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*

July 27, 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 worker and firm outcomes. We find that supervisor heterogeneity in performance ratings is positively correlated with subordinate pay and pay for performance, with supervisor pay, and with objective team-level performance measures. Our evidence suggests that workers benefit from being assigned to higher rating supervisors, but that higher raters also manage to elicit higher output from subordinates. Firms seem to be partially informed about the differences in ratings behavior across supervisors.

*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

Subjective performance evaluations are an ubiquitous and controversial component of the modern workplace. Supervisors are asked to rate their team-members and these ratings are used to compensate employees, to allocate workers to tasks, and to determine who gets promoted. However, supervisor evaluations are inherently subjective, so that supervisors might differ widely in how they evaluate equivalent behavior. These idiosyncratic differences might in turn significantly impact employees earnings, career progression, and work satisfaction. The presence of supervisor ratings heterogeneity will also constrain firms when they design performance management systems. Despite its importance for workers and firms, little is known about the extent and nature of ratings heterogeneity across supervisors. Unanswered questions include: How much heterogeneity is there in ratings behavior? How and how much does this heterogeneity affect worker outcomes? What is the nature of the heterogeneity in supervisor rating behavior? Do supervisors simply differ in how lenient and generous they are?[1] Or do differences in ratings behavior reflect differences in management style that have real impacts on output? Finally, is the firm informed about differences across supervisors and how does the firm respond to their presence?

We turn to exceptionally rich data from the performance system of a Scandinavian service sector firm to answer these questions. We find substantial ratings heterogeneity across supervisors which correlates positively with subordinate outcomes, supervisor outcomes, and objective measures of team performance. Subordinates of a high rater are paid more, are more likely to be promoted, and have more stable jobs at the firm. From self-reports, we also know that they are more satisfied at their work and with their immediate supervisors. Teams supervised by higher raters tend to perform better on other objective performance indicators and higher raters tend to earn more, suggesting that they are also more valued by the firm. The evidence thus suggests both that employees benefit from working for higher raters and that managerial ability and rater generosity correlate.

We develop a simply analytic framework to interpret these novel facts. Within the framework, we explore the nature of the ratings heterogeneity across supervisors: are ratings differences across supervisors driven by differences in managerial ability or by differences in leniency bias? And, the framework allows us to investigate the extent to which firms are informed about this heterogeneity.

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 which is not directly observed by the firm and the supervisors.

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

Supervisors observe worker output and report on this output to the firm. However, supervisors 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 affects 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 depend on how informed firms are about the differences between supervisors.

We can use this model to rationalize our empirical results. We find that team performance and supervisor earnings are higher when teams are supervised by higher raters. We explain this finding within our model by assuming that ratings style across supervisors is related to managerial ability and that firms are at least somewhat informed about this heterogeneity. However, the evidence from quitting behavior and worker satisfaction surveys suggests that employees do earn economic rents from working for "higher raters". In the context of our model, we conclude that firms are not fully informed about this heterogeneity.

The remainder of the paper proceeds as follows. Section 2 introduces the firm and the data at our disposal. In Section 3 we present the model and show what it implies for how career outcomes and performance are related to rater heterogeneity. Section 4 presents the empirical analysis. Section 5 discusses and investigates in reduced form 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.

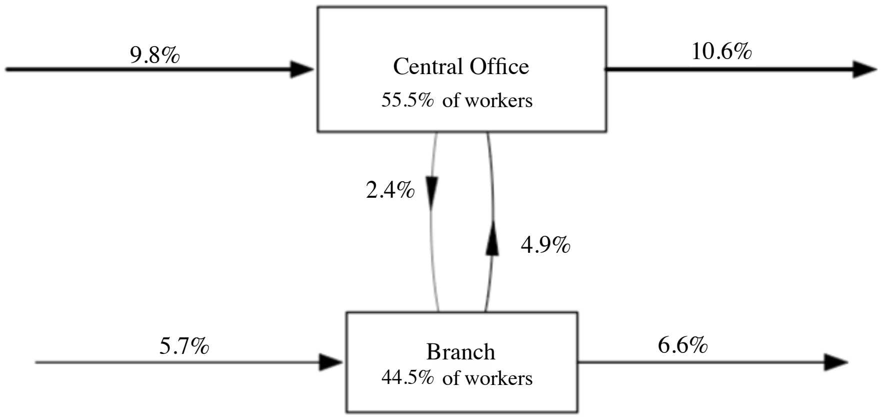
2 Firm and Data

2.1 Firm Overview

We rely on personnel data from a large Scandinavian service sector firm. 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. In 2013, there were 269 branches and the median branch had 15 employees.

The firm has 11 identifiable job levels (see appendix figure A1). Workers in the central corporate office have a variety of functions and there are more high level jobs (level 11). In contrast, the typical branch has 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 sometimes a trainee (level 1). The plurality

Figure 1: Flows across Central Office and Branch Network
in % of employees in CO or B respectively



of workers enter the firm at a low level (1-3) and there is a consistent stream of promotions across all levels, with roughly 10% of workers promoted out of a given level in a given year and very few demotions. We pay particular attention to the branch network because we have access to objective (financial and performance based) performance measures of the branch performance. We have no such measures for the central office.

Our sample comprises all employees engaged in domestic activities between 2004 and 2014.[2] 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.[3] 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).

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 as well as for the estimation sample, which restricts to the

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

[3]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.

60% of worker-years where both a pay and performance measure are available.[4]We report earnings (and its components) relative to average per capita earnings in this country. On average, individuals at this firm earn 185% of the national average, which reflects the fact that this the workforce is highly skilled and that this firm is known to be an attractive employer. 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. Comparing supervisors and workers, we observe that supervisors are on average only about a year older than the average employee (45.2 vs 44 years), and have one-and-a-half years more firm tenure (19.7 vs 18.1 years). The long tenure of average employees reflects the fact that quit and layoff rates are quite low. In the full sample, the total separation rate (quits+layoffs) is 8.4%.[5]

[4]We further restrict to the sample of workers and supervisors for whom we can estimate performance rating fixed effects, omitting workers with less than two years of data and supervisors with less than two subordinates. We drop a small number of additional supervisors who do not yield independent varation once worker fixed effects and controls are included in the empirical specification.

[5]The quit and layoff rate in the estimation sample is much lower since many of those separating from the firm do not receive a rating in the year they leave.

Table 1: Summary Statistics

	Full Sample (N=136,286)		Estimation Sample (N=77,077)	
	Mean	St Dev	Mean	St Dev
Outcomes:				
With performance	0.714	0.452	1	0
Performance	3.48	0.658	3.51	0.665
Earnings[1]	1.73	1.17	1.85	1.04
Wages[1]	1.63	0.69	1.74	0.55
Bonuses[1]	0.10	0.70	0.10	0.69
Bonus received	0.307	0.461	0.310	0.462
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	0.162
Layoffs	0.018	0.133	0.005	0.071
Controls:				
Full-time	0.836	0.370	1	0
Tenure	17.5	13.5	18.1	13.3
Age	43.5	11.3	44.0	10.7
In Branch	0.445	0.497	0.443	0.497
Female	0.519	0.500	0.439	0.496
Supervisor Female	0.278	0.448	0.270	0.444
Supervisor age	44.2	10.4	45.2	8.0
Supervisor tenure	19.6	11.8	19.7	11.6

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%	3.00%	49.33%	41.28%	6.27%
52.44%			47.56%	

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.

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. 90% of ratings are either a 3 or a 4,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 lose 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 allows us to interpret linear regression coefficients as marginal effects on the probability of receiving a "passing grade". For these reasons, we build our empirical investigation around this pass-fail metric.[6]

Our data also contains two measures of branch performance. The first reports how a branch ranks on a set of Key Performance Indicators (KPIs) within a group of peer branches, defined by the firm. The KPIs include measures of financial performance of the branches, as well as other metrics (for example, customer satisfaction). The set of KPIs 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 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 a 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 access to employee job satisfaction surveys for the years 2004 to 2010. These surveys include questions 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

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

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 related to the supervisor. 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 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, naturally, the data is not linked do to employer or supervisor data.. Interestingly, employers - including our firm - usually contract with outside consulting companies to conduct employee satisfaction surveys. These consulting firms then report back to the firm averages at the branch or business unit level. By collecting the data at arms-length, the firms hope to induce truthful reporting by employees. 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 itself was not able to make this link.

In summary, we have unusually rich panel data with information on the vertical and horizonal structure of the firm, the careers of individuals, the performance evaluations received and the identities of the raters, measures of branch-level performance and survey responses from worker satisfaction surveys. We know of no equivalent data-set in the literature.

2.2 Variation in Performance Measures

There is substantial systematic variation in the incidence of passing grades across supervisors and workers. To estimate 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 $\phi_{s(i,t)}$, as well as time-varying worker controls (X_{it}) and supervisor controls ($Y_{s(i,t),t}$):[7]

$$p_{it} = \alpha_i + \phi_{s(i,t)} + \beta' X_{it} + \gamma' Y_{s(i,t)t} + \epsilon_{it}^p \tag{1}$$

We can estimate equation (1) as is, but we are also concerned about the variation in the estimated fixed

[7] The worker controls (X_{it}) include indicators for 5-year age and tenure groups, full-time status, gender, job level; supervisor controls (Y_{st}) include indicators for 5 year groups in age, gender, and job level of the supervisor. 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.

Table 3: Variance-Covariance Matrix of Ratings Components

Panel A: Unadjusted

	α_r	ϕ_r	ε_r	Std Dev.
α_r	0.104			0.322
ϕ_r	-0.018	0.035		0.187
ε_r	0	0	0.142	0.377

Panel B: Adjusted

	α_r	ϕ_r	ε_r	Std Dev.
α_r	0.074			0.272
ϕ_r	-0.011	0.023		0.153
ε_r	0	0	0.124	0.352

Notes: Reported are the Second Moments of the ratings associated with workers and supervisors as well as adjusted moments using the methodology outline in appendix A1. The last column reports the standard deviation of the reported variables. α are worker fixed effects in ratings, ϕ are supervisor fixed effects and ε are partial residuals (after conditioning out controls).

effects induced by estimation error and the fact that the estimation error across α and ϕ will be correlated.[8] To account for this sampling variation, we adapt the approach of Card, Heining, and Kline (2013). See appendix A.1 for more details.

Table A2 presents the variance-covariance matrix of α, ϕ, and ϵ for both unadjusted and adjusted estimates in panels A and B, respectively. The adjustment for sampling error has a modest effect on the moments, typically reducing their magnitude by roughly a third. We find that there is substantial variation in ϕ_s across supervisors. Using the adjusted moments in panel B, a one-standard deviation increase in ϕ_s amounts to a 15.3 percentage point (32%) increase in the probability of receiving a passing grade. Thus a move from the 10th to the 90th percentile in the distribution of ϕ_s, assuming that ϕ_s is normally distributed, is associated with a 39 percentage point increase in the probability of receiving a passing grade. The heterogeneity at the worker level is even larger - a standard deviation in α_i amounts to a 27.2 percentage point increase (57%) in the probability of receiving a passing grade. Finally, there is substantial idiosyncratic variation in ratings, holding constant these fixed effects and a rich set of time-varying controls.

Identifying these fixed effects relies on variation generated by worker mobility across supervisors. In our data, employees typically change supervisors at least once over their tenure at the firm and often more. Similarly, supervisors manage many different employees over time, with some employees joining or leaving their teams almost every year.[9] However, we are natiurally worried that worker sorting across supervisors drives the estimated heterogeneity in ratings. To understand the problem, we test for symmetry in the

[8] Because our panel is relatively short (11 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 cannot simply use the estimated fixed effects $\left(\hat{\alpha}_i^r, \hat{\phi}_s^r\right)$ to characterize the true variation.

[9] Over the period 2004 to 2014 the average employee had 2.94 different supervisors. Employees who were with the firm throughout the entire period had on average 4.31 different supervisors. The average supervisor manages 9.76 (s.d. of 10.16) employees in a given year, and 21.48 different employees over the full time period 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. The interconnectedness in the firm is in fact so large that the largest connected set covers the entire firm.

impact of moving to a different supervisor, following Card, Heining and Kline (2013).[10] We find that the increase in ratings associated with moving from a low- to a high-rater is remarkably similar to the decrease associated with moving from a high- to low-rater. This symmetry is reassuring that sorting of workers across supervisors is not too problematic.

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.

3 Model[11]

Basic Setup

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 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 a random time-varying component $\varepsilon^q_{i,t}$. This time-varying component is normally distributed with mean 0 and variance σ^2_q and 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 report a rating r to the firm. Below we will introduce heterogeneity across supervisors along two dimensions: (a) heterogeneity

[10] We split supervisors in half, based on their average propensity to pass subordinates. We plot average ratings of workers that move across different combinations of supervisors in appendix figure A2. We find that, regardless of origin, workers that move to a high rater see ratings increases, and workers that move to a low- rater see ratings decreases, both of similar magnitude, and similar to the average difference in ratings across high- and low-raters, 0.27.

[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] Variables that vary across supervisors are indexed s or s(i,t), where $s(i,t)$ indexes the supervisor that individual i is assigned to in period t. Unless necessary for clarity, we do suppress individual and time subscripts but tend to retain the supervisor subscript to indicate that the variable varies across supervisors.

in managerial ability, which impacts the worker's cost of effort (μ_s), and (b) heterogeneity 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 workers are assigned to and the wage function. This wage function depends on the known characteristics of supervisors and and the rating a worker receives.

2. Workers are matched to supervisors, observe the parameter μ_s that parameterizes managerial ability, exert effort e and produce q.

3. Supervisors observe q and provide ratings r.

4. Workers are paid according to their contracted wage function.

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

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

Able 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.

Supervisors trade off the conflicting goals of being lenient and reporting truthfully on their employee's productivity. We embed this trade-off in supervisor preferences:

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

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 supervisors report the observed output q plus a supervisor specific parameter $\beta_s=\frac{\tilde{\beta}_s}{\tilde{\gamma}_s}$ which we call the supervisor bias". It measures the strength of the motive to report favorably relative to the motive to report truthfully.

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

[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).

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.

We now consider the contracts firms and workers enter. We limit ourselves to a static set-up with common information among agents (workers, supervisors, and 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 pay-off relevant aspects of the employment relationship including the assignment $\{\mu_s, \beta_s\}$ as well as the 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 worker type α_i and supervisor characteristics $\{\mu_s, \beta_s\}$. Thus, we consider contracts of the form $w_{i,s,t} = a_{i,s} + b_{i,s} r_{i,t}$.[14] In addition, we assume whatever is needed to ensure that the exponent in equation (3) is normally distributed conditional on the available information *both* at the contracting stage and when workers choose 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

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

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^*:

$$e^* = b_{i,s} \mu_s \tag{8}$$

[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]What exactly we need to assume varies with what we assume about the information agents have. 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.

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.

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:

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

The optimal piece-rate b maximizes the sum of the expected profit and the certainty equivalent after substituting the optimal effort (eq.8):[18]

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

This results in the standard solution familiar from the literature:

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

Competition in the labor market implies that profits from any worker-supervisor pair are zero:

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

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

[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.

the supervisor β_s. Consequently, the effort choice e^* does not vary with β_s either. Rearranging the certainty 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 surplus across worker-supervisor pairs and 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

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

This expression cannot generally be signed because increases in μ_s induce two countervailing effects. On one hand, the costs of providing any given effort level declines with μ_s which tends to lower compensation. One the other hand, when μ_s increases, the optimal piece rate increases as well and so does the risk borne

[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.

by workers. To induce workers to bear this risk, compensation will have to increase. 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 ($b > \frac{1}{2}$). When incentives are high ($b > \frac{1}{2}$), much effort is provided and better managers reduce the efforts costs born by workers significantly. Therefore wages decline with μ_s 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 how wages vary with 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 so that there are no complementarities between α and (μ_s, β_s). Thus, both positive and negative assortative matching are entirely consistent with our set-up.

To summarize, when we assume that (μ_s, β_s) are 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 also note at this point that when (μ_s, β_s) are known, then the surplus going to the employee does not vary with the supervisor type since firms extract the entire surplus for each employee. Thus, we expect workers to be indifferent to their supervisor assignment - a testable prediction in our data because we obtain direct evidence on work satisfaction through employee surveys and because we observe quit rates of employees.

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 concerns hidden effort e. Thus, information about (μ_s, β_s) is commonly shared 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, managerial ability μ_s is not known to anybody. However, employees observe μ_s after having been assigned to a supervisor and before choosing effort. As before, the optimal level of effort conditional on the piece rate b is (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 $\left(\mu_s^E, \beta_s^E, \alpha\right)$: $w\left(r; \mu_s^E, \beta_s^E, \alpha\right) = a(\mu_s^E, \beta_s^E, \alpha) + b(\mu_s^E, \beta_s^E, \alpha)r$.

We again 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, firms compete in the market for workers and supervisors so 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 (eq. eq. (11)): $b = \frac{\mu_s}{\mu_s+\psi\sigma_q^2}$. 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, the piece rate declines in $\theta_\mu\sigma_\mu^2$ which measures differences in managerial ability that are unobserved during the contracting stage. This is because after being assigned to a supervisor the worker observes μ_s and can then "game" the performance system by exerting more effort when μ_s is low and less when it is high. 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 does not depend on β_s^E since it only enters the workers certainty equivalent through the expected wage. And, as before, we competition in the labor market implies that differences in α_i are paid to workers. Thus expected employee compensation at the contracting stage is additively separable between α and a function that depends on μ_s^E only.

$$E\left[w_i|\alpha,\beta_s^E,\mu_s^E\right] = \alpha + h\left(\mu_s^E\right) \tag{16}$$

Again, we can't sign how expected employee compensation relates to μ_s^E. As before, expected output net of worker compensation does not vary with β_s^E and increases in μ_s^E. Thus, earnings of the supervisor are independent of β_s^E and increase 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. As researchers, 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.

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

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

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.

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: $\frac{\partial \mathbf{E}[\mathbf{w}\|\phi_\mathbf{s}]}{\partial \phi}$	0	$\neq 0^*$
	Piece rate: $\frac{\partial \mathbf{b}}{\partial \phi}$	0	> 0
	Productivity: $\frac{\partial \mathbf{E}[\mathbf{q}\|\phi_\mathbf{s}]}{\partial \phi}$	0	> 0
	Supervisor Wages: $\frac{\partial \mathbf{w}}{\partial \phi}$	0	> 0
Uninformed Firms ($\theta_\mu = \theta_\beta = 1$)	Wages: $\frac{\partial \mathbf{E}[\mathbf{w}\|\phi_\mathbf{s}]}{\partial \phi}$	> 0	> 0
	Piece rate: $\frac{\partial \mathbf{b}}{\partial \phi}$	0	0
	Productivity: $\frac{\partial \mathbf{E}[\mathbf{q}\|\phi_\mathbf{s}]}{\partial \phi}$	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.*

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 4 summarizes what these four different sets of assumptions imply for the compensation of workers and supervisors and the expected productivity of workers.

Table 4 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 table 4 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 Section outlined predictions from our interpretative model that allow distinguishing between heterogeneity across managers in ability (μ_s) and leniency (β_s) as well as how much the firm knows about this heterogeneity.

4.1 Empirical Methods

Central to our empirical analysis is identifying the heterogeneity in ratings behavior observed across supervisors and how it affects worker and supervisor outcomes. Here, we illustrate how we measure the relationship between ratings heterogeneity and one such component, worker earnings.

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 . For now, we also make the subscripts explicit.

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

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

The basic approach to obtain the second moments of $(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w)$ and $(\varepsilon_{it}^r, \varepsilon_{it}^w)$ 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 variation in the estimated fixed effects induced by estimation error and the fact that the estimation error across the different unobserved effects will not be orthogonal to each other.[23] We adapt the approach of Card, Heining, and Kline (2013) to a setting with a stacked system of equations.

To generate the varaince-covariance matrix of the unobserved effects in table 3, we followed a similar approach to Card, Heining, and Kline (2013), using the variance-covariance matrix of the estimates of the fixed effects to shrink the variance-covariance matrix. See appendix A.1 for more detail.To test the comparative statics generated by the model, we would like to estimate equations like (18), where α, ϕ, and $\epsilon^p$are defined as in equation 1 as the worker fixed effects, supervisor fixed effects and idiosyncratic error term in the ratings regression. We augment

$$outcome_{it} = b_0 + b_1\alpha_i + b_2\phi_{s(i,t)} + b_3\epsilon_{it}^p + \beta' X_{it} + \gamma' Y_{s(i,t)t} + \epsilon_{it}^o \tag{18}$$

As noted above, the problem with estimating the fixed effects in ratings is that the components will be measured with error that is likely correlated. This in turn biases estimates of $b_1, b_{2,}$ and b_3. .

Also, a much less computationally intensive method is to use a split-sample approach. We can estimate $\alpha' s$ and $\phi' s$ using half of our worker-year observations, then estimate a second set using the other half of the sample. For workers and supervisors that are observed in both samples, estimates of their fixed effects should be highly correlated. At the same time their measurement error will be uncorrelated. We can thus instrument for the fixed effects estimated on one half of the sample with those estimated on the other half. To maximize the overlap of workers and supervisors across the two samples, we split the sample into even and odd years. Because of the low turnover in our sample, we can use 95% of our worker-year observations, i.e., 95% of our observations are to workers and supervisors observed in both even and odd years.

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

[23]Because our panel is relatively short (11 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 $\left(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w\right)$.

Table 5: Log(Earnings) Components on Ratings Components

	OLS (1)	CHK correction (2)	IV (3)
Supervisor Ratings effect (φ)	0.103*** (0.014)	0.113*** (0.0027)	0.139*** (0.024)
Worker ratings effect (α)	0.095*** (0.003)	0.109*** (0.0015)	0.114*** (0.005)
Pass Residual (ε)	0.020*** (0.001)	0.025** (0.0014)	(na)

Notes: Column 1 presents the regression of log earnings on the supervisor effects and the wage effects in ratings. Column (2) presents coefficients based on the estimator in Card, Heining, and Kline (2013). Column (3) estimates supervisor and worker effects in even and odd years, separately, and uses estimates in even years as instruments for estimates in odd years and vice versa. All regressions control for time-varying worker and supervisor controls. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

4.2 Results

4.2.1 Employees' Earnings increase with ϕ_r

Table 5 shows results all three approaches for the outcome log(earnings) of the worker. We find that working for a high-rating supervisor is associated with substantially higher earnings. For example, in column 1, we find that moving from a supervisor who never passes subordinates to one who passes all of them increasing earnings by about 10%. A more reasonable move would be to a supervisor who rates one standard deviation (0.153) higher, which raises earnings by 1.6%. We find fairly similar effects across specification, though the unajusted estimates (column 1) are slightly smaller in magnitude, as we would expect if the error is interpreted as "measurement error". For example, using our split sample IV, we find that a one standard deviation higher rater increases earnings by 2.1%.

We also find that worker effects positively correlated with earnings. A one standard deviation higher α is associated with earnings increases of 2.5-3%. And, having an idiosyncratically high rating period also give workers a small earnings boost. Both results are reassuring that subjective performance ratings contain some information content.[24]

Appendix table A1 provides robustness checks for these results. Column 1 replicates the OLS specification from table 5 while column 2 restricts the sample to just the branches – where job functions are more uniform – and column 3 restricts to the sample of observations for which we have objective branch ratings – which will be important below for testing the model predictions. Even though the size of the coefficients varies across the regression samples, we find qualitatively consistent estimates. Finally, columns 4-6 separate earnings into

[24]We cannot identify the effect of the ϵ in ratings in our split sample IV strategy. Because of the incidental parameters problem, the ϵ will be correlated with worker and supervisor effects within either sample.

its components: log(wages), the probability of receiving a bonus, and log(bonus) conditional on receiving one. We find impacts for all components and our estimates suggest that nearly half the impact of ϕ on earnings operates through base wages, while the remaining majority operates through bonuses.

We therefore find strong evidence that having a high-rating supervisor has positive, significant effects on earnings. This result is inconsistent with the hypothesis that supervisor heterogeneity stems solely from leniency bias that the employer is perfectly informed about (see table 4). A perfectly informed firm would instead undo any impacts of supervisor leniency on wages. Instead, the result is consistent with either that are uninformed about leniency, or that supervisor heterogeneity stems primarily from supervisor ability and firms are informed about this.

4.2.2 Piece Rates Increase with ϕ_r

If supervisor heterogeneity is driven primarily by ability, rather than leniency, the model predicts that the piece rate component of pay will be larger for higher raters. This is because higher ability supervsiors lower the marginal cost of effort for the worker and linear piece rates should be declining in the cost of ramping up effort.

In Table 6 we augment equation 18 by allowing an interaction betwen ϕ_r and overall performance in a given period. This measures whether individual earnings are more responsive to performance measures when the supervisor is a higher rater. We indeed find that overall earnings and bonus, conditional on receiving one, are more strongly related to own performance for individuals assigned to higher raters. Though we do not find effects for the probability of receiving a bonus. Overall, this finding is consistent with the hypothesis that there is heterogeneity in managerial ability that the firm is informed about.

4.2.3 Branch Level Productivity Is Possitively Correlated with ϕ_r

We now consider how supervisor and worker heterogeneity correlate with objective performance measures. For the years 2007-2010, we have access to a correlate of objective productivity for a subset of the branches. This correlate is an annual ranking of branches within a set of peer branches along a number of key performance indicators (KPIs).

We regress these rankings, in various functional forms, on the average employee and supervisor ratings effects within the branch-year.[25] We can also correct for measurement error in the fixed effects by instrumenting for the branch averages obtained in even years with those obtained in odd years and vice versa.[26]

[25] If there is only one supervisor in a given branch-year, as is often the case, the average supervisor effect is fixed effect for that supervisor. In cases where there is more than one rater, the average supervisor fixed effect is obtained by averaging across supervisors, weighted by the number of subordinates each rated this period.

[26] We use the entire dataset in odd and even years, respectively, to estimate fixed effects from each sample, at the worker-year level. The first-stage of the IV is natrually estimated on the same sample and at the same level as the second-stage: branch-years

Table 6: Supervisor Heterogeneity and Pay-for-Performance

Dependent Variable	Log Earnings	Log Bonus	Pr Bonus
	(1)	(2)	(3)
Supervisor FE (ϕ)	0.059***	0.454***	0.171***
	(0.009)	(0.124)	(0.020)
Worker FE (α)	0.095***	0.518***	0.228***
	(0.003)	(0.028)	(0.007)
Pass Residual (ε)	0.021***	0.209***	0.096***
	(0.001)	(0.015)	(0.005)
ϕ*Pass	0.089***	0.458***	0.001
	(0.022)	(0.126)	(0.024)
Observations	77,077	23,864	77,077
Partial R-squared	0.819	0.630	0.333

Notes: OLS results. See table 5.

For precision, we weight by the number of employees in the branch-year.

Results are reported in table 7 for unadjusted (panel A) and for adjusted (panel B). In either case, we find that branch-years with higher raters perform significantly better. Using our adjusted estimates, we find that a branch with a one standard deviation higher ϕ has a 0.04 higher (7.7%) inverse rank score, is 4 percentage points (13%) more likely to be ranked among the top 5 branches in the peer group, and 5.6 percentage points (12%) more likely to be ranked in the top half. The unadjusted estimates are statistically sginficant for all variables but the probability of being in the top branch and the adjusted estimates are always larger in magnitude and usually statistically significant.

The model predicts that if leniency (β_s) drives 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) drives supervisor heterogeneity then we objective performance will be positively related to supervisor effects, regardless of whether the firm is informed or not. The results in table 7 support the latter view, even though the standard errors are fairly large. We are awaiting additional data on the performance of branches and hope that this will allow us to draw sharper conclusions.

4.2.4 Supervisor Pay Increases in Supervisor Heterogeneity

The fourth comparative static relates $\hat{\phi}_s$ to supervisor pay. We regress components of supervisor pay on their own ratings fixed effect, as well as the average worker fixed effect of their subordinates. We also adjust for sampling error using the split sample IV strategy.[27] Observations are weighted by the number of

with KPIs.

[27] As in the branch-year regressions, we obtain supervsior and worker fixed effects for the full odd- and even-year samples. We then instrument for supervisor effects and the average worker effect to a given supervisor in a given year using the estimates

Table 7: Branch KPI Performance on Ratings Components

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)
Panel A: Unadjusted				
Branch-Year-Level Average:				
Supervisor FE (φ)	0.164**	0.0361	0.190**	0.256**
	(0.0653)	(0.0515)	(0.0913)	(0.114)
Worker FE (α)	0.0138	0.0408	0.0859	0.0368
	(0.0523)	(0.0434)	(0.0756)	(0.108)
Panel B: Adjusted with IV				
Branch-Year-Level Average:				
Supervisor FE (φ)	0.266**	0.0362	0.263	0.365**
	(0.129)	(0.116)	(0.218)	(0.179)
Worker FE (α)	0.00370	0.0317	0.0623	0.0454
	(0.0797)	(0.0566)	(0.113)	(0.160)
Observations	781	781	781	781

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 branch's KPI ranking in that year divided by the number of branches it is ranked against. In Panel B, we estimate supervisor and worker fixed effects on odd and even years separately. We instrument for the branch-year averages in odd years with those obtained in even years and vice versa. instruments for branch-level average supervisor and worker effects with Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Supervisor Outcomes on Ratings Components

Dependent Variable:	(1) Log (Earnings)	(2) Pr(bonus)	(3) Log (Bonus)
Panel A: Unadjusted			
Supervisor FE (φ)	0.234***	0.120***	0.718***
	(0.0540)	(0.0238)	(0.169)
Average Worker FE (α)	0.219***	0.145***	0.705***
	(0.0342)	(0.0194)	(0.107)
Observations	8,436	8,436	4,982
Panel B: Adjusted with IV			
Supervisor FE (φ)	0.336***	0.170***	1.036***
	(0.0783)	(0.0353)	(0.252)
Average Worker FE (α)	0.246***	0.168***	0.742***
	(0.0445)	(0.0241)	(0.147)
Observations	8,131	8,131	4,820

Notes: Observations are at the supervisor-year level, weighted by number of subordinates with non-missing pay and performance variables. Outcomes are supervisor pay variables in the given year. In Panel B, we estimate supervisor and worker fixed effects on odd and even years separately. We instrument for the supervisor effect and supervisor-year-level average worker effects in odd years with those obtained in even years and vice versa. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

subordinates to a given supervisor at time t.

Results are reported in table 8. We find that supervisors earnings are strongly positively influenced by their ratings style (as well as the quality of the team they are supervising). This is true both when we look at log earnings and the probability of receiving a bonus and the size of the bonus conditional on receiving one. These results are thus again consistent with the hypothesis that heterogeneity in ratings behavior correlate with differences in managerial ability across supervisors, and that firms are informed about these differences. If instead ratings heterogeneity were driven by leniency, and firms were informed, we would next expect to see these differences show up in ratings.

4.3 Discussion

In section 2.2 we demonstrated that there is significant heterogeneity across supervisors in ratings style. One standard deviation in the supervisor heterogeneity in ratings style is associated with about a 15 percentage point increase in the pass rate. In this section, we showed that heterogeneity in ratings has real consequence for the earnings of individuals and the supervisors themselves, as well was for the firm, reflected in KPI rankings across branches. When evaluating this evidence in the context of our model, we conclude that heterogeneity in ratings at least partially reflects differences in managerial ability and that firms are at least

from the opposite subsample.

partially informed. [28]

5 Careers

Though a bit outside the scope of our model, we find it relevant to ask about the careers of workers inside the firm. Can a subordinate of a high-rater expect to see benefits in addition to a small pay increase? We first examine mobility patterns as a function of ratings components.

In table 9 columns (1) and (2) estimate equation (18) using the probability of quitting or being laid off as dependent variables, respectively. We find that separation rates are lower for those working for higher raters, regardless of whether these seperations are initiated by the firm or by the employee.

Columns (3) and (4) show that, conditional on staying at the firm in two adjacent years, a promotion across years is more likely to take place, and a demotion less likely, if the worker is supervised by a higher rater. In Column (5), we show that, conditional on staying in the job level, workers are more likely to stay with their supervisor if that supervisor is a high rater.

We thus find that workers matched to higher raters enjoy more stable jobs. They also have higher promotion probabilities, in addition to their higher earnings (table 5). We find this body of results as suggestive that workers actually gain rents from working with higher raters, rather than being kept at their participation constraint as in the principal-agent model presented above. The most direct evidence on this margin is shown in column (6). From a set of worker satisfaction surveys, administered by a third party, we have information about how employees feel about their immediate supervisors.[29] Taking the average response for this module, we find that subordinates tend to be more satisfied with their supervisors when their supervisors are higher raters.

All these findings are consistent and support the notion that individuals earn economic rents when working for higher raters. The firm does not fully extract all rents due to working for higher raters. This is either because the firm is not fully informed about the supervisor heterogeneity, or because it shares the surplus generated by the high-rating supervisors with the subordinates.

How large are the financial returns to working for a higher rater? We consider the following thought experiment: how does an increase in $\phi_{s(i,t)}$ affect the present discounted value of earnings keeping all other supervisor effects in all other periods constant? That is, we consider the effect of being assigned to a

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

[29]The questions are about (i) "The professional skills of the immediate supervisor" (ii) "The leadership skills of the immediate supervisor" (iii) "My immediate superior is energetic and effective" (iv) "My immediate superior gives constructive feedback on my work" (v) "My immediate superior delegates responsibility and authority so I can complete my work effectively" (vi) "My immediate superior helps me to develop personally and professionally" (vii) "What my immediate superior says is consistent with what he/she does". The answers are coded on a stanine scale and we construct an index of the satisfaction with the supervisor by summing the responses.

Table 9: Worker Outcomes and Ratings Components

Dependent Variables:	(1) Layoff	(2) Quit	(3) Promotion	(4) Demotion	(5) Same Supervisor	(6) Bottom Up Evaluation
Supervisor FE (ϕ)	-0.005***	-0.007*	0.044***	-0.006**	0.032	0.158***
	(0.002)	(0.004)	(0.009)	(0.002)	(0.027)	(0.048)
Worker FE (α)	-0.005***	-0.015***	0.095***	-0.018***	0.064***	0.165***
	(0.001)	(0.002)	(0.004)	(0.001)	(0.008)	(0.018)
Pass Residual (ε)	-0.001*	-0.003**	0.055***	-0.006***	0.004	0.067***
	(0.001)	(0.002)	(0.004)	(0.001)	(0.006)	(0.011)
Observations	77,077	77,077	74,602	74,602	65,415	67,832
R-squared	0.005	0.040	0.121	0.013	0.033	0.019

Notes: Columns 1-2 estimate the probability that the worker is laid off or quit in t+1 for all workers observed in t. Columns 3 and 4 estimate the probability of a promotion and demotion for workers observed in two adjacent years. Column 5 estimates the probability that the worker is with the same supervisor in the next year among those observed in adjacent years in the firm at the same level of the hierachy (i.e., not promoted or demoted). Column 6 reports the worker's self-reported satisfaction of their supervisor. All regressions include time-varying worker and supervisor controls. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

supervisor with a higher $\phi_{s(i,t)}$ without allowing for this change in the assignment to affect the generosity of any other supervisors over an employee's career.

Estimates of the financial returns based on a static regression such as those reported in Table 5 will deviate substantially from the effect we are trying to capture. Multiplying the regression coefficient 0.103 in a static regression with the standard deviation of ϕ_s from Table 3 (0.153) provides such a simple static estimate and it suggests an increase in log earnings of about 1.5% associated with a one-standard deviation increase in $\phi_{s(i,t)}$. However, this estimate is likely to deviate substantially from the full effect of an increase in $\phi_{s(i,t)}$ on the PDV of earnings holding the other supervisor effects constant. On one hand, it will underestimate the economic rents because it considers the impact on current year earnings only. On the other hand, the unobserved effects $\phi_{s(i,t)}$ will correlate over time since supervisor relationships do last for longer than just one period, potentially inducing an upward bias in the naive static estimate of 1.5%

A more sophisticated approach estimates a dynamic equation relating current wages to several lagged supervisor effects:

$$W\left(l, \phi^t, e_t\right) = \quad exp \quad \left(g_1\left(l_{it}\right) + h_1\left(X_{i,t}\right) + \Sigma_{\tau=0}^{k}\beta_\tau\phi_{s(i,t-\tau)} + \Sigma_{\tau=0}^{k}\theta_\tau\varepsilon_{i,t-\tau} + e_{i,t}\right) \tag{19}$$

By including multiple lags in $\phi_{s(i,t)}$, this specification accounts for the persistence in supervisorary relations. It also controls for multiple lags in the ratings residual $\varepsilon_{i,t}$ and for constant and time-varying controls

$X_{i,t}$. These controls include the individual ratings effect α_i^r as well as controls for age, experience, job tenure.

At this point, we face a choice of whether to include job level effects in eq. (19) or not. Without job level controls we might surrfer from omitted variable bias since supervisor types ϕ_s might vary systematically over the job hierachy. In Table 7, columns 1 and 2 we report the regression coefficients on $\phi_{s(i,t)}$ when we include (col. 1) and when we exclude (col. 2) job level controls from the basic static regression. The significant differences suggest indeed that omitting job levels results in substantial bias.

Including job levels however implies that relying on eq. (19) will omit the dynamic effects ϕ_s has on log earnings because it affects the probability of promotions and demotions. To capture this effect, we also estimate an equation predicting promotions:

$$P\left(Promoted_t | X_{i,t}, l_{i,t}, \phi^t, e_t\right) = g_2\left(l_{i,t}\right) + h_2\left(X_{i,t}\right) + \Sigma_{\tau=0}^{k}\gamma_\tau\phi_{s(i,t-\tau)} + \Sigma_{\tau=0}^{k}\xi_\tau\varepsilon_{i,t-\tau} \tag{20}$$

Here l_{it} is a vector of job level dummies, $X_{i,t}$ a set of controls. The remainder of eq. (20) allow for lags in $\phi_{s(i,t)}$ and $\varepsilon_{i,t}$. Table 7, col. 6 reports estimates of this equation including 4 lags in supervisor unobserved effects. These estimates suggest that having a more generous supervisor increases the probability of being promoted in the current period by 4.0%, but does not have a positive effect in subsequent periods. For that reason, we estimate equation 20 allowing only the current supervisor heterogeneity and unobserved ratings effects to affect the promotion probability.

We combine the estimates of equations (19) and (20) to estimate how $\phi_{s(i,t)}$ on average affects earnings in future periods. To simplify the analysis, we do not account for the possibility of multiple promotions and abstract from demotions. The latter are relatively rare overall, while we have found no long run effects of $\phi_{s(i,t)}$ on promotion probabilities, suggesting that omitting multiple promotions will have a relatively small effect on our calculations.

Table 10: Earnings Dynamics and Supervisor Heterogeneity

Dependent Variable	Log Earnings	Log Earnings	Log Earnings	Log Earnings	Promotions	Promotions
Supervisor FE (φ)	(1)	(2)	(3)	(4)	(5)	(6)
Contemporaneous φ	0.228***	0.103***	0.086***	0.032**	0.043***	0.040**
	(0.029)	(0.014)	(0.029)	(0.013)	(0.009)	(0.018)
Lag 1			0.050***	0.017*		0.020
			(0.018)	(0.009)		(0.022)
Lag 2			0.053***	0.015*		-0.012
			(0.017)	(0.009)		(0.023)
Lag 3			0.045***	0.017**		-0.018
			(0.016)	(0.009)		(0.020)
Lag 4			0.059***	0.023**		0.018
			(0.018)	(0.010)		(0.016)
Job Level FE		X		X	X	X
Observations	77,077	77,077	22,569	22,569	77,077	22,569
Partial R-squared	0.434	0.818	0.436	0.822	0.116	0.073

Notes: The table reports regressions of log earnings or promotion events on contemporanous supervisor effects (φ), worker unobserved effects (α), and residuals (ε) from the ratings equation (1). All regressions contain the same number of lags in (ε) as in (φ) and control for the same set of controls as in the basic specification reported in table 5B, with the exception that job level controls for either the worker or the supervisor are only included where indicated. Significance levels are represented using stars: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Thus, we calculate the % impact of $\phi_{s(i,,t)}$ on earnings at t+k on an individual at job-level $l_{i,t}$ using estimates of ϕ_k and $(g_1(l_{i,t}+1) - g_1(l_{i,t}))$ from eq. 19 and of γ_0 from eq. (20):

$$\frac{1}{W_{t+k}} \frac{dW_{t+k}}{d\phi_{s(i,t)}} = \phi_k + \gamma_0 \left(g_1 \left(l_{i,t} + 1 \right) - g_1 \left(l_{i,t} \right) \right) \tag{21}$$

For our estimate of the impact on the PDV of earnings, we average (21) across the distribution of job levels in our sample for k>0. We then generate the discounted PDV assuming an average remaining career of 20 years and a discount rate of 5%. The parameter estimates needed to perform this calculation are taken from Table 7, col. 4 and 5. We also require the distribution of individuals across job levels and the relationship between job levels and log earnings (the function $g_1(l)$) for these calculations. Due to confidentiality issues, we are unable to report $g_1(l)$ but we can report the average earnings differential across job levels in the firm, which is 14.8%.

We estimated (Table 4) that $\sigma_\phi = 0.15$. With this value, we have a PDV of 2.6% of average annual earnings associated with a one-period, one standard standard deviation increase in ϕ given the assumptions made above. The direct wage effect (holding the job level constant) amounts to 1.5% of this, while the return associated with being promoted to a higher job level accounts for 1.1%. These calculations assume that the direct wage effect reverts to 0 after 4 periods. Table 7 does not indicate any diminishing effect over time, so a reasonable alternative assumption would be to assume that an of 0.02 of ϕ_s persists after the fourth period. Under this assumption, the estimated overall return from a one-standard deviation in ϕ increases to 4.4% of annual earnings. If instead we assume that both the direct earnings effect and the promotion effect dissipate after 5 periods, then we obtain an economic value of σ_ϕ of 1.8%. Overall, we find these effects to be quite large.[30]

6 Conclusion

To come...

References

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

[30]Consider for instance that employee-supervisor relations on average last for 2.5 years. Consider then the value to an individual to be assigned a supervisor at the 95th as opposed to the 5th percentile of the earnings distribution. Under normality, this amounts to 4 standard deviations. The differential between these two assignments in the PDV of earrnings amounts on average to about 2.5*4*2.6%= 26% of an annual salary.

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

A Appendix

A.1 Adjusting Ratings Fixed Effects for Correlated Measurement Error

To obtain worker and supervisor fixed effects in ratings, we would like to estimate equation (1) shown above. However, fixed effects obtained from this regression will suffer from correlated measurement error. To adjust the variance-covariance matrix of these estimates, we follow a similar methodology to Card, Heining, and Kline (2013).

$$r_{it} \quad = \quad \alpha_i^r + \phi_s^r + \varepsilon_{it}^r \tag{22}$$

Equation (22) expresses ratings fixed effects. The unobserved persistent effects (α_i^r, ϕ_s^r) absorb individual and supervisor persistent differences in ratings and are allowed to correlate freely with each other. By construction, they are uncorrelated with the unobservable (ε_{it}^r). We assume (for now) that the unobservables are uncorrelated over time and across employees.

The basic approach is to estimate the fixed effects regression above and then use to obtain the second moments of (α_i^r, ϕ_s^r) and (ε_{it}^r) is to estimate fixed effect regressions on the system (**??**) and then use the estimated fixed effects and their correlation structure to determine the correlation structure in $(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w)$.

Given N worker-year observations, stacking equations for ratings and log wages yields the following system of $2 * N$ equations:

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

The (α_r, α_s) are N^*-vectors containing the individual effects, where 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. The N-by-N* design matrix D identifies the employees and the N-by-S design matrix F identifies the supervisors. $(r, w, \varepsilon_r.\varepsilon_w)$ are N-vectors.

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:

$$V_{\hat{\xi}} = \left(Z'Z\right)^{-1} Z'\Omega Z \left(Z'Z\right)^{-1} \tag{23}$$

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 assume that the unobservables are uncorrelated across time and 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.

. For an unbiased estimate $\hat{\xi}$ of ξ and any matrix A there is a simple expression for the expectation of the quadratic form $E\left[\hat{\xi}'A\hat{\xi}\right]$:

$$E\left[\hat{\xi}'A\hat{\xi}\right] = \xi'A\xi + tr\left(AV_{\hat{\xi}}\right) \tag{24}$$

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

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

where Q is the demeaning matrix[31], an idempotent matrix. Defining $A_{D\alpha_r} = \begin{bmatrix} D'QD & 0 \\ 0 & 0 \end{bmatrix}$ comformable with ξ, we get $\hat{\sigma}^2_{D\alpha_r} - \frac{1}{N^*-1}tr(A_{D\alpha_r} * V_{\hat{\xi}}) \rightarrow \sigma^2_{D\alpha_r}$. We proceed in the same manner for the other variances and covariances required.

The correction in equation 24 centers on $V_{\hat{\xi}}$, which of course depends directly on what we assume about the error structure Ω. So far we proceeded as if $(\varepsilon_r, \varepsilon_w)$ are uncorrelated across individuals and 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 patternr 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.[32] 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

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

[32] 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 $\hat{\xi}$.

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.

HERECentral to our empirical analysis is identifying the heterogeneity in ratings behavior observed across supervisors and how it affects worker and supervisor outcomes. Here, we illustrate how we measure the relaitonship between ratings heterogeneity and one such component, worker earnings.

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 . For now, we also make the subscripts explicit.

$$\begin{aligned} r_{it} &= \alpha_i^r + \phi_s^r + \varepsilon_{it}^r \\ w_{it} &= \alpha_i^w + \phi_s^w + \varepsilon_{it}^w \end{aligned} \tag{25}$$

The unobserved persistent effects $(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w)$ absorb individual and supervisor persistent differences in ratings and wages respectively and are allowed to correlate freely with each other. By construction, they are uncorrelated with the unobservables $(\varepsilon_{it}^r, \varepsilon_{it}^w)$. However, they are allowed to correlate with each other, as implied by the theory.[33] We assume (for now) that the unobservables $(\varepsilon_{it}^r, \varepsilon_{it}^w)$ are uncorrelated over time and across employees. Let Ω denote the 2-by-2 variance covariance matrix of $(\varepsilon_{it}^r, \varepsilon_{it}^w)$.

The basic approach to obtain the second moments of $(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w)$ and $(\varepsilon_{it}^r, \varepsilon_{it}^w)$is to estimate fixed effect regressions on the system (25) 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 variation in the estimated fixed effects induced by estimation error and the fact that the estimation error across the different unobserved effects will not be orthogonal to each other.[34] We adapt the approach of Card, Heining, and Kline (2013) to a setting with a stacked system of equations.

Given N worker-year observations, stacking equations for ratings and log wages yields the following system of $2 * N$ equations:

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

[33]Because they correlate with each other, we estimate the above eq. (25) jointly rather than separetely.

[34]Because our panel is relatively short (11 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 $\left(\alpha_i^r, \alpha_i^w, \phi_s^r, \phi_s^w\right)$.

The (α_r, α_s) are N^*-vectors containing the individual effects, where 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. The N-by-N* design matrix D identifies the employees and the N-by-S design matrix F identifies the supervisors. $(r, w, \varepsilon_r . \varepsilon_w)$ are N-vectors.

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:

$$V_{\hat{\xi}} = (Z'Z)^{-1} Z'\Omega Z (Z'Z)^{-1} \tag{26}$$

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 assume that the unobservables are uncorrelated across time and 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.

. For an unbiased estimate $\hat{\xi}$ of ξ and any matrix A there is a simple expression for the expectation of the quadratic form $E\left[\hat{\xi}'A\hat{\xi}\right]$:

$$E\left[\hat{\xi}'A\hat{\xi}\right] = \xi'A\xi + tr\left(AV_{\hat{\xi}}\right) \tag{27}$$

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

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

where Q is the demeaning matrix[35], an idempotent matrix. Defining $A_{D\alpha_r} = \begin{bmatrix} D'QD & 0 \\ 0 & 0 \end{bmatrix}$ conformable 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 proceed in the same manner for the other variances and covariances required.

The correction in equation 27 centers on $V_{\hat{\xi}}$, which of course depends directly on what we assume about the error structure Ω. So far we proceeded as if $(\varepsilon_r, \varepsilon_w)$ are uncorrelated across individuals and time. Fur-

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

Table A1: Log(Earnings) Components on Ratings Components – Robustness

Dependent Variable:	(1) All	(2) Log(Earnings) Branches	(3) KPI years	(4) Log(wages)	(5) Pr(bonus)	(6) Log(bonus)
phi	0.103***	0.035***	0.028***	0.046***	0.172***	0.770***
	(0.014)	(0.006)	(0.008)	(0.006)	(0.019)	(0.112)
alpha	0.095***	0.089***	0.077***	0.073***	0.228***	0.503***
	(0.003)	(0.002)	(0.004)	(0.002)	(0.007)	(0.027)
epsilon	0.020***	0.015***	0.011***	0.012***	0.096***	0.194***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.005)	(0.015)
Observations	77,077	34,145	7,840	77,077	77,077	23,864
R-squared	0.818	0.904	0.909	0.887	0.333	0.629

Notes: Column 1 presents the regression of log earnings on the supervisor effects and the wage effects in ratings. Columns 2 and 3 restrict the second stage regressions to observations in the branch system and observations with an associated KPI ranking, respectively. Columns 4-6 present results for three additional dependent variables, the log of base wages, whether or not the worker received the bonus in the given year, and the log size of the bonus conditional on receiving one. All regressions control for time-varying worker and supervisor controls. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

thermore, 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 patternr 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.[36] 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.

A.2 Additional Figures and Tables

[36]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 $\hat{\xi}$.

Figure A1: Flows Across Job Level

in % of employees in job levels respectively

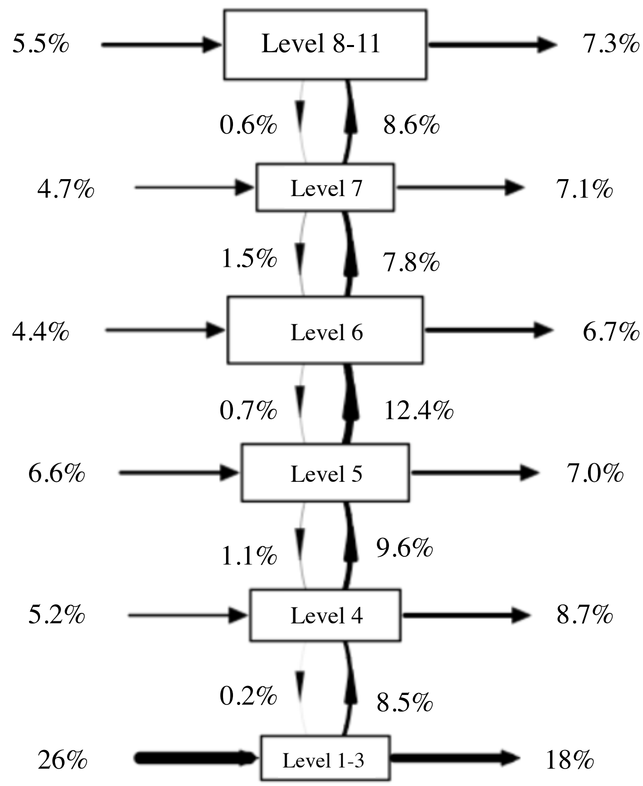


Table A2: Log(Earnings) Components on Ratings Components, Adjusted

Dependent Variables:	Log Wages (1)	Supervisor Wage Effect (φ) (2)	Individual Wage Effect (α) (3)	Wage residual (ε) (4)
Supervisor Ratings effect (φ)	0.113***	0.092***	0.021***	0
	(0.0027)	(0.0012)	(0.0023)	(na)
Worker ratings effect (α)	0.109***	-0.004	0.113***	0
	(0.0015)	(0.0007)	(0.0013)	(na)
Pass Residual (ε)	0.025**	0	0	0.025***
	(0.0014)	(na)	(na)	(0.00083)

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. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

Figure A2: Mean Performance of Supervisor Changers, by Supervisor Effect at Origin and Destination