Supervisors and Performance Management Systems

Work in Progress - Please request new version from authors before citing.

Anders Frederiksen, Lisa Kahn, and Fabian Lange

Aarhus University, Yale School of Management, and McGill University*

March 25, 2016

Abstract

Supervisors occupy central roles in production and performance monitoring in a firm. We study how supervisor heterogeneity in performance evaluations affects career and firm outcomes using data on a 360 degree performance system of a Scandinavian service sector firm. We find a large amount of heterogeneity in performance ratings associated with supervisors. We write down a principal-agent model where supervisor heterogeneity can come in the form of real differences in the ability to elicit output from subordinates or from differences in a taste for leniency when rating subordinates. Within the context of this model, we investigate the nature of supervisor heterogeneity and the degree to which firms are informed about this heterogeneity by relating supervisor heterogeneity in ratings to subordinate pay and pay for performance, objective performance measures, and supervisor pay. We find that worker pay and promotions are positively affected by a supervisor's propensity to rate highly, but that objective output is not related to the supervisors rating behavior. This suggests that supervisor heterogeneity is primarily driven by differences in leniency, about which the firm is relatively uninformed. Our research is the first to document the important variation in ratings across supervisors and to show that it is positively related to worker outcomes.

*Anders Frederiksen, Aarhus University, Department of Business Development and Technology, Birk Center Park 15, 7400 Herning, Denmark, Lisa Kahn Yale School of Management, PO Box 208200, New Haven, CT, 06520, Fabian Lange, McGill University, Department of Economics, Leacock Building, 855 Sherbrooke Street West, Montreal Quebec H3A2T7, Canada (fabolange@gmail.com)

1 Introduction

Modern cooperations increasingly rely on performance management systems that call upon supervisors to rate the performance of their subordinates. Supervisor ratings are used to compensate employees, to allocate workers to tasks, and to determine who gets promoted. Supervisors are however not passive instruments that provide unbiased measures of employee performance. Instead, as we show in this paper, idiosyncratic differences in rating behavior across supervisors are prevalent in modern performance management systems. The sources of these idiosyncratic differences are not well understood, but they are likely to have important consequences. From the perspective of an employee, having a biased supervisor is likely to affect earnings, career outcomes, and work satisfaction. From the perspective of the firm, idiosyncracies of ratings behavior might constrain performance management systems. In this paper, we explore the degree of heterogeneity in supervisor ratings behavior and how important this heterogeneity is in shaping employee's careers. Furthermore, we strive to understand the nature of the heterogeneity in supervisor rating behavior that we document in the data. Do supervisors differ in their ratings behavior because some are simply more lenient and generous than others? Alternatively, do the differences in ratings reflect differences in management style that have real impacts on output? Finally, how informed is the firm about these heterogeneities and how does the firm respond to their presence?

In this paper, we develop an analytic framework that allows investigating the nature of heterogeneity in ratings behavior across supervisors and the extent to which firms are informed about this heterogeneity. Supervisors may rate differently because of differences in managerial ability, reflected in output, or because they differ in how lenient they are. [1]We show how to operationalize this framework using data from the performance system of a Scandinavian service sector firm. We find that supervisor heterogeneity is correlated with subordinate outcomes: workers assigned to higher raters are subsequently paid more, are more likely to be promoted, and have more stable jobs at the firm. However, we also find that supervisor heterogeneity in ratings is uncorrelated with objective productivity measures at the team level. Our results thus suggest that supervisors differ from each other in their tendency to report positively as opposed to accurately on their workers performance and that firms find it difficult to undo these biases when setting pay and determining promotions.

Previous research has shown the importance of managerial heterogeneity in production (Bertrand and Schoar 2003, Lazear, Shaw, and Stanton 2015). We contribute to this literature by exploring the central role managers play in performance evaluation systems. These systems shape workers' careers in firms, as

[1]Guilford (1954) introduced leniency bias to describe stable differences across raters in how they rate others unrelated to productive differences among ratees.

employers use evaluations to set pay and to learn about worker quality. [2] We document that raters differ significantly in how lenient they are when rating their subordinates and we show that this heterogeneity matters for outcomes employees care about such as earnings, promotions, and lay-offs.[3] We also show that objective performance at the team-level does not vary with the rating style of supervisors. Supervisors thus affect the outcomes of workers at the firm in ways that are not related to either the innate productivity of employees or the objective outcomes of the teams supervised by the raters.

To interpret the data, we develop an simple behavioral model of behavior of supervisors, subordinates, and the firm. In particular, we follow a long tradition in personnel economics and postulate that the central human resource challenge facing the firm is to incentivize workers to exert effort (Holmstrom 1979, Holmstrom and Milgrom 1987, Lazear 2000). The three actors in our model are the workers without supervisory function, the supervisors, and the firm. Workers choose to exert effort, but the chosen effort level is hidden from the firm and the supervisors. Supervisors observe worker output and report on this output to the firm. However, supervisors do not report truthfully on worker performance - rather they face a trade-off between reporting truthfully and reporting favorably about their team members. In our model, supervisors differ along two dimensions. First, they differ in how much weight they place on reporting truthfully as opposed to favorably. Second, they differ in their managerial ability, which we model as differences in the marginal costs of exerting effort on the part of their subordinates. Given this set-up, we consider the optimal linear compensation contracts of workers as well as salary contracts for supervisors. Our model is set up in a way that allows us to ask how the optimal contracts would depend on how informed firms are about the differences between supervisors.

Our dataset is especially rich, relative to other personnel datasets such as that used in the canonical work by Baker, Gibbs, and Holmstrom (1994a, 1994b). We test the model using a long panel (2004-2014) with data on worker outcomes, the identity of supervisors that rate and manage them, and information on team outcomes. The first main piece of evidence is that "high raters" (i.e., supervisors who give high ratings to their employees, on average) are associated with better outcomes for subordiantes in terms of pay, bonuses, promotions and job stability. The second piece of evidence is that branches managed by "high raters" do not achieve higher scores on Key Performance Indicators (KPI) describing branch performance. If high raters also tended to manage more effectively, then the branches they manage would outperform those of other managers. Our data do not support this. Furthermore, the compensation and career outcomes of

[2] On the importance of employer learning in the evolution of workers' careers see Gibbons and Katz (1996), Alontji and Pierret (2001), Lange (2007), Kahn and Lange (2014).

[3] Kane et al (1995) present evidence for leniency bias as defined by Guilford among small samples of police officers, nurses, and social workers. Our work goes beyond Kane et al. (1995) in documenting the role of rater heterogeneity for employee outcomes and careers as well as the performance system of a large firm. And, we are able to relate the rating heterogeneity to outcomes directly of interest to the firm, enabling us to distinguish between leniency bias and potential heterogeneity in managerial ability across supervisors.

supervisors are only weakly associated with their proclivity to rate their subordinates highly. If managers differed primarily in their managerial effectiveness and firms were informed about these differences, then we would expect supervisors compensation to correlate highly with their rating behavior. The data however shows that career outcomes of supervisors and their compensation are unrelated to their tendency to rate their subordinates highly. We thus believe that our data broadly supports an interpretation according to which (a) the heterogeneity in supervisor rating behavior is largely associated with leniency as opposed to managerial effectiveness and (b) the firm is uninformed about the differences in reporting behavior across supervisors.[4]The remainder of the paper proceeds as follows. We introduce the firm and the data at our disposal in the following Section (Section 2). Section 3 then describes the model and its implications for how career outcomes and performance are related to supervisor heterogeneity. Section 4 presents the empirical analysis. Section 5 contains a discussion and reduced form investigation of a number of hypothesis related to the dynamics of performance measurement. In particular we ask whether there is evidence that the firm is learning about the heterogeneity of their supervisors. Section 6 concludes.

2 Firm and Data

2.1 Firm Overview

We rely on personnel data from a large Scandinavian service sector firm.[5] Our sample comprises all employees engaged in domestic activities between 2004 and 2014. In our data we have 22,688 unique employees with a total of 136,286 employee-year observations. Table 1 provides summary statistics for the full sample (all reported monetary figures are in 2010 US dollars). On average, workers earn \$78,139 ~~US dollars~~ with a standard deviation of \$51,487. In the data it is possible to distinguish between base pay and annual bonus. Roughly 30 percent of the workers receive a bonus and the bonus pool is close to 20 percent of the wage pool. In our sample, 51.9 percent are women, the average age is 43.5 years and tenure is 17.5 years. We also observe that 83.6 percent are working full time.

The firm is divided into an extensive branch network and a central corporate office (see Figure 1). The branches comprise 44 percent of workers. Across branches jobs are comparable and involve close client contact. Workers in the central corporate office have a variety of functions and there are more high level jobs (level 11). These differences are reflected in the compensation structures across branches and corporate functions in that both average compensation and the variance in compensation is higher in the corporate

[4]An informed firm would strive to eliminate differences in outcomes across workers that are associated with supervisors as they do not reflect differences in individual skill or effort. Of course it is also possible that the firm ignores these biases for reasons outside of our model.

[5]The firm is a market leader within the domestic market. It also has some international activities, but we focus on the domestic workforce here.

functions. In 2013, there were 269 branches and the median branch had 15 employees. The typical branch had a branch manager (level 9), a deputy branch manager (level 7), 5-7 senior workers in client-facing roles (levels 6), and 5-7 junior workers in client-facing roles (levels 4-5) and some times a trainee (level 1). Because of the qualitative differences between branch and corporate jobs in this firm and because of the availablity of objective (financial and performance based) branch-level performance measures we will in part of our analysis pay particular attention to the branch network.

Just prior to the period covered by our data, the firm developed a performance management system. Each worker receives a rating that is meant to describe their aggregate performance. It ranges from 1 (unsatisfactory) to 5 (outstanding). In 2004, when our data begins, the system was still being rolled out, and 42 percent of the employees received performance ratings. In the following years, the system continued to spread so that by 2008 the system covered almost 82 percent of the employees and the coverage stayed at that level or slightly above throughout the remainder of the sample period.[6] In the branch network, ratings are typically given by the branch manager, but we also observe that deputy branch managers rate employees. In corporate functions, employees are typically rated by the worker with the highest job level within a given function. Overall the typical manager is rating 10 employees as the average span of control in the firm is 9.76 (s.d. 10.16).

The distribution of performance scores is shown in table 2. The lowest rating of 1 is rarely given and only 3 percent receive the second lowest rating of 2. The clear majority receive the ratings 3 and 4, with more than 50 percent of all employees receiving a 3. Only 5.7 percent of employees are rewarded the highest rating of 5. This range of ratings, as well as the effective range (of 3 to 5) is common among subjective performance systems, as shown in Frederiksen, Lange, and Kriegel (2013). Because most ratings are either a 3 or a 4, we loose little information by using a "pass-fail" performance metric, which equals 1 if the rating is 4 or 5 and zero otherwise. The "pass-fail" performance metric makes it easy to interpret linear regression coefficients in that they represent marginal effects on the probability of receiving a "passing grade". For these reasons, we will build our empirical investigation around this pass-fail metric.[7]

Our data also contains two measures of branch performance. The first measure of performance reports how a branch ranks on a set of Key Performance Indicators (KPIs) within group of peer branches. The Key Performance Indicators (KPIs) include measures of financial performance of the branches, as well as other metrics (for example, customer satisfaction). The set of KPI changes from year to year as the firm's focus evolves. Branches are placed into peer groups based on size and customer base, and these peer groups vary from year-to-year. The average peer group has 17 branches. These branch rankings, which we hereafter

[6] There is no systematic variation in who gets rated when we look at full-time vs. part-time employees, corporate vs. branch employees or across job levels. Hence, we are not worried about any systematic reasons for missing ratings.

[7] Results using the entire scale are available upon request and qualitatively and quantitatively very similar.

term "KPI rankings", are available from 2007-2010. The second measure of branch level performance reflects the branches' financial development between January in year t and t+1. We have succesfully obtained this information for the year 2013 (i.e. the development in performance between jan. 2013 and jan. 2014). This measure is constructed such that at score of 100 implies no change in financial performance between the two years. A score of 110 implies a 10 percent improvement. Among the 160 branches for which we have financial performance information the average score is 102.6.

In addition to supervisor ratings we have obtained access to employee job satisfaction surveys for the years 2004 to 2010. These surveys contain information about the employees' perceptions of supervisors' performances. Employeers are asked 7 questions: 1) The professional skills of my immediate superior, 2) The leadership skills of my immediate superior, 3) My immediate superior is energetic and effective, 4) My immediate superior gives constructive feedback on my work, 5) My immediate superior delegates responsibility and authority so I can complete my work effectively, 6) My immediate superior helps me to develop personally and professionally, and 7) What my immediate superior says is consistent with what he/she does. These questions are answered on a 10-point scale and we use the average across the seven questions as the employees overall assessment of his/hers immediate superior. The minimum score is 1 (low) and the maximum score is 10 (high). On average employees rate their supervisors at 8.164 with a standard deviation of 1.373.

It is unusual to have employee satisfaction data merged with the personnel files. Supplements to surveys such as the National Longitudinal Survey of Youth (NLSY), the German Socio-Economic Panel (GSOP), and the British Household Panel Survey (BHPS) sometimes do contain employee satisfaction data, but such data is typically unavailable to supervisors in companies. Firms usually contract with outside consulting companies to conduct employee satisfaction surveys. These consulting firms then report back to the firm averages at the branch/business unit level. The reason why firms collect the data at arms-length is to minimize bias because employees might be reluctant to truthfully report concerns unless the data is anonymized. As researchers we have been able to obtain the survey data at the individual level and to merge it onto the personnel records. Hence, we know how a given employee evaluates his/hers superior, even though the firm and the supervisor did not have access to this information at the individual level.

2.2 Variation in Performance Measures

There is substantial systematic variation in the incidence of passing grades across supervisors and workers. To illustrate this variation, we specify the event that individual, i at time t "passed" her performance review ($p_{it} = 1$). We relate this event to an individual effect α_i, a supervisor effect ϕ_s, as well as time-varying worker controls (X_{it}), a vector of time-varying supervisor controls (Y_{st}):[8]

$$p_{it} = \alpha_i + \phi_s + \beta' X_{it} + \gamma' Y_{st} + \epsilon_{it} \quad (1)$$

In equation (1), the unobservable is denoted ϵ_{it}. We describe below in more detail how we estimate the variation in α_{it} and ϕ_{st}, but note for the time being simply that these estimates account for sampling variation following Card, Heining, and Kline (2013).

Estimating this regression requires the assignment of employees to supervisors to vary substantially over time. In our data, employees typically change supervisors repeatedly. Similarly, supervisors manage many different employees over time, with some employees joining or leaving their teams almost every year. Over the period 2004 to 2014 the average employee had 2.94 different supervisors (s.d. of 1.71). If we look at the employees who were with the firm throughout this period we find that they on average had 4.31 different supervisors (s.d of 1.59). The average supervisor manages 9.76 (s.d. of 10.16) employees in a given year. On average, supervisors manage a total of 21.48 different employees (s.d. of 37.28) while they are recorded as supervisors in our data. Those individuals who were supervisors throughout the entire sample period on average managed a total of 50.18 different employees.[9]

The variation in the probability of passing across supervisors in the data is substantial. We find that a one-standard deviation difference in ϕ_s amounts to a 14.18% increased probability of receiving a passing grade. Assuming that ϕ_s is normally distributed, we can also illustrate the variation in supervisor effects by noting the change in the probability of receiving a passing grade associated with moving from the 10th to the 90th percentile in the distribution of ϕ_s. Such a change, all else equal, is associated with a 36% increase in the probability of receiving a passing grade. The variation in the supervisor effects that we observe in the data is thus substantial.[10]

In the next section, we develop a model that offers two possible explanations for the systematic variation in performance evaluations across supervisors and empirical predictions to separate the two.

[8] The worker controls (X_{it}) include cubics in age and tenure, and indicators for full-time status, gender, job level; supervisor controls (Y_{st}) include a cubic in age of the supervisor and gender and job level of the supervisor dummies. We also control for business unit indicators (whether the worker is in a branch or the specific function in headquarters), and year fixed effects. The latter help control for differences in usage of performance ratings as they become more common in the firm.

[9] The interconnectedness in the firm is in fact so large that the largest connected set covers the entire firm.

[10] The variation in α_i is even larger. We find that a standard deviation in α_i amounts to a 26.5% increase in the passing probability. Moving from the 10th to the 90th percentile of the distribution of α_i amounts to an increase in the passing probability of 67%.

3 Model[11]

Supervisors play a crucial functional role in performance management systems. Besides managing and supervising teams, they report on the performance of their subordinates. Firms rely on these reports to set pay and to determine promotions. Naturally, supervisors may differ in both their rating behavior and in their ability to manage employees. Firms then face the problem of how to design performance management systems in the face of this heterogeneity. We analyze how heterogeneity in managerial ability and supervisor rating behavior affects data generated by performance systems under different assumptions on how well informed firms are about the heterogeneity across supervisors.

Let the marginal product of an employee not in a supervisory role (a "worker") be $q_{i,t}$. As expressed in equation 2, we assume that this marginal product ("output") depends on effort ($e_{i,t}$) which is not directly observed by her supervisor or by the firm. Worker productivity also depends on the productive type α_i, and random time-varying luck $\varepsilon^q_{i,t}$, distributed normally with mean 0, and variance σ^2_q. This time-varying components is independent of $(e_{i,t}, \alpha_i)$.

$$q_{i,t} = e_{i,t} + \alpha_i + \varepsilon^q_{i,t} \tag{2}$$

Firms do not directly observe $q_{i,t}$ but supervisors do.[12] Having observed q, supervisors choose to report r to the firm. Supervisors differ along two dimensions: (a) heterogeneity in managerial ability, which impacts the worker's cost of effort (μ_s), and (b) heterogeneity in rating behavior, whereby supervisors differ in their willingness to tradeoff a truthful rating with a more generous one (β_s).

The timing of the model is as follows:

1. Workers and firms sign contracts that specify the type of supervisors individuals are matched with and the wage function. This wage function specifies how compensation depends on the supervisor characteristics and the rating that a worker receives.

2. Workers are matched to supervisors, exert effort e and produce q.

3. Supervisors observe q and provide subjective evaluations r.

4. Based on this rating, workers are paid according to the wage function that formed part of the contract.

[11] As we lay-out the model, we will focus on its implications and the intuitions embodied in it without presenting derivations in detail. Many results follow immediately from known results in the literature (see for example Holmstrom (1979) and need not be rederived here. A somewhat more formal treatment of the arguments is provided in the appendix.

[12] During the remainder of this paper, we suppress individual and time subscripts unless required for understanding. However, we do subscript variables that vary across supervisors with s.

Workers have Constant Absolute Risk Aversion (CARA) preferences over effort e and wages w:[13]

$$v(w,e) = -exp\left(-\psi\left(w - \frac{1}{2\mu_s}e^2\right)\right) \tag{3}$$

These preferences depend on the managerial ability of a worker's supervisor in that good managers reduce the marginal cost of effort. We parametrize this idea using μ_s so that better supervisors have higher μ_s. Workers choose effort to maximize eq. (3) taking as given the wage contract and μ_s. All else equal, workers for better supervisors will exert more effort.

We also allow for heterogeneity in reporting behavior. Some supervisors are more lenient than others, but all prefer to truthfully report on the performance of employees. This implies supervisors trade off the conflicting goals of being lenient and reporting accurately on their employee's productivity. We embed this trade-off in supervisor preferences:

$$u(w_s, q, r) = w_s + \tilde{\beta}_s r - \frac{\tilde{\gamma}_s}{2}(r-q)^2 \tag{4}$$

Here the parameters $\left(\tilde{\beta}_s, \tilde{\gamma}_s\right)$ allow for heterogeneity across supervisors in how they trade off leniency against accuracy. Supervisors choose a report r trading off the desire to be nice against the distaste for deviating from true worker output, q. The result, shown in equation (5), is that supervisor reports r sum the observed output q and $\beta_s = \frac{\tilde{\beta}_s}{\tilde{\gamma}_s}$. The supervisor specific parameter β_s measures the strength of the motive to report favorably relative to the motive to report truthfully. We will call this parameter the "supervisor bias".

$$r = q + \frac{\tilde{\beta}_s}{\tilde{\gamma}_s} = q + \beta_s. \tag{5}$$

Substituting (2) in (5) and denoting by e_s the equilibrium effort level that team members of the supervisor s exert, we get:

$$r = \alpha_i + (e_s + \beta_s) + \varepsilon_{it}^q = \alpha_i + \phi_s + \varepsilon_{it}^q \tag{6}$$

Variation in ratings attributable to the supervisor is summarized byϕ_s. As discussed above, this variation can arise either because supervisors differ in their managerial quality μ_s or because they differ in their bias β_s.

[13]The functional form assumptions embodied in equation (3) keep the problem tractable. By assuming CARA, we abstract from income effects that might otherwise affect the trade-off between effort and risk. Quadratic effort costs result in linear first order conditions for effort and thus result in closed form solutions. Below, we make assumptions that ensure that wages are normally distributed conditional on worker choices and information. Combined with the exponential form in (3) this allows exploiting known results on expectations of log-normally distributed random variables (deGroot reference).

We now consider the contracts firms and workers enter. We limit ourselves to a static set-up, but we consider different assumptions on how informed the firm is about the differences between supervisors. We do require that during the contracting stage information is common among the agents (workers, supervisors, and the firm). However, we consider the possibility that information on the supervisor types $\{\mu_s, \beta_s\}$ is imperfect. The contracts workers, supervisors, and firms enter into specify all aspects of the employment relationship that are relevant for pay-offs. For workers this implies that they consist of assignments $\{\mu_s, \beta_s\}$ as well as a mapping of observed ratings to wages.

We make a number of assumptions to keep the analysis tractable. First, as is common in the literature, we restrict attention to wage contracts that are linear in the ratings. The parameters of these wage contracts are allowed to vary with each individual and supervisor assignment. Thus, we consider contracts of the form $w_{i,s,t} = a_{i,s} + b_{i,s} r_{i,t}$.[14] In addition, we make the necessary assumptions that ensure that the exponent in equation (3) is normally distributed conditional on the available information *both* at the contracting stage and when individuals decide on effort.[15] This allows us to use well-known results on the expectation of log normal random variables (deGroot (1970)) to represent worker preferences using the certainty equivalent. That is, we can express the participation constraints as

$$E[w - \frac{1}{2\mu_s} e^{*2} | I_C] - \frac{1}{2} \psi var \left(w - \frac{1}{2\mu_s} e^{*2} | I_C \right) \geq \underline{u(\alpha)} \qquad (7)$$

where I_C represents the information available during the contracting stage and e^* is the optimal effort level chosen by the worker.[16] Workers observe μ_s when choosing this effort and face a linear wage. Maximizing equation (3) subject to the linear contract delivers the optimal effort choice e^*:

$$e^* = b_{s,i} \mu_s \qquad (8)$$

We now solve for the optimal terms $(a_{i,s}, b_{i,s})$ of the wage contract. The solution depends on what firms and workers know about supervisors.

[14] In a closely related setting with normal signals and with preferences of the type provided in Holmstrom and Milgrom (1987) find that the optimal contract does take the linear form. We suspect but have not proven that our setting could be specialized further to map into **?** and that linear contracts are therefore at least conceivably optimal. For now, I think that exercise is besides the point.

[15] The exact assumptions required vary with assumptions maintained about the information avvailable to agents in the model. We generally need that output noise is normally distributed. If (μ_S, β_s) are known this will suffice. If (μ_S, β_s) are *partially* unknown, then the heterogeneity in supervisor types (μ_S, β_s) conditional on expectations of agents needs to be normally distributed.

[16] The outside opportunity $\underline{u}(\alpha)$ depends on the productive type of the worker, since firms compete for workers and are symmetrically informed about the type of workers.

3.1 The Informed Firm and the Performance Management System

We begin by assuming that firms and workers are perfectly informed about the supervisors and workers types : μ_s, β_s and α_i. Firms offer workers both an assignment to a supervisor with characteristics (μ_s, β_s) and a wage contract that maps observed signals r onto wages. The terms of the wage contract are allowed to vary with $I_C = \{\mu_s, \beta_s, \alpha_i\}$.[17] Thus, wage

contracts are:

$$w = a(\mu_s, \beta_s, \alpha) + b(\mu_s, \beta_s, \alpha) r$$

Substituting the optimal effort e^* from eq. (8) into the certainty equivalent (7) and simplifying, we obtain the participation constraint:

$$a + b(\alpha + \beta_s) + \frac{1}{2}b^2\mu_s - \frac{\psi}{2}b^2\sigma_q^2 \geq \underline{u}(\alpha) \tag{9}$$

It is straightforward to show that the piece-rate b maximizes the sum of the expected profit and the certainty equivalent subject to workers choosing effort optimally (eq.8).[18] Thus, the optimal piece rate solves

$$b_s^* = \underset{\{b\}}{argmax}\left\{\alpha + b\mu_s - \frac{b^2}{2}\left(\mu_s + \psi\sigma_q^2\right)\right\} \tag{10}$$

This results in the standard solution familiar from the literature:

$$b_s^* = \frac{\mu_s}{\mu_s + \psi\sigma_q^2} \tag{11}$$

Because firms compete for workers, they will make zero profits for any worker-supervisor pair. This zero profit condition amounts to

$$\alpha + b\mu_s - a - b(\alpha + \beta_s + b\mu_s) - w_s(\mu_s, \beta_s) = 0 \tag{12}$$

where $w_s(\mu_s, \beta_s)$ is the wage paid to a supervisor with characteristics (μ_s, β_s).

Consider now how the compensation of employees and supervisor varies with (α, β_s, μ_s).

We begin with β_s. From eq (11), we have that the optimal piece-rate does not depend on the generosity of the supervisor β_s. Consequently, the effort choice e^* does not vary with β_s either. Rearranging the certainty

[17] In Section 3.2, we consider firms that are imperfectly informed about supervisor heterogeneities (μ_s, β_s).

[18] For this, set up the profit maximization of the firm subject to the Participation constraint. The first order condition with respect to the intercept can be used to show that the Lagrange multiplier on the participation constraint equals 1, from which the statement in the text follows.

equivalent in eq (7) to isolate expected compensation we have $E[w|I_C] = \underline{u(\alpha)} + \frac{1}{2\mu_s}e^{*2} + \frac{1}{2}\psi var(w|I_C)$. All terms on the right hand side are independent of β_s, implying that expected compensation of employees will not vary with β_s. The reason is that the firm extracts the entire surplus using base compensation $a(\mu_s, \beta_s, \alpha_i)$ - workers with more generous supervisor will simply see their base compensation reduced.[19] Since effort and employee compensation do not differ with β_s, neither does the revenue net of the employees compensation that a firm can make from a given assignment of an employee to a supervisor β_s. Thus, supervisor compensation will not vary with β_s either.

Continuing with α_i, we note that competition for workers ensures that the expected compensation of workers increases one-for-one with α_i. It is obvious that supervisor compensation will not vary with α_i.

Consider now μ_s. From equation (11), we have that the optimal loading increases in μ_s. To determine the effect on average compensation, consider the certainty equivalent after substituting the expected wage of an employee:

$$E[w|\alpha, \mu_s, \beta_s] - \frac{1}{2\mu_s}e^2 - \frac{\psi}{2}b^2\sigma_q^2$$

Since the entire surplus is extracted from workers we obtain

$$\frac{d\left(E[w|\alpha, \mu_s, \beta_s] - \frac{1}{2\mu_s}e^2 - \frac{\psi}{2}b^2\sigma_q^2\right)}{d\mu_s} = 0$$

Workers maximize the certainty equivalent by choice of e. We can thus apply the envelope condition and ignore any variation in effort in response to variation in μ_s. However, as μ_s varies, so will the piece-rate b (see eq. 11).[20] Thus, we obtain

$$\begin{aligned}
\frac{d(E[w|\alpha, \mu_s, \beta_s])}{d\mu_s} &= \frac{\partial\left(\frac{1}{2\mu_s}e^2\right)}{\partial\mu_s} + \frac{\partial\left(\frac{\psi}{2}b^2\sigma_q^2\right)}{\partial b}\frac{\partial b}{\partial\mu_s} &&= -\frac{1}{2\mu_s^2}e^2 + \psi\sigma_q^2 b\frac{\partial b}{\partial\mu_s} \\
&= -\frac{1}{2}b^2 + b\left(\frac{\psi\sigma_q^2}{\mu_s+\psi\sigma_q^2}\right)^2 &&= -\frac{1}{2}b^2 + b(1-b)^2 \\
\Rightarrow sign\left(\frac{d(E[w|\alpha, \mu_s, \beta_s])}{d\mu_s}\right) &= &&sign\left(-\frac{1}{2}b^2 + b(1-b)^2\right)
\end{aligned}$$

This expression cannot generally be signed. When incentives are low-powered ($b < \frac{1}{2}$), total pay increases in μ_s, while the opposite is true when incentives are high-powered ($b > \frac{1}{2}$). The costs of providing any given effort level declines with μ_s which tends to lower compensation. At the same time, when μ_s increases,

[19]Note also that with informed firms, workers will also not receive any non-pecuniary benefits from working for more lenient supervisors. This is because the firm would extract any non-pecuniary benefits using the intercept of expected compensation. This provides an additional approach to testing for how informed the firm is about heterogeneity across supervisor by examining whether voluntary mobility (within the firm or quits) of employees varies with the supervisor effects.

[20]The piece rate is not chosen to maximize the certainty equivalent, so no envelope condition applies here.

the optimal piece rate increases as well and so does the risk borne by workers. This will tend to increase compensation since workers need to be induced to bear the increase in risk. When incentives are high ($b > \frac{1}{2}$), much effort is provided. Thus, better managers reduce the efforts costs born by workers significantly when incentives are high. Therefore wages for workers with better managers decline if incentives are high. When incentives are low, effort provision is low and little is gained in terms of reducing effort costs by working for a better manager. Thus, pay increases with μ_s when incentives are low ($b < \frac{1}{2}$) because workers need to be compensated for the extra risk they bear.

Regarding the compensation of the supervisor, note that the surplus generated by any supervisor-worker match increases in μ_s. As firms compete for supervisors, any differences in the surplus across μ_s are paid to the supervisor. Thus the compensation of the supervisor increases in her managerial ability: $\frac{\partial w_s(\mu_s)}{\partial \mu_s} > 0$.

We have so far considered the problem of how wages depend on supervisor and worker heterogeneity without considering the problem of assigning workers to supervisors. Since worker type α enters additively in the production function and does not affect the risk-effort trade-off as summarized by worker preferences (3), we have no predictions for how α and (μ_s, β_s) are assigned to each other. Both positive and negative assortative matching are entirely consistent with this set-up.

To summarize, when we maintain the assumption that (μ_s, β_s) is known, then we have the following predictions of how wage contracts and output relate to (α, β_s, μ_s) :

1. The optimal piece-rate $b(\mu_s, \beta_s, \alpha)$ is independent of (α, β_s) and increases in μ_s.
2. The average compensation received by employees increases one-for-one in α and is independent of supervisor generosity β_s. It is not possible to sign the relation between average compensation of employees and μ_s.
3. Expected output $E[q|\mu_s, \beta_s, \alpha]$ increases in μ_s and α and is independent of β_s
4. Earnings of supervisors $w_s(\mu_s, \beta_s)$ are independent of β_s and increase in μ_s

We will now consider the case when the firm is only partially informed about the heterogeneity in (μ_s, β_s).

3.2 The Partially Informed Firm and the Performance Management System

So far we assumed that (μ_s, β_s) are known to the firm. Next, we analyze contracts when firms and workers are only partially informed about supervisor types. We continue to assume that the only information asymmetry in the model is about the hidden effort e. Thus, we assume that all agents in the economy share the same information about (μ_s, β_s) during the contracting stage. We proceed in much the same fashion as when analyzing the problem faced by the informed firm.

To begin, assume that (μ_s, β_s) are independent normally distributed random variables with variances σ_β^2 and σ_μ^2. Firms and employees hold beliefs $\left(\beta_s^E, \mu_s^E\right)$ about the supervisor characteristics such that

$$\begin{aligned} \beta_s &= \beta_s^E + \varepsilon_\beta \\ \mu_s &= \mu_s^E + \varepsilon_\mu \end{aligned}$$

Let the errors $(\varepsilon_\beta, \varepsilon_\mu)$ also follow a normal distribution and be independent of each other. We parametrize the share of total variation in β and μ unknown to firms as θ_β and θ_μ so that

$$\begin{aligned} \sigma_\beta^2 &= var\left(\beta_s^E\right) + var\left(\varepsilon_\beta\right) &= \left(1-\theta_\beta\right)\sigma_\beta^2 + \theta_\beta \sigma_\beta^2 \\ \sigma_\mu^2 &= var\left(\mu_s^E\right) + var\left(\varepsilon_\mu\right) &= \left(1-\theta_\mu\right)\sigma_\mu^2 + \theta_\mu \sigma_\mu^2 \end{aligned}$$

During the contracting stage, the marginal cost of effort is not known to anybody. However, employees observe the marginal cost of effort after having been assigned to a supervisor and before they decide on their optimal effort level. The optimal level of effort conditional on the piece rate b is obtained in the same manner as before (see eq. 8): $e^* = b\mu_s$.

During the contracting stage, the parties share information on $\left(\mu_s^E, \beta_s^E\right)$. A work contract consists of an assignment of a worker α_i to a supervisor with $\left(\mu_s^E, \beta_s^E\right)$ and a wage contract that depends on $\left(\mu_s^E, \beta_s^E, \alpha\right)$: $w\left(r; \mu_s^E, \beta_s^E, \alpha\right) = a(\mu_s^E, \beta_s^E, \alpha) + b(\mu_s^E, \beta_s^E, \alpha)r$.

As before, we can use the employee's certainty equivalent to write the participation constraint:

$$a + b\left(\alpha_i + \beta_s^E\right) + b^2 \frac{\mu_s^E}{2} - \frac{\psi}{2}\left(b^2\left(\theta_\beta \sigma_\beta^2 + \sigma_q^2\right) + \frac{b^4}{4}\theta_\mu \sigma_\mu^2\right) \geq \underline{u}\left(\alpha\right) \tag{13}$$

The firm problem is still to maximize profits from any given worker-supervisor pair:[21]

$$\Pi\left(\mu_s^E, \beta_s^E, \alpha\right) = \underset{\{a,b\}}{Max}\left\{\alpha + b\mu_s^E - a - b\left(\alpha + \beta_s^E + b\mu_s^E\right) - w_s\left(\beta_s^E, \mu_s^E\right)\right\} \tag{14}$$

s.t. the participation constraint (13).

And, as before, competition in the labor market for workers and supervisors will ensure that expected profits conditional on $\left(\alpha, \beta_s^E, \mu_s^E\right)$ will equal zero.

Wage contracts between partially informed firms and employees

[21] We have already imposed the optimal effort choice $e = b\mu_s^E$.

The optimal loading is implicitly determined by the FOC of eq. 14:

$$\mu_s^E = b\left(\mu_s^E + \psi\left(\theta_\beta \sigma_\beta^2 + \sigma_q^2 + b^2 \frac{\theta_\mu \sigma_\mu^2}{2}\right)\right) \tag{15}$$

The RHS of this expression increases monotonically in b and there is thus a unique loading that solves the firms problem.

It is instructive to compare (15) with the optimal loading of the informed firm: $b = \frac{\mu_s}{\mu_s + \psi \sigma_q^2}$ stated in eq. (11). Besides replacing μ_s^E with μ_s, there are two differences. First, the signal becomes less informative as the share of the variation in β_s that is unknown to the firm increases. Thus, the optimal loading declines in $\theta_\beta \sigma_\beta^2$. Second, $\theta_\mu \sigma_\mu^2$ measures differences in managerial ability that are unobserved by both workers and the employer during the contracting stage. However, once the worker has been assigned to a supervisor she observes the marginal cost of effort μ_s associated with this supervisors. At that point, she will exploit this additional information and will "game" the performance system in the sense of supplying disproportionally more effort in low marginal cost stages than in high marginal cost states of the world. Therefore, the usefulness of setting incentives using performance signals declines in $\theta_\mu \sigma_\mu^2$ and so does the optimal loading.

As before, firms extract any surplus from workers during the contracting stage. Again, expected compensation will be independent of β_s^E since β_s^E only enters the workers certainty equivalent through the expected wage. And, as before, we have that competition for employees implies that productive differences across employees are paid to workers so that we have

$$a = a_0\left(\mu_s^E\right) + \alpha - b\beta_s^E \tag{16}$$

It is again not possible to sign the relationship between average employee compensation and μ_s^E. Furthermore, as before, we have that expected output net of the wage for the employee is independent of β_s^E and increases in μ_s^E. Thus, earnings of the supervisor are independent of β_s^E and increases in μ_s^E. Thus, we have the following results that are analogous to those stated at the end of Section 3.1:

1. The optimal piece rate $b\left(\mu_s^E, \beta_s^E, \alpha\right)$ is independent of $\left(\beta_s^E, \alpha\right)$ and increases in μ_s^E.

2. Expected compensation increases one-for-one in α and is independent of β_s^E. It is not possible to sign the relationship between expected compensation of the employee and μ_s^E.

3. Expected output $E\left[q|\mu_s^E, \beta_s^E, \alpha\right]$ increases in μ_s^E and α and is independent of β_s^E.

4. Earnings of supervisors $w_s\left(\mu_s^E, \beta_s^E\right)$ are independent of β_s^E and increase in μ_s^E.

These results mirror those in the previous section. We also have an additional result on the relation between the piece rate and the unobserved variation in supervisor heterogeneity.

5. The optimal piece rate declines in $\theta_\beta\sigma_\beta^2$ and $\theta_\mu\sigma_\mu^2$.

Besides these results, we can ask how employee and supervisor salaries as well as output depend on those components not observed by the firm. This question is empirically of interest because we have access to a panel of ratings and pay. We thus have an information advantage relative to the firm when it is setting pay. Furthermore, it is conceivable that firms do not use the available data optimally. Firms might therefore act as if they are uninformed about (β_s, μ_s) even though they might have inferred (β_s, μ_s) from the available data.

Thus, consider what predictions are obtain for how wages of an employee vary with $\left(\beta_s, \mu_s, \beta_s^E, \mu_s^E\right)$:

$$
\begin{aligned}
w\left(\beta_s, \mu_s, \beta_s^E, \mu_s^E, \alpha\right) = &\quad a_0\left(\mu_s^E\right) + \alpha - b\beta_s^E + b\left(\beta_s + b\mu_s\right) \\
= a_0\left(\mu_s^E\right) + \alpha_i + b\varepsilon_\beta + b^2\mu_s &\quad = a_0\left(\mu_s^E\right) + \alpha + b\theta_\beta\beta_s + b^2\mu_s + b\epsilon_\beta
\end{aligned}
$$

where we substitute the linear projection of $\varepsilon_\beta = \frac{cov(\varepsilon_\beta, \beta_s)}{var(\beta_s)}\beta_s + \epsilon_\beta = \frac{cov\left(\varepsilon_\beta, \beta_s^E + \varepsilon_\beta\right)}{var(\beta_s)}\beta_s + \epsilon_\beta = \theta_\beta\beta_s + \epsilon_\beta$. And, we have that a workers output is given by

$$
q = b\mu_s + \alpha + \varepsilon^q
$$

These two equations show how expected output and wages vary with $\left(\beta_s, \mu_s, \beta_s^E, \mu_s^E\right)$ in the partially informed firm:

1. Expected compensation increases in β_s, where the coefficient on β_s is given by the product of the optimal piece-rate multiplied by the proportion of the variation of supervisor heterogeneity that is unknown to the firm.

2. Output does not vary with β_s, but does vary with μ_s.

3.3 A 2-by-2 Matrix to Distinguish Types of Heterogeneity and How Informed the Firm is

Above we analyzed a structure that allows for different assumptions of how supervisors differ from each other and how informed the firm is about the types of supervisors employed. Supervisors could differ in their ability to manage their employees as well as in their bias. And, firms could differ in how informed they are about the differences between supervisors. Depending on the assumptions made, we obtain different predictions that we can test in the firm data available to us.

At this point, we find it useful to consider extreme assumptions on the source of heterogeneity and the information available to firms in order to build intuition about how the fundamentals of the model map into the data on ratings, compensation, and output. In particular, we will consider the situation where firms are perfectly informed ($\theta_\beta = \theta_\mu = 0$) or completely ignorant ($\theta_\beta = \theta_\mu = 1$). And, we will distinguish the case when supervisors differ primarily in how lenient they are $\left(\sigma_\beta^2 > 0, \sigma_\mu^2 = 0\right)$ from the case when supervisors differ primarily in their ability to elicit effort from their team members $\left(\sigma_\beta^2 = 0, \sigma_\mu^2 > 0\right)$. Combining, we obtain 4 different sets of assumptions on how supervisors differ from each other and how informed the firm is.

Recall, empirically we will strive to measure the heterogeneity ϕ_s in ratings associated with supervisors using the panel of performance ratings and the supervisor identifiers included in the data. We will then related worker and supervisor compensation as well as a measure of expected productivity of workers in a given team to ϕ_s. Table 3 summarizes what these four different sets of assumptions imply for the compensation of workers and supervisors and the expected productivity of workers.

Table 4: Model Predictions

Information \ Heterogeneity		Leniency $\left(\sigma_\beta^2 > 0, \sigma_\mu^2 = 0\right)$	Effectiveness $\left(\sigma_\beta^2 = 0, \sigma_\mu^2 > 0\right)$
Fully Informed Firms ($\theta_\mu = \theta_\beta = 0$)	Wages: $\frac{\partial \mathbf{E}[\mathbf{w}\mid\phi_\mathbf{s}]}{\partial \phi}$	0	$\neq 0^*$
	Piece rate: $\frac{\partial \mathbf{b}}{\partial \phi}$	0	> 0
	Productivity: $\frac{\partial \mathbf{E}[\mathbf{q}\mid\phi_\mathbf{s}]}{\partial \phi}$	0	> 0
	Supervisor Wages: $\frac{\partial \mathbf{w}}{\partial \phi}$	0	> 0
Uninformed Firms ($\theta_\mu = \theta_\beta = 1$)	Wages: $\frac{\partial \mathbf{E}[\mathbf{w}\mid\phi_\mathbf{s}]}{\partial \phi}$	> 0	> 0
	Piece rate: $\frac{\partial \mathbf{b}}{\partial \phi}$	0	0
	Productivity: $\frac{\partial \mathbf{E}[\mathbf{q}\mid\phi_\mathbf{s}]}{\partial \phi}$	0	> 0
	Supervisor Wages: $\frac{\partial \mathbf{w}}{\partial \phi}$	0	0

**The model does not make a clear prediction about the relationship between employee wages and* ϕ_s.

The above table reveals that the four different set of assumptions can indeed be distinguished.

It is intuitive that informed firms will undo any differences between supervisors in how lenient they are. Thus, wages of workers and supervisor, productivity and piece rates will not vary with ϕ_s if it reflects only differences in leniency. By contrast, the informed firm will be very responsive to differences in the managerial effectiveness of supervisors. Thus, supervisor wages, piece rates, productivity and potentially average employee compensation will vary with effectiveness of the supervisor when firms are well informed. Assuming that firms are perfectly informed, we can thus determine whether supervisors differ primarily in leniency or in managerial effectiveness by testing whether supervisor and employee compensation, productivity, and piece rates co-move with ϕ_s.

By contrast, if firms are uninformed, then the piece rates and the wages of supervisors will not vary across supervisors, regardless of why supervisors differ from each other (leniency or effectiveness). However, if firms are uninformed, we will find that employee wages will vary with ϕ_s, regardless whether it reflects leniency or managerial effectiveness. However, if the firm is uninformed, then expected productivity will only vary with ϕ_s if it indeed represents differences in managerial effectiveness μ_s.

Inspection of the above table reveals that observing how employee compensation varies with ϕ_s is particularly important to distinguish informed from uninformed firms if the main source of heterogeneity across supervisors is how lenient they are toward their team members. In uninformed firms, such variation increases average compensation of workers since the firm can not undo this variation. The informed firm by contrast will simply undo this source of variation. Similarly, observing how productivity varies with ϕ_s is necessary to distinguish between heterogeneity in leniency β_s and effectiveness μ_s if firms are uninformed.

Overall, we have developed a structure that allows for two fundamentally distinct interpretations of supervisor heterogeneity. We can distinguish between these sources of heterogeneity and can also empirically test how well informed the firm is about the supervisor heterogeneity within this structure.

4 Testing the Model

The previous sections analyzed the implications of heterogeneity across managers in ability (μ_s) and leniency (β_s). We showed how one can distinguish between the source of heterogeneity (μ_s or β_s) as well as the amount of information held by the firm by exploring how total pay, pay-for-performance, actual productivity, and supervisor pay vary with the empirical heterogeneity in ratings behavior observed across supervisors.

4.1 Empirical Methods

4.1.1 Estimating the Second Moment of Worker and Supervisor Effects in Ratings and Earnings

Central to our empirical analysis is identifying the heterogeneity in ratings behavior observed across supervisors and how it affects compensation. From equation 6 we have defined this heterogeneity as $\phi_s = e_s(\mu_s) + \beta_s$. Table 4 summarized what the model implies about how ϕ_s relates to (1) total pay of subordinates, (2) the pay for performance component of subordinate pay, (3) actual productivity, and (4) total pay of supervisors.

Consider then the following set of equations relating ratings and log wages to individual and supervisor effects. For the purpose of exposition we suppress time-varying controls so that one should think of the dependent variables as residuals from a regression on a set of time-varying controls. We do make the subscripts explicit for now.

$$\begin{aligned} r_{it} &= \alpha_i^r + \phi_s^r + \varepsilon_{it}^r \\ w_{it} &= \alpha_i^w + \phi_s^w + \varepsilon_{it}^w \end{aligned} \tag{17}$$

The $(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w)$ absorb individual and supervisor persistent differences in ratings and wages respectively. They are allowed to correlate freely. By construction, the unobservables $(\varepsilon_{it}^r, \varepsilon_{it}^w)$ are uncorrelated with the persistent differences. However, they are allowed to correlate with each other, as implied by the theory.[22] We assume (for now) that the errors are uncorrelated across individuals and time. Let Ω denote the 2-by-2 variance covariance matrix of $(\varepsilon_{it}^r, \varepsilon_{it}^w)$.

The first and primary goal of the empirical analysis is to estimate how the ratings and wage components attributable to workers and supervisors covary. That is, we strive to estimate the 2nd moments of $(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w)$ as well as $(\varepsilon_{it}^r, \varepsilon_{it}^w)$. In a second step, we will then relate output at the branch level to the variation in ratings associated with individuals and supervisors. The basic approach is to estimate fixed effect regressions on the system (17) and then use the estimated fixed effects and their correlation structure to determine the correlation structure in $(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w)$. To do so, we need to account for the sampling variation in the estimated fixed effects and the fact that the sampling error in the estimated fixed effects are not orthogonal from each other.[23]

Consider then stacking the N equations referring to ratings and the N equations related to log wages. The resulting system is

$$\begin{pmatrix} r \\ w \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix}' \begin{pmatrix} \alpha_r \\ \alpha_w \end{pmatrix} + \begin{pmatrix} F & 0 \\ 0 & F \end{pmatrix}' \begin{pmatrix} \phi_r \\ \phi_w \end{pmatrix} + \begin{pmatrix} \varepsilon_r \\ \varepsilon_w \end{pmatrix}$$

where D refers to the design matrix identifying the individual employee and F refers to the design matrix identifying the supervisors. $(r, w, \varepsilon_r . \varepsilon_w)$ are N-vectors. The (α_r, α_s) are N^*-vectors containing the individual

[22]Because they correlate with each other, we estimate the above eq. (17) jointly rather than separetely.

[23]Because our panel is relatively short (10 years at most), we face an incidental parameter problem in that the number of observation per employee and supervisors is relatively small and fixed. Thus, we can not use the second moments of the estimated fixed effects$\left(\hat{\alpha}_i^r, \hat{\alpha}_i^w, \hat{\phi}_s^r, \hat{\phi}_s^w\right)$ directly to estimate the second moment matrix of $\left(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w\right)$.

effects. N^* is the number of different employees in our data. The (ϕ_r, ϕ_w) are S-vectors of supervisor effects where S is the number of supervisors in the data.

Let $y = (r, w)$, $Z = (I \otimes D, I \otimes F)$, $\xi = (\alpha_r, \alpha_w, \phi_r, \phi_w)'$ and $\epsilon = (\varepsilon_r, \varepsilon_w)'$. Then

$$\hat{\xi} = (Z'Z)^{-1} Z'y$$

These estimates are unbiased but inconsistent in (N^*, S)since the number of time-periods per worker is fixed (and small). The variance-covariance matrix for the estimated fixed effects $\hat{\xi}$ is:

$$V_{\hat{\xi}} = (Z'Z)^{-1} Z'\Omega Z (Z'Z)^{-1} \tag{18}$$

We consistently estimate Ω using the within-transformation (within individuals and supervisors) of the dependent variables y to difference out the fixed effects that are not consistently estimated and exploit the independence assumption across i and t. For now we abstain from allowing for more dependencies across time or individuals, but it is possible to allow for more general error structures. Estimating Ω means estimating the variances of ε_r and ε_w, as well as the covariance of both.

To estimate the second moment matrix of $\xi = (\alpha_r, \alpha_w, \phi_r, \phi_w)$, we adapt the approach of Card, Heining, and Kline (2013) to a setting with a stacked system of equations. For an unbiased estimate $\hat{\xi}$ of ξ and any matrix A there is a simple expression for the expectation of the quadratic form $E\left[\hat{\xi}' A \hat{\xi}\right]$:

$$E\left[\hat{\xi}' A \hat{\xi}\right] = \xi' A \xi + tr\left(A V_{\hat{\xi}}\right) \tag{19}$$

By choosing A appropriately, we can let $\xi' A \xi$ equal the quantity that we want to estimate and then use $E\left[\hat{\xi}' A \hat{\xi}\right] - tr\left(A V_{\hat{\xi}}\right)$ as an estimator of $\xi' A \xi$. For instance, consider estimating the variance of α_r in our sample:

$$\sigma^2_{D\alpha_r} = \frac{1}{N^* - 1} \alpha_r' D' Q D \alpha_r$$

where Q is the demeaning matrix[24], an idempotent matrix. Defining $A_{D\alpha_r} = \begin{bmatrix} D'QD & 0 \\ 0 & 0 \end{bmatrix}$ comformable with ξ, we get $\hat{\sigma}^2_{D\alpha_r} - \frac{1}{N^*-1} tr(A_{D\alpha_r} * V_{\hat{\xi}}) \to \sigma^2_{D\alpha_r}$. We can proceed in the same manner for the other variances and covariances required.

The correction in equation 19 centers on $V_{\hat{\xi}}$, which of course depends directly on assumption made regarding the error structure Ω. So far we proceeded as if $(\varepsilon_r, \varepsilon_w)$ are uncorrelated across individuals and

[24] $Q = (I - i(i'i)^{-1} i')$

time. Furthermore, we assumed that $(\varepsilon_r, \varepsilon_w)$ is identically distributed across individuals. The former assumption is unlikely to hold and the later is in fact ruled out by the fact that r is a limited dependent variable. In future revisions, we will have to allow for heteroskedasticity in ε_r . In addition, we plan to allow for more complex temporal pattenrs of dependence in $(\varepsilon_r, \varepsilon_w)$. Our results so far should therefore be considered as preliminary with the caveat that the current specification of Ω is unlikely to describe the data generating process well.

Table 4 contains the estimated variance-covariance matrix of ξ obtained in this manner as well as our estimate of Ω. [25] These estimates inform us about the amount of heterogeneity in ratings associated with supervisors and individual employees. And, we can use these estimates to construct the regressions relating log wages itself as well as the components of log wages (α_w, ϕ_w) to the employee and supervisor effects in ratings. Based on this regression, we can then evaluate the predictions on log earnings summarized in Table 3.

4.1.2 Relating Ratings Effects to Productivity and Supervisor Earnings.

Besides estimating the distribution of the fixed effects ξ, we also regress other outcomes on linear combinations of the vector ξ. For example, we relate branch level outcomes on average worker and supervisor effects at the branch level. And, we relate supervisor outcomes to measures of team performance averaged within their team. Our estimates are based on regressions of outcomes on linear combinations of the estimated fixed effects $\hat{\xi}$, correcting for the measurement error using $V_{\hat{\xi}}$, our estimate of the estimation error of $\hat{\xi}$ from eq. 18.

Denote the outcome of interest by k_b where b indexes the unit of observation. This could for instance be a branch-year or a supervisor, depending on the outcome considered. The equation we ~~wish t~~ estimate is of the form

$$k_b = \beta' (H_b \xi) + \varepsilon_b \tag{20}$$

where ε_b is an unobservable with $E[\varepsilon_b | H_b \xi] = 0$ and $H_b \xi$ is the linear combination pertaining to observation b that produces the independent variable of interest. H_b has p rows, where p is the number of independent variables in equation 20. For instance, $H_b \xi$ might generate the average worker fixed effect α_i^r across all employees working in branch b at a given time as well as the average ϕ_r across all supervisors working in the branch at that time. Let $\theta_b = H_b \xi$ and $\hat{\theta}_b = H_b \hat{\xi}$. Then $\hat{\theta}_b = \theta_b + e_b^\theta$ with $e_b^\theta = H_b \left(\hat{\xi} - \xi\right)$, $E\left[e_b^\theta | \theta_b\right] = 0$ and $V_{b,\hat{\theta}} = H_b V_{\hat{\xi}} H_b'$. We can then use $V_{b,\hat{\theta}}$ to obtain a correction for the measurement error in $\hat{\theta}$ induced by

[25]For comparision, Appendix Table @ shows $E\left[\hat{\xi}\hat{\xi}'\right]$, the sample covariation in $\hat{\xi}$. This sample covariation in the estimated effects can be thought of as a naive estimator that does not account for the estimation error in $\hat{\xi}$.

the estimation error in $\hat{\xi}$.

Denote by N_b the number of observations for equation (20), by K the N_B dimensional row vector collecting the dependent variables, by Θ the N_B-by-p matrix with θ_b in row b and by ϵ_b then N_b vector that collects the unobservables ε_b. This allows us to stack equation 20 and write:

$$K = \Theta\beta + \epsilon$$

Let $\hat{\Theta}$ be the N_B-by-p matrix analogous to Θ constructed using the estimates $\hat{\theta}$. And, define $E_\theta = \hat{\Theta} - \Theta$. Using $\hat{\theta}_b$ as the independent variables in an OLS regression results in

$$\hat{\beta}_{OLS} = \left(\hat{\Theta}'\hat{\Theta}\right)^{-1}\left(\hat{\Theta}'K\right) = \left(\hat{\Theta}'\hat{\Theta}\right)^{-1}\hat{\Theta}'\Theta\beta + \left(\hat{\Theta}'\hat{\Theta}\right)^{-1}\hat{\Theta}'\epsilon$$

As the number of observations N_b increases,[26]

$$\hat{\beta}_{OLS} \to E\left[\hat{\theta}_b\hat{\theta}_b'\right]^{-1} E\left[\theta_b\theta_b'\right]\beta$$

This expression is of course the standard expression for the probability limit of the OLS estimator when the independent variables are measured with error. $E\left[\theta_b\theta_b'\right]$ describe the variation in independent variable measured without error and $E\left[\hat{\theta}_b\hat{\theta}_b'\right]$ the measurement error in the indepedent variables measured with error. The product $E\left[\hat{\theta}_b\hat{\theta}_b'\right]^{-1} E\left[\theta_b\theta_b'\right]$ is the analog to the signal to noise ratio for more than one variable.

We propose as a corrected estimator

$$\hat{\beta}_c = \left(\hat{\Theta}'\hat{\Theta} - \frac{1}{N_B}\Sigma_b V_{b,\hat{\theta}}\right)^{-1}\left(\hat{\Theta}'\hat{\Theta}\right)\hat{\beta}_{OLS} \tag{21}$$

For this estimator we have

$$\begin{aligned}\hat{\beta}_c &= \left(\hat{\Theta}'\hat{\Theta} - \frac{1}{N_B}\Sigma_b V_{b,\hat{\theta}}\right)^{-1}\left(\hat{\Theta}'\Theta\beta + \hat{\Theta}'\epsilon\right) \\ &\to \left(E\left[\hat{\theta}_b\hat{\theta}_b'\right] - plim\left(\frac{1}{N_B}\Sigma_b V_{b,\hat{\theta}}\right)\right)^{-1} E\left[\theta_b\theta_b'\right]\beta\end{aligned}$$

Now, since $E\left[\hat{\theta}_b\hat{\theta}_b'\right] = E\left[\left(\theta_b + e_\theta\right)\left(\theta_b + e_\theta\right)'\right] = E\left[\theta_b\theta_b'\right] + E\left[e_\theta e_\theta'\right]$ and $plim\left(\frac{1}{N_B}\Sigma_b V_{b,\hat{\theta}}\right) = E\left[e_\theta e_\theta'\right]$ we get $\hat{\beta}_c \to \beta$.

[26] $\frac{1}{N_B}\hat{\Theta}'\hat{\Theta} \to E\left[\hat{\theta}_b\hat{\theta}_b'\right]$ where the expectation is taken with respect to the distribution governing the random vectors for each branch b. We also have that that $\frac{1}{N_b}\Theta' E_\theta \to 0$ since the $E_\theta$is the estimation error of θ_b which is uncorrelated with θ_b across the branches. Similar reasoning implies that $\frac{1}{N_B}\hat{\Theta}'\Theta \to E\left[\theta_b\theta_b'\right]$ and $\frac{1}{N_B}\hat{\Theta}'\epsilon \to 0$.

In summary, we use our estimate of the first stage estimation error in $\hat{\xi}$ to obtain estimates of the estimation error in $\hat{\theta}$. This estimation error is akin to measurement error for$\hat{\theta}$ as a measure of θ. However, since we know the variance-covariance matrix of this measurement error, we can correct for it using the estimator in eq. (21).

4.2 Results

4.2.1 Supervisors differ substantially in their ratings behavior.

To begin consider how much supervisor heterogeneity there is in ratings behavior. Table 4 reports the second moments for all components of the log wage and ratings equations (17). we find that a standard deviation in ϕ_r amounts to a 14.8 percentage point increase in the probability of receiving a passing grade (4 or 5 out of 5).[27] We judge this to be indicative of substantial heterogeneity in rating styles across supervisors. The heterogeneity at the individual level is even larger - a standard deviation in α_r amounts to a 26.2 percentage point increase in the probability of receiving a passing grade.

4.2.2 Employees Earnings increase with ϕ_r

Using the estimates in Table 4, we can generate the regressions of log wages as well as its componenents $(\alpha_w, \phi_w, \varepsilon_w)$ on the ratings components $(\alpha_r, \phi_r, \varepsilon_r)$. Table 5-A lists the coefficients of these regressions. Each column of Table 5-A refers to a different dependent variable, be it log earnings or its components capturing individual heterogeneities α_w, supervisor heterogeneity ϕ_w or the unobservable ε_w.[28] We note that employees' compensation increases in their supervisors ratings component ϕ_r. Quantitatively, a one standard deviation increase in ϕ_r predicts about a 0.7% increase in earnings. A move from the 5th to the 95th percentile in the distribution of ϕ_r thus predicts an increase in log earnings of about 3%. Going back to the implications drawn from our simple model and summarized in Table 3, we observe that this finding rules out the joint hypothesis that supervisor heterogeneity is unrelated to productivity differences and that firms are perfectly informed about the heterogeneity in ratings behavior across supervisors.

We also find (col. 2) that worker heterogeneity in wages tends to be related to worker heterogeneity in ratings, but not to supervisor heterogeneity in ratings. Similarly, we find (col. 3) that supervisor heterogeneity in wages tends to be positively related to supervisor but not employee heterogeneity in ratings. These findings are suggestive that the wage variation is indeed driven (in a causal sense) by variation in the ratings rather than by sorting of individuals across supervisors. For instance, had we observed that

[27] Remember that the unconditional probability of receiving a passing grade is 45.6%.

[28] The regression coefficientss reported in Table 5-A are obviously just transformations of the second moment matrices in Table 4. Nevertheless, but they provide a useful way of summarizing the relationships in the data.

the heterogeneity in ratings across supervisors largely increased the individual effects associated with the individual employees, then this would be hard to reconcile with a model structure like the one presented in Section 3 in which ratings differences across supervisors generate compensation differences.

Table 5-B presents analog to Table 5-A obtained by directly regressing log earnings and its components on the supervisor and individual fixed effects obtained from the ratings regressions without correcting for the incidental parameter problem as described in Section 4.1.1. Comparing Table 5-A with table 5-B we find that the estimates corrected for the small sample problem are smaller for the worker fixed effects which are typically estimated using fewer observations than the supervisor effects. However, the estimates are not systematically different for the regression coefficients on ϕ_r . Quantitatively, the regression coefficients are quite close for all regression estimates. This suggests to us that we can in reduced form explore whether the log earnings relationship with the different components of the ratings equation is robust using simple regressions of log earnings on the fixed effects from the ratings equation. This allows us to forego the quantitatively costly step of constructing the estimates following Card et al. (2012) for each robustness check we might want to consider .

Table 5-C shows such reduced form estimates. Column 2 shows estimates when splitting the sample period into 2004-2008 and 2009-2014 so as to estimate the fixed effects using the first period and then regressinglog earnings on these effects using the second period. Column 3 and column 4 rely on observations in the branches (column 3) and in branches with KPI (column 4) only. Even though the size of the coefficients varies across the regressions samples, we find qualitatively consistent estimates in all subsamples.

4.2.3 Piece Rates Decline (not Increase) with ϕ_r

Table 6 presents the interaction of ϕ_r with the individual performance in a given period. This measures whether individual earnings are more responsive to performance measures conditional on supervisor heterogeneity. This specification suggests, if anything, that they are less sensitive, contradicting the prediction of the model based on heterogeneity in managerial productivity that we outlined in Section 3.

4.2.4 Branch Level Productivity Does Not Vary with ϕ_r

We now consider how our estimates of the supervisor and worker heterogeneity correlate with objective performance measures. For the years 2007-2010, we have access to a correlate of objective productivity for branches in a subset of our data in the form of a KPI ranking. [29] These rankings represent how a branch ranks relative to a set of peers in a given time-period along a number of key performance indicators.

[29] We are in the process of collecting additional data related to objective performance at the branch level.

We use these rankings as the outcome variable in equations (20). Our independent variables are the averages across employee fixed effects α_r in each branch as well as averages across supervisor effects ϕ_r in each branch. These independent variables are of course based on the estimate fixed effects which are estimated with error. We use the estimator (21) to correct for the bias arising from using independent variables that are estimated with error. We experiment with a number ofdifferent specifications for the dependent variable to ensure our results are not sensitive to functional form.

Before considering how average worker and supervisor effects are related to the KPI rankings across branches, we ask how the average subjective rating of all workers in a branch is related to the KPI ranking. This regression is informative for whether subjective or objective measures are related across branches in the first place. Regressions are weighted by the number of workers in the branch b, in year t with non-missing pay and performance data.

Panel A of table 7 summarizes these results. For all measures, subjective performance and KPI ranking are positively correlated. When an entire branch moves from failing to passing, the inverse rank score increases by 0.124, or by roughly two ranking spots (column 1); the probability of being the top branch in the peer group increases by 9 ppts (column 2); the probability of being in the top 5 branches increases by 22 ppts (column 3); the probability of being in the top half of the branches in the peer group increases by 19 ppts. All these correlations are highly significant. This is thus reassuring that subjective performance and objective performance are designed to pick up the same thing.[30]

The second regression equation is a direct test of our model. The explanatory variables are branch-time averages of the typical variables used. $\bar{\hat{\alpha}}_i$ is the average propensity to pass among all workers in the branch in that year; $\bar{\hat{\phi}}_s$ is the average propensity of supervisors to pass employees among supervisors employed at the branch in that year; $\bar{\hat{\epsilon}}_{ist}$ is the average idiosyncratic component of performance among all workers in the branch in that year. If there is only one supervisor giving ratings at the branch level, as is often the case, $\bar{\hat{\phi}}_s$ is the supervisor fixed effect for that supervisor. In cases where there is more than one rater, $\bar{\hat{\phi}}_s$ is the average fixed effect across raters, weighted by the number of subordinates they rated this period.

The model predicts that if leniency (β_s) is the primary driver of supervisor heterogeneity then objective performance will be unrelated to the supervisor effect because then supervisors do not influence actual productivity. If instead manager ability (μ_s) is the primary driver of supervisor heterogeneity then we should see a positive relationship between objective performance and supervisor effects, regardless of whether the firm is informed or not.

Panel B of table 7 shows the relationship between KPI ranking and the components of performance. We

[30]We have also explored the relationship between branch performance and performance of the highest-ranking person in the branch and obtained similar results.

find that the primary driver of the relationship between subjective and objective performance measures is via the idiosyncratic shock to performance (ϵ) and not through the worker or supervisor effects. For each KPI variable this is the only coefficient that is statistically significant and it is also an order of magnitude larger than the other coefficients.

Standard errors are fairly large so that we cannot rule out even a fairly sizable positive relationship between supervisor effects and objective performance. However, the evidence so far on this dimension supports leniency, rather than ability, as being the primary driver of the supervisor effects. We are awaiting additional data on the performance of branches and hope that this will allow us to draw sharper conclusions.

4.2.5 Supervisor Pay Does Not Increase in Supervisor Heterogeneity

The fourth comparative static relates $\hat{\phi}_s$ to supervisor pay. We estimate the following regression:

$$pay_{st} = \beta_0 + \beta_1 \hat{\bar{\alpha}}_{t} + \beta_2 \hat{\phi}_s + \beta_3 \hat{\bar{\epsilon}}_{ist} + \beta' \bar{X}_{it} + \gamma' Y_{st} + \nu_{bt}$$

We regress pay measures for a supervisor, s, in year, t, on that supervisor's propensity to pass subordinates ($\hat{\phi}_s$). We control for average quality of workers in the pool that year, the average idiosyncratic performance of subordinates this period, average time-varying characteristics of subordinates and time-varying supervisor characteristics. We use the corrected estimator eq. (21) as described above. Standard errors are clustered at the supervisor level, the level of variation underlying the key explanatory variable. Observations are weighted by the number of subordinates to a given supervisor in time t.

If supervisor heterogeneity is driven primarily by ability, and firms are informed about this ability, then supervisors should capture it in their pay ($\beta_2 > 0$). If firms are uninformed, or if supervisor heterogeneity is primarily driven by leniency, then high-rating supervisors will not be paid differently compared to low-rating supervisors ($\beta_2 = 0$).

Results are reported in table 8. We find that supervisors earnings are positively influenced by average team quality and the avearage idiosyncratic performance of the team, but is not affected by the supervisor effect. This is the case both for log(earnings) and for log(bonuses). Hence, our results are in line with leniency and an uninformed firm as the explanation for supervisor heterogeneity.

4.3 Discussion

To summarize we find that for supervisors who give higher ratings, holding constant the ability of their workers:

1. their workers receive more in total earnings

2. their workers do not have higher pay for performance components
3. their branch-level objective performance is not higher
4. they themselves receive marginally higher pay.

Returning to table 4, our model would rationalize this set of results as primarily supporting the bottom-left quadrant: manager heterogeneity is driven by leniency which the firm is uninformed about. When workers receive higher ratings, the firm does not know whether true performance or leniency is driving these ratings so they reward them with higher pay and do not adjust the strength of incentives. But since differences performance ratings are driven largely by leniency and not actual ability, these differences will not show up in objective measures of productivity. We do find small positive impacts on pay of the supervisors themselves. These effects are only marginal and could again reflect the fact that firms are uninformed. [31]

5 Reduced Form Exploration: Learning about Supervisors?

We interpret the results above to be broadly consistent with supervisors that differ in how lenient they are in rating their team-members and a firm that is imperfectly informed about this heterogeneity across supervisors. Informed firms would undo the variation in leniency across supervisors when setting employee pay. This is not what we observe in the earnings regressions. Rather, individuals earnings increase in ϕ_r.

One might expect firms to learn over time about the heterogeneity in supervisor ratings. If so, the relationship between earnings of employees and supervisor heterogeneity should be weaker for supervisors that are well known to the firm. We investigate this hypothesis next. In particular, we interact our estimate of supervisor heterogeneity $\hat{\phi}_r$ with the length of time that a supervisor has been rating individuals. Table 9 reports these specifications. Consistent with the notion that the firm learns to distinguish between more or less lenient raters, we find that a negative interaction of $\hat{\phi}_r$ with rater experience. The effect is however not rapid. Based on the linear interaction reported in Table 9, it takes about 20 years of observing a supervisor rating employees for the relation between $\hat{\phi}_r$ and log earnings to decline to zero.

6 Conclusion

In this paper, we use personnel data from a Scandinavian firm in the service sector to examine how important supervisor heterogeneity in ratings behavior is in setting pay. We find large systematic differences in average supervisor ratings using the variation in ratings within workers across supervisors. The heterogeneity across

[31] Though not shown, we also find that managers who rate their subordinates more highly receive higher ratings by their own supervisors.

supervisors is related to heterogeneity in pay across workers, but it does not predict supervisor pay nor objective performance indicators that vary across branches. Our results are preliminary in that we are currently working on allowing for richer error specifications. Our results might also evolve as additional data on objective performance at the branch level performance becomes available.

Overall, our results are consistent with variation in supervisor ratings that does not reflect differences in productivity associated with the supervisor and a firm that is not well informed about supervisor heterogeneity. Our finding that supervisor heterogeneity in ratings does not affect supervisor earnings nor performance at the branch level supports the notion that supervisor heterogeneity is not associated with differences in productivity across supervisors. Our finding that firms are not able to fully undo the ratings heterogneity across supervisors when setting individual pay suggests that the firm is not perfectly informed about the heterogeneity across supervisors in ratings behavior. We also find that the relation between supervisor heterogeneity and compensation weakens the longer a supervisor has been in a supervisory role, suggesting that maybe the firm learns about supervisor heterogeneity as supervisor tenure increases.

References

Altonji, Joseph and Charles Pierret (2001), "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 113: pp. 79-119.

Baker, George, Michael Gibbs, and Bengt Holmstrom (1994a), "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, CIX: pp. 881-919.

Baker, George, Michael Gibbs, and Bengt Holmstrom (1994b), "The Wage Policy of a Firm," *Quarterly Journal of Economics*, CIX: pp. 921-955.

Bertrand, Marianne, and Antoinette Schoar (2003), "Managing with style: the effect of managers on firm policies," *Quarterly Journal of Economics*, 118(4): pp. 1169-1208.

Card, David, Joerg Heining, and Patrick Kline (2013), "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128(August): pp. 967-1015.

Farber, Henry. and Robert Gibbons (1996), "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111: pp. 1007-1047.

Frederiksen, Anders, Fabian Lange, and Ben Kriechel (2013), "Performance Evaluations and Careers: Similarities and Differencs across Firms," mimeo.

Gibbons, Robert and Michael Waldman (1999), "A theory of wage and promotion dynamics inside firms," *Quarterly Journal of Economics*, 114(4): pp. 1321-58.

Gibbons, Robert and Michael Waldman (2006), "Enriching a Theory of Wage and Promotion Dynamics

Inside Firms, *Journal of Labor Economics*, Vol 24: pp. 59-107.

Guilford, J.P. (1954) “Psychometric Methods”. New York: McGraw-Hill.

Holmstrom, Bengt (1979), “Moral Hazard and Observability, *The Bell Journal of Economics*, Vol 10(1), pp. 74-91.

Holmstrom, Bengt and Paul Milgrom (1987), “Aggregation and Linearity in the Provision of Intertemporal Incentives,” *Econometrica*, Vol 55(2), pp. 303-328.

Jovanovic, Boyan (1979), Job Matching and the Theory of Turnover, *The Journal of Political Economy*, 87(October), pp. 972-90.

Kahn, Lisa and Fabian Lange (2014), “Employer Learning, Productivity and the Earnings Distribution: Evidence from Performance Measures,” *Review of Economic Studies*, 81(4) pp.1575-1613.

Lange, Fabian (2007), “The Speed of Employer Learning”, *Journal of Labor Economics*, Vol 25(1).

Lazear, Edward (2000), “Performance Pay and Productivity,” *The American Economic Review*, Vol 90(5), pp. 1346-1361.

Lazear, Edward, Kathryn Shaw, and Christopher Stanton (2015), “The Value of Bosses,” *Journal of Labor Economics*, Vo. 33(4): pp. 823-861.

Oyer, Paul and Scott Schaefer (2011), “Personnel Economics: Hiring and Incentives,” in the *Handbook of Labor Economics*, 4B, eds. David Card and Orley Ashenfelter, pp. 1769-1823

Table 1: Summary Statistics

	Full Sample (N=136,286)		Estimation Sample (N=78,859)	
	Mean	St Dev	Mean	St Dev
Outcomes:				
With performance	0.714	0.452	1	0
Performance	3.48	0.658	3.51	0.668
Earnings[1]	1.73	1.18	1.87	1.10
Wages[1]	1.63	0.70	1.76	0.58
Bonuses[1]	0.1	0.70	0.11	0.71
Bonus received	0.294	0.456	0.312	0.463
Wage growth in pct.	0.023	0.130	0.022	0.065
Promotions	0.099	0.299	0.107	0.309
Quits	0.066	0.249	0.027	0.163
Layoffs	0.018	0.133	0.005	0.071
Controls:				
Full-time	0.836	0.370	1	0
Tenure	17.5	13.5	18.0	13.3
Age	43.5	11.3	44.0	10.7
In Branch	0.445	0.497	0.436	0.496
Female	0.519	0.500	0.439	0.496
Supervisor Female	0.278	0.448	0.265	0.441
Supervisor age	44.2	10.4	44.4	9.9
Supervisor tenure	19.6	11.8	19.4	11.8

Note: The Full Sample consists of all observations between 2004-2014 with either a wage or a performance measure. The Estimation Sample consists of all individuals with performance measures, working full-time, observed at least twice in the data for whom we can estimate double fixed effects specifications.

1) Earnings, Wages, and Bonuses data are reported relative to average earnings, wages, and bonuses in the country.

Table 2: Performance Distribution

Fail			Pass	
1	2	3	4	5
0.11%	2.98%	49.04%	41.40%	6.47%
52.13%			47.87%	

This table is based on the estimation sample which consists of those 78,859 individuals with 2 or more ratings for whom we can estimate fixed effects.

Table 4 Panel A Variance-Covariance Matrix of Worker and Supervisor Effects in Ratings and Log Earnings

	α_r	α_w	ϕ_r	ϕ_w	Std Dev.
α_r	0.069				0.262
α_w	0.008	0.008			0.091
ϕ_r	-0.009	-0.002	0.020		0.142
ϕ_w	-0.001	-0.001	0.002	0.003	0.050

Table 4 Panel B Variance Covariance Matrix of Unobservables in Rating and Log Earnings Equation

	ε_r	ε_w	Std Dev.
ε_r	0.124		0.352
ε_w	0.003	0.007	0.082

Notes: Reported are the Second Moments of the ratings and log earnings components associated with workers and supervisors as well as the residuals from eq. (17) using the methodology outlined in Section 4.1.1. The last column reports the standard deviation of the reported variables.

Table 5-A: Log(Earnings) Components on Rating Effects

Dependent Variables:	Log Wages (1)	Wage Effect (φ) (2)	Wage Effect (α) (3)	Wage residual (ε) (4)
Supervisor Ratings effect (φ)	0.049***	0.090***	-0.041***	0
	(0.002)	(0.001)	(0.002)	(na)
Worker ratings effect (α)	0.108***	-0.009***	0.117***	0
	(0.001)	(0.001)	(0.001)	(na)
Pass Residual (ε)	0.024***	0	0	0.024***
	(0.005)	(na)	(na)	(0.005)

Notes: Column 1 presents the regression of log wages on the supervisor effects and the wage effects in ratings. Columns (2)-(4) present the regressions of each of the components of the log wage equation on the components of the ratings equations. The regression coefficients are obtained using the variance-covariance matrix (table 4) based on the estimator in Card, Heining, and Kline (2012). By construction the residuals are orthogonal to the worker and supervisor effects.

ble 5-B: Unadjusted Impact of Performance Rating Components on Log(Earnings) Compone

Dependent Variables:	Log Wages (1)	Wage Effect (φ) (2)	Wage Effect (α) (3)	Wage residual (ε) (4)
Supervisor Ratings effect (φ)	0.047***	0.067***	-0.020***	0
	(0.002)	(0.001)	(0.002)	(na)
Worker ratings effect (α)	0.085***	-0.006***	0.091***	0
	(0.001)	(0.001)	(0.001)	(na)
Pass Residual (ε)	0.028***	0	0	0.028***
	(0.001)	(na)	(na)	(0.001)
Observations	78,659	78,659	78,659	78,659

Notes: Column 1 presents the regression of log wages on the supervisor effects and the wage effects in ratings. Columns (2)-(4) present the regressions of each of the components of the log wage equation on the components of the ratings equations. The ses regressions are obtained by regressing log earnings and the estimated effects frmo the wage

Table 5-C: Robustness Checks

Dependent Var: Log(Earnings)	(1) Full Sample	(2) Split Sample	(3) Branches Only	(4) KPI Only
Supervisor Ratings effect (φ)	0.047***	0.070***	0.032***	0.039***
	(0.002)	(0.004)	(0.003)	(0.006)
Worker ratings effect (α)	0.085***	0.079***	0.082***	0.072***
	(0.001)	(0.004)	(0.001)	(0.003)
Pass Residual (ε)	0.028***	0.045***	0.021***	0.019***
	(0.001)	(0.003)	(0.001)	(0.003)
Observations	78,659	15,111	34,323	7,896
R-squared	0.045	0.036	0.092	0.085

Notes: Column 1 presents the regression of log earnings on the supervisor effects and the wage effects in ratings using the Full estimation sample. It corresponds to column (1) of Table 5-B. Columns (2)-(4) present log earnings regressions on sub-samples. Column (2) presents regressions when the fixed effects are based on 2004-2008 and the regression is performed on 2009-2014. For columns (3) and (4) the fixed effects in the ratings equation are estimated on the full sample, but the regressions are performed on individuals in the branches only (column (3)) or individuals in branches with

Table 6: Supervisor Effects and Pay for Performance

Dependent Variable	Log Earnings
	(1)
Supervisor FE (φ)	0.048***
	(0.003)
Worker FE (α)	0.087***
	(0.002)
Pass Residual (ε)	0.029***
	(0.001)
φ*Pass	-0.049***
	(0.005)
Observations	78,659
Partial R-squared	0.046

Notes: See table 5. Significance levels are represented using stars: ***

Table 7: Supervisor Effects and Branch-Level Productivity (KPI's)

Dependent Variable: (mean)	(1) Inverse Rank Score (0.53)	(2) Pr(Top) (0.06)	(3) Pr(Top 5) (0.30)	(4) Pr(Top half) (0.46)	(5) Financial performance (102.9)
Panel A: Average Performance					
Average Pass Rate	0.124***	0.086**	0.218***	0.193**	1.766
	(0.046)	(0.039)	(0.075)	(0.081)	(2.047)
Panel B: Average Components of Performance					
Branch-Level Average:					
Supervisor FE (φ)	0.028	0.025	-0.001	0.102	2.281
	(0.081)	(0.068)	(0.130)	(0.141)	(3.306)
Worker FE (α)	0.017	0.083	0.025	-0.003	0.449
	(0.079)	(0.067)	(0.127)	(0.138)	(3.497)
Pass Residual (ε)	0.242***	0.118*	0.429***	0.392***	2.194
	(0.075)	(0.064)	(0.121)	(0.131)	(3.227)
Observations	766	766	766	766	156

Notes: Observations are at the branch-year level, weighted by number of workers with non-missing pay and performance variables. Inverse rank score is -1 times the branche's KPI ranking in that year divded by the number of branches it is ranked against. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Supervisor Effects and Supervisor Pay

Dependent Variable:	Supervisor Log Earnings	Supervisor Log Bonus
	(1)	(2)
Supervisor FE (φ)	0.011	0.059
	(0.019)	(0.070)
Team-Average Worker FE	0.103***	0.307***
	(0.0201)	(0.0663)
Team-Ave Pass Resid	0.025***	0.078
	(0.009)	(0.049)
Observations	9,444	6,816
Partial R^2	0.018	0.007

Table 9: Supervisor Experience and Ratings Heterogeneity

Dependent Variable: Log Earnings	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ϕ_r	0.041***	0.041***	0.052***	0.040***	0.041***	0.052***	0.040***	0.041***	0.051***	0.051***
	(0.003)	(0.008)	(0.005)	(0.003)	(0.003)	(0.011)	(0.008)	(0.008)	(0.005)	(0.011)
α_r	0.086***	0.086***	0.086***	0.083***	0.087***	0.086***	0.083***	0.087***	0.085***	0.085***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
ε_r	0.028***	0.028***	0.028***	0.028***	0.035***	0.028***	0.028***	0.035***	0.035***	0.035***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)
Rater Tenure / 100	0.146***	0.146*	0.142***	0.144***	0.145***	0.142*	0.144*	0.145*	0.141***	0.141*
	(0.027)	(0.084)	(0.027)	(0.027)	(0.027)	(0.084)	(0.083)	(0.084)	(0.027)	(0.083)
ϕ_r * Rater Tenure / 100			-0.306***			-0.306			-0.284***	-0.284
			(0.102)			(0.278)			(0.107)	(0.277)
α_r * Rater Tenure / 100				0.094			0.094		0.047	0.047
				(0.057)			(0.097)		(0.060)	(0.088)
ε_r * Rater Tenure / 100					-0.169***			-0.169***	-0.170***	-0.170***
					(0.055)			(0.061)	(0.055)	(0.061)
SE clustered at Supervisor Level?		Yes				Yes	Yes	Yes		Yes
Observations	78,859	78,859	78,859	78,859	78,859	78,859	78,859	78,859	78,859	78,859
R-squared	0.046	0.046	0.046	0.046	0.046	0.046	0.046	0.046	0.046	0.046

Notes: Observations are at the individual-year level. When indicated, observations are clustered by supervisor. Regressions control for the set of controls described identical to Table 5. Rater Tenure represents the number of years individuals have been in rating employees. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.