Incentive Effects of Bonus Payments: Evidence from an International Company

Axel Engellandt
(University of Basel)

and

Regina T. Riphahn
(University of Basel, IZA, DIW)

May 28, 2004

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 evaluations are more heterogeneous,
iii) 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 (i) and (iii) are supported, and that (ii) 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.

Keywords: bonus, merit pay, pay for performance, worker effort, social status, mobility, absenteeism, overtime
JEL Code: J33, M12, J24, J41, M50, C25

Correspondence to
Regina T. Riphahn
WWZ - University of Basel
Post Box 517
CH - 4003 Basel
Phone 0041 - 61 - 267 3367
Fax: 0041 - 61 - 267 3351
regina.riphahn@unibas.ch

We thank the company which provided the data and Michael Gerfin and participants of the Ausschuss für Bevölkerungsökonomie for helpful comments.

Despite many wide-ranging claims about their supposed importance, there has been little empirical assessment of incentive provision for workers.
(Prendergast, 1999, p.7)

1. Introduction

Does performance related pay help to increase worker effort? Most economists and managers believe so. Yet the empirical evidence is amazingly sparse even five years after Prendergasts' 1999 survey of the field. Given the increasing reliance of employers on merit pay this is surprising.

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)

[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 as measured in four subsequent individual surveys. 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 significant predictors of attitudes toward profit sharing.

conclude "The appraisal/merit component (...) clearly had deleterious effects on staff members' organizational commitment and attitudes towards supervision, particularly among high 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. 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. As workers' assignment to departments is determined by central organizational units and not by the individual, selection problems should not affect the results.

This identification strategy is most similar to the one used by Kahn and Sherer (1990), who applied panel data describing a firm's 92 managers. First, they predict for each manager the marginal return in terms of bonus and merit pay to a unit of subjectively perceived performance level. Then the authors regress subsequent performances on the predicted returns. They find that managers with a high bonus impact of performance have higher subsequent performance levels.

Interestingly, this result does not hold with respect to merit pay. The authors conclude that only differences in bonuses 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 carefully handle the issues of causality, and to control for unobserved effects. Our analysis does not just look at managers who received much attention in the pay-performance literature but at all workers from a Swiss unit of the 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 one-time bonus payments. In addition, worker effort responds positively to the availability of earnings mobility as represented by flexible evaluations: workers in departments where performance ratings vary over time provide more effort than workers in departments where individual ratings hardly change.

Next we describe the institutional setting in the firm that provided our data, and section three discusses theory and hypotheses. Section four describes the data and empirical approach before section five presents the results and section six concludes.

---

[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, a measurement strategy that is appropriate for their household panel data but not for data from one given company.

2. Institutional Background

The personnel data we received from an international company cover one organizational unit in Switzerland between 1999 and 2002. In this time our unit employed on average about 6,425 individuals for most of whom we have repeated observations.[5]

The firm is organized in departments, of which over the years we observe a total of 78 different ones in the full sample. 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 2 - 5,000 Swiss Francs (about 700-1,700 EURO) which is granted for special achievements.[6] This bonus payment is at the discretion of supervisors who on average 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 frequency of bonus payments varied with a mean of 4 per 100 employees in the bottom and a mean of 13 bonuses per 100 employees in the top half of the distribution of bonus paying departments.

The other is a more complex merit 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

[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 two key variables were unavailable (department number and performance rating).

[6] Depending on employees' salary the bonus amounts to between 10 and 100 percent of a gross monthly salary.

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 then cross-checked by the supervisors' managers. The absolute amount of the merit payout depends not only on individual performance, and individual base salary, but is additionally influenced by the performance of the entire division.[8]

In 2002 the previously required distribution of performance ratings assigned in every department was abandoned but management clearly communicated the expectation that future 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 substantively by these reforms we do not pay much attention to the changes.[10] Performance oriented pay was extended to all employees already in the mid 1990s, several years before our data was gathered.

As a share of base salary costs the expenditures for surprise bonus payments amount to 0.3 percent and those for evaluation based merit pay remain at about 15 percent. The company prides itself as paying performance related remunerations and overall salaries well above industry

---

[7] The salary shares 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.

[8] The merit based payout is determined as the product of individual performance ratings (e.g. 120 percent), division performance (e.g. 105 percent) and a fixed salary rate of e.g. 10 percent for the lowest hierarchical levels (up to 20 percent for the highest). This person receivs merit pay of 1.2 * 1.05 * 0.1 = 12.6 percent of the base salary. For managers with identical personal and divisional ratings the merit payout amounts 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.

averages. It also offers an employee stock ownership plan and a stock option plan for executives.

3. Theory and Hypotheses

Within the institutional framework described above we test whether performance related pay enhances worker effort. The literature shows that beneficial effects on worker behavior are not at all granted and that they may vary depending on the mode of performance related remuneration.[11] Baron and Kreps (1999, p.261) list determinants of the effectiveness of performance related pay: (i) the saliency of a payment matters and surprise remunerations receive more employee attention than expected payments. (ii) Workers may reduce effort when they perceive inequities because differences in base salaries result from events in the past. (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 formula determining performance pay allows for substitution effects between individual and divisional performance. Additionally, the transparency of an evaluation system and its reliability from the workers' viewpoint affect its effectiveness.

	Surprise Bonus		Evaluation System	
Saliency	+	high	○	limited
Related to past performance	+	no	○	depends on application
Time & risk preference	+	payout immediately	+	payout immediately
Correlation to individual effort	+	depends on application	○	limited by effect of division result
Transparency	○	limited by subjective judgement	+	clearly structured system
Reliability	○	no claims to bonus	+	but depends on subjective judgement

Note: + indicates beneficial and ○ the lack of beneficial payment system characteristics.

In the above table the advantages and disadvantages of the two performance pay systems

[11] Kahn and Sherer (1990) found the manangers 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. However, these results are generated under quite restrictive assumptions.

in our firm are presented. Whereas surprise boni are more salient, less related to past performances and - in contrast to evaluation based pay - not systematically affected by division performance, the evaluation system seems more reliable as it is based on annual evaluations with clearly stated criteria. The different patterns of strengths and weaknesses do not generate clear predictions as to which of the two systems may generate stronger incentive effects. In our empirical work we evaluate instead the effectiveness of performance pay by testing three hypotheses: First, 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 (H1).

Since departments differ systematically in their utilization of the evaluation based system this generates heterogeneity in the extent to which employees' performance affects their remuneration. If performance pay enhances work 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 indicators:

The first measure describes the cross-sectional variance of performance ratings by 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 second hypothesis (H2).

The second measure covers the variability of person-specific evaluations 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 their rating next year as they closely match actual performance. We hypothesize that performance incentives are higher in the second scenario where every employee is evaluated anew every year as opposed to a situation where established judgements predominate over time (H3).

4. Empirical Approach and Data Description

Empirical Approach: To test our three behavioral hypotheses we proceed in three steps: In step 1 we investigate the correlation between worker effort and performance pay using descriptive statistics on average effort in departments with high and low performance incentives. Since effort is viewed as a response to incentives we relate effort in year t to incentive measures as of the previous year t-1.

In step 2 of the analysis 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 to departments (j) over several periods (t), we estimate the following baseline specification:

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

Here Y represents an indicator of worker i's effort in period t, I measures the performance pay incentives following our three hypotheses, and X provides background indicators describing worker and department. δ controls for department type fixed effects, ν reflects annual 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 overtime, which at times takes on negative values, and apply a tobit model for the always positive number of days of absence.

Step 3 of the analysis consists of robustness tests where we investigate whether the same results obtain for different subsamples, under differently defined dependent and independent variables and with alternative lag structures.

Dependent Variables: In order to test whether bonus payments, cross-sectional and dynamic variability of departmental performance evaluations affect worker behavior we need indicators of worker effort. We apply two proxy measures which are used frequently in the literature: The

first describes workers' days of absence due to health problems.[12] The second indicates how many hours of overtime an individual accumulated at the end of the year.[13] Swiss 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. The data documents that numerous workers accumulate more than 120 hours of overtime which in the end is a gift to the employer. Overtime within the 120 hours limit can be interpreted as 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 can be used as an indicator of motivation and effort.

Explanatory Variables: As our three key indicators of performance pay incentives we consider the department specific number of bonus payments per year and employee, the standard deviation of the department's performance ratings, 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. Since the average change in ratings between subsequent periods in a given department would equal zero we apply a different indicator, which is calculated in two steps: First we obtain for every person the difference in performance ratings for the years t-2 and t-1. Then we calculate the standard deviation of these person specific changes by department: The higher the average change in ratings the larger the standard deviation of this measure.

---

[12] For other studies using absenteeism as an indicator of effort see e.g. Treble 2001, or Ichino and Riphahn 2003, Engellandt and Riphahn 2004.

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

Our control variables (X) describe the individual worker and this workers' 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 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 without and with missing or extreme performance ratings (values 0 and 1 and beyond 200), which reflect specific 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 those 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.[14] 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

[14] To generate the lagged standard deviation in department rating changes which is a regressor for absences and overtime in period t, we need the difference between ratings in periods t-1 and t-2.

observed over the two periods of 2001 and 2002.[15]

5. Results

Step 1: The results of the first descriptive step of our analysis are presented in Table 2. It shows the average effort outcomes for individuals in departments where performance incentives are above or below department averages. We would expect less effort, i.e. more days of absence and fewer overtime hours in departments with few bonus payments, with small standard deviations of annual performance ratings, and with small intertemporal rating changes. However, not all of the tables' 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 3: 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 1: 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 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.

Step 2: 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 variance 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

[15] 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.

all specifications.

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 standard deviation of ratings in both Tables indicates a correlation between incentive and effort outcome that is contrary to hypothesis 2: 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.[16] The surprising positive correlation between intertemporal rating variability and absence days in Table 2 changes sign when controls for other performance pay mechanisms are considered: Holding other incentive mechanisms constant, departments with higher intertemporal rating variability now experience (insignificantly) lower absence rates.

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 effects is considered, in specification 3 we add four measures describing departmental characteristics, and in specification 4 a year dummy and department type effects are considered. However, the nature of the correlation between the incentive and the 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.

[16] The same holds when a linear estimator replaces the Tobit estimator.

Therefore the conclusions based on the first column appear to be robust across specifications: The data support the hypotheses regarding the effort enhancing effects of bonus payments (H1) and of the variability in individual ratings over time (H3). In contrast, the hypothesis that effort is higher in departments that evaluate workers on a wider scale is rejected.[17]

What may explain these surprisingly different incentive effects of rating dispersion and rating flexibility over time? The coefficient estimates could suffer from simultaneity bias if supervisors chose ratings and bonus generosity *in response to* observed effort e.g. as a discipline device, instead of as suggested so far as incentives for future behavior. In that case our results were misleading. Since our data lacks instruments we cannot perform a proper endogeneity test. 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 Granger causality tests which inform about intertemporal correlation patterns between effort and incentive measures. If supervisors set their ratings independent of overall department outcomes we should not find significant correlation patterns between lagged 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 as a measure of effort. 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 dispersions in

[17] These results also are robust to considering only one of the three incentive measures in the specification. Only the effect of the dispersion rating change in the absence equation turns insignificantly positive when the other two incentive indicators are not considered in the model. This corresponds to the bivariate correlation pattern observed in Table 2 discussed above.

current ratings. This supports the hypothesis of simultaneity bias for the rating dispersion outcome.

Table 4b presents the evidence from Granger causality tests. Using average values for the effort measures by department and year we apply linear regressions with random effect controls to the following model for each of the j = 1, 2, 3 incentive indicators (I_{jt}) separately for the two effort indicators absences and overtime (k=1,2)

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

Table 4b presents the estimates of γ_{jk} for each of the six models. We find again that (only) the 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 in the model and to the choice of the estimator.[18]

While the evidence is weak it 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. Suppressing surprise bonus payments in response to low effort similarly would not be effective.

Therefore we perform the robustness tests in the third step of our analysis only for the two indicators which appear to be exogenous to indicators of past efforts. 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

[18] 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 the 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.

expected direction, whereas as before the estimates in the absenteeism model are mostly statistically insignificant with five of eight having the expected sign.

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 worked by 11.4 percent. Adding a standard deviation to the average of the observed bonus frequency yields an increase in overtime work by 9.6 percent.[19]

Robustness tests confirmed the insignificant correlation of absenteeism with incentive measures throughout. Therefore we abstain from presenting 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). This may be due to the limited employment protection enjoyed by Swiss workers.

Step 3: In the last step of our analysis we investigate, whether the results obtained for the overtime indicator are robust to changes in sample and specification. We reestimated the model of column 4 in Table 5a 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 3 the main conclusions regarding the direction of the effects are

[19] The predicted difference in overtime between the 25th and 75th percentile of the distribution of incentive measures amounts to 13.4 percent and 11 percent for rating flexibility and bonus payments, respectively.

astonishingly robust to these different groupings of the sample.[20]

The only subgroups for which we find systematically different response patterns are managers vs. regular employees (see Table 6a). The 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 is an interesting result, it adds to the literature on managerial effort which focuses on top managers and the effects of 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 expect the reduced response intensity among managers.

As a second type of robustness check we modified the definition of our incentive measures. Fundamentally, we are testing whether past experience of department policies affect future behavior. As workers at times are reassigned to new departments we can specify our incentive measure 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 employees just joined. As a robustness test we therefore 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. With redefined incentive indicators the estimated coefficients does not change in sign or significance. Therefore measurement error 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 robustness 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

[20] Results are not provided to save space, but are available from the authors.

distributions. The last three columns in Table 6c shows the results of applying these measures to specification 4 of Table 5a. With one exception the results are highly robust to these specification changes. Only for the 90-10 differential do we obtain a different effect of rating flexibility which so far we cannot explain.

Our final set of robustness checks accounts for the potential omission of relevant variables at the department or individual level by considering fixed effects estimates. The coefficients obtained from fixed effect estimations of the full model are presented in Table 6c and convincingly confirm the evidence presented so far: More flexible assignment of ratings over time and a higher frequency of bonus payments are correlated with higher effort.

In addition to these results obtained at the level of individual effort outcomes we ran estimations to explain the average level of overtime (and absences) per department. The average effort outcomes were regressed on department incentive measures and on the available departmental characteristics. All results show positive coefficients of the incentive measures, however in this sample with 44 different departments they remain statistically insignificant.[21]

Overall, the analysis allows us to draw the following conclusions: Effort seems to be higher in departments paying more bonuses per employee and in departments where person-specific ratings are more flexible over time, supporting hypotheses 1 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 2.

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 a international company

[21] We performed these robustness checks for the absence measure as well which yielded generally beneficial incentive effects that, however, could not be estimated precisely.

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, the dispersion in department specific rating changes, and the frequency of surprise bonus payments in a department. We hypothesize positive effort effects of all three incentive instruments.

The results support the hypothesis that worker effort responds to bonus payments that are directly related to past performance. 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. Thus earnings mobility generates the expected behavioral response. Interestingly, the effectiveness of the two instruments appears to be of comparable magnitude. An increase in the incentive measure 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. In principle, such an effect may be due to the limited valuation that workers attach to the payoff of the evaluation system, to the complexity and low transparency of the payment scheme, or to the dissatisfaction with supervisors' performance ratings. Alternatively, however, the estimates of the effect of rating dispersion on subsequent worker effort may suffer from simultaneity bias. This could occur if supervisors respond in their rating behavior to the effort observed in the department. In order to gauge whether such an effect might exist in our data 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 Granger causality tests. The evidence on effort - incentive correlations suggests a pattern of reversed causality between effort and incentives for

the indicator of rating dispersion. Since such simultaneity bias does not appear for the other two incentive indicators we focused the analysis on these two measures.

We confirm our findings of positive incentive effects for these two measures in numerous robustness tests which investigated different subsamples, alternative definitions of the incentive indicators as well as different estimators. Even when controlling for department or individual fixed effects did the significant correlation between incentives and effort hold up. This suggests that the incentive-effort correlation works well for the incentive schemes considered here. Thus both types of performance pay applied in the company provide effective incentives for employee productivity independent of their weaknesses discussed above. The results support the importance of flexible individual performance evaluations, which open opportunities for upward mobility.

We confirm the prior literature with respect to the incentivating effect of bonus payments, while our findings on the importance of flexible opportunities for (upward) mobility are new to the literature. While we find clear support for the effectiveness of surprise bonus payments as applied in our firm this does not necessarily imply that a substantial increase in the frequency or size of bonus payments would continue to improve productivity effects to the same extent.

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 Ressource 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.

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( ), 481-507.

Murphy, Kevin J., 1999, Executive compensation, in: Ashenfelter O. and David Card (1999) *Handbook of Labor Economics* vol. 3, 2485-2563.

Prendergast, Candice, 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	7.360	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	0.075	0.067
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	0.210
Job level = 3 (0/1)	0.104	0.305
Job level = 4 (0/1)	0.123	0.329
Job level = 5 (0/1) *(reference group*)	0.167	0.373
Job level = 6 (0/1)	0.129	0.336
Job level = 7 (0/1)	0.070	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	0.240
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	6.350	1.640
Number of employees (in thousand)	0.264	0.219
Year Dummies		
Year = 2001 (0/1)	0.427	0.495
Year = 2002 (0/1) *(reference group)*	0.573	0.495
Department Type		
Research department (0/1) *(reference group)*	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. This implies that department characteristics do not present averages at the department level but 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	5.90
> 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 in one-sided tests, respectively.

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**	-1.056**	-1.294**	-1.351**
	(0.205)	(0.202)	(0.201)	(0.203)
Std. Deviation of Rating Change	0.685**	0.640**	0.466**	1.267**
	(0.157)	(0.155)	(0.154)	(0.179)
Bonus Payments	0.361**	0.121*	0.635**	0.412**
	(0.070)	(0.072)	(0.079)	(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**	0.210*	0.146	0.117
	(0.120)	(0.117)	(0.119)	(0.120)
Std. Deviation of Rating Changes	-0.031	-0.088	-0.131	-0.017
	(0.091)	(0.089)	(0.090)	(0.107)
Bonus Payments	-0.361**	-0.051	0.006	-0.020
	(0.043)	(0.044)	(0.048)	(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 Granger Causality

	Std. Dev. of Rating in t	Std. Dev. of Rating Change in t	Bonus Payments in t
	Coef. (abs. t-value)	Coef. (abs. t-value)	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 (0.104)	-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 effect, and the year fixed effect. ** and * indicate statistical significance at the 5 and 11 percent level.

Table 5a Baseline Results: Random Effects Linear Regression of Overtime Hours

	1	2	3	4
Indicators of Performance Pay Incentives in t-1:				
Std. Deviation of Rating Change	0.340**	0.299**	0.054	0.856**
	(0.143)	(0.141)	(0.141)	(0.169)
Bonus Payments	0.317**	0.067	0.568**	0.341**
	(0.070)	(0.072)	(0.078)	(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)	374.282**	261.932**	219.598**	200.813**

Table 5b Baseline Results: Random Effects Tobit Regression of Days of Absence

	1	2	3	4
Indicators of Performance Pay Incentives in t-1:				
Std. Deviation of Rating Changes	0.082	-0.023	-0.088	0.016
	(0.083)	(0.082)	(0.082)	(0.102)
Bonus Payments	-0.345**	-0.041	0.015	-0.012
	(0.043)	(0.043)	(0.047)	(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)	259.086**	195.86**	195.49**	195.50**

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 Managers and Regular Employees

	Employee	Manager
Indicators of Performance Pay Incentives in t-1:		
Std. Deviation of Rating Changes	1.114**	0.210
	(0.186)	(0.357)
Bonus Payments	0.404**	0.082
	(0.089)	(0.172)

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**	0.0269**	1.201**	-0.186**
	(0.193)	(0.005)	(0.221)	(0.061)
Bonus payments	0.373**	0.366**	0.361**	0.416**
	(0.852)	(0.080)	(0.080)	(0.080)

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

	Overtime	
	Department FE	Individual FE
Indicators of Performance Pay Incentives in t-1:		
Std. Deviation of Rating Changes	0.938 **	1.856**
	(0.263)	(0.297)
Bonus payments	0.542**	0.590**
	(0.102)	(0.128)

Note: The tables present estimated coefficients and standard errors in parentheses. 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 5a. The standard errors in the estimations with department fixed effects in Table 6d are robust at the individual level to account for heteroskedasticity.