot perform proper endogeneity tests. Instead, we follow two alternative avenues to gauge the plausibility of simultaneity bias: First, we investigate the correlation patterns of effort outcomes and incentives over time. A negative correlation between average effort and subsequent rating dispersion by department would be suggestive of a supervisor's response to effort. Second, we apply simplified simultaneity tests which inform about intertemporal correlation patterns between departmental effort and incentive measures. If supervisors set individual ratings independent of overall department outcomes we should not find significant correlation patterns between lagged departmental effort and current incentive measures.

The correlation patterns between effort outcomes and incentive measures are depicted in Table 4a. Simultaneity bias would be likely if we found significant negative correlations between lagged effort and contemporaneous incentive measures. Table 4a yields little systematic evidence for absences. However, with respect to overtime work we find indeed significant "discipline device" type correlations for the dispersion of ratings but not for the other incentive indicators: Low levels of past overtime work are correlated with high dispersion in current ratings. This supports the hypothesis of simultaneity bias for the rating dispersion outcome.

In order to measure department-specific effort we use average effort values by department and year and apply linear regressions with random effects to the following model for each of the $j = 1, 2, 3$ incentive indicators (I_{jt}) and separately for the two ($k=1,2$) effort indicators absences and overtime (Y_{kt-1}):

$$I_{jt} = \alpha_{jk} + \beta_{jk} I_{jt-1} + \gamma_{jk} Y_{kt-1} + \delta_{jk} + v_{jkt} + \epsilon_{jkt}$$

Table 4b presents the estimates of γ_{jk} for each of the six models. We find again that rating

dispersion appears to be correlated with lagged effort outcomes in the case of overtime work. This result is robust to controls for additional explanatory variables and to the choice of the estimator.²²

This evidence supports the view that in contrast to rating flexibility and bonus payments the incentive indicator "rating dispersion" is likely to be endogenous and affected by simultaneity bias. Given the complexity of the instrument "dispersion of changes in individual ratings" it seems plausible that this is not a disciplinary tool handily applied by supervisors. Similarly, suppressing surprise bonus payments in response to low effort would not appear to be a plausible supervisor response.

Therefore we perform the robustness tests only for the two incentive measures which appear to be exogenous to past effort. Table 5 presents the estimates of the original model, when the potentially endogenous indicator for the lagged standard deviation of ratings is omitted. The results for the remaining incentive indicators are nearly unchanged: for the overtime model we find clear and significant incentive effects in the expected direction, whereas the estimates in the absenteeism model are statistically insignificant.²³ The impact of rating flexibility and of bonus payments on the number of overtime hours is substantial: an increase in the dispersion of rating flexibility by one standard deviation from the mean raises overtime hours by 11.4 percent. Adding a standard deviation to the average of the observed bonus frequency yields an increase in overtime hours by 9.6 percent; providing one more bonus per year per 100 employees yields an average increase in the number of overtime hours by 0.34 or about 20 minutes.

²² Certainly one would wish to control for additional lagged values in the model, however since our panel is so short and several indicators are not available for the first year of the data (e.g. bonus payment, overtime, absences) we cannot generate second lags for all of the incentive measures. A regression of rating dispersion on lagged overtime controlling for two lags of rating dispersion also yielded a significant negative coefficient for the overtime measure.

²³ We reestimated columns 1 through 3 of Table 3 with very similar results, which we do not present to save space. Also, we estimated the models without random effects but with standard errors clustered by individual. The results were unaltered.

Robustness tests confirmed the insignificant correlation of absenteeism with incentive measures throughout. Therefore we do not present further results for this indicator of worker effort. While the lack of responsiveness of absenteeism as a measure of effort would be surprising in many labor markets this is less so for Switzerland: in prior research using representative Swiss labor market data we found that workers do not modify their absence behavior in response to behavioral incentives (see Engellandt and Riphahn 2004).

Finally, we investigate, whether the incentive effects found so far are robust to changes in sample, variable definition, and specification. We reestimated the overtime model of Table 5 for various subsamples to investigate whether certain employee subgroups respond differently to incentives provided by performance-based pay. The full sample was split by workers' sex, we estimated separately for those with performance ratings above and below the median, for those working in large and in small departments, and for those with tenure of more and less than five years. While in some cases the standard errors of our estimates differed from those presented in Table 5 the main conclusions regarding the direction of the effects are robust to these different groupings of the sample.²⁴

There are only types of subsamples for which we find systematically different - and interesting - response patterns: managers vs. regular employees and department types (see Table 6a). The former results suggest that the overall responses of managers to pay for performance incentives is - if at all significant - of a smaller magnitude than that of other employees. This result adds to the literature on managerial effort which generally focuses on top managers and their responses to stock and option holdings (Murphy 1999). However, if overtime hours and absence days provide a poorer proxy for the effort of managers than for other workers we would actually expect the reduced response intensity among managers.

The last three columns in Table 6a yield that among the three types of departments in our

²⁴ These results are not provided to save space, but are available from the authors.

company the incentive mechanisms yield the largest marginal effects among production workers. A potential explanation for the effectiveness of the incentive measures for this group of employees might be that supervisors here have particularly objective performance measures which are easily observable and reflect individual effort more closely than the measures available in other departments. Argument (iv) in our theoretical section predicted a higher effectiveness of incentive mechanisms when they are closely determined by individual effort - this could apply to production workers in particular. For employees in the research departments a more intense application of performance incentives appears to reduce overtime. It is plausible that for scientists both, the incentive mechanism and the effort measure are not really applicable: the effort of researchers may not be reflected in the number of overtime hours worked and in these occupations individual performance is particularly difficult to evaluate. Therefore a deviation from our hypothesized patterns of behavioral response is not surprising. Finally, employees in administrative jobs do respond positively to stronger incentives even though only the bonus effect is measured precisely. This pattern seems plausible given that a large share of the employees grouped in this type of department are managers, which we discussed above.

As a second robustness check we modified the definition of our incentive measures. Fundamentally, we are testing whether past experience of department policies affects future behavior. As workers at times are reassigned to new departments, we can specify the indicator of department policies in two alternative ways: so far, we depicted e.g. the bonus payments which occurred last year in the department where an individual works today. One could argue as well that it is the individual experience with bonus incentives that matters as opposed to the history of the department which a new employee just joined. We redefined our incentive measures to reflect the average of last year's rating and bonus experience of all employees' currently working in department j , independent of where this experience was gathered. Even with redefined incentive indicators the estimated coefficients do not change in sign or significance. Therefore

the definition of our incentive measure does not seem to bias our results (see first column in Table 6b).

There are many ways to measure dispersion and so far we only looked at the standard deviation. To investigate the sensitivity of the results to this choice we applied three alternative measures of the dispersion in rating changes: The variance, the average deviation from the median, and the difference between the 90th and 10th percentile of the department-specific distributions. The last three columns in Table 6b show the results of applying these measures to our baseline specification. With one exception the results are robust to these specification changes. Only for the 90-10 differential do we obtain an unexplainable different effect of rating flexibility.²⁵

Since we know that workers forgo any overtime hour worked that extends the balance on their overtime account beyond 120 hours at the end of the month, we coded a third indicator of worker effort that does not just measure the number of overtime hours worked, but that indicates whether the worker worked for free by running up an overtime balance of more than 120 hours. This was the case for about three percent of our observations. We fitted a Probit estimator on the indicator of unpaid overtime work using our baseline specification of Table 5. The results in Table 6c support the previous findings, with positive impacts of both, high individual variation in performance ratings and a large number of bonus payments on individuals' propensity to provide more than 120 hours of overtime work.

In principle, individuals might endogenously move between departments in response to departmental performance pay policies. While employees do move (in our data about 7 percent of the observations changed departments), in private conversation the human resources management of our company considered it extremely unlikely that moves occur in response to

²⁵ We repeated the baseline regressions dropping the observations with the top and bottom 5 percent values of the independent variable. This did not change the baseline results. Therefore the different effects of the 90-10 differential is not due to a small number of outlier observations.

departmental performance pay policies. Nevertheless, as one additional robustness test we dropped those individuals from the sample, who were observed to move between departments. If our results were due to endogenous sorting of workers to departments we would expect a clear drop in the coefficient estimates when the "sorted workers" are omitted from the sample. The estimation results for both dependent variables, i.e. overtime hours and the indicator of more than 120 overtime hours are presented in Table 6d. Our results are robust to this change in sample, where the marginal effects of the incentive measures now even slightly increased in the 120 hours model. This suggests that endogenous department changes are an unlikely explanation of our findings.

Our final set of robustness checks accounts for the potential omission of relevant variables at the department and individual level by considering fixed effects estimators. The results are presented in Table 6e and confirm prior findings. More flexible assignment of ratings over time and a higher frequency of bonus payments are correlated with higher effort. The coefficients in all three columns are identified based both on changes in departmental policies over time (within effects) and on individuals changing departments. The difference between the three columns relates to the type of unobserved heterogeneity that is controlled for: in columns one and three department fixed effects, such as department culture and supervisor characteristics are controlled for. In columns two and three person-specific heterogeneity such as intrinsic motivation and leisure preference are accounted for. The signs of our key results are unchanged when fixed effects are controlled for. Interestingly, we obtain a much larger marginal effect of rating flexibility in column two, which does not control for department-specific heterogeneity. This suggests that department-specific heterogeneities cause a downward bias in the estimation when they are not controlled for, i.e. unobserved supervisor effects might reduce the effectiveness of incentive mechanisms.

Overall, the analysis allows us to draw the conclusion that effort seems to be higher in

departments where person-specific ratings are more flexible over time and which provide more bonuses per employee, supporting hypotheses 2 and 3. Since we cannot reject the possibility of simultaneity bias in the coefficient estimates of rating dispersion we do not have unbiased estimates of the incentive effect addressed by hypothesis 1.

6. Conclusions

This study measures the effectiveness of performance pay as an incentive mechanism to generate worker effort. We applied panel data on 6,425 employees in an international company and investigated how their effort as measured by overtime hours and (non-)absence days responded to three types of incentives: the dispersion in department-specific performance ratings (H1), the flexibility of individual ratings over time in a given department (H2), and the frequency of surprise bonus payments in a department (H3). We hypothesize positive effort effects for all three incentive instruments.

The results support the hypothesis that worker effort responds positively to surprise bonus payments (H3). Similarly, workers provide more effort if their supervisors reevaluate their performance anew from year to year as opposed to leaving individual positions unchanged over time (H2). Thus, the "contestability of ratings" generates the expected behavioral response. The effectiveness of the two instruments appears to be of comparable magnitude: an increase in incentive intensity by one standard deviation beyond the mean is correlated with an increase in overtime by 11 percent in the case of rating flexibility and by almost 10 percent in the case of bonus payments.

Surprisingly, the dispersion in departmental performance ratings is significantly negatively related to subsequent worker effort (H1). Our estimate may suffer from simultaneity bias if supervisors responded in their rating behavior to the effort observed in the department. In order to gauge the relevance of this explanation we compared the bivariate correlation patterns

between rating dispersion and worker effort and found significant negative correlations between the average level of overtime work provided in a department and *subsequent* rating dispersion. This surprising correlation was confirmed in simultaneity tests and suggests a pattern of reverse causality between effort and incentives for the indicator of rating dispersion. Since such simultaneity bias does not appear for the other two incentive indicators our analysis focused on these two measures.

We confirm our findings of positive incentive effects for these two measures in numerous robustness tests which look at different subsamples, alternative definitions of the incentive indicators as well as different estimators. Interestingly, the effectiveness of these incentive mechanisms vary with the extent to which worker output is observable. We find significant correlations between incentives and effort even when controlling for department or individual fixed effects. Thus both types of performance pay applied in the company, surprise bonuses and performance oriented pay, provide effective incentives for employee productivity. While our findings on the incentive effects of bonuses confirm prior findings, this is the first study to provide evidence regarding the importance of flexible individual performance evaluations.

Bibliography

- Aggarwal, Rajesh H. and Andrew A. Samwick, 2003, Performance incentives within firms: The effect of managerial responsibility, *The Journal of Finance* 58(4), 1613-1649.
- Armstrong-Stassen, Marjorie, Arthur Wolfram, and Lorraine D'Souza, 1993, The influence of individual characteristics, plan characteristics and work-related factors on salaried employees' attitudes towards profit sharing: A Canadian case, *The International Journal of Human Resource Management* 4(4), 933-943.
- Baron, James N. and David M. Kreps, 1999, *Strategic Human Resources. Frameworks for General Managers*, John Wiley and Sons, New York et al.
- Booth, Alison L. and Jeff Frank, 1999, Earnings, productivity, and performance-related pay, *Journal of Labor Economics* 17(3), 447-463.
- Cable, John and Nicholas Wilson, 1989, Profit-sharing and productivity: An analysis of UK engineering firms, *Economic Journal* 99(236), 366-375.
- Cable, John and Nicholas Wilson, 1990, Profit-sharing and productivity: Some further evidence, *Economic Journal* 100(401), 550-555.
- Drago, Robert, 1991, Incentives, pay, and performance: a study of Australian employees, *Applied Economics* 23(9), 1433-1446.
- Engellandt, Axel and Regina T. Riphahn, 2004, Temporary Contracts and Employee Effort, forthcoming: *Labour Economics*.
- Groves, Theodore, Yongmiao Hong, John McMillan, and Barry Naughton, 1994, Autonomy and incentives in Chinese state enterprises, *Quarterly Journal of Economics* 109(1), 183-209.
- Ichino, Andrea and Regina T. Riphahn, 2004, The effect of employment protection on worker effort. A comparison of absenteeism during and after probation, *CEPR Discussion Paper No. 3847*, London, forthcoming: *Journal of the European Economic Association*.
- Jones, Derek C. and Takao Kato, 1995, The productivity effect of employee stock-ownership plans and bonuses: Evidence from Japanese panel data, *American Economic Review* 85(3), 391-414.
- Kahn, Lawrence and Peter D. Sherer, 1990, Contingent pay and managerial performance, *Industrial and Labor Relations Review* 43(3), 107-S - 120-S.
- Landers, Renée M., James B. Rebitzer, and Lowell J. Taylor, 1996, Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms, *American Economic Review* 86(3), 329-348
- Lee, Michael Byungnam and Yinsog Rhee, 1996, Bonuses, unions, and labor productivity in South Korea, *Journal of Labor Research* 17(2), 219-238.
- Marsden, David and Ray Richardson, 1994, Performing for pay? The effects of 'Merit Pay' on motivation in the public service, *British Journal of Industrial Relations* 32(2), 243-261.
- Marsden, David, Stephen Frenche, and Kubo Katsuyuki, 2001, Does performance pay de-motivate, and does it matter? *London School of Economics - Centre for Economic Performance Discussion Paper No. 503*.

- Morton, Peter, 1998, Annual bonuses, employment and productivity change in Taiwan's manufacturing sector, *Journal of Economic Behavior & Organization* 37(4), 481-507.
- Murphy, Kevin J., 1999, Executive compensation, in: Ashenfelter O. and David Card (1999) *Handbook of Labor Economics* vol. 3, 2485-2563.
- Prendergast, Canice, 1999, The provision of incentives in firms, *Journal of Economic Literature* 37(1), 7-63.
- Rehbinder, Manfred, 2002, *Schweizerisches Arbeitsrecht*, 15.ed., Stämpfli Verlag, Bern.
- Schwab, Donald P. and Craig A. Olson, 1990, Merit pay practices: Implications for pay-performance relationships, *Industrial and Labor Relations Review* 43(3), 237-S - 255-S.
- Sousa-Poza, Alfonso and Alexandre Ziegler, 2003, Asymmetric information about workers' productivity as a cause for inefficient long worker hours, *Labour Economics* 10(6), 727-747.
- Taylor, Paul J, and Jon L. Pierce, 1999, Effects of introducing a performance management system on employees' subsequent attitudes and effort, *Public Personnel Management* 28(3), 423-452.
- Treble, John, 2001, Productivity and effort: The labor-supply decisions of late Victorian coalminers, *Journal of Economic History*, 61(2), 414-438.

Table 1 Descriptive Statistics

Variable Group and Description	Mean	Standard Deviation
Dependent Variables		
Overtime hours	23.843	43.583
Days of absence	7.36	18.249
Indicators of Performance Pay Incentives in t-1		
Standard deviation of department ratings	17.303	2.482
Standard deviation of department rating changes	16.492	3.184
Bonus payments per 100 employees	7.485	6.691
Individual Characteristics		
Age	43.408	9.593
Age squared	1976.302	821.142
Male (0/1)	0.622	0.485
Married (0/1)	0.656	0.475
Job level = 1 (0/1)	0.007	0.083
Job level = 2 (0/1)	0.046	0.21
Job level = 3 (0/1)	0.104	0.305
Job level = 4 (0/1)	0.123	0.329
Job level = 5 (0/1) (<i>reference group</i>)	0.167	0.373
Job level = 6 (0/1)	0.129	0.336
Job level = 7 (0/1)	0.07	0.256
Job level = 8 (0/1)	0.102	0.302
Job level = 9 (0/1)	0.107	0.309
Job level = 10 (0/1)	0.061	0.24
Job level = 11 or beyond (0/1)	0.084	0.277
Department Characteristics		
Average age of employees	43.408	1.672
Share of male employees	0.622	0.159
Average job level of employees	6.35	1.64
Number of employees (in thousand)	0.264	0.219
Year Dummies		
Year = 2001 (0/1)	0.427	0.495
Year = 2002 (0/1) (<i>reference group</i>)	0.573	0.495
Department Type		
Research department (0/1) (<i>reference group</i>)	0.562	0.496
Administrative department (0/1)	0.112	0.315
Production department (0/1)	0.326	0.469

Note: The table describes all variables at the level of N = 8,872 observations in the analysis file. Department characteristics are weighted by the number of employees per department.

Table 2 Effort Outcomes by Incentive Intensity

Average characteristics of employees in departments with	Overtime Hours in t	Days of Absence in t
Standard deviation of department ratings in t-1		
≤ Mean	27.24	6.94
> Mean	21.04 **	7.71 *
Standard deviation of department rating changes in t-1		
≤ Mean	22.93	5.9
> Mean	24.23 *	7.98 **
Bonus payments per employee in t-1		
≤ Mean	23.43	8.22
> Mean	24.73 *	5.54 **

Note: The mean incentive intensities are calculated based on comparisons across departments and years (N=76), while the average overtime and absence outcomes reflect the characteristics of the 8,872 person-year observations. ** and * indicate whether the differences in mean effort outcomes across departments below and above the median are significantly different at the 5 and 10 percent level respectively, in one-sided tests.

Table 3a Random Effects Linear Regression of Overtime Hours

	1	2	3	4
Indicators of Performance Pay Incentives in t-1:				
Std. Deviation of Ratings	-1.061** (0.205)	-1.056** (0.202)	-1.294** (0.201)	-1.351** (0.203)
Std. Deviation of Rating Change	0.685** (0.157)	0.640** (0.155)	0.466** (0.154)	1.267** (0.179)
Bonus Payments	0.361** (0.070)	0.121* (0.072)	0.635** (0.079)	0.412** (0.081)
Individual Characteristics (14)	-	yes	yes	yes
Department Characteristics (4)	-	-	yes	yes
Year Fixed Effect (1)	-	-	-	yes
Department Type Fixed Effects (2)	-	-	-	yes
Variance of Random Effect (μ)	399.857**	290.247**	240.595**	222.683**

Table 3b Random Effects Tobit Regression of Days of Absence

	1	2	3	4
Indicators of Performance Pay Incentives in t-1:				
Std. Deviation of Ratings	0.367** (0.120)	0.210* (0.117)	0.146 (0.119)	0.117 (0.120)
Std. Deviation of Rating Changes	-0.031 (0.091)	-0.088 (0.089)	-0.131 (0.090)	-0.017 (0.107)
Bonus Payments	-0.361** (0.043)	-0.051 (0.044)	0.006 (0.048)	-0.020 (0.049)
Individual Characteristics (14)	-	yes	yes	yes
Department Characteristics (4)	-	-	yes	yes
Year Fixed Effect (1)	-	-	-	yes
Department Type Fixed Effects (2)	-	-	-	yes
Variance of Random Effect (μ)	258.61**	195.73**	195.38**	195.43**

Note: The tables present estimated coefficients and standard errors in parentheses. All models are estimated on 8,872 person-year observations. ** and * indicate statistical significance at the 5 and 10 percent level. The individual and department characteristics contain those listed in Table 1. The number of estimated parameters for each group of indicators is provided in parentheses. The last rows of the tables present estimates of the variance of the unobserved individual effects μ . The asterisks indicate that the variance estimates were significantly different from zero at the 1 percent level in all models.

Table 4a Bivariate Correlation Patterns: Incentive Measures and Effort Outcomes

	Std. Dev. of Rating in t	Std. Dev. of Rating Change in t	Bonus Payments in t
Average Overtime in:			
t-2	-0.286 *	-0.023	-0.182
t-1	-0.244 **	-0.051	-0.171
t	-0.291 **	-0.011	-0.045
Average Absences in:			
t-2	-0.298	-0.237	-0.026
t-1	0.115	-0.173	-0.173
t	-0.011	-0.199	-0.199 *

Note: The figures present bivariate correlation coefficients between the incentive measures observed in department j in period t and the effort outcomes observed in this department two and one period before and contemporaneously. Figures are based on 76 department year observations. ** and * indicate statistical significance at the 5 and 10 percent level.

Table 4b Linear Random Effects Regressions Testing for Simultaneity

	Std. Dev. of Rating in t Coef. (abs. t-value)	Std. Dev. of Rating Change in t Coef. (abs. t-value)	Bonus Payments in t Coef. (abs. t-value)
Average Overtime in t-1	-0.030 * (1.61)	-0.012 (0.50)	-0.0004 (0.55)
Average Absence in t-1	0.077 (0.93)	0.143 (0.10)	-0.003 (1.01)

Note: The regressions use 76 observations at the department-year level. Besides the lagged effort measure the regressions control for the lagged value of the dependent variable (i.e. the respective incentive measure), average characteristics of employees in the department (average level, average age, share males), the department type fixed effects, and the year fixed effect. ** and * indicate statistical significance at the 5 and 11 percent level.

Table 5 Baseline Results: Random Effects Estimation of Overtime Hours and Days of Absence

	Overtime Hours Linear Regression	Days of Absence Tobit
Indicators of Performance Pay Incentives in t-1:		
Std. Deviation of Rating Changes	0.856** (0.169)	0.016 (0.102)
Bonus Payments	0.341** (0.081)	-0.012 (0.049)
Individual Characteristics (14)	yes	yes
Department Characteristics (4)	yes	yes
Year Fixed Effect (1)	yes	yes
Department Type Fixed Effects (2)	yes	yes
Variance of Random Effect (μ)	200.813**	195.5**

Note: The tables present estimated coefficients and standard errors in parentheses. All models are estimated on 8,872 person-year observations. ** and * indicate statistical significance at the 5 and 10 percent level. The individual and department characteristics contain those listed in Table 1. The number of estimated parameters for each group of indicators is provided in parentheses. The last rows of the tables present estimates of the variance of the unobserved individual effects μ . The asterisks indicate that the variance estimates were significantly different from zero at the 1 percent level in all models.

Table 6a Random Effects Linear Regression of Overtime Hours for Regular Employees and Managers and by Department Type

	Type of Worker		Department Type		
	Employee	Manager	Production	Research	Administration
Indicators of Performance Pay Incentives in t-1:					
Std. Deviation of Rating Changes	1.114** (0.186)	0.210 (0.357)	2.506** (0.375)	-0.767** (0.235)	0.200 (0.283)
Bonus Payments	0.404** (0.089)	0.082 (0.172)	0.786** (0.190)	-0.498** (0.122)	0.531** (0.168)
Number of observations	8,163	746	2,893	4,987	992

Table 6b Random Effects Linear Regression of Overtime Hours with Alternative Definitions of Departmental Incentive Characteristics

	Indicators Based on Individual Experience	Alternative Dispersion Measures		
		Variance	Median	90-10
Indicators of Performance Pay Incentives in t-1				
Std. Deviation of Rating Changes	1.087** (0.193)	0.0269** (0.005)	1.201** (0.221)	-0.186** (0.061)
Bonus Payments	0.373** (0.852)	0.366** (0.080)	0.361** (0.080)	0.416** (0.080)

Table 6c Random Effects Probit Regression of More than 120 Overtime Hours

	1	2	3	4
Indicators of Performance Pay Incentives in t-1:				
Std. Deviation of Rating Change	0.062** (0.013)	0.050** (0.009)	0.051** (0.011)	0.060** (0.013)
Bonus Payments	0.008* (0.005)	-0.008* (0.004)	0.015** (0.005)	0.009* (0.005)
Individual Characteristics (14)	-	yes	yes	yes
Department Characteristics (4)	-	-	yes	yes
Year Fixed Effect (1)	-	-	-	yes
Department Type Fixed Effects (2)	-	-	-	yes

Table 6d Random Effects Regressions of Overtime without Observations of Individuals who Changed Departments

	Overtime Hours	More than 120 hours
Indicators of Performance Pay Incentives in t-1:		
Std. Deviation of Rating Changes	0.807** (0.203)	0.063* (0.015)*
Bonus Payments	0.337** (0.093)	0.011* (0.006)*
Number of observations	7,220	7,220

Table 6e Linear Regression of Overtime Hours with Controls for Fixed Unobserved Heterogeneity at the Department and Individual Level

	Overtime		
	Department FE	Individual FE	Dep. and Ind. FE
Indicators of Performance Pay Incentives in t-1:			
Std. Deviation of Rating Changes	0.938** (0.263)	1.856** (0.297)	0.914** (0.323)
Bonus Payments	0.542** (0.102)	0.590** (0.128)	0.599** (0.130)

Note: The tables present estimated coefficients and standard errors in parentheses. If not indicated otherwise, all models are estimated on 8,872 person-year observations, in Table 6a 8,163 employee and 746 manager observations are used. ** and * indicate statistical significance at the 5 and 10 percent level. The models use the same specification as in column 4 in Table 3a and 5. The standard errors in the estimations with department fixed effects in Table 6d are robust at the individual level to account for heteroskedasticity.