NBER WORKING PAPER SERIES

SUPERVISORS AND PERFORMANCE MANAGEMENT SYSTEMS

Anders Frederiksen
Lisa B. Kahn
Fabian Lange

Working Paper 23351
http://www.nber.org/papers/w23351

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2017, Revised June 2019

We are grateful for helpful comments from seminar participants at the GAPE conference at Aarhus University, SOLE annual conference, Richmond Federal Reserve, University of Calgary, University of Edinburgh, Stockholm School of Economics, IZA Bonn, University of Tennessee, University of California, Riverside, MIT Sloan, University of Albany, Rensselaer Polytechnique, Syracuse University, Vanderbilt, Society of Labor Economics, University of Illinois, Queens, NBER Summer Institute, Zurich University, and Royal Holloway, London. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Supervisors and Performance Management Systems
Anders Frederiksen, Lisa B. Kahn, and Fabian Lange
NBER Working Paper No. 23351
April 2017, Revised June 2019
JEL No. J24,M5

ABSTRACT

We study how heterogeneity in performance evaluations across supervisors affects employee and supervisor careers and firm outcomes using data on the performance system of a Scandinavian service sector firm. We show that supervisors vary widely in how they rate subordinates of similar quality. To understand the nature of this heterogeneity, we propose a principal-agent model according to which supervisors can differ in their ability to elicit output from subordinates or in their taste for leniency when rating subordinates. The model also allows for variation in how informed firms are about this heterogeneity. Within the context of this model, we can discern the nature of the heterogeneity across supervisors and how informed firms are about this heterogeneity by relating estimated supervisor heterogeneity in ratings to worker, supervisor, and firm outcomes. We find that subordinates matched to a high-rating supervisor are paid significantly more and their pay is more closely aligned with performance. We also find that higher raters themselves are paid more and that the teams managed by higher raters perform better on objective performance measures. This evidence suggests that supervisor heterogeneity stems, at least in part, from real differences in managerial ability and that firms are at least partially informed about these differences. We conclude by quantifying how important heterogeneity in supervisor type is for workers' careers. For a typical worker, matching to a high rater (90th percentile) relative to a low rater (10th percentile) for just one year results in an increase in the present discounted value of earnings equivalent to 6-12 percent of an annual salary.

Anders Frederiksen
Aarhus University
Department of Business Development
and Technology
Birk Centerpark 15
7800 Herning
Denmark
afr@btech.au.dk

Lisa B. Kahn
Department of Economics University
of Rochester
280 Hutchison Rd
P.O. Box 270156
Rochester, NY 14627
and NBER
lisa.kahn@rochester.edu

Fabian Lange
Department of Economics
McGill University
855 Sherbrooke Street West
Montreal QC H3A, 2T7 and
NBER
fabian.lange@mcgill.ca

1 INTRODUCTION

Subjective performance evaluations are a ubiquitous and controversial feature of the modern workplace. Firms use these evaluations as indicators of worker performance and skills. They affect employee compensation, task assignment, promotions and retention (Frederiksen, Lange, and Kriechel, 2017). However, ratings are also affected by the identity of the rater: the worker's supervisor. For one, performance evaluations are inherently subjective, so supervisors might differ widely in how they rate equivalent behavior. Furthermore, supervisors have been shown to differ in their ability to manage subordinates, thus affecting how their subordinates perform on the job (Bertrand and Schoar, 2003; Lazear, Shaw, and Stanton, 2015). These differences in the ability to manage will plausibly influence the performance ratings subordinates receive. Little is known, however, about the extent and nature of ratings heterogeneity across supervisors, the degree to which firms are informed about such heterogeneity, and how it impacts workers' careers.

If supervisors give different ratings for the same underlying performance, then this will undermine the performance management system. It will burden workers with unneccesary risk and limit firms ability to use performance evaluations for setting incentives. As a consequence, firms may desire to counteract any heterogeneity with forced curves or other rules restricting the discretion supervisors have when rating subordinates. However, such policies may unintentionally interfere with how supervisors manage their teams if heterogeneity in ratings instead stems from real differences in a manager's ability to elicit output.

In this paper, we strive to estimate the magnitude and understand the nature of the heterogeneity in subjective ratings across supervisors. Using an exceptionally rich data set containing the performance management system of a Scandinavian service sector firm, we uncover substantial heterogeneity in ratings across supervisors: we estimate that a worker receives a 30 percent boost in ratings when assigned to a one-standard-deviation higher-rating supervisor. This heterogeneity is economically important: being assigned to a high rater (at the 90th percentile of the ratings distribution) for just one year is associated with an increase in the present discounted value of lifetime earnings at the firm equivalent to 6 to 12 percent of annual earnings, relative to being assigned to a low rater (10th percentile). The evidence strongly suggests that supervisors have important impacts on workers' careers inside the firm.

We develop a simple analytic framework to guide our empirical analysis. This framework allows supervisors to differ both in managerial ability and in their preferences for leniency when giving ratings. Furthermore, the model allows the degree to which firms are informed about differences across supervisors to vary. In the context of this model, we interpret how supervisor heterogeneity in ratings correlates with outcomes of subordinates, supervisors, and teams inside the firm. Both subordinates and supervisors earnings are higher and teams perform better on objective output metrics when the supervisors are high raters.

Because of these findings, we conclude that heterogeneity in ratings is driven primarily by real differences in managerial ability that firms are at least partially informed about.

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. Neither firms nor supervisors directly observe the effort workers exert. Supervisors observe worker output and report on this output to the firm. Supervisors differ along two dimensions. First, they vary in how much weight they place on reporting ratings truthfully as opposed to favorably. We refer to this heterogeneity as "leniency bias."[1] Second, they differ in their managerial ability, which affects their subordinates marginal costs of effort (or, equivalently, the output per unit effort). Given this setup, we consider the optimal linear compensation contracts of workers as well as salary contracts for supervisors. Our model also allows us to ask how the optimal contracts depend on how informed firms are about the differences between supervisors. This model yields comparative statics that we can take to the data to disentangle whether ratings heterogeneity is largely driven by leniency or ability and whether firms are largely informed or uninformed about such heterogeneity.

In our empirical analysis, we first estimate how much dispersion there is in ratings across supervisors using the observed dispersion in supervisor fixed effects from a regression of performance ratings on supervisor and worker fixed effects, as well as controls. This estimate adjusts for the well-known correlated measurement error problem inherent in double fixed effects models (see Andrews, Gill, Schank, and Upward (2008), Card, Heining, and Kline (2013), and Kline, Saggio, and Solvsten (2018)). We then estimate how rater heterogeneity correlates with outcomes of subordinates, supervisors, and teams. In this analysis, we use a variety of approaches to account for measurement error in the key explanatory variable, rater heterogeneity: we use a split-sample instrumental variables approach that is robust to misspecifying the contemporaneous error structure as well as estimates based on either the fixed effects directly or the bias correction discussed above.[2]

We find that subordinates of high raters are paid more than subordinates matched to low raters. This finding could be explained, in the context of our model, as being driven by heterogeneity in managerial

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

[2] Attemps to estimate the variation in unobserved heterogeneity in wages across workers and firms using fixed effects estimates commonly run into the problem that the latter are estimated with error (e.g., Abowd, Kramarz, and Margolis, 1999). In our analysis, we estimate the variation in worker and supervisor effects in ratings (the unobserved effects) using the heteroskedasticity robust estimator of Kline, Saggio, and Solvsten (2018). We then worker outcomes such as earnings onto the unobserved worker and supervisor effects. To correct for bias here, we adapt the Andrews et al. (2008) approach (which requires homoskedasticity) to a stacked system of equations with a double fixed effect structure. This allows us to estimate a bias-adjusted variance-covariance matrix of the unobserved effects in performance ratings and earnings, which we then use to compute regression coefficients.

ability, or by heterogeneity in leniency about which the firm is uninformed. However, we also find that teams managed by high raters tend to outperform those managed by low raters on a set of objective criteria available at the branch level; we find a similar effect when we examine individual financial performance measures available for a small subsample. These findings are consistent with heterogeneity in ability across managers - more capable managers lower the effort costs and increase the output of their team members. Two further findings corroborate this interpretation. First, higher raters earn more themselves, suggesting they are more valued by the firm. Second, pay of subordinates working for higher raters tends to be more closely aligned with their performance, as implied by our model if high raters are also better managers. Finally, workers matched to higher raters self-report that they are more satisfied with their immediate supervisors and we find suggestive evidence that they are less likely to change supervisors or quit the firm, though these latter results are less robust. This suggests workers benefit from being matched to a high rater, even though they also exert more effort.

Within the context of our model, our empirical findings have a consistent and clear interpretation: higher raters tend to be better managers and the firm has some but not perfect information on who the better managers are. That higher raters are better managers explains why their teams perform better on objective criteria. Furthermore, subordinates of better managers/higher raters tend to exert more effort, which explains why they are paid more. When the firm is at least partially informed about who the better managers are, they reward better managers with higher compensation. In addition, they optimally expose subordinates of better managers to stronger incentives since better managers reduce the marginal cost of effort of their subordinates. We also find evidence that suggests that employees earn economic rents from working for higher raters: their jobs are more stable and they report higher work satisfaction when working for high raters. This leads us to conclude that the firm is not fully informed about the ratings heterogeneity across supervisors, since it would otherwise extract these rents. Consistent with this notion, we find that rents attenuate for supervisors with higher tenure, about whom the firm presumably has more information.

We go beyond the confines of our static model to quantify how much rather heterogeneity affects the careers of workers. We allow differences across supervisors to dynamically impact worker pay, both directly and through promotion probabilities. We find that assignment to a higher rater has lasting positive effects on individual compensation. This arises because the effects on pay persist for some time and because being matched with a high rater increases the odds of a promotion. We thus conclude that better managers have large and real impacts on the careers of their subordinates: for a typical worker, matching to a high rater (90th percentile) relative to a low rater (10th percentile) for just one year results in an increase in the present discounted value of earnings equivalent to 6 to 12 percent of an annual salary.

Our work contributes to several important literatures in personnel and labor economics. The literature on productivity effects of managers predominantly studies upper management and CEOs (Bennedsen, Perrez-Gonzales, Wolfenzon 2007; Bertrand and Schoar, 2003; Kaplan, Klebanov, and Sorensen 2012). Ours is one of the few papers to explore productivity effects of supervisors lower in the firm hierarchy. Lazear, Shaw, and Stanton (2015) exploit the daily rotation of line managers to estimate how productivity of subordinates in a low-skilled service task (transactions per hour) varies across these managers.[3] Consistent with their work, we find that supervisors differ in their ability to elicit output from subordinates. In contrast to their setting, we study workers performing complex tasks for whom objective measures of performance are intrinsically hard to come by. To do so, we must estimate a model of behavior when information is imperfect. Our analysis exploits both objective and subjective measures of productivity, as well as worker and supervisor pay and career outcomes within the firm. We conclude that subjective evaluations and objective performance are closely related and that the firm is at least partially informed about the differences in productivity across supervisors. Our paper thus complements Lazear, Shaw, and Stanton (2015) in finding large productivity differences across supervisors in a very different setting than the simple service sector jobs they consider. We go beyond their analysis and provide an approach for understanding variation in manager behavior in a more typical setting where objective performance metrics are difficult to craft and firms instead rely on subjective ratings. Our analysis sheds light on the crucial role lower and middle managers play in the wide-spread and growing use of subjective ratings systems.

We contribute to a small literature in economics on the role and use of subjective performance measures by directly addressing whether the key subjective component of ratings, the supervisor effect, contains bias.[4] The question of bias in subjective evaluations has been taken up in an extensive literature in personnel psychology. This literature, however, rarely goes beyond documenting the presence of bias and tends to think of the firm as passive in the face of any reporting biases. Our approach is economic in the sense that we allow the firm to actively respond to the presence of bias in subjective ratings in designing its performance systems. Integrating the behavioral responses of the various actors improves our understanding of performance management inside the firm.

Even though we allow for bias related to supervisors in subjective evaluations, our approach emphasizes that subjective evaluations are informative about differences in skills across workers. This is important for an influential literature in labor and personnel economics that emphasizes the importance of employer

[3] Bloom and Van Reenen (2007) show substantial heterogeneity in management practices across firms. More recently, Hoffman and Tadelis (2018) find evidence that manager people skills are an important driver of subordinate retention.

[4] Concerns over how to interpret subjective performance ratings go back to Medoff and Abraham (1980, 1981), at least. An overview of this literature can be found in Frederiksen, Lange, and Kriechel (2018). They summarize empirical patterns in the data on subjective performance evaluations from six of the more prominent papers in this literature, including Baker, Gibbs, and Holmstrom (1994a, 1994b), Flabbi and Ichino (2001), Dohmen (2004), Gibbs and Hendriks (2004), Frederiksen and Takats (2011), and Frederiksen (2013). Theoretical papers on the topic include Tirole (1986), Milgrom (1988), Prendergast and Topel (1993, 1996), and MacLeod (2003).

learning in the labor market, but abstracts away from how this learning takes place.[5] Despite the presence of supervisor bias, firms can learn about worker productivity using subjective performance evaluations even when, as it is true in most modern workplaces, good objective measures of individual performance are not available.

Overall, our paper demonstrates that rater heterogeneity is an important feature of the employment relationship at this firm and has sizable impacts on the careers and outcomes of employees and supervisors, as well as for the firm itself. Rater heterogeneity cannot simply be interpreted as differential leniency bias. Instead, it is part and parcel of differential ability of managing and eliciting effort from subordinates. This finding is true in the firm we study and naturally may depend on the setting, but the concept that managerial heterogeneity in ratings should be taking seriously and can be diagnosed with observable data is novel and important. On a practical level, thisour findings suggest caution in addressing rater heterogeneity using practices such as forced scales or disincentivizing deviations from rating norms. Such practices might well interfere with the ability of supervisors to effectively manage their teams.[6]

The remainder of the paper proceeds as follows. Section 2 introduces the firm and the data at our disposal and presents new stylized facts on heterogeneity across managers in subjective performance ratings. In Section 3 we develop the model and show what it implies for how earnings and performance are related to rater heterogeneity. Section 4 contains the empirical analysis proposed by our model. Section 5 investigates the dynamic effects of supervisors on pay. Section 6 concludes.

2 FIRM AND DATA

2.1 Firm Overview

We rely on personnel data covering the domestic workforce of a large Scandinavian service sector firm between 2004 and 2014. The performance management system was introduced just prior to 2004, when our data begins. At any given time, the firm employed about 13,000 workers. This number fluctuates slightly over the years but does not exhibit a discernible trend. The workforce at this firm is highly educated and the firm is known to be an attractive employer. We briefly summarize the firm and data here, and provide more detail in Appendix A.

The firm consists of a central corporate office and an extensive branch network with several hundred

[5]See Altonji and Pierret (2001), Farber and Gibbons (1996), Gibbons and Waldman (1999, 2006), Lange (2007), and Kahn and Lange (2014).

[6]For theoretical work on the trade-off between rules and discretion inside firms, see Bolton and Dewatripont (2012), Dessein (2002), Alonso and Matouschek (2008), Aghion and Tirole (1997), and Li, Matouschek, and Powell (2017). On the empirical side, Hoffman, Kahn, and Li (2018) find that managers perform worse than an algorithm when choosing hires in a low-skilled service sector setting. This contrasts our own finding and suggests the value of managerial subjective assessments varies across settings.

branches.[7] The workforce is roughly equally split across the central office and the branch network. Tasks within the branch network are fairly uniform and involve close contact with clients, while workers in the central office serve a variety of functions. Branches vary in size but the median employee in the branch network works in a branch with 17 employees. Our analysis covers all employees of the firm, regardless of whether they work in a branch or in the central office.

There are 11 identifiable job levels. The typical branch has a branch manager (levels 9–11), a deputy branch manager (levels 8–9), 6–9 senior (level 6) and 5-7 junior (levels 3-5) workers in client-facing roles, and sometimes a trainee (level 1). In the central office, the distribution of jobs tends somewhat more towards higher level jobs. Churn is fairly low at this firm and there is some movement between the branch network and the central office.

Our data contains compensation measures, positions within the firm, and demographics. In addition, as part of the performance management system, each worker receives a rating from a supervisor. We observe these ratings, as well as a link to the supervisor responsible for the rating. The rating is meant to describe overall performance in a given year. In the branches, ratings are typically provided by the branch manager or by a deputy branch manager. In the central office, the titles are different, but ratings usually come from the immediate supervisor.

Compensation decisions at this firm are made roughly once a year – timed to follow the performance review period – and involve input from actors all along the hierarchy. Workers receive salaries and are eligible for bonuses. In any given year, only a fraction of employees will however receive a bonus (see Table 2.) The overall pay pool is set at the top of the firm. It is then broken down to divisions below, cascading down to managers at lower levels of the hierarchy. Typically a pay pool will be set for 10-15 employees either in a branch or subdivision of corporate. From there, managers have a fair amount of discretion to allocate both raises and bonuses from these set pay pools. The pay pool for a given branch or division is set based on historical patterns, financials, other performance indicators, union negotiations, and macroeconomic conditions. Lower level managers also provide input into the process of determining the pay pool for their units. For example, a manager might argue that unit A outperformed unit B and request a larger pool for the former. Managers also report upwards that individual workers or teams require larger raises/bonuses, for retention reasons, because they performed extraordinary tasks, or because a promotion cannot immediately be accommodated.

This compensation system might be described as a flexible hybrid between a top-down and bottom-up system. The top-down structure allows upper level management to keep control of the total wage bill. At the same time, lower level managers do report on the conditions and needs of their units and influence how

[7] Upon request of the firm, we can not disclose the exact number.

pay is broken down within the firm. The magnitude of their influence can vary across branches, divisions and job levels. Such a system strikes us as common across a wide range of firms.[8] Importantly, managers retain some ability to give one worker a bit more without necessarily taking away from another team member. The firm is aware that rigid constraints on compensation choices for small groups of workers would make cooperation in teamwork settings difficult.

2.2 Estimation Sample and Summary Statistics

Our estimation sample consists of 85,269 full-time worker-year observations. Appendix A.2 details our sample restrictions. The most important is the requirement that an observation has a performance rating. About a quarter of workers lack a performance measure, largely because the performance system took a few years to be fully rolled out. In 2004, 43 percent received performance ratings but the system spread rapidly; by 2008, 83 percent of the employees were covered. The coverage stayed at that level or slightly above throughout the remainder of the sample period (through 2014). Workers are also less likely to have a performance review during their first and last years at the firm, simply because they may not be present during the review period.[9]

This estimation sample serves to identify the supervisor effects central to our analysis. Within this sample, we have 77,682 observations with a compensation measure (compensation is not available in our final year of data, 2014), stemming from 14,214 unique workers. We also take advantage of information from worker satisfaction surveys, financial performance, and branch-level objective performance measures. As detailed below, availability of these variables is often limited to subsets of the data, which implies that the number of observations sometimes varies across empirical specifications.

The performance ratings range from 1 (unsatisfactory) to 5 (outstanding). The distribution of the performance score is shown in Table 1. As is common among companies using subjective evaluations (Frederiksen, Lange, and Kriechel, 2018), the ratings are concentrated in a small subset of the support: 91 percent of ratings are either a 3 or a 4.[10] For this reason, our empirical investigation is built around a "pass-fail"

[8] For example, in a university setting, the provost will set the pay pool for divisions, such as Arts and Sciences, and then deans will distribute this pay pool to academic departments, where finally a chair might have some discretion in allocating raises and bonuses within their department. In this process, there is feedback up the hierarchy as well. For example, members of the individual academic departments will have an easier time assessing performance of its professors, and determining who might need a retention raise or who should be brought up for a promotion case, etc. This feedback would naturally impact the overall compensation received by members of the department.

[9] There is some systematic variation in who receives ratings in that more stable workers (e.g., those with higher tenure and those outside of the lowest job levels) are more likely to be rated. However, after controlling for year effects, remaining observables such as tenure or job level have little power in predicting whether an individual will be missing a performance rating.

[10] The firm does not restrict the distribution of performance ratings a manager can give, but it does encourage supervisors to use the full scale, and holds training meetings every so often to help supervisors calibrate their ratings. The distribution of ratings at this firm is consistent with that observed in other firms we are familiar with (see Frederiksen, Lange, and Kriechel (2018)).

Table 1: Performance Distribution

	Fail			Pass	
Rating	1	2	3	4	5
Distribution	0.001	0.031	0.508	0.402	0.059
Sum	0.539			0.461	

Note: This table is based on the estimation sample consisting of 85,269 observations.

performance metric, which equals 1 if the rating is 4 or 5 and zero otherwise. This mapping allows us to interpret linear regression coefficients as marginal effects of the probability of receiving a "passing grade."[11]

Table 2 provides summary statistics for the estimation sample. We report earnings (and its components) relative to average per capita earnings in this country. Earnings average 185 percent of the national mean, consistent with this being a sought-after firm with skilled workers. Roughly 30 percent of the workers receive a bonus, and the bonus pool has historically been close to 6 percent of the wage pool.

Next, 83 percent of workers remain in the sample in the next year, and most of these also stay in the same business unit (defined as either the branch or the function in the central office).[12] Of those employees present in consecutive years, 11 percent are promoted and 1 percent are demoted annually. Finally, 1 percent of workers are laid off in the next two years.[13] Supervisor relationships are somewhat sticky; 65% of employees who work at the firm in consecutive years and are neither promoted nor demoted keep the same supervisor. Overall, about 50 percent of workers remain with the same supervisor from one year to the next.[14] We describe this mobility in more detail in Appendix A.3.

Our data contain two measures of objective performance. During 2007–2010, we have rankings of branches within peer-groups defined by the firm. The rankings are based on a set of Key Performance Indicators (KPIs) and 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. We call these branch rankings "KPI rankings" hereafter. As reported in Table 2, the average rating (ranking divided by number of branches in the peer group) was 0.53.

Our second measure of objective performance refers to financial performance of a subset of individual employees working in the branch network. We cannot reveal the precise content of these financial measures,

[11] We adopt this terminology for ease of exposition. Naturally, the firm does not report to workers that they have "failed" their review, and their interpretation may be more nuanced.

[12] About half of exits from the sample are due to quits or layoffs and half are workers who temporarily do not meet our criteria for having non-missing performance. Among the latter are many who are exiting the firm in the following business year just prior to receiving a performance rating.

[13] We take a two-year perspective because workers are less likely to receive a performance review in their last year at the firm. Worker exit rates are higher in the unrestricted sample, which includes workers who do not receive ratings. We believe this is because, as noted, more stable workers are somewhat more likely to receive ratings.

[14] None of these variables are defined in the last year of data, since they are right censored.

Table 2: Summary Statistics

	Estimation Sample		
	Mean	Std. Dev.	N
Outcomes:			
Pass	0.46	0.50	85,269
Earnings[1]	1.85	1.05	77,682
Received bonus	0.31	0.46	77,682
Bonuses (including zeros)[1]	0.10	0.69	77,682
Stay in Sample[2]	0.83	0.37	77,682
Stay in business unit[2]	0.72	0.45	77,682
Stay with supervisor[2]	0.53	0.50	77,682
Promotion[3]	0.11	0.31	75,197
Demotion[3]	0.01	0.11	75,197
Two-year layoff rate[2]	0.010	0.098	69,527
KPI Rating	0.53	0.28	7,871
Financial performance	-0.074	0.126	2,502
Bottom-Up Evaluation	4.72	1.00	74,993
Controls:			
Full-time	1.00	0.00	85,269
In Branches	0.44	0.50	85,269
Age	44.03	10.67	85,269
Tenure	17.98	13.29	85,269
Female	0.44	0.50	85,269
Supervisor Age	45.20	7.96	85,269
Supervisor tenure	19.60	11.57	85,269
Supervisor female	0.27	0.45	85,269

Note: The summary statistics are reported for the sample used to estimate the fixed effects in the ratings equation (see section 2). Not all variables are available for all observations in this sample. "Pass" is our constructed performance measure that equals 1 if the subjective performance evaluation was 4 or 5, and equals 0 if it was 1, 2, or 3. Stay in firm, in business unit, with supervisor, promotion and demotion refer to any change in the worker's status over the next year. Business unit is the branch or function in the central office. KPI rating is the branch-level ranking divided by the number of peer branches in the comparison set. Financial performance is the year-over-year growth rate of the individual's financial portfolio. Bottom-up evaluation is the average of seven questions workers answer regarding their satisfication with their supervisors. Responses range from 1 to 10; we average answers on all responses and norm the variable to have a standard deviation of 1. "In Branches" equals 1 if the worker was in the branch network and 0 if in the central corporate office.

1) Divided by average earnings in the country. Income variables not available in last year of data, 2014. 2) Restricted to not right-censored obs, excluding the last year of data (last two years for layoff rate). "Stay in sample" denotes the probability of being retained in the estimation sample in the following year. By far the most common reason for leaving the sample is to leave the firm within 2 years. 3) Restricted to not right-censored obs that did not quit or get laid off in respective year.

but one way to think about them is the following: Employees in client-facing roles administer a portfolio of clients and over the year these portfolios produce returns. We have information on these returns for the years 2014 and 2015. The measure we use is the year-over-year growth rate of the value of the portfolio. We refer to these measures as "financial performance" hereafter. In these years, the average growth rate was -0.07, though this is only for a subset of the firm's overall financial performance.

We also have access to job satisfaction surveys. These surveys include questions about the employees' perceptions of supervisor performance.[15] 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 and the maximum score is 10. Our outcome measure, hereafter "bottom-up evaluations", takes an average across these questions, normed to have a standard deviation of 1. The average of this measure is 4.7.[16]

In our analysis, we control for both worker and supervisor characteristics. Supervisors are on average only about one year older than the average employee (45.2 vs. 44 years), and have one and a half more years of tenure in the firm (19.6 vs. 18 years).

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

2.3 Variation in Performance Measures

We now demonstrate that supervisors differ substantially in how they rate their subordinates. In equation 1, the indicator variable p_{it} denotes whether individual i at time t receives a 4 or 5 on his or her performance evaluation. We relate this event to an individual effect α_i, a supervisor effect $\phi_{s(i,t)}$, as well as time-varying

[15] The employees are asked to respond to 7 items: 1) The professional skills of my immediate superior, 2) The leadership skills of my immediate superior, 3) My immediate superior is energetic and effective, 4) My immediate superior gives constructive feedback on my work, 5) My immediate superior delegates responsibility and authority so I can complete my work effectively, 6) My immediate superior helps me to develop personally and professionally, and 7) What my immediate superior says is consistent with what he/she does.

[16] It is unusual to have employee satisfaction data merged with personnel files (Frederiksen, 2017). Employers — including our firm — usually contract with outside consulting companies to conduct employee satisfaction surveys. This is done with the primary purpose of maintaining the employees' anonymity. By collecting the data at arm's length, the firms hope to induce truthful reporting by employees. The consulting firms then typically report to the firm the average employee satisfaction scores at the branch/unit/department level. As researchers we were able to obtain individual survey responses and merge them onto the personnel records. Hence, we know how a given employee evaluates his or her superior, even though the firm itself was not able to make this link. Supplements to surveys such as the National Longitudinal Survey of Youth (NLSY), the German Socio-Economic Panel (GSOEP), and the British Household Panel Survey (BHPS) sometimes do contain employee satisfaction data, but, naturally, such data is not linked to employer or supervisor data.

worker controls (X_{it}) and supervisor controls ($Y_{s(i,t),t}$):[17]

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

Estimating such a double fixed effects model requires sufficient variation generated by worker mobility across supervisors. In our data, workers frequently move between supervisors. In the unbalanced 2004–2014 panel, the average employee had 3.3 (s.d. of 1.5) different supervisors. Only 10% of those observed for at least 2 years are rated by the same supervisor during their entire time at the firm. Employees who were with the firm throughout all of 2004-2014 had on average 4.25 different supervisors. Similarly, supervisors manage many different employees over time, with some employees joining or leaving their teams almost every year. The average supervisor manages 10.38 (s.d. of 6.74) employees in a given year and 27 different employees over the full time period they are recorded as supervisors in our data. In fact, the workforce in this firm is so well connected that the largest connected set covers the entire firm. This firm is thus characterized by frequent moves between workers and supervisors, helpful for estimating the fixed effects that we require (see Appendix A.3).

Even though supervisor moves are common, the double fixed effects specifications requires an exogeneity assumption regarding this mobility. In particular, we are worried that sorting based on time-varying performance might bias our estimates of equation 1. Following Card, Heining, and Kline (2013), we present an event study of performance for workers who change supervisors (figure 1) to help evaluate whether non-random sorting is present in our data.

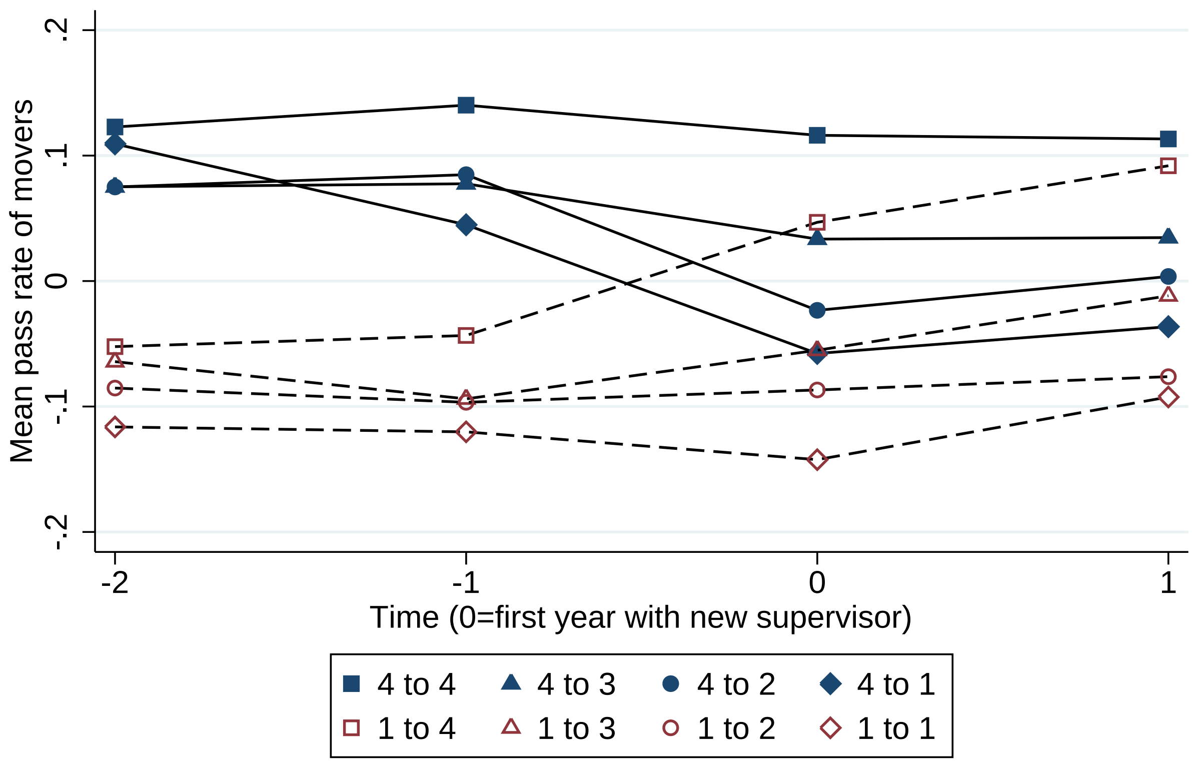
We split the set of supervisors into quartiles based on their average propensity to pass subordinates. We use pass probability of co-workers (the leave-out mean) to avoid selecting on a worker's own performance. We then plot average ratings of workers in the two years before and two years after they move across supervisors as a function of origin and destination supervisor category.[18] For simplicity, the figure focuses on people leaving quartile 4 supervisors (those with the highest likelihood of passing their subordinates) and quartile 1 supervisors (those with the lowest likelihood).

Figure 1 makes several important points. First, in the two years before moving, workers currently matched to a high rater have similar (high) performance, regardless of whether they are about to move to a different

[17]The controls include for the worker (X_{it}) indicators for five-year age and tenure groups, gender, and job level. For the supervisor ($Y_{s(i,t),t}$) the controls include indicators for five-year age groups, gender and job level. We do find some variation in average ratings across supervisor characteristics. In particular, older supervisors and supervisors that are higher up the job hierarchy tend to give higher ratings. We also control for an indicator for whether or not the worker is in a branch network, and year fixed effects. The latter help control for differences in usage of performance ratings as they become more common in the firm.

[18]There are about 1,500 to 2,000 worker-year observations in each mobility pair. The pass probabilities used to define types of moves and plotted in the figure have been residualized on the same controls specified above, except for the worker and supervisor fixed effects.

Figure 1: Mean Pass Rate of Supervisor Changers, by Pass Propensity of Co-Workers at Origin and Destination Supervisor



Notes: Figure shows mean pass rate of workers who change supervisors. Supervisors are classified into quartiles based on their propensity to pass co-workers (leave-out mean). Pass probabilities are residualized on our standard controls.

high rater or to a lower rater. Similarly, workers matched to a low rater have fairly similar (low) performance regardless of their destination. This lack of pre-trend in performance alleviates some of the concern about worker sorting. For instance, we do not observe that workers who are already on a decline move towards lower raters.

Second, transitioning across supervisor type has important consequences for performance. Workers moving across supervisor type experience large changes in performance, while workers who change supervisors within quartile experience little change in performance. That is, average ratings fan out for the different types of moves only after the move occurs.

Third, the effects of mobility on performance are symmetric across different types of moves and of roughly similar magnitude. A worker moving from a high to a low rater experiences a sizable drop in pass probability, while a worker moving from a low to a high rater experiences a sizable increase of comparable magnitude.

The three findings documented in Figure 1 suggest that the basic specification with additively separable worker and supervisor effects characterized the data well.

We would now like to estimate the variation in the unobserved effects $\{\alpha_i, \phi_s\}$ using the variation in the estimated fixed effects $\left\{\hat{\alpha}_i, \hat{\phi}_s\right\}$ from equation 1. However, we run into a well-known incidental parameters problem. The time dimension of the panel is fixed and relatively short (11 years at most) so that we have only a few observations to estimate each employee and supervisor fixed effect – the average supervisor rates in 3.7 years and the average worker is present in 7.3 years. These fixed effects are unbiased but inconsistent estimates of the unobserved effects. The variance of the fixed effects will therefore overstate the variance in the unobserved effects because it contains an estimation error. Also, the estimation error can be expected to correlate across worker and supervisor effects.[19]

Andrews, Gill, Schank, and Upward (2008) show how to address this problem by adjusting the variance-covariance matrix of the estimated fixed effects using the variance-covariance matrix of the estimation error for these same fixed effects. In double-fixed effect models, this adjustment will tend to reduce the size of the estimated variances compared to a naive estimator. The Andrews et al. (2008) approach however requires that the unobservables are homoskedastic, an assumption that is necessarily violated in our context since our outcomes are binary. We follow Kline, Saggio, and Solvsten (2018) to estimate the variation in the unobserved effects while allowing for heteroskedasticity.[20] Hereafter, we refer to the estimates obtained with this procedure as the "bias-adjusted estimates".

[19] This correlated estimation error will likely be negative. To see this, note that the model is saturated in worker and supervisor effects so the predicted value from the fixed effect regression necessarily goes through the sample mean for each worker and supervisor. If a worker effect is estimated with positive error, the supervisor effect will tend to be estimated with negative error to bring the predicted values for observations associated with that worker back through the mean.

[20] See also Card, Heining, and Kline (2013) and Gaure (2014). To arrive at our estimates we rely on the matlab code helpfully provided by Kline, Saggio, and Solvesten (2018) at https://github.com/rsaggio87/LeaveOutTwoWay.

Table 3, Panel A shows both unadjusted (column 1) and bias-adjusted (column 2) estimates of the second moments of $\alpha, \phi,$ and ϵ^p. The adjustment for sampling error has a fairly strong effect on the moments, reducing their magnitude by about a third.[21] Either way, we find that ϕ_s varies substantially across supervisors. Using the bias-adjusted moments in column 2, we find that the variance of ϕ_s is 0.019. This means that a one-standard-deviation increase in the supervisor ratings effect amounts to a 13.7 percentage point (30%) 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 35 percentage point increase in the probability of receiving a passing grade. The heterogeneity at the worker level is even larger — a standard deviation increase in α_i amounts to a 25.9 percentage point increase (56%) in the probability of receiving a passing grade.

We also find substantial idiosyncratic variation in ratings, holding constant these fixed effects and a rich set of time-varying controls. This residual variance is an input into the bias adjustment and must also be adjusted. We use a within-transformation of the error term, demeaning by team (worker-supervisor pairs) to obtain an unbiased and consistent estimate of this variance. This differences out unobserved effects that are not consistently estimated. In practice, this adjustment has only a small effect on our estimate of the variance in the idiosyncratic component of ratings.

Finally, we estimate the covariance between worker and supervisor effects to be quite small (-0.0093 using the bias adjustment). Thus any systematic worker sorting across supervisors based on fixed performance differences is likely small.[22] This is plausible given the high degree of churn between workers and supervisors described above, and the rich set of controls also included in the performance regression.

Panel B of Table 3 reports the variation in ratings explained by the different components in equation 1. We provide R-squares and F-statistics for: the controls alone, controls plus worker fixed effects, controls plus supervisor fixed effects, and the full model. We find that both sets of fixed effects explain a substantial amount of the variation in ratings. Given that we have significantly more worker effects than supervisor effects, the R-squared with the worker effects (column 2) is quite a bit larger than that with supervisor effects (column 3). However, the F-statistics show that per fixed effect the variation explained by each set is nearly identical. Furthermore, from column 4, we observe that the F-statistics on the combined restrictions is nearly identical to those in columns 2 and 3. This means that the variation explained by the two sets of fixed effects is close to orthogonal, which follows from the finding in Panel A that the estimated effects are

[21] The problem for which we need to adjust arises because of estimation error in the fixed effects. An ad-hoc, non-technical approach to this problem is to estimate the variance of the unobserved effects employing only fixed effects from workers and supervisors that we observe a minimum number of times. We implement this approach allowing the minimum cut-off on the number of observations to increase. As the cut-off increases, we find that the variances of the fixed effects converge towards those found using the Kline, Saggio, and Solvesten (2008) approach. This increases our confidence that the bias-adjustment succeeds in identifying the actual variation in unobserved effects across workers and supervisors in this firm.

[22] Naturally the covariances between the unobserved effects and the error term are 0.

Table 3: Variances of Ratings Components

Panel A: The Second Moments of the Ratings Components		
	(1) Unadjusted	(2) Bias Adjusted
Var(supervisor ratings effects) (ϕ)	0.029	0.019 (0.001)
Var(worker ratings effects) (α)	0.129	0.067 (0.001)
Var(pass residual) (ε)	0.113	0.178
Covariance(α,ϕ)	-0.014	-0.008 (0.001)
Sample size	85,269	84,690

Notes: See section 2.3. Column 1 reports unadjusted estimates from equation 1, a regression of receiving a performance rating equaling 4 or 5 on worker fixed effects (α), supervisor fixed effects (ϕ), and controls; ε are the residuals. Bias adjusted estimates (with standard errors in parentheses) correct those in column 1 based on leave-one-out estimates of variance of the unobservables ε following Kline, Saggio, and Solvsten (2018). The KSS procedure relies on 579 fewer observations since the leave out variances can only be estimated on a subset of the data. The controls include indicators for worker five-year age and tenure groups, gender, job level, and whether the worker is in the branch network, supervisor five-year age groups, gender, and job level, as well as year fixed effects.

Panel B: Explained Variation in Fixed Effects Regression				
	(1)	(2)	(3)	(4)
Specification:	Controls only	Controls + Worker Fixed Effects (α)	Controls + Supervisor Effects	Full Model
R-Square	0.144	0.500	0.234	0.544
F-Stat on controls	239.78	147.63	121.40	89.88
(r = number of restrictions)	(r = 60)	(r = 59)	(r = 58)	(r = 57)
F-stat on worker fixed effects		3.561 (r = 14,213)		3.264 (r = 14,213)
F-stat on supervisor fixed effects			3.556 (r = 2,729)	2.398 (r = 2,729)
F-stat on combined fixed effects				3.533 (r = 16,939)
Observations	85,269	85,269	85,269	85,269
Degrees of freedom	85,208	70,996	82,481	68,269

Notes: There are 60 control variables (some of which, like gender, drop out with the inclusion of fixed effects), 14,214 worker fixed effects and 2,730 supervisor fixed effects. Column 1 reports a regression of performance (pass-fail) on just the controls (listed in the note to panel A). The column 2 regression includes controls and worker fixed effects. The column 3 regression includes controls and supervisor fixed effects. The column 4 regression includes controls, worker effects and supervisor effects. We report the R-squared and F-statistics for subsets of included variables, as well as degrees of freedom associated with each regression. All F-statistics are significant at all conventional significance levels.

only weakly correlated. Overall, we find that both worker and supervisor effects have substantial explanatory power for ratings.

3 MODEL

In the previous section, we saw that there is substantial heterogeneity in ratings behavior of supervisors. In this section, we develop a model with testable predictions that allows us to distinguish between the sources of ratings heterogeneity and to determine how informed the firm is about such differences across supervisors. We allow for two types of heterogeneity. First, supervisors might differ in terms of their leniency bias: observing the same performance, some supervisors are simply more inclined to give high ratings than others. Second, supervisors may differ in managerial ability: some supervisors elicit higher performance from their workers. These two hypotheses have differential implications for worker, supervisor, and firm outcomes that can be tested using our detailed data on subjective and objective performance as well as data on worker and supervisor career outcomes. See Appendix B for more detail and proofs of the propositions.

3.1 Basic Setup

We consider a static model where the marginal product of an employee, i, who is not in a supervisory role (a "worker") is q_i. As expressed in equation 2, we assume that this marginal product ("output") depends on effort e_i, which is not directly observed by the supervisor or by the firm. Worker productivity also depends on the worker's productive type α_i and a random component ε_i^q. This component is normally distributed with mean 0 and variance σ_q^2 and is independent of (e_i, α_i). For simplicity, we assume that α_i is observed by all parties (workers, supervisors, the current firm, and prospective firms).[23]

$$q_i = e_i + \alpha_i + \varepsilon_i^q \tag{2}$$

The firm does not directly observe q_i; however, the supervisor assigned to worker i (denoted by the subscript s) does. Having observed q_i, supervisors report a rating r_i to the firm. Below we introduce two dimensions of heterogeneity across supervisors: (a) heterogeneity in supervisors' abilities (μ_s), which impacts the workers' costs of effort, and (b) heterogeneity in the supervisors' willingness to report generously on worker performance (β_s). From now on we suppress individual subscripts unless necessary for clarity. We retain the supervisor subscripts to indicate that a variable varies across supervisors.

[23]In this static setup, imperfect information about α_i is simply absorbed in the noise term ε_i^q. As long as the noise surrounding α_i is uncorrelated with the other elements of the model it has no implications for the derived results. For a dynamic setting with career concerns, see Gibbons and Murphy (1992).

The timing of the model is as follows:

1. Workers and the firm sign contracts that specify the wage function contingent on known supervisor (s) characteristics.
2. Workers match to supervisors, observe their supervisor's type, exert effort e, and produce q.
3. Supervisors observe q and provide ratings r.
4. Workers are paid according to their contracted wage function.

As is common in the literature, we assume that workers have Constant Absolute Risk Aversion (CARA) preferences with a coefficient of absolute risk aversion ψ, and that their preferences $U(.)$ are additively separable in wages and effort cost $c(e)$. Equation 3 shows the cost of effort function:

$$c(e) = -\frac{1}{2\mu_s}e^2 \tag{3}$$

Able supervisors reduce the marginal cost of effort and μ_s parameterizes this idea: better supervisors have higher μ_s. Workers take μ_s as given when they choose their effort level. All else equal, workers for better supervisors will exert more effort. We term μ_s "managerial ability."[24]

Supervisors have preferences for accuracy in reporting ($\tilde{\gamma_s}$) and they differ in terms of their preferences for leniency $\left(\tilde{\beta}_s\right)$, which leads to a trade-off between these conflicting goals:

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

Firms compete for supervisors in a competitive labor market. In expectation, any realized supervisor-employee match therefore needs to earn zero-profit. Thus, the compensation of supervisors w_s will be equal to the value of the expected output of the match net of the compensation going to the employee.

Maximizing supervisor utility with respect to r yields:

$$r = q + \tfrac{\tilde{\beta}_s}{\tilde{\gamma_s}} = q + \beta_s. \tag{5}$$

[24]The above formulation normalizes the marginal product of effort in equation (2) to one and allows the marginal costs of effort in equation (3) to vary across supervisors. An observationally equivalent formulation would normalize the marginal cost of effort and allow for variation in the marginal product of effort across supervisors. What is important is only the ratio of the marginal product to the marginal cost of effort so that it is irrelevant whether we allow for heterogeneity across supervisors in eqs. (2) or (3).

Hence, supervisors report observed output q plus a supervisor-specific parameter $\beta_s = \frac{\tilde{\beta}_s}{\tilde{\gamma}_s}$ which we refer to as "leniency bias" as it measures the strength of the motive to report favorably relative to the motive to report truthfully.

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

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

The parameter ϕ_s summarizes how ratings vary with the supervisor. As discussed above, this variation can arise either because supervisors differ in their managerial ability μ_s or because they differ in their leniency β_s.[25]

We now consider contracts that specify all payoff-relevant aspects of the employment relationship, including the assignment (μ_s, β_s) and the mapping of observed ratings to wages. At the contracting stage, agents (workers, supervisors, and the firm) share information about supervisor types, though this information may be imperfect.[26] We discuss the empirical implications of variation in worker ability α, supervisor leniency β_s, and managerial ability μ_s using two propositions. The first presents results for the case where agents are perfectly informed about (μ_s, β_s) and the second for the case when agents are only imperfectly informed.[27]

As is common in the literature, we restrict attention to wage contracts that are linear in the ratings. Thus, we consider contracts of the form $w_i = a_{is} + b_{is} r_i$.[28] The parameters (a_{is}, b_{is}) of these wage contracts are allowed to vary with information on worker and supervisor types available at the contracting stage. The term b_{is} represents all components of pay that covary with contemporaneous performance. We hereafter refer to b_{is} as the piece-rate, following a common practice in the literature on linear pay-for-performance schemes.

We assume that the firm competes for workers and supervisors in a perfectly competitive market so that outside options equal expected productivity and compensation is set to make agents indifferent across firms. We assume subordinate ratings do not directly enter into supervisor pay.

[25]Eq. 6 retains the individual index i to emphasize the connection to the double-fixed effect specification estimated above. We also retain the index in the discussion of the wage contract that follows to be clear about how individual variation across types α_i affect contracting.

[26]Regarding the assignment of workers to supervisors, we note that 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, in equilibrium any assignment is viable and both positive and negative assortative matching are entirely consistent with our set-up.

[27]While we allow for imperfect information about supervisor type, we assume this information is common to all market participants so that supervisors are paid their expected marginal product. This deviates from an important literature on asymmetric learning whereby the incumbent firm retains an information advantage over competing firms (Greenwald (1986), Gibbons and Katz (1991), Acemoglu and Pischke (1998), Schonberg (2007), Pinkston (2009), Kahn (2013), Waldman (1984)). However, in these models, worker pay is still correlated with their ability, so we believe our assumption does not affect the qualitative implications of the model.

[28]In a closely related setting with normal signals and with preferences of the type provided, Holmstrom and Milgrom (1987) find that the optimal contract does take the linear form.

3.2 The Informed Firm

We begin by assuming that firms (both the current employer and competing firms), supervisors, and workers are perfectly informed about (μ_s, β_s). The firm offers workers 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 $(\mu_s, \beta_s, \alpha_i)$. Thus, wage contracts are:

$$w = a\left(\mu_s, \beta_s, \alpha_i\right) + b\left(\mu_s, \beta_s, \alpha_i\right) r$$

Proposition 1 states properties of the wage contract and how expected compensation of employees and supervisors vary with $(\mu_s, \beta_s, \alpha_i)$.

Proposition 1. *Under perfect information about supervisor and worker types $(\mu_s, \beta_s, \alpha_i)$:*

1. The optimal piece rate is given by $b_s^* = \frac{\mu_s}{\mu_s + \psi\sigma_q^2}$;

2. Expected output increases one-for-one with α_i, does not vary with β_s, and increases with μ_s;

3. Expected compensation of workers increases one-for-one with α_i and does not vary with β_s. It increases with μ_s iff $b < \frac{1}{2}$.

4. Expected compensation of supervisors does not vary with α_i or β_s, and increases with μ_s;

5. Workers do not earn economic rents; that is, worker surplus $S = U(w - c(e)) = 0$.

The optimal piece rate is familiar to students of principal agent models. Greater uncertainty σ_q^2 or risk aversion ψ lowers the piece rate as the firm shields the employee from risk. On the other hand, if the marginal cost of effort declines (μ_s increases), then the piece rate increases as the trade-off between effort provision and risk improves.

Expected effort and output thus increase in μ_s because effort becomes less costly on the margin and because the piece rate increases and thus induces higher effort. Furthermore, the surplus from a worker-supervisor match increases in μ_s because, holding effort constant, the cost of effort declines in μ_s. Since firms compete for supervisors, supervisor compensation must also increase in μ_s. By definition output, q, increases one-for-one with worker ability, α, and, since firms compete for workers, so does worker compensation.

The finding that may be least intuitive is the last part of point 3, which establishes that there is no global relationship between worker compensation and marginal cost of effort μ_s. Two countervailing effects bear on expected compensation when μ_s increases. On one hand, the cost of providing any given effort level declines in μ_s. This will lower compensation, since firms will use the intercept of the wage equation to extract all surplus from employees. On the other hand, the optimal piece rate increases in μ_s and so does the risk borne by workers. Thus, compensation will have to increase on average to induce workers to bear this risk. When incentives are low-powered ($b < \frac{1}{2}$), then little effort is provided and consideration of risk dominates that of effort cost and total pay increases in μ_s. The opposite is true when incentives are high-powered ($b > \frac{1}{2}$) and workers exert a lot of effort. In that case, better managers reduce the effort cost born by workers significantly and wages decline with μ_s.

Proposition 1 states that neither output nor compensation vary with β_s when firms are perfectly informed. The intuition is straightforward. Optimal risk sharing induces the firm to remove any source of variation from employee compensation unless it can be used to incentivize effort. Since β_s does not enter into the effort cost function and does not correlate with the signal noise, the firm will neutralize any variation in β_s when setting employee compensation. This also implies that effort choice and expected output are independent of β_s and so the surplus obtained from a given employee does not vary with β_s. Therefore supervisor compensation does not vary with β_s either.

We also note that when (μ_s, β_s) are known, the surplus going to the employee does not vary with the supervisor type since, as we have just noted, worker pay does not vary with β_s and the firm sets pay as a function of μ_s to extract the entire surplus for each employee (point 5). Thus, we expect workers to be indifferent to their supervisor assignment.

3.3 The Partially Informed Firm

We now consider the situation when agents are imperfectly informed. To begin, assume that (μ_s, β_s) are independent normally distributed random variables with variances σ_β^2 and σ_μ^2. To capture the idea that agents are imperfectly informed we assume that firms (both the current employer and competing firms) and employees hold beliefs $\left(\beta_s^E, \mu_s^E\right)$ about the supervisor characteristics such that

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

where the errors $(\varepsilon_\beta, \varepsilon_\mu)$ follow a normal distribution and are independent of each other.[29] We parameterize the share of total variation in β and μ unknown to agents as θ_β and θ_μ so that

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

[29] The normality assumptions ensure that the exponential in the utility function is normally distributed both before and after the contracting stage, and we can thus use standard techniques to solve the worker's problem.

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 $\left(\mu_s^E, \beta_s^E, \alpha\right)$:

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

However, we also assume that 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 thus: $e^* = b\mu_s$.

Proposition 2 now establishes properties of the wage contract and expected compensation when information about types is imperfect. We distinguish in this proposition between the effects of variation across supervisors that is known to firms $\left(\beta_s^E, \mu_s^E\right)$ and variation in (β_s, μ_s) that is partially unknown to the firm.

Proposition 2. *Under imperfect information about supervisor type (μ_s, β_s):*

1. *The optimal piece rate is the unique implicit solution to $\mu_s^E = b_s \left(\mu_s^E + \psi\left(\theta_\beta \sigma_\beta^2 + \sigma_q^2 + b_s^2 \frac{\theta_\mu \sigma_\mu^2}{2}\right)\right)$;*

2. *Expected output conditional on $\left(\mu_s^E, \beta_s^E, \alpha\right)$ does not vary with β_s^E and increases with μ_s^E. Expected output conditional on (μ_s, β_s, α) does not vary with β_s and increases with μ_s. Both increase one-for-one in α_i;*

3. *Expected compensation of workers conditional on $\left(\mu_s^E, \beta_s^E, \alpha\right)$ does not vary with β_s^E . The relationship with μ_s^E cannot be globally signed. Expected compensation of workers conditional on (μ_s, β_s, α) increases with β_s. Its relationship with μ_s also cannot be globally signed. Both increase one-for-one with α_i;*

4. *Expected compensation of supervisors conditional on $\left(\mu_s^E, \beta_s^E, \alpha\right)$ does not vary with α or β_s^E but increases with μ_s^E. Expected compensation of supervisors conditional on (μ_s, β_s, α) does not vary with α or β_s but increases with μ_s.*

5. *Worker surplus $S = U(w - c(e))$ does not vary with μ_s^E and β_s^E but increases in μ_s and β_s.*

The intuition for how outcomes vary with $\left(\mu_s^E, \beta_s^E, \alpha\right)$ is directly analogous to the variation in outcomes with (μ_s, β_s, α) when there is full information.

It is instructive to compare the piece rates under partial and full information. Besides replacing μ_s with μ_s^E, there are two differences. First, the signal becomes less informative, and thus the optimal loading declines, as the share of the variation in β_s that is unknown to the firm $(\theta_\beta \sigma_\beta^2)$ increases. Second, the piece rate declines in the share of variation in managerial ability that is unknown during the contracting stage $(\theta_\mu \sigma_\mu^2)$. This is because after the contract is entered into and workers are assigned to supervisors, workers observe the actual effort cost μ_s. At that point, they can "game" the performance system by exerting more effort when μ_s is high and less when it is low. Therefore, the usefulness of setting incentives using performance signals declines in $\theta_\mu \sigma_\mu^2$ and so does the optimal loading.

Table 4: Model Predictions

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

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

A second notable difference is that the firm is only able to neutralize the variation in β_s that it is informed about. Thus, it will absorb any variation in β_s^E when compensating workers to remove any risk that is not of use in setting incentives. However, expected compensation will increase with β_s. And, workers earn rents that are increasing in both β_s and μ_s.

Finally, expected output does of course still increase in μ_s but not in β_s. Workers observe a lower cost of effort, even when the firm only imperfectly observes this, and work harder.

3.4 Leniency Bias vs. Managerial Ability — Perfect vs. Imperfect Information?

Our primary goal is to identify the source of heterogeneity in supervisor ratings, ϕ_s, and whether or not firms are informed about such heterogeneity. From the ratings equation (6), above, it follows immediately that panel data on performance ratings alone does not allow to separate heterogeneity in managerial ability and leniency bias. However, propositions 1 and 2 provide diverging predictions for how output and compensation vary with β_s and μ_s for both fully informed and imperfectly informed firms, respectively. These allow us to identify the sources of heterogeneity and whether the firm is informed or not.

It is useful to consider extreme cases to build intuition about how the fundamentals of the model map into the data on ratings, compensation, and output. In particular, we contrast firms that are perfectly informed ($\theta_\beta = \theta_\mu = 0$) with firms that are completely ignorant ($\theta_\beta = \theta_\mu = 1$). We also distinguish the case when supervisors differ solely in how lenient they are $\left(\sigma_\beta^2 > 0,\ \sigma_\mu^2 = 0\right)$ from the case when supervisors differ solely in their ability to elicit effort from their team members $\left(\sigma_\beta^2 = 0, \sigma_\mu^2 > 0\right)$. Table 4 summarizes these four cases and what they imply for the relationships between supervisor heterogeneity in ratings, ϕ_s, and compensation and productivity.

Table 4 reveals that the data indeed allows us to differentiate between the four cases.

To start, we observe that if the firm is informed, then heterogeneity in leniency does not covary with any of the outcomes we consider (top left quadrant). Compensation contracts in this case are structured to simply undo the heterogeneity in leniency for both workers and supervisors. Effort is not directly affected by leniency, and therefore productivity and supervisor earnings are likewise unaffected. We also note that variation in leniency that the firm is uninformed about (bottom left quadrant) does not affect the incentive component of the contract. It therefore does not lead to variation in productivity or supervisor earnings even if the firm is uninformed of it. However, if the firm does not know who the lenient managers are, then assignment to a lenient manager entails rents to the worker and higher earnings.

By contrast, variation in managerial ability generally leads to increases in productivity. Should the firm know about the variation in managerial ability across supervisors (top right quadrant), then supervisor wages and employee piece rates will both increase in the ability of the supervisor, but the worker will not be able to earn any surplus from this variation in ability. If instead, the firm is uninformed about who the good supervisors are (bottom right quadrant), then workers earn higher wages and receive rents from working for better supervisors.

Combined, these differing predictions allow us to test the model and we turn to this task now.

4 TESTING THE MODEL

The model's predictions contingent on the nature of supervisor heterogeneity (ability and leniency) and the information structure are listed in Table 4. In this section, we empirically evaluate these predictions using detailed personnel data. Specifically, we explore the relationship between supervisor ratings heterogeneity (estimated in section 2.3) and wages, piece rate strength, objective output (as measured by KPI rankings and financial performance), worker surplus (as measured by job stability and bottom-up evaluations), as well as supervisor pay and mobility outcomes.

4.1 Wages

A key comparative static from Table 4 is the relationship between supervisor ratings heterogeneity (ϕ_s) and worker wages. We evaluate this relationship using the following model:

$$log(w_{it}) = \beta_0 + \beta_1 \phi_{s(i,t)} + \beta_2 \alpha_i + \beta_3 \epsilon^p_{it} + \beta' X_{it} + \gamma' Y_{s(i,t)t} + \nu_{it} \tag{7}$$

where the dependent variable $log(w_{it})$ is log earnings for a worker i in year t. The unobserved supervisor effects in performance are captured by ϕ_s, worker effects in performance are denoted α_i, and the idiosyncratic performance shock is denoted ϵ_{it}^{p}. We also include the rich set of controls for supervisor and worker characteristics $\left(X_{it}, Y_{s(i,t)t}\right)$ applied when estimating equation 1. These absorb systematic variation in performance and pay that is outside the scope of the model (for example, job function). Finally, we assume that the error term, ν_{it}, is uncorrelated with the variables preceding it.

We use three strategies to estimate the parameters $(\beta_1, \beta_2, \beta_3)$ in equation 7 and report these in Table 5. We first present results using a naive strategy: simply regress log earnings on the fixed effects $\left(\hat{\phi}_s, \hat{\alpha}_i, \hat{\epsilon_{it}}\right)$ obtained from the fixed effects specification of equation 1 in Section 2.3. We cluster standard errors by supervisor, the level of variation underlying our main dependent variable. Results are summarized in columns 1-3 of Table 5, labelled "OLS".

In column (1) we find a sizable and statistically significant relationship between ϕ_s and log earnings. Our estimate implies that moving from a supervisor who never passes subordinates to one who passes all of them increases earnings by about 10 percent. In Section 2.3, we found the bias-adjusted standard deviation of ϕ_s to be 0.137. Thus, a move from a 10th percentile rater to one at the 90th percentile of ϕ_s is associated with an increase in earnings of about 3.3 percent. We also find that worker effects correlate positively with earnings. A one-standard-deviation higher α is associated with earnings increases of 2.6 percent. Finally, having an idiosyncratically high rating in the current period (ϵ) gives workers a positive but modest earnings boost.

In column (2) we add business unit fixed effects as additional controls.[30] One may be worried that differences across units due to, say, size or client base give rise to a positive correlation between ratings and earnings. We do not have enough mobility of supervisors across business units to separately identify unit fixed effects from supervisor fixed effects in ratings. However, including business unit fixed effects in the earnings regression takes the conservative approach of assigning any common component in ratings to the unit itself, and not, say, to unusually better or worse idiosyncratic supervisors or workers. The coefficient on ϕ_s falls by about half with the inclusion of this control, but remains significant at the one percent level. This specification also helps rule out a reverse causality story: that business units that tend to receive large salary pools must give out high ratings to rationalize spending of the salary pool. The fact that our results hold within branch fixed effects suggests this is not the case.[31]

Up to this point, we have leveraged worker switches across supervisors to estimate supervisor heterogeneity in performance ratings (section 2.3), but we are not explicitly using this variation to understand the impact of such heterogeneity on worker earnings. The worker fixed effects specification in column (3)

[30] We include a separate fixed effect for each branch as well as each function in the central corporate office.

[31] Though not shown, we also find similar results when we control for branch-by-year fixed effects, estimated off of location-years with at least two supervisors. This helps alleviate concerns that idiosyncratic productivity shocks at the unit or location level drive our results.

identifies the coefficient on ϕ explicitly from workers who switch supervisors. Here we find the point estimate is quite a bit attenuated, to 0.025, but still significant at the one percent level. This attenuation is partly because we must identify over 14,000 worker fixed effects, which absorbs quite a lot of the variation in log earnings. But also, effects in columns 1 and 2 might indeed be at least in part driven by sorting as opposed to causal effects of supervisors. However, even here, a move from a 10th to 90th percentile rater is associated with an earnings increase of about 0.9 percent. That is non-trivial and, as we discuss in section 5, can compound over time.[32]

Another way to see the effect of switching supervisors is with an event-style analysis, summarized in figure 2. Here we regress log(earnings) on the change in ϕ associated with the supervisor switch interacted with indicators for event time before or after the switch. The omitted category is the year before the supervisor switch. Regressions also control for worker fixed effects and typical controls. We find that the impact of a change in ϕ becomes apparent only after the worker actually makes the switch and impacts earnings at about the magnitude of column 3 in table 5.

The naive estimator in columns 1-3 has the virtue of being extremely transparent, but unfortunately is biased for the reasons discussed before: the worker and supervisor fixed effects in ratings are contaminated by correlated measurement error. Our second strategy is therefore to pursue an instrumental variables approach. We split the sample into two separate periods and obtain two distinct sets of estimates for $\alpha's$ and $\phi's$, one from each subsample. These two sets of fixed effects will be highly correlated because they are estimates of the same underlying unobserved effects. At the same time, the estimation errors across the two sets of estimates are uncorrelated. We can thus correct for the incidental parameter problem by instrumenting fixed effects estimated from one subsample with the fixed effects from the other subsample, and vice versa.

Our preferred way of splitting the sample is by even and odd years because it maximizes the overlap of workers and supervisors across the two samples.[33] Because of the low turnover in our sample, we retain almost all observations (74,641 out of 77,692) when requiring this overlap. We hereafter term this the split-sample IV estimation.[34] Results are reported in columns 4-5 of Table 5. This approach allows us to estimate β_1 and β_2. However, we cannot estimate β_3, the coefficient on the ratings residual. The reason is that the error term from one subsample is uncorrelated with the unobserved effects (α and ϕ) in that subsample as

[32]In Appendix A.3, we present estimates using only individuals after they have had a supervisor switch and restrict even further to specific supervisor moves that are plausibly more exogenous. Rather than leveraging the change in outcomes as a function of the change in ϕ, as in the worker fixed effects specifications, this strategy explores the effect of a plausibly exogenously allocation of ϕ on worker outcomes. Our results (see Table A4) are consistent across such moves. Most notably, we find very similar results when we restrict ourselves to individuals assigned to a new supervisor whose supervisor in the previous period left the firm.

[33]We have experimented with splitting the sample in other ways — for instance, into an early and late period (pre- and post-2009). The results are fully consistent with those reported here but typically the overlap in the samples is much smaller and the estimates are therefore noisier.

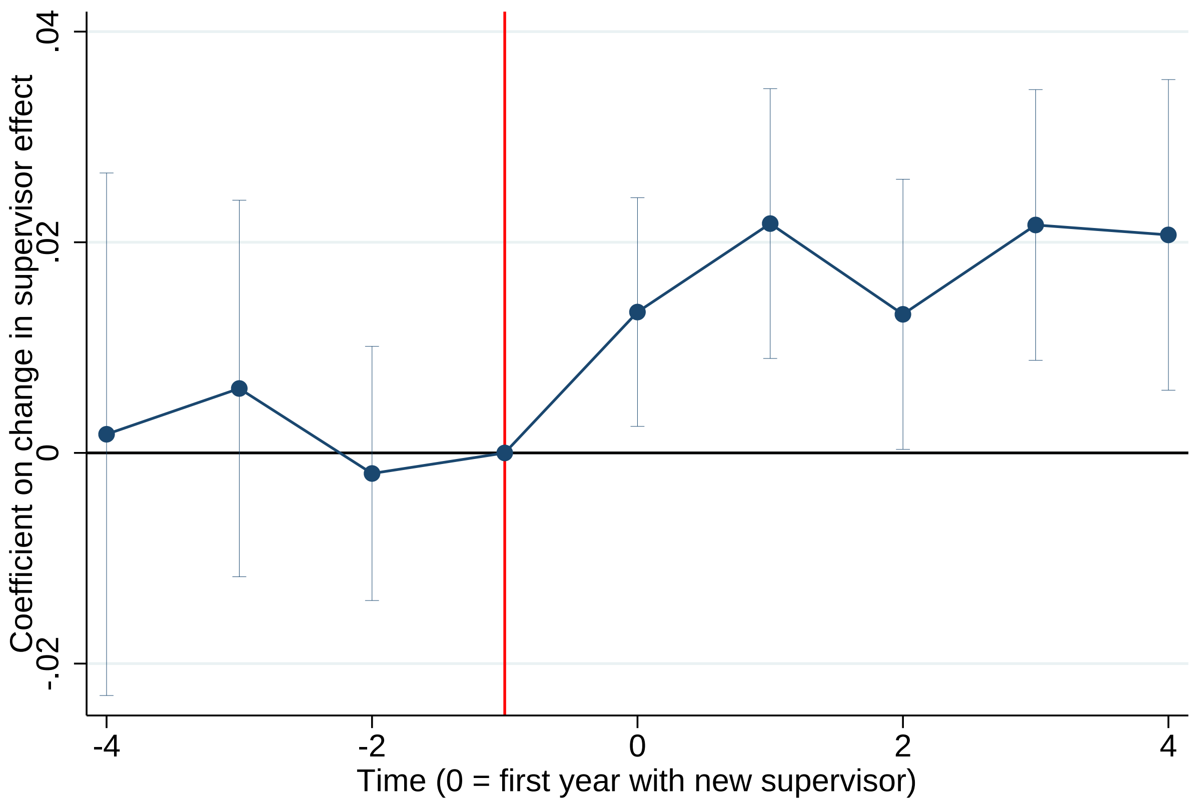
[34]The first stage regressions, using α's and ϕ's estimated on odd years to predict those estimated on even years and vice versa, are highly predictive. The F-statistic on the instruments is 186 for predicting ϕ and 3,572 for predicting α.

Table 5: Log(Earnings) and Ratings Components

	Dependent Variable: Log(Earnings)						
	OLS			Split sample IV		Bias correction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supervisor ratings effect (φ)	0.095***	0.054***	0.025***	0.117***	0.054**	0.114***	0.063***
	(0.013)	(0.009)	(0.004)	(0.023)	(0.023)	(0.0027)	(0.0024)
Worker ratings effect (α)	0.098***	0.093***		0.117***	0.117***	0.109***	0.096***
	(0.003)	(0.003)		(0.004)	(0.004)	(0.0015)	(0.0013)
Pass residual (ε)	0.021***	0.021***	0.020***	(na)	(na)	0.025***	0.024***
	(0.001)	(0.001)	(0.001)			(0.0014)	(0.0012)
Business Unit FEs		X			X		X
Worker FEs			X				
Observations	77,682	77,682	77,682	74,641	74,633	77,583	77,583
R-Squared	0.818	0.856	0.955	0.814	0.852		

Notes: Columns 1-3 present OLS regressions of log earnings on ratings components. Columns 4-5 estimate supervisor and worker effects in even and odd years, separately, and use estimates in even years as instruments for estimates in odd years and vice versa. Columns 6-7 presents coefficients based on the estimator in Andrews et al. (2008). Where indicated, we include business unit fixed effects (separate indicators for each branch as well as each function within the central corporate office). All regressions also include controls listed in Table 3. Standard errors in columns 1-5 are clustered by supervisor. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

Figure 2: Log(Earnings) Event Study



We regress log(earnings) on the change in the supervisor ratings effect between time -1 and 0 interacted with lags and leads of event ti with -1 as the omitted category. Regression includes full controls and worker fixed effects. 90% confidence intervals are also indicated.

well as the unobserved effects in the other subsample. Consequently the first stage will fail when attempting to instrument for $(\hat{\epsilon^{p}_{it}})$ in one subsample with α's and $\phi's$ obtained from the other subsample. The split-sample IV also does not lend itself well to the worker fixed effects specification because the first stage with worker fixed effects is not conceptually correct. We want to predict one noisy estimate of ϕ with another noisy estimate of ϕ and in any given year, we would not have within-worker variation in these.

Our third strategy is to expand the Andrews et al. (2008) correction – our conceptual framework for obtaining bias-adjusted estimates of variances in ratings components (section 2.3) – to a joint system of two double fixed effects regressions (one for ratings and one for earnings).[35] Once the second moment matrices of the unobserved effects are obtained they can be transformed into implied regression coefficients. Results are reported in columns 5-6 of Table 5. This estimator is computationally intensive and we therefore only implement it for worker earnings (and not the other dependent variables explored below). The methodology also requires strong distributional assumptions regarding the error terms that are not immediately applicable for some of our outcome variables such as those aggregated to the supervisor or branch level. However, this methodology does have the advantage that it can be applied to almost the entire estimation sample, not only the one consisting of workers and supervisors present in both even and odd years. It only requires that we be able to estimate worker and supervisor fixed effects in log(earnings), which applies for 77,583 observations.[36]

Across all specifications, we find that working for a high-rating supervisor is associated with substantially higher earnings. The unadjusted estimates are a bit smaller in magnitude, compared to the split sample IV and bias corrections, as we would expect if the estimation error is interpreted as "measurement error." Recall from Table 4 that these results are consistent with either heterogeneity being driven primarily by supervisor ability, or by supervisor leniency if firms are uninformed about supervisor heterogeneity. By contrast, the informed firm would undo any variation driven by leniency in compensation. Our findings in Table 5 thus reject the joint hypothesis that (1) the heterogeneity in ratings across supervisors is driven by leniency bias and (2) the firm is informed about this heterogeneity.

4.2 Piece Rates

A key difference between high-ability supervisors and lenient supervisors in our model is that high-ability supervisors lower the marginal cost of effort for workers. Consequently, informed firms will raise piece rates for subordinates who are matched to better managers while piece rates will not vary across supervisors that differ only in their leniency bias. Hence, one way to disentangle supervisor ability from leniency is to

[35] We have not succeeded in adapting Kline, Saggio, and Solvsten (2018) to the two-equation setting and thus rely on Andrews et al. (2008). For ratings, adjusted estimates are very similar using either the Andrews et al. or the Kline et al. approach.

[36] Recall, our estimation sample is already restricted to observations for whom we can identify supervisor and worker fixed effects in ratings. However, earnings data are not available in our last year of data, and this restriction results in the loss of identification for a small number of worker and supervisor fixed effects.

Table 6: Pay-for-Performance and Ratings Components

Dependent variable:	Log(Earnings)			Pr(Received a Bonus)			Log(Bonus)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Supervisor ratings effect (φ)	0.067***	0.042***	0.018***	0.167***	0.170***	0.068***	0.463***	0.362***	0.037
	(0.009)	(0.009)	(0.005)	(0.020)	(0.019)	(0.021)	(0.110)	(0.093)	(0.090)
Worker ratings effect (α)	0.108***	0.097***		0.238***	0.235***		0.583***	0.552***	
	(0.004)	(0.003)		(0.008)	(0.007)		(0.032)	(0.026)	
Pass residual (ε)	0.031***	0.025***	0.023***	0.092***	0.086***	0.090***	0.269***	0.234***	0.245***
	(0.003)	(0.002)	(0.002)	(0.006)	(0.006)	(0.006)	(0.022)	(0.020)	(0.027)
φ*Pass	0.075***	0.032**	0.019*	-0.003	-0.041*	-0.045	0.415***	0.181*	0.225**
	(0.018)	(0.013)	(0.010)	(0.026)	(0.024)	(0.028)	(0.117)	(0.093)	(0.113)
Business Unit FEs		X			X			X	
Worker FEs			X			X			X
Observations	77,682	77,682	77,682	77,682	77,682	77,682	24,001	24,001	24,001
R-squared	0.819	0.856	0.955	0.334	0.373	0.523	0.629	0.739	0.897

Notes: OLS results. See table 5. Regressions include all controls specified in table 3, as well as business unit or worker fixed effects where indicated. Standard errors are clustered by surpervisor. φ*Pass is the interaction of the supervisor fixed effect and the pass-fail performance rating.

determine if variable pay components are sensitive to supervisor heterogeneity.

To shed light on this relationship, we augment our earnings regression 7 by interacting supervisor heterogeneity (ϕ_s) with performance ($pass$) in a given period. The interaction measures whether performance ratings have a stronger effect on earnings when the supervisor is a higher rater. We also take a more direct approach and use as dependent variables the probability of receiving a bonus and the log of the size of the bonus, conditional on receiving one. For these models we only present OLS results because we do not know how to adapt the split-sample IV and or the bias correction method to identify the coefficient on $\phi * pass$. Table 6 contains the results.

We find that total earnings and bonuses, conditional on receiving one, are more strongly related to the worker's performance when assigned to a high rater. All else equal, passing the performance review is worth an additional 7 percent in wages when working for a high rater (the coefficient 0.075 on $\phi * Pass$ in column 1). The main effect of passing the performance review can be obtained by summing coefficients on all the components: 0.20 ($= 0.067 + 0.108 + 0.031$). Therefore, matching with a high rater increases the strength of pay for performance by about a third. The magnitudes on both the main effects and the interaction between ϕ and pass fall with the inclusion of business unit fixed effects in column 2 or worker fixed effects in column 3, but the story is still qualitatively similar: workers earn an extra 2-3 percent when they pass their performance review if they also work with a higher rater, or about 20 to 50 percent of the main effect of passing (the sum of the coefficients on ϕ, α, and ϵ). Thus the difference in incentive strength across raters is primarily not due to differences across units in general performance, norms, or job functions and holds up to variation that explicitly takes advantage of supervisor switches. From columns 4-9, much of the effect is due to the size of the bonus, conditional on receiving one. This is consistent with our understanding that supervisors have discretion over allocating salaries, but especially bonuses.

These findings are consistent with the hypothesis that supervisor heterogeneity is driven by heterogeneity in managerial ability that the firm is informed about.

4.3 Productivity

As pointed out above, our model implies that more able supervisors lower the marginal cost of effort and raise productivity. In contrast, lenient supervisors do not alter productivity. These associations hold irrespective of whether the firm is informed about supervisor heterogeneities. Hence, as long as we have an independent measure of productivity (separate from performance ratings), we can use the correlation between supervisor heterogeneity and productivity to disentangle supervisor ability from leniency.

We have access to two such measures of objective performance. During the years 2007–2010 the company ranked branches within a set of peers along a number of Key Performance Indicators (KPIs) that reflect financial outcomes, customer relations, etc.[37] For the years 2014 and 2015 we have information on individual financial performance. This latter metric is available for about half of the workers in the branches, primarily among senior workers with client facing roles. For both KPI performance and financial performance we investigate their relation to worker and supervisor fixed effects using OLS and our split-sample IV approach. For power reasons, we use fixed effects estimated on the sample as a whole, not the restricted sample where objective performance measures are available. For the financial performance regressions, it is worth noting that our personnel records end in 2014, while the financial performance measure covers the years 2014 and 2015. Hence, in practice, we regress our financial performance measures on the fixed effects associated with the supervisor the employee was assigned to in 2014 , and, as usual, cluster standard errors by supervisor.[38]

Table 7 presents the OLS (Panel A) and the split-sample IV (Panel B) estimates for both sets of measures. The KPI regressions (columns 1–4) relate the branch rankings to averages of employee and supervisor fixed effects within the branch-year.[39] Analogous to the split-sample IV on individuals, we correct for estimation error by instrumenting for the branch averages based on the average α's and ϕ's at the branch-level from even years with those from the odd years and vice versa.[40] In these aggregated regressions, we control for a limited set of variables, either averaged to the branch-year level or at the individual level.[41] For the financial performance regressions (column 5), the fixed effects pertain to the individual workers and their supervisors

[37]We have reestimated all results presented in this paper on the subsample restricted to branches and years where KPIs are available, and found them to generally be quite robust to this sample restriction.

[38]This is likely to induce some downward bias because supervisors change over time. The degree of downward bias will depend on how persistent ϕ is. At the branch level, we observe that $corr(\phi_t, \phi_{t+1}) = 0.852$ and $corr(\phi_t, \phi_{t+2}) = 0.766$.

[39]If there is only one supervisor in a given branch-year, as is often the case, the average supervisor effect is the ratings 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.

[40]The first-stage of the IV is naturally estimated on the same sample and at the same level as the second-stage: branch-years for branches with KPI data.

[41]These include year effects, the average worker age, tenure, and share female, as well as the average of each job-level indicator.

and we include our typical individual-level controls (see Table 3). We include branch size as a control in all specifications to account for different patterns in productivity based on local demand, the client base, etc... . Unfortunately, we do not have enough variation to allow for branch fixed effects given that we only have data from 2007 to 2010 and that average supervisor effects by branch vary slowly. .[42] Also, our measures themselves control for baseline heterogeneity since KPI rankings are relative to a peer group of branches that the firm defines, and financial performance is the individual's year-over-year growth rate in their portfolio.

Our general finding from Table 7 is a positive relationship between higher rating supervisors and objective performance of subordinates and branches. We explore a range of functional forms for the KPI rankings. Using our IV estimates, we find that a branch with a one-standard-deviation higher ϕ has a 0.05 higher inverse rank score (-1 times the branch's ranking divided by the number of branches in the peer group), or 9 percent, is 1.9 percentage points (31 percent) more likely to be the top-ranked branch, is 6 percentage points (20 percent) more likely to be ranked among the top 5 branches in the peer group, and 6.4 percentage points (13 percent) more likely to be ranked in the top half. These magnitudes are economically large. We also find positive effects for individual financial performance. IV estimate implies that a one-standard-deviation higher rating supervisor is associated with a 1 percentage point (13 percent) larger growth rate.

The results in Table 7 indicate a positive relationship between supervisor heterogeneity and performance. Unfortunately, we only have performance measures for a small number of years and branches, which at times challenges the statistical significance of the point estimates. The evidence we do provide, however, suggests that there is a positive relation between supervisors ratings heterogeneity and team performance. This conclusion is strengthened by the fact that we have two distinct performance measures that occur at different points in time. Hence, our results support the hypothesis that manager ability (μ_s), rather than leniency bias (β_s), drives supervisor heterogeneity.[43]

4.4 Supervisor Outcomes

The fourth comparative static relates supervisor heterogeneity in ratings to the supervisors' own pay. Supervisor compensation would not correlate with supervisor heterogeneity if firms were uninformed about ϕ_s. Nor would firms compensate supervisors for being more lenient. Only if supervisor heterogeneity reflects managerial ability about which the firm is informed will ϕ_s and supervisor compensation correlate positively.

To investigate this relationship we regress supervisor outcomes on their own ratings fixed effect, as well

[42]Similarly the table does not include worker fixed effects specifications. These are not apt for the KPI regressions that are aggregated to the branch-year level, and, with only two years of individual financial performance data, we do not allow enough variation for identifying worker fixed effects in column 5.

[43]The estimated impacts of α on objective performance in Table 7 are statistically insignificant. However, the 95 percent confidence intervals for the point estimates typically include large positive effects. The results are thus consistent with worker quality that correlates positively with branch performance, though too noisy to be conclusive.

Table 7: Objective Performance and Ratings Components

	Branch KPI Rankings				Individual Financials
	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Inverse Rank Score	Pr(Top)	Pr(Top 5)	Pr(Top half)	Year-over-year growth rate
(mean)	(-0.53)	(0.06)	(0.30)	(0.48)	(-0.074)
Panel A: OLS					
Supervisor ratings effect (φ)	0.181**	0.089	0.216*	0.255*	0.045**
	(0.082)	(0.068)	(0.131)	(0.144)	(0.020)
Worker ratings effect (α)	0.023	0.042	0.084	0.056	0.008
	(0.073)	(0.061)	(0.117)	(0.128)	(0.009)
Pass residual (ε)					0.001
					(0.007)
Observations	781	781	781	781	2502
R-squared	0.032	0.033	0.037	0.019	0.07
Panel B: Split-Sample IV					
Supervisor ratings effect (φ)	0.332**	0.130	0.405*	0.441*	0.065*
	(0.144)	(0.118)	(0.230)	(0.250)	(0.034)
Worker ratings effect (α)	-0.039	0.009	-0.023	-0.016	-0.002
	(0.108)	(0.089)	(0.171)	(0.188)	(0.012)
Observations	781	781	781	781	2466
R-squared	0.003	0.025	0.017	0.011	0.063

Notes: Columns 1-4 are estimated using data from 2007-2010 at at the branch-year level; performance components are the branch-year averages. Column 5 is estimated using worker-level data on a subset of employees for years 2014-15. Here, the performance components are also at the individual level. 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. Regressions include controls for branch size. Column 5 includes all controls listed in table 3. Columns 1-4 include year effects, branch size, and the branch-year averages of worker age, tenure, share female, and job level dummies. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

as the average worker fixed effect for the group of subordinates the supervisor rated in that year. We present OLS and split-sample IV results.[44] These regressions control for the characteristics of the supervisor and the average characteristics of the group of workers being supervised, as well as branch size. Standard errors are clustered by supervisor.

Results are reported in Table 8. Supervisor earnings are strongly positively correlated with their own ratings style (as well as the quality of the team they supervise). This is true for log earnings overall and for the size of the bonus conditional on receiving one. For example, using the IV estimates, we find that supervisor earnings increase by 2 percent for each standard deviation in ϕ_s. Much of this increase comes through an increase in the size of the bonus received, conditional on receiving one. Furthermore, supervisors who are high raters are substantially more likely to pass the performance review they receive from their own supervisor. A one-standard deviation higher rater is 6 percentage points more likely to pass their own performance review.

Supervisor earnings also positively correlate with the quality of the team they supervise, α. This result is intriguing, even though our model cannot rationalize it. One possibility is that perhaps the firm cannot perfectly separate the ability of supervisors from the ability of workers.

We find little evidence that ratings heterogeneity correlates with mobility in either direction. The point estimates on promotion probability are positive and large, but those on staying with the firm or in the business unit are negative. Standard errors related to these outcomes are simply too large to say anything definitive.

Consistent with our earlier findings, the positive relationship between supervisor ratings behavior and their own compensation and ratings provides support for the hypothesis that supervisor heterogeneity reflects ability differences that the firm is informed about rather than differences in leniency.

4.5 Worker Surplus and the Information Structure

The last comparative static we consider is whether worker surplus is related to supervisor heterogeneity. This relationship is particularly informative about the information structure. In our model, fully informed firms will always hold workers to their participation constraint, eliminating any variation in surplus resulting from supervisor characteristics. Evidence that worker surplus increases in supervisor heterogeneity indicates that the firm is not fully informed about differences across supervisors in ϕ_s.

We use worker mobility and worker satisfaction surveys to look for evidence regarding rents associated

[44] As in the branch-year regressions of table 7, we include branch size but not branch or worker fixed effects. For the IV specifications, we obtain supervisor 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 from the opposite subsample.

Table 8: Supervisor Outcomes and Ratings Components

Dependent variable:	Log(earnings)	Pr bonus	Log(bonus)	Pass	Stay in Firm	Stay in Unit	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: OLS				
Supervisor ratings effect (φ)	0.105***	-0.015	0.263***	0.391***	-0.059*	-0.054	0.017
	(0.022)	(0.033)	(0.091)	(0.044)	(0.035)	(0.038)	(0.025)
Worker ratings effect (α)	0.097***	0.069**	0.316***	0.463***	0.027	0.014	0.030
	(0.019)	(0.031)	(0.078)	(0.043)	(0.035)	(0.038)	(0.028)
Pass residual (ε)	0.010	0.007	0.007	0.321***	-0.011	-0.008	0.004
	(0.011)	(0.025)	(0.056)	(0.031)	(0.030)	(0.035)	(0.023)
Observations	8,513	8,513	5,017	9,473	8,513	8,513	8,305
R-Squared	0.771	0.572	0.650	0.136	0.066	0.074	0.127
			Panel B: Split sample IV				
Supervisor ratings effect (φ)	0.125***	-0.070	0.386**	0.447***	-0.038	-0.029	0.056
	(0.037)	(0.053)	(0.151)	(0.073)	(0.067)	(0.070)	(0.040)
Worker ratings effect (α)	0.119***	0.045	0.350***	0.544***	0.007	0.012	0.077*
	(0.033)	(0.046)	(0.125)	(0.066)	(0.054)	(0.060)	(0.044)
Observations	8,269	8,269	4,875	9,155	8,269	8,269	8,084
R-Squared	0.769	0.576	0.650	0.127	0.056	0.071	0.128

Notes: Observations are at the supervisor-year level. φ is how the supervisor rates their own subordinates, and α and ε are the averages of the subordinates rated that 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. Outcomes are supervisor pay, performance, and mobility variables in the given year. Pass is whether the supervisor passed their own performance review; promotion probability is restricted to observations that did not leave the firm in the next year. Controls are listed in table 3; worker controls are the average for characteristics. We also include the average branch size of subordinates rated in the given year in all specifications. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

with ϕ_s. Worker satisfaction surveys provide direct evidence on how workers perceive their supervisors. These surveys are taken by an independent consulting firm and anonymized before being returned to the firm, which should encourage workers to report their answers truthfully. Fortunately for us, we have access to the survey responses prior to anonymization. For the other outcomes, workers can, to some extent, influence their assignments across units within the firm, and can likely control whether they stay with the firm.

Table 9 presents OLS (panel A) and split-sample IV estimates (panel B). For each outcome we present our base-line specification as well asseparate specifications with branch and worker fixed effects. The latter leverages only worker switches for identifying the coefficient on ϕ. It turns out that the results are robust across these specifications, but we should note that estimates become unstable and imprecisely estimated when we use the supervisor leave-out mean in performance ratings as the key regressor.

Column 1 shows that workers are more likely to stay in the firm in the next year if working for a higher-rating supervisor.[45] This effect is marginally significant in the base specification and becomes larger in magnitude and significance when we add business unit fixed effects in column 2 or worker fixed effects in column 3. Columns 4-6 show that workers are more likely to stay in the same business unit in the next year when assigned to a higher rater. In columns 7-9, workers are not any more likely to stay with their particular supervisor, though, as we have noted, many supervisor switches are driven by moves of the supervisors themselves, which would be outside the control of the worker.

Finally, the results in columns 10-12 are based on data from the employee job satisfaction survey. The dependent variable is the average across seven survey questions relating to the supervisor, normed to have a standard deviation of 1. The results show that subordinates tend to be more satisfied with their supervisors when their supervisors are higher raters. While this effect is statistically significant, its economic importance is modest as an assignment to a one-standard-deviation higher rater is associated with a modest 0.034 increase in the bottom-up rating a worker ascribes to his or her supervisor (the mean of that variable is 4.7). The magnitude is fairly similar in the worker fixed effects specification, which explicitly leverages supervisor switches and also controls for differences in average ratings behavior across workers.

Together, these findings indicate that workers earn rents when assigned to high rating supervisors, even though it is difficult to ascertain the magnitude of these rents. Nevertheless, while the evidence presented so far suggests that supervisor heterogeneity reflects ability differences that the firm is informed about, the results in Table 9 suggest that the firm is not perfectly informed about such differences.[46]

If the firm is unable to perfectly distinguish between worker and supervisor effects, it may still learn

[45] We have estimated the probabilities of quit and layoff separately and find that being matched to a higher rater has similar negative impacts on both, though estimates are noisier than the combined probability shown in the table.

[46] Table 9 also reveals that workers with higher α may earn rents; they are less likely quit, more likely to stay with their current supervisor, and report being more satisfied at their job.

Table 9: Do Workers Value High Raters?

Dependent variables:	Stay in Firm			Stay in Unit			Stay with supervisor			Bottom-up evaluation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: OLS												
Supervisor FE (ϕ)	0.023*	0.040***	0.038**	0.033*	0.054***	0.049**	0.014	0.013	0.006	0.136***	0.096**	0.102**
	(0.014)	(0.014)	(0.018)	(0.018)	(0.018)	(0.023)	(0.026)	(0.029)	(0.027)	(0.042)	(0.045)	(0.043)
Worker FE (α)	0.062***	0.063***		0.076***	0.080***		0.080***	0.080***		0.141***	0.140***	
	(0.005)	(0.005)		(0.006)	(0.006)		(0.008)	(0.008)		(0.016)	(0.017)	
Pass residual (ε)	0.008**	0.009**	0.010**	0.004	0.004	0.005	0.009	0.010	0.009	0.053***	0.049***	0.056***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.010)	(0.010)	(0.010)
Observations	77,682	77,682	77,682	77,682	77,682	77,682	77,682	77,682	77,682	74,993	74,993	74,993
R-Squared	0.044	0.070	0.277	0.073	0.121	0.277	0.042	0.073	0.241	0.021	0.044	0.351
Panel B: Split sample IV												
Supervisor FE (ϕ)	0.045*	0.097**		0.047	0.093*		0.039	0.026		0.235***	0.163	
	(0.026)	(0.038)		(0.036)	(0.049)		(0.050)	(0.073)		(0.078)	(0.116)	
Worker FE (α)	0.068***	0.068***		0.099***	0.101***		0.110***	0.112***		0.162***	0.164***	
	(0.008)	(0.008)		(0.010)	(0.010)		(0.012)	(0.012)		(0.024)	(0.025)	
Observations	74,641	74,633		74,641	74,633		74,641	74,633		71,835	71,828	
R-Squared	0.039	0.061		0.068	0.117		0.035	0.066		0.020	0.044	
Business Unit FEs		X			X			X			X	
Worker FEs			X			X			X			X

Notes: Column 1-3 estimate the probability that the worker stayed in our sample between t and t+1; columns 4-6 estimate the probability that the worker stayed in the same branch or business unit (if in corporate); columns 7-9 estimate the probability that the worker stayed with the same supervisor; columns 10-12 reports the worker's self-reported satisfaction of their supervisor. All regressions include time-varying worker and supervisor controls (see Table 3), and, where indicated, branch or worker fixed effects. Standard errors clustered by supervisor. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

over time. In that case, benefits accruing to a worker who is matched to a high rater should attenuate with supervisor tenure. We explore this in Table 10. We specify tenure as the number of years the supervisor has been rating subordinates. We augment our main specification (equation 7) to include an interaction between ϕ and an indicator equaling 1 if the supervisor has above median tenure, defined as the length of time they have been rating subordinates.[47]

Beginning with log(earnings) in columns 1, we consistently find that the benefit of a high rating supervisor on worker earnings is smaller when the supervisor has more tenure. Magnitudes are especially large for the split-sample IV, where we find that offsets are about half to three-quarters of the main effect.[48] Effects for the remaining variables are noisy. However, the mobility variables (stay in firm, stay in unit, and stay with supervisor) all have the opposite sign of the main effect ϕ, and magnitudes are sizable. This is suggestive of the same attenuation as seen in earnings, but of course not conclusive.

Overall, Table 10 provides suggestive evidence that the benefits workers experience when associated with a high rater are indeed partially undone for supervisors with more tenure. This is consistent with the idea that there are rents associated with being matched to certain supervisors, but that these rents are driven by imperfect information. However, the bottom-up evaluations indicate that subordinates still enjoy working for a better manager, even if they accrue fewer economic rents. Of course the evidence is far from conclusive, given the large standard errors.

[47] We also include the main effect for the supervisor tenure variable, which is essentially subsumed in our controls. We have explored a range of different functional forms, all yielding consistent results. We also find similar results when defining tenure as the overall length of time the supervisor has been with the firm.

[48] We instrument for $\phi * tenure$ with the interaction of ϕ in the alternating even or odd year times the tenure variable.

Table 10: Worker Outcomes, Ratings Components and Supervisor Tenure

Dependent variables:	Log(Earnings)	Stay in Firm	Stay in Unit	Stay with Supervisor	Bottom-Up Evaluations
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Supervisor FE (ϕ)	0.102***	0.056***	0.064***	0.039	0.111**
	(0.012)	(0.018)	(0.022)	(0.030)	(0.046)
Worker FE (α)	0.098***	0.062***	0.075***	0.080***	0.141***
	(0.003)	(0.005)	(0.006)	(0.008)	(0.016)
Pass residual (ε)	0.021***	0.009**	0.005	0.009	0.053***
	(0.001)	(0.004)	(0.005)	(0.006)	(0.010)
ϕ*Above median tenure	-0.016	-0.089***	-0.084**	-0.066	0.051
	(0.017)	(0.025)	(0.034)	(0.045)	(0.064)
Observations	77,682	77,682	77,682	77,682	74,993
R-squared	0.818	0.045	0.073	0.042	0.022
Panel B: Split sample IV					
Supervisor FE (ϕ)	0.146***	0.058	0.079	0.081	0.231**
	(0.028)	(0.040)	(0.051)	(0.067)	(0.105)
Worker FE (α)	0.118***	0.068***	0.099***	0.111***	0.162***
	(0.004)	(0.008)	(0.010)	(0.012)	(0.024)
ϕ*Above median tenure	-0.066*	-0.033	-0.075	-0.093	-0.004
	(0.036)	(0.056)	(0.081)	(0.104)	(0.140)
Observations	74,641	74,641	74,641	74,641	71,835
R-squared	0.814	0.039	0.069	0.035	0.020

Notes: See tables 5 and 9. We augment our main regression equation with an interaction between the supervisor ratings effect (ϕ) and an indicator equaling 1 if the length of time the supervisor has been giving ratings is above median. We also include the main effect of the supervisor tenure variables, which are essentially subsumed in our other controls. All regressions include time-varying worker and supervisor controls (see Table3). Standard errors clustered by supervisor. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

4.6 Discussion

We have presented results on how the heterogeneity in ratings associated with supervisors ϕ_s relates to outcomes for employees, supervisors, and the firm. We found that (1) individual earnings increase with ϕ_s, (2) piece rates increase in ϕ_s, (3) team productivity as measured by the KPI ranking increases in the average ϕ_s within a branch, (4) individual financial performance increases in ϕ_s, (5) supervisor pay increases in ϕ_s, and (6) workers appear to earn (moderate) rents from being matched to higher raters, that are attenuated for raters with longer tenures about whom the firm is presumably more informed. These findings speak to the nature of the heterogeneity associate with supervisors and whether the firm is informed about this heterogeneity. Consulting Table 4, our evidence suggests that supervisor heterogeneity in ratings is driven mainly by differences in managerial ability and that the firm is partially informed about this heterogeneity.

There are three pieces of evidence that support the interpretation that heterogeneity in managerial ability drives at least some of the variation in ratings across supervisors. First, we find that objective performance increases when individuals or teams are managed by a high rater (Table 7), which directly supports the managerial ability hypothesis. Second, high-rating supervisors earn significantly higher salaries (Table 8) suggesting that firms value them, as would be the case when high raters are also better managers. Third, subordinates of higher raters tend to face stronger incentives (Table 6), which is rationalized in our model by the fact that better managers lower the marginal cost of worker effort (an equivalent assumption is that better managers increase output per additional unit of effort).

The observation that the strength of incentives for employees and that average compensation of supervisors vary with ϕ_s suggests that the firm is informed about the heterogeneity in ratings styles across supervisors. However, it seems a priori plausible that firms will not be perfectly informed. This notion is consistent with the observation that subordinates earn (moderate) economic rents when working for higher rating supervisors. The perfectly informed firm would extract all rents from its employees by adjusting their base salaries to place them on their participation constraints, and we do see evidence consistent with this behavior for supervisors with longer tenures. The firm also appears to reward supervisors for the fixed quality of their subordinates (Table 8), which may also be indicative of a lack of ability to perfectly discern what drives performance.

Of course, outside our model there are other reasons why firms might share rents with higher raters or with employees. This firm may purposely do a better job fostering a feeling of satisfaction for desirable workers and supervisor-worker matches.

Finally, while these results have a clear interpretation within the context of our model, one could write down other models of compensation and bonuses that might generate similar predictions. Regardless of the model used to interpret these results, we have shown that there is substantial heterogeneity in performance ratings across supervisors and that this heterogeneity is indeed associated with heterogeneity in objective output. Firms should therefore think twice before imposing forced curves or other rules that limit the variation in subjective performance ratings as it may undermine supervisors' ability to manage.

5 HOW INFLUENTIAL ARE HIGH RATERS FOR CAREERS?

In Section 4.1, we established that working for a high-rating supervisor is associated with higher contemporaneous earnings. Next, we consider how longer-term career outcomes vary with supervisor type. This requires us to think about dynamic effects in relation to ϕ_s and thus forces us to step outside of the static model presented in Section 3. In particular, we are interested in how ratings affect earnings in subsequent years, even after a worker has left the high-rating supervisor. This could manifest because pay raises are persistent but also because high-rating supervisors may affect the progression of a worker along the job hierarchy.

We begin by estimating the persistence of ϕ_s on pay. We base our estimates on the following dynamic equation relating current log earnings to several lagged supervisor effects:

$$w\left(l, \phi^t, e_t\right) = \quad g_1\left(l_{it}\right) + h_1\left(X_{i,t}\right) + \Sigma_{\tau=0}^{k}\beta_\tau \phi_{s(i,t-\tau)} + \Sigma_{\tau=0}^{k}\theta_\tau \varepsilon_{i,t-\tau} + e_{i,t} \tag{8}$$

Equation 8 includes k lags in supervisor effects as well as the contemporaneous value $\phi_{s(i,t)}$.[49] These lags allow ϕ_s to influence earnings for up to k periods. Estimates from equation 8 do not represent the full dynamic effects of being assigned a higher rater (ϕ_s) for two reasons. First, we control for job-level effects (l_{it}) to account for any variation in ratings style across job levels. However, part of the effect of ratings heterogeneity on future earnings arises through promotions and we will explore that effect below. Second, we control for ratings type of supervisors in other periods. This removes any effect of the current supervisor that can be attributed to persistence in the supervisor match. Estimates of β_τ thus yield the impact of a one-time match to a higher rater τ periods ago on earnings today over and above any promotion effects and effects attributable to persistence in supervisor ratings styles.

Results are summarized in Table 11. Column 1 replicates the earnings effect from Table 5, the impact of ϕ_s on contemporaneous earnings. Once we include lagged supervisor effects in the regression, the sample size naturally begins to drop. To understand any differences across samples, column 2 shows the main earnings specification from column 1, restricted to workers who are present for at least five periods in the firm, that

[49] Equation 8 also includes controls for k lags in the ratings residual $\varepsilon_{i,t}$, for α_i, and for the typical constant and time-varying controls $X_{i,t}$.

Table 11: Earnings Dynamics and Supervisor Heterogeneity

Dependent variable		Log earnings	
	(1)	(2)	(3)
Supervisor FE (ϕ):			
Contemporaneous ϕ	0.095***	0.068***	0.029**
	(0.013)	(0.017)	(0.014)
Lag 1 ϕ			0.017*
			(0.009)
Lag 2 ϕ			0.014
			(0.009)
Lag 3 ϕ			0.015*
			(0.009)
Lag 4 ϕ			0.021**
			(0.010)
Non-missing lags		X	X
Observations	77,682	22,609	22,609
R-squared	0.818	0.821	0.822

Notes: The table reports regressions of log earnings on contemporanous supervisor effects (ϕ), worker unobserved effects (α), and residuals (ε) from the ratings equation (1). Lag 1 ϕ is the ϕ associated with the supervisor the worker was matched to in t-1. Lag 2 ϕ is for t-2, etc. All regressions contain the same number of lags in (ε) as in (ϕ) and control for the same set of controls as in the main specification reported in table 5. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

is, with at least four lags in supervisor effects.[50] The coefficient is a bit smaller in magnitude for the sample of more stable workers, 0.068 compared to 0.095, but still qualitatively similar.

Column 3 presents results including all four lags of supervisor effects. The coefficient on the contemporaneous earnings effect drops to 0.029. This is because part of the supervisor effect on earnings comes through persistence of supervisors across periods (as shown at least qualitatively in Table 9). Furthermore, we find that impacts of supervisor's ϕ on pay are quite persistent. The coefficients on the lagged ϕ range between 0.014 and 0.021. This suggests that roughly half to two-thirds of the contemporaneous pay increase associated with having a high-rating supervisor persists several periods later. While the effect of being assigned to a higher rater is somewhat transitory, the large fraction that remains could indicate direct salary impacts that amortize over time (as opposed to effects driven solely by transient bonuses), or perhaps because of lasting effects on human capital.

These coefficients are estimated holding constant job level. This means that they do not include any impact of ϕ_s on earnings through promotions and demotions. We do not estimate regressions omitting job level controls because supervisor types vary systematically across job levels — higher raters tend to be further

[50] Balancing the desire to understand the lag structure of earnings effects against the need to maintain sample sizes in a panel of only 11 years, we choose to focus on 5 years of lags. Appendix Table A5 explores robustness to more and less restricted samples, based on number of available lags, and we find results to generally be quite similar, quantitatively.

Table 12: Worker Outcomes and Ratings Components

Dependent variables:	(1) Promotion	(2) Demotion	(3) Layoff
		Panel A: OLS	
Supervisor FE (ϕ)	0.045***	-0.005*	-0.008***
	(0.010)	(0.003)	(0.003)
Worker FE (α)	0.096***	-0.019***	-0.010***
	(0.004)	(0.001)	(0.002)
Pass residual (ε)	0.056***	-0.006***	-0.001
	(0.004)	(0.001)	(0.001)
Observations	75,197	75,197	69,527
R-squared	0.121	0.014	0.008
		Panel B: Split sample IV	
Supervisor FE (ϕ)	0.035*	-0.009	-0.013**
	(0.019)	(0.005)	(0.005)
Worker FE (α)	0.105***	-0.024***	-0.014***
	(0.006)	(0.002)	(0.002)
Observations	72,289	72,289	66,654
R-squared	0.118	0.013	0.008

Notes: Columns 1 and 2 estimate the probability that the worker was promoted or demoted between t and t+1 for those observed in adjacent years in the firm. Column 3 estimates the probability that the worker was laid off by t+2 for all workers observed in t, excluding the last two years of data where t+2 outcomes cannot be observed. All regressions include time-varying worker and supervisor controls. (See table 5). Standard errors clustered by supervisor. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

up the hierarchy. Instead, to account for how raters affect earnings through mobility in the job hierarchy, we also estimate equations predicting mobility at the firm.

We find that ϕ_s does indeed accelerate movement up the job hierarchy. Table 12 shows that a higher ϕ_s makes promotions more likely (column 1) and negative career moves in the form of demotions (column 2) and layoffs (column 3) less likely.[51] Using the split-sample IV estimates, we find that a one-standard-deviation higher rater increases the probability of promotion by 0.5 percentage points (5 percent). It decreases the probability of a demotion by 0.1 percentage point (13 percent, though this effect is not statistically significant) and a layoff by 0.2 percentage point (19 percent), for these already rare outcomes. Thus faster progression through the job hierarchy may be an important channel through which a high-rating supervisor raises earnings in the long-run.

Next, we engage in the following thought experiment: how does an increase in ϕ_s in one period affect the

[51] Columns 1 and 2 estimate the probability of promotion or demotion between years t and $t+1$ for workers present in both years. Column 3 estimates the probability of a layoff between t and $t+2$; we estimate the two-year rather than the one-year layoff probability because ratings are less likely to be taken in the final year.

present discounted value (PDV) of earnings, keeping all other supervisor effects in all other periods constant? This incorporates three components: (1) the persistence of the contemporaneous impact of ϕ_s on pay, (2) the impact of ϕ_s on promotions in the current and subsequent periods, and (3) the impact of promotions on pay. We consider these three components separately, rather than estimating the full earnings stream associated with a given ϕ_s because this allows us to control for job level in (1) while still allowing job level to impact earnings. As explained above, this is important because supervisor heterogeneity ϕ_s varies systematically across the job hierarchy. To simplify the analysis, we abstract from demotions and firm exit, both fairly rare events.[52]

The rate of impact of a given supervisor effect in period t ($\phi_{s(i,t)}$) on earnings in period $t+k$ is given in equation 9. It equals the persistent component of the within job-level pay effect, β_k (from equation 8, above), plus the impact of $\phi_{s(i,t)}$ on the probability of promotion, γ, times the average pay increase associated with a promotion ($g_1(l_{i,t}+1) - g_1(l_{i,t})$).

$$\frac{1}{W_{t+k}}\frac{dW_{t+k}}{d\phi_{s(i,t)}} = \beta_k + \gamma\left(g_1(l_{i,t}+1) - g_1(l_{i,t})\right) \tag{9}$$

To aggregate these over time, we obtain $\frac{1}{W_{t+k}}\frac{dW_{t+k}}{d\phi_{s(i,t)}}$ for each $k \leq 20$ (assuming careers last another 20 years) and aggregate them using a discount rate of 5 percent. For lags $k \leq 4$, the parameter estimates needed to perform these calculations are taken from Table 11. For lags $k > 4$, we make two different assumptions about β_k. First, we conservatively set β_k in all future periods $k > 4$ to zero, since we have not estimated these effects. However, Table 11 does not indicate any diminishing effect over time, within the four estimated lags, so a reasonable alternative assumption is a permanent $0.02 * \phi_s$ impact on wages for $k > 4$.

We obtain γ, the impact on promotion probability, from the split-sample IV estimate in Table 12. Since our calculation ($\frac{1}{W_{t+k}}\frac{dW_{t+k}}{d\phi_{s(i,t)}}$) depends on the job level of an individual, we average the promotion gains ($g_1(l+1) - g(l)$) using the observed distribution of wages and of workers across job levels. In our data, the average earnings increase associated with moving up adjacent job levels is 16 percent.[53] We allow this impact of promotion on earnings to persist for the full 20 periods. When we estimate dynamic promotion equations, we find that the contemporaneous promotion effect is persistent. In unreported results, we see no evidence that workers assigned to low raters catch up in terms of promotions and also no evidence that a one-time assignment to a high-rater results in multiple promotions.

Using our estimated standard deviation of ϕ_s, 0.137 (Table 3), we determine that a one-period, one-standard-standard deviation increase in ϕ_s is associated with an increase in the PDV of earnings of 2.2 to

[52]We also abstract away from any path dependence in ϕ. We do find that the correlation between the supervisor ratings effect in t and that in $t+1$ is 0.75 overall and 0.25 when there is a supervisor change between periods. Assigning causality to this correlation and taking it into account would raise the PDV of a one-period assignment to a high rater.

[53]Due to confidentiality issues, we are unable to provide the disaggregated inputs to this estimate.

4.7 percent of average annual earnings, corresponding to the more and less conservative assumptions on the persistence of β_k for $k > 4$. The direct wage effect, β_k, amounts to 1.2 to 3.8 percentage points of this, while the return associated with being promoted to a higher job level accounts for the remainder. If instead we assume, more conservatively, that the promotion effect dissipates after five periods, then the PDV estimates are 0.6 percentage points smaller.

These effects are quite large. When comparing workers assigned to supervisors at the 90th and 10th percentiles of the ratings distribution, the former can expect an increase in the PDV of earnings equivalent to 6 to 12% of an annual salary.

6 CONCLUSION

In this paper we provide evidence that supervisors differ widely in their ratings behavior. A worker matched to a one standard deviation higher rater is 13.7 percentage points (30 percent) more likely to receive a passing score (a performance score in the upper half of the performance scale). To understand this variation, we provide a theoretical framework that allows for two sources of heterogeneity in ratings behavior: leniency bias and managerial ability. We also allow the degree to which firms are informed about the heterogeneity to vary.

Within the context of this model, we conclude that differences in managerial ability are an important component of the heterogeneity in supervisors' ratings behavior.[54] This conclusion is based on the empirical finding that worker pay, pay for performance, supervisor pay, and individual and team-level objective performance measures are all increasing in the supervisor's propensity to give passing ratings to subordinates. Workers also appear to enjoy working with higher raters since they are less likely to voluntarily move away from them (by quitting or switching supervisors) and give them better ratings on bottom-up evaluations. This suggests that firms are unable to fully extract the surplus produced in the match between a worker and a high-rating supervisor, possibly because they are not fully informed about the heterogeneity in supervisor's ratings behavior. Consistent with the latter, we find that this surplus is attenuated for supervisors with higher tenure, about whom the firm is presumably more informed.

These results all have a clear interpretation within the context of our model. However, one could develop other models of compensation and bonuses that would generate the same predictions. For example, if bonuses are distributed based on a threshold rule, rather than linearly, a lenient supervisor will cause workers to exert more effort if they are close to the threshold margin. Or, supervisors may differ in their propensity to make subordinate pay vary with performance; those applying stronger incentives should get more output out of

[54]We can not rule out that leniency bias contributes to the heterogeneity in supervisor' ratings behavior, but we can rule out that heterogeneity in leniency bias alone sustains the variation in ratings across supervisors.

their workers and give them higher ratings. We do not know enough about how bonuses are set inside this firm to speak to these hypotheses. It may also be that a lenient supervisor generates a "warm glow" among his or her team that in and of itself generates higher output. Disentangling these and other stories is beyond the scope of this paper. Instead, our goals have been to (1) highlight the surprising and sizable variation in ratings across supervisors, and, (2) disciplining ourselves to one plausible model, which allowed us to dig deeper into the nature and information structure of this heterogeneity. Regardless of our model we can conclude that heterogeneity in ratings is indeed reflected in objective output measures suggesting that how supervisors rate and manage their employees interact in important ways.

Subjective performance reviews are controversial because workers may worry they are vulnerable to managerial biases. As a result, firms may desire to impose rules designed to correct for biases.[55] They might, for instance, force supervisors to grade their employees on a curve. However, our work cautions against such practices. At the firm we study, supervisor heterogeneity in ratings reflects, at least in part, real differences in the ability to elicit output from subordinates. Hence, firms should exercise care when they consider introducing forced curves or other guidelines restricting supervisors in how they can rate their subordinates.

References

[1] Abowd, John, Francis Kramarz, And David N. Margolis (1999), "High Wage Workers And High Wage Firms," *Econometrica*, 67(2): Pp. 251-333.

[2] Acemoglu, Daron, and Jorn-Steffen Pischke (1998), "Why Do Firms Train? Theory and Evidence," *Quarterly Journal of Economics*, 113(1): pp. 79-119.

[3] Aghion, Philippe, and Jean Tirole (1997), "Formal and Real Authority in Organizations," *The Journal of Political Economy*, 105(1): pp. 1-29.

[4] Alonso, Ricardo, and Niko Matouschek (2008), "Optimal Delegation," *The Review of Economic Studies*, 75(1): pp. 259-293.

[5] Altonji, Joseph, and Charles Pierret (2001), "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116(1): pp. 313-350.

[55] During the hiring process, there is evidence that reducing discretion among hiring managers can improve decision making. In low-skilled settings, Autor and Scarborough (2008) show that job testing can improve hiring decisions and Hoffman, Kahn, and Li (2018) show that firms may do better by removing some discretion of hiring managers and relying more on a high-quality job test.

[6] Martyn. Andrews, L. Gill, T. Schank, and R. Upward (2008). "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society (A)*, 171(3): pp. 673-697.

[7] Autor, David, and David Scarborough (2008), "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments," *Quarterly Journal of Economics*, 123(1): pp. 219-277.

[8] Baker, George, Michael Gibbs, and Bengt Holmstrom (1994a), "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, 109(4): pp. 881-919.

[9] Baker, George, Michael Gibbs, and Bengt Holmstrom (1994b), "The Wage Policy of a Firm," *Quarterly Journal of Economics*, 109(4): pp. 921-955.

[10] Bennedsen, Morten, Francisco Perez-Gonzalez, and Daniel Wolfenzon (2007) "Do CEOs Matter?" Working Paper No. 13-2007, Copenhagen Business School.

[11] 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.

[12] Bloom, Nick, and John Van Reenen (2007), "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Journal of Economics*, 122(4): pp. 1351-1408.

[13] Bolton, Patrick, and Mathias Dewatripont (2010), "Authority in Organizations," in *The Handbook of Organizational Economics*, Robert Gibbons and John Roberts (eds.). Princeton, NJ: Princeton University Press.

[14] Card, David, Joerg Heining, and Patrick Kline (2013), "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128(3): pp. 967-1015.

[15] DeGroot, Morris H. (1970), "*Optimal Statistics Decisions*," Hoboken, NJ: John Wiley and Sons, Inc.

[16] Dessein, Wouter (2002), "Authority and Communication in Organizations," *Review of Economic Studies*, 69: pp. 811-838.

[17] Dohmen, Thomas (2004), "Performance, Seniority, and Wages: Formal Salary Systems and Individual Earnings Profiles," *Labour Economics,* 11(6): 741-763.

[18] Farber, Henry, and Robert Gibbons (1996), "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111(4): pp. 1007-1047.

[19] Flabbi, Luca, and Andrea Ichino. (2001), “Productivity, Seniority and Wages: New Evidence from Personnel Data,” *Labour Economics,* 8(3): 359-387.

[20] Frederiksen, Anders (2013), “Incentives and Earnings Growth,” *Journal of Economic Behavior and Organization*, 85(1): 97-107.

[21] Frederiksen, Anders (2017), “Job Satisfaction and Employee Turnover: A firm-level perspective,” *German Journal of Human Resource Management — special issue on personnel economics*, forthcoming.

[22] Frederiksen, Anders, and Elod Takáts (2011), “Promotions, Dismissals and Employee Selection: Theory and Evidence,” *Journal of Law, Economics and Organization*, 27(1): 159-179.

[23] Frederiksen, Anders, Fabian Lange, and Ben Kriechel (2018), “Performance Evaluations and Careers: Similarities and Differences across Firms.”” *Journal of Economic Behavior and Organization*, Vol. 134, February (2018), pp. 408-429.

[24] Gaure, Simen (2014) “Correlation Bias Correction in Two-way Fixed-Effects Regression”, Stat 3(1): pp. 379-390.

[25] Gibbons, Robert, and Kevin J. Murphy (1992), “Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence,” *Journal of Political Economy*, 100(3): pp. 468-505.

[26] Gibbons, Robert, and Lawrence Katz (1991), “Layoffs and Lemons,” *Journal of Labor Economics*, 9(4): pp. 351-380.

[27] Gibbons, Robert, and Michael Waldman (1999), “A Theory of Wage and Promotion Dynamics Inside Firms,” *Quarterly Journal of Economics*, 114(4): pp. 1321-58.

[28] Gibbons, Robert, and Michael Waldman (2006), “Enriching a Theory of Wage and Promotion Dynamics Inside Firms,” *Journal of Labor Economics*, 24(1): pp. 59-107.

[29] Gibbs, Michael, and Wallace Hendricks (2004), “Do Formal Salary Systems Really Matter?,” *Industrial and Labor Relations Review,* 58(1): pp. 71-93.

[30] Greenwald, Bruce C. (1986), “Adverse selection in the labour market.” *The Review of Economic Studies,* 53(3): pp. 325-347.

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

[32] Hoffman, Mitchell, Lisa B. Kahn, and Danielle Li (2018), “Discretion in Hiring,” *Quarterly Journal of Economics*, 133(2): pp.765-800.

[33] Hoffman, Mitchell and Steven Tadelis (2018), "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis," mimeo, UC Berkeley.

[34] Holmstrom, Bengt (1979), "Moral Hazard and Observability," *The Bell Journal of Economics*, 10(1): pp. 74-91.

[35] Holmstrom, Bengt, and Paul Milgrom (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55(2): pp. 303-328.

[36] Kahn, Lisa B. (2013), "Asymmetric Information Between Employers," *American Economic Journal: Applied Economics* 5(4): pp. 165-205.

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

[38] Kaplan Steven N., Mark M. Klebanov, and Morten Sorenson (2012). "Which CEO Characteristics and Abilities Matter?" *Journal of Finance* 67(3): 973-1007.

[39] Kline Patrick, Raffaele Saggio, and Mikkel Sølvsten (2018). "Leave-out estimation of variance components". arXiv preprint arXiv:1806.01494, 2018.

[40] Lange, Fabian (2007), "The Speed of Employer Learning", *Journal of Labor Economics*, 25(1): pp. 1-35.

[41] Lazear, Edward (2000), "Performance Pay and Productivity," *The American Economic Review*, 90(5), pp. 1346-1361.

[42] Lazear, Edward, Kathryn Shaw, and Christopher Stanton (2015), "The Value of Bosses," *Journal of Labor Economics*, 33(4): pp. 823-861.

[43] Li, Jin, Niko Matouschek, and Michael Powell (2017), "Power Dynamics in Organizations," *American Economic Journal: Microeconomics,* forthcoming.

[44] MacLeod, W. Bentley (2003), "Optimal Contracting with Subjective Evaluation," *American Economic Review,* 93(1): pp. 216-240.

[45] Medoff, James L., and Katharine G. Abraham (1980), "Experience, Performance, and Earnings," *Quarterly Journal of Economics,* 85(4): pp. 703-736.

[46] __________, (1981), "Are Those Paid More Really More Productive?" *Journal of Human Resources*, 16(2): pp.186-216.

[47] Milgrom, Paul R. (1988), "Employment Contracts, Influence Activities, and Efficient Organization Design" *Journal of Political Economy,* 96(1): pp. 42-60.

[48] Pinkston, Joshua C. (2009), "A Model of Asymmetric Employer Learning with Testable Implications," *Review of Economic Studies*, 76: pp. 367-394.

[49] Prendergast, C., Topel, R.H., (1993) "Discretion and Bias in Performance Evaluation," *European Economic Review,* 37(2-3): pp. 355-365.

[50] __________. 1996. "Favoritism in Organizations," *Journal of Political Economy,* 104(5): pp. 958-978.

[51] Schonberg, Uta (2007), "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 25: pp. 651-692.

[52] Tirole, Jean (1986), "Hierarchies and Bureaucracies: on the Role of Collusion in Organizations," *Journal of Law, Economics and Organization,* 2(2): pp. 181-214.

[53] Waldman, Michael (1984), "Job Assignments, Signalling, and Efficiency," *RAND Journal of Economics*, 15(2): 255-267.

A Data Appendix

A.1 Firm Structure

The firm consists of an extensive branch network, as well as a central corporate office. We take advantage of both.

Figure A1 depicts the flow of employees between the branch network and the central office as well as flows for employees entering and exiting the firm. Churn is fairly low at this firm: about 10 percent of employment at the central office enters and exits each year, while roughly 6 percent of employment in the branch network enters and exits. There is also some movement between the branch network and the central office.

Table A1 provides summary statistics separately for the branches and the central function. Salaries are on average higher for employees in central functions. This is especially true for the bonus pool which is much larger for workers in the central functions. Other variables are fairly similar across the two samples, though we do not have KPI ratings or financial performance for workers in the central office.

Figure A2 provides some detail on the flows across different job levels in the firm. There is plenty of mobility up the hierarchy of the firm, and less mobility down, consistent with other personnel datasets (Baker, Gibbs, and Holmstrom 1994a and 1994b). Workers enter and exit from all levels of the hierarchy, though the lowest levels have much more churn.

Figure A3 zooms in on the branch network and provides both these same flows across job levels within the branch system, and also describes how workers move between the branches and the central corporate office. Promotion and demotion rates are relatively similar within the branch system, compared to the firm as a whole. There is also a modest amount of movement to and from the central office (bottom arrows at each level).

A.2 Data Construction

Performance reviews take place in March of a given year, and are meant to evaluate performance over the preceding 12 months. We associate performance in a given period with earnings over the 12 months immediately following that period. For example, performance in 2012 is the review corresponding to the period April 2011-March 2012, and earnings in 2012 are the sum of earnings from April 2012 to March 2013. This implies that in the last sample year (2014) we observe performance data but not compensation measures.

In the performance management system, we begin with a sample of 153,984 employee-year observations. After dropping 5,677 observations for which we lack basic control variables we retain 148,107 employee-year observations from roughly 24,000 unique workers. Next, we drop 24,055 observations that are part-time and 3,068 employees that are low-level staff such as cleaners or apprentices. From there, 32,554 observations lack a performance measure. As noted in the text, this is largely because the performance system took a few years to be fully rolled out. In 2004, 43 percent of the sample received performance ratings but the system spread rapidly; by 2008, 83 percent of the employees were covered. The coverage stayed at that level or slightly above throughout the remainder of the sample period. The next most important reason why observations lack performance ratings is in their first and last year at the firm employees are less likely to be present during the performance review period.[56] Dropping those without a performance measure leaves us with 88,430 observations.

In Section 2.3, we described the double fixed effects regression used to understand heterogeneity across supervisors and employees in the performance ratings that they give and receive, respectively. To create our estimation sample, we drop 3,161 observations for whom we cannot econometrically identify their employee fixed effect or associated supervisor fixed effect. The bulk of these dropped observations (2,765) come from employees who are only in the firm for a single year and thus the employee fixed effect absorbs the entire variation. The other 396 observations are dropped because some supervisors do not have enough variation once worker fixed effects, supervisor controls, and worker controls are included in the empirical specification.

These restrictions result in an estimation sample of 85,269 worker-year observations, corresponding to 57.6 percent of the full data set and 70.5 percent of the full-time sample. Within this sample, we have 77,682 observations with a compensation measure – in our last year of data, 2014, we have only performance measures and not yet a full year of compensation data.

A.3 Worker mobility

For identifying supervisor and worker fixed effects it ratings, we require that the temporary variation of performance is exogenous to the matching of workers to supervisors. This is more likely to be the case if there is no explicit process in place matching workers at particular junctions of their careers to specific supervisors. It is also more likely the case if there is a lot of mobility of workers and supervisors in the firm that is unrelated to individual variation in performance. From conversations with the firm, we know that there is no fixed procedure matching workers to supervisors nor are there any a priori reasons why temporarily low (or high)-performing workers might be matched with specific supervisors. In addition, we note that moves across supervisors are very common in this firm. In this subsection, we describe this mobility in more detail.

[56] There is some systematic variation in who receives ratings in that more stable workers (e.g., those with higher tenure and those outside of the lower job levels) are more likely to be rated. However, once the system becomes stable in 2008, observables such as tenure or job level have very little power in predicting whether an individual will be missing a performance rating.

Table A2 describes employee moves in the firm conditional on being observed in adjacent periods (t and $t+1$). Column 1 gives the distribution of moves across job levels and business units.[57] Column 2 gives the probability that workers with a given type of move (defined by the row) changed supervisors. Columns 3-5 restrict to observations where a supervisor switch actually occurs. Column 3 gives the distribution of supervisor switches across worker transition type. Column 4 gives the average change in ϕ (supervisor ratings effect) upon supervisor switch, while column 5 gives the average change in ϵ (the transitory component of ratings), conditional on the type of worker transition defined in the row.

The first thing to note is that, even while 77% of workers remain in the same position, the remainder move across job levels and/or business units. Further, such moves typically entail a supervisor switch. For instance, 66% of workers making a lateral move to another business unit switch supervisors as well. If they are also promoted, then they switch supervisors 85% of the time. Even for workers who do not move job levels or business units, we observe that supervisor switches are relatively common (30%). Because the majority of individuals remain at their position in any given year, supervisor switchers are much more likely to come from this group (64%, column 3). The next table explores these changes, showing they are primarily driven by different types of supervisor moves.

The remaining columns of Table A2 describe the change in performance components following a supervisor switch. In column 4, we show that workers do not face a systematic change in ϕ when switching supervisors. The average change overall (bottom row) of -0.006 is only 4 percent of a standard deviation in ϕ (see Table 3). This, to some extent, helps to allay concerns about endogenous sorting. If workers seek supervisors based on their ratings behavior, we would expect to see more systematic changes in ϕ. The final column explores changes in ϵ, the transient component of performance ratings. If the firm sorted workers to new supervisors at a moment when they expected the worker's performance to change, that should show up in the epsilon. Yet, for the most part, we see only small magnitudes here, especially compared to the standard deviation of ϵ, 0.35. The one exception is we see negative effects for workers newly promoted. This is consistent with some regression to the mean after an unusually positive performance generated a promotion.

Table A3 narrows in on the group of workers who remain in the same business unit and job level but still switch supervisors. In this table, we describe the destination of the outgoing supervisor and the change in performance components following the switch. Conditional on employees remaining in the same position in the firm, many supervisor switches are due to variation in the position of the supervisors. That is, for half of these switches (one minus 0.48 in the 6th row, column 1), the old supervisor leaves their current situation. These moves are primarily driven by supervisors exiting the sample (23 percent), but also come from lateral moves to other business units (12 percent), and internal promotions (8 percent). Smaller fractions of outgoing

[57]We define a business unit move as a move to another branch or to another function within corporate.

supervisors are demoted or promoted across business units. Again for most of these moves the changes in ϕ and ϵ are small in magnitude.[58]

The table probably understates the firm's need to make reassignments following such moves because there might be ripple effects within a unit. For example, someone leaving the business unit might necessitate reorganizing all the teams within.

The general message of Tables A2 and A3 is that movement of employees and supervisors across job levels and across business units constantly requires reshuffling of teams and supervisory relationships. The frequency of this reassignment should alleviate concerns that endogenous mobility threatens our inference. We can take this a step further by focusing on moves that are more plausibly unrelated to worker trends in performance. In Table A4, we present log(earnings) results, akin to those in Table 5, restricting the sample to those associated with various supervisor switches. Column 1 reproduces the baseline model relating log(earnings) to ratings components (column 1 of Table 5). Column 2 restricts the sample to observations immediately following a supervisor switch and the worker stays put (the sample in Table A3). Column 3 further restricts to situations where the supervisor switched because of a move across levels, units or exit. Finally, column 4 is restricted to a sample of supervisor exits. Reassuringly, we find fairly similar results across these subsamples, chosen so that the supervisor switch is, as best we can tell, outside the control of the worker.

[58]For supervisor demotions within the same business unit, subordinates experience a relatively large increase in ϵ of 0.063. This again could be regression to the mean, as a supervisor may have been demoted for unusually bad team performance.

B Model Details

This appendix fills in details related to the model. We restate some of the material developed in the paper itself. Many results follow immediately from known results in the literature (see for example Holmstrom [1979]) and in those cases we do not present detailed derivations.

B.1 The Basic Setup

As above and repeated here in equation 10, we assume that employee output, q, depends on effort, e, productive type, α, and a random component ε^q . ε^q is normally distributed with mean 0 and variance σ_q^2 and independent of (e, α).

$$q = e + \alpha + \varepsilon^q \tag{10}$$

We assume that the firm observes neither effort, e, nor output, q, supervisors observe q but not e, and both parties observe α.

Workers have CARA preferences $v\left(w, e\right) = -exp\left(-\psi\left(w - c\left(e\right)\right)\right)$, with a coefficient of absolute risk aversion ψ. Their preferences are additively separable in wages and effort cost $c(e)$, defined as:

$$c(e) = -\frac{1}{2\mu_s} e^2 \tag{11}$$

The parameter μ_s parameterizes the notion of heterogeneity in managerial ability: better supervisors have higher μ_s and reduce the marginal cost of effort.

Having observed q, supervisors report a rating r to the firm. 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 in equation 12:

$$u(w_s, q, r) = w_s + \tilde{\beta}_s\left(r - q\right) - \frac{\tilde{\gamma}_s}{2}\left(r - q\right)^2 \tag{12}$$

Supervisors will choose r to maximize their utility, resulting in the following reporting function:

$$r = q + \frac{\tilde{\beta}_s}{\tilde{\gamma}_s} = q + \beta_s. \tag{13}$$

The timing of the model is as follows:

1. Workers and firms sign contracts that specify the known characteristics of the supervisors that workers

are assigned to and the linear wage function. $w_{i,s} = a_{i,s} + b_{i,s}r_i$. Here we explicitly index the contract terms with both i and s, since they can depend on both the worker and the supervisor.

2. Workers meet with the supervisors they are assigned to, exert effort e, and produce q.

   (a) When we allow for incomplete information about supervisor types at the contracting stage, we assume that workers observe the actual managerial ability μ_s upon matching with their supervisors and before deciding upon effort.

3. Supervisors observe q and provide ratings r.

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

B.2 The Informed Firm and Proposition 1

We begin by assuming that firms and workers are perfectly informed about the supervisors and workers types : (μ_s, β_s, α).

Thus, wage contracts are:

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

Substituting (10) into (13) and denoting by e_s the equilibrium effort that subordinates of supervisor s exert, we get:

$$r = \alpha + (e_s + \beta_s) + \varepsilon^q = \alpha + \phi_s + \varepsilon^q \tag{14}$$

The parameter ϕ_s summarizes the variation in ratings that can be attributed to the supervisor.

The only uncertainty faced by workers at the contracting stage is about ε^q, which is normally distributed. We use well-known results on the expectation of log normal random variables (deGroot, 1970) to represent worker preferences using the certainty equivalent and express the participation constraints as follows, where I_C represents the information available during the contracting stage and e^* is the optimal effort level chosen by the worker.

$$E[w - \frac{1}{2\mu_s}e^{*2}|I_C] - \frac{1}{2}\psi var\left(w - \frac{1}{2\mu_s}e^{*2}|I_C\right) \geq \underline{u(\alpha)} \tag{15}$$

Maximizing worker expected utility subject to the linear contract delivers the optimal effort choice e^*:

$$e^* = b_s\mu_s \tag{16}$$

Worker type α enters additively in the production function and does not affect the risk-effort trade-off.

There is thus no advantage from assigning particular workers to particular supervisors. Thus, in equilibrium any assignment is viable and both positive and negative assortative matching are entirely consistent with our set-up.

Substituting the optimal effort e^* from eq. 16 into the certainty equivalent (15) and simplifying, we obtain the participation constraint:

$$a_{is} + b_s\left(\alpha + \beta_s\right) + \frac{1}{2}b^2\mu_s - \frac{\psi}{2}b_s^2\sigma_q^2 \geq \underline{u}\left(\alpha\right) \tag{17}$$

We next reproduce Proposition 1 from above, followed by the derivation.

Proposition. *Under perfect information about supervisor and worker types* $(\mu_s, \beta_s, \alpha_i)$*:*

1. The optimal piece rate is given by $b_s^* = \frac{\mu_s}{\mu_s+\psi\sigma_q^2}$*;*

2. Expected output increases one-for-one with α_i*, does not vary with* β_s*, and increases with* μ_s*;*

3. Expected compensation of workers increases one-for-one with α_i*, does not vary with* β_s*, and increases with* μ_s *iff* $b < \frac{1}{2}$*;*

4. Expected compensation of supervisors does not vary with α_i *or* β_s*, and increases with* μ_s*;*

5. Workers do not earn economic rents; that is, worker surplus $S = U(w - c(e)) = 0$*.*

The optimal piece rate b_s maximizes expected profit subject to the worker's participation constraint after substituting in the optimal effort (eq. 16). Simplifying yields the following maximization problem for the firm's choice of b_s:[59]

$$b_s^* = \underset{\{b\}}{argmax}\left\{\alpha + b_s\mu_s - \frac{b_s^2}{2}\left(\mu_s + \psi\sigma_q^2\right)\right\} \tag{18}$$

This results is the standard solution familiar from the literature and stated in point 1 of the proposition:

$$b_s^* = \frac{\mu_s}{\mu_s + \psi\sigma_q^2} \tag{19}$$

Substituting the optimal effort (equation 16) and piece rate (equation 19) into the output equation 10 results in $E\left[q|\alpha,\mu_s,\beta_s\right] = \alpha + E\left[e|\mu_s,\beta_s\right] = \alpha + b_s\mu_s = \alpha + \frac{\mu_s}{\mu_s+\psi\sigma_q^2}\mu_s$. This establishes point 2: expected output increases one–for-one with α, does not vary with β_s and increases with μ_s.

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

$$\alpha + b\mu_s - a_{is} - b_s\left(\alpha + \beta_s + b\mu_s\right) - w_s\left(\mu_s, \beta_s\right) = 0 \tag{20}$$

[59]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.

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

For expected compensation of workers (point 3), note that solving equation 20 implies that the firm will set worker pay so that their certainty equivalent exactly equals the outside option: $E[w|I_C] = \underline{u(\alpha)} + \frac{1}{2\mu_s}e^{*2} + \frac{1}{2}\psi var(w|I_C)$. From the equation 16, the optimal effort choice does not vary with the generosity of the supervisor β_s, so none of the terms in expected compensation 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 supervisors simply see their base pay reduced. Competition also implies that expected compensation increases one-for-one with α.

To determine the effect on average compensation, we set the derivative of the certainty equivalent with respect to μ_s equal to zero since we know the entire surplus is extracted from workers:

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

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_s (see eq. 19).[60] Thus, we obtain

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

Expected worker compensation is thus increasing in μ_s when $b_s < \frac{1}{2}$ and is otherwise decreasing.

Regarding the compensation of the supervisor (point 4), note that the zero profit condition (equation 20) implies that worker wages will be set at their outside option. Since effort and worker compensation do not vary with β_s, neither does the surplus across worker-supervisor pairs. Thus supervisor compensation will not vary with β_s either. Furthermore, worker ability, α_i, is given entirely to the worker so it will not enter the supervisor's pay. In contrast, 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: $\frac{\partial w_s(\mu_s)}{\partial\mu_s} > 0$.

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

B.3 The Partially Informed Firm and Proposition 2

To capture the partial lack of information in a tractable manner we assume that (μ_s, β_s) are independent normally distributed random variables with variances σ_β^2 and σ_μ^2 and we assume that agents 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 parameterize 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, uncertainty now includes uncertainty about the signal noise ε^q as well as (μ_s, β_s). A contract is now an assignment to $\left(\beta_s^E, \mu_s^E\right)$ and a linear wage contract specifying the relation between reported ratings and compensation conditional on the assignment.

Given the distributional assumptions made and using the CARA preferences, we can rewrite the participation constraint using the certainty equivalent which now reads:

$$a + b\left(\alpha_i + \beta_s^E\right) + b^2 \frac{\mu_s^E}{2} - \frac{\psi}{2}\left(b^2\left(\theta_\beta \sigma_\beta^2 + \sigma_q^2\right) + \frac{b^4}{4}\theta_\mu \sigma_\mu^2\right) \geq \underline{u}\left(\alpha\right) \tag{21}$$

This certainty equivalent depends on how much is unknown about (μ_s, β_s) which is parameterized by $\theta_\beta \sigma_\beta^2$ and $\theta_\mu \sigma_\mu^2$. The unknown variation in β_s and μ_s represents risk from the point of view of the worker since it will affect her compensation and effort costs. The certainty equivalent (21) accounts for this risk.

Upon meeting a supervisor, employees observe the marginal cost of effort μ_s. As before, we can solve for the optimal effort choice, which again is $e = b_s \mu_s$. The firm's problem is to maximize expected profits from any given worker-supervisor pair, which reads:

$$\Pi\left(\mu_s^E, \beta_s^E, \alpha\right) = \underset{\{a,b\}}{Max}\left\{\alpha + b\mu_s^E - a_i - b_s\left(\alpha + \beta_s^E + b_s \mu_s^E\right) - w_s\left(\beta_s^E, \mu_s^E\right)\right\} \tag{22}$$

s.t. the participation constraint (21).

And, as before, firms compete in the market for workers and supervisors so that in equilibrium expected profits conditional on $\left(\alpha, \beta_s^E, \mu_s^E\right)$ equal zero.

We can now derive the implications of Proposition 2, which we repeat here.

Proposition. *Under imperfect information about supervisor type (μ_s, β_s):*

1. The optimal piece rate is the unique implicit solution to $\mu_s^E = b_s\left(\mu_s^E + \psi\left(\theta_\beta \sigma_\beta^2 + \sigma_q^2 + b_s^2 \frac{\theta_\mu \sigma_\mu^2}{2}\right)\right)$;

2. Expected output conditional on $\left(\mu_s^E, \beta_s^E, \alpha\right)$ does not vary with β_s^E and increases with μ_s^E. Expected output conditional on (μ_s, β_s, α) does not vary with β_s and increases with μ_s. Both increase one-for-one in α_i;

3.Expected compensation of workers conditional on $\left(\mu_s^E, \beta_s^E, \alpha\right)$ does not vary with β_s^E . The relationship with μ_s^E cannot be globally signed. Expected compensation of employees conditional on (μ_s, β_s, α) increases with β_s. Its relationship with μ_s also cannot be globally signed. Both increase one-for-one with α_i;

4. Expected compensation of supervisors conditional on $\left(\mu_s^E, \beta_s^E, \alpha\right)$ does not vary with α or β_s^E but increases with μ_s^E. Expected compensation of supervisors conditional on (μ_s, β_s, α) does not vary with α or β_s but increases with μ_s;

5. Worker surplus $S = U(w - c(e))$ does not vary with μ_s^E and β_s^E but increases in μ_s and β_s.

The optimal loading is implicitly determined by the FOC of eq. 22:

$$\mu_s^E = b_s\left(\mu_s^E + \psi\left(\theta_\beta \sigma_\beta^2 + \sigma_q^2 + b_s^2 \frac{\theta_\mu \sigma_\mu^2}{2}\right)\right) \tag{23}$$

The right-hand side of this expression increases monotonically in b and there is thus a unique loading that solves the firm's problem (point 1). Furthermore, as is apparent from equation 23, the optimal piece rate declines in $\theta_\beta \sigma_\beta^2$ and $\theta_\mu \sigma_\mu^2$.

We can still write expected output as $q = b\mu_s + \alpha + \varepsilon^q$ (where $b\mu_s$ is still the optimal effort choice). And this still increases one-for-on with α, does not vary with β_s (or β_s^E), and is increasing in μ_s (and μ_s^E). This establishes point 2.

For expected compensation of workers, we can again rely on similar arguments for Proposition 1. As before, the firm extracts any surplus from workers during the contracting stage. Again, competition in the labor market implies that expected compensation increases one-for-one withα. And, as before, expected compensation does not depend on the known variation in leniency bias β_s^E. This is because it enters the workers participation constraint (eq. 21) only through the expected wage. The firm can extract any variation in β_s^E using the intercept of the wage contract and thus make the expected wage independent of β_s^E.

We thus rewrite expected compensation as (24), which is additively separable in α and a function that depends on μ_s^E only, and the pay for performance piece (a function of optimal effort and the unexpected ratings boost due to leniency).

$$\begin{aligned} E\left[w|\alpha, \beta_s^E, \mu_s^E, \beta_s, \mu_s\right] = & \quad \alpha + h\left(\mu_s^E\right) + b\left(\left(\beta_s - \beta_s^E\right) + b\mu_s\right) \\ = \alpha + h\left(\mu_s^E\right) + b\varepsilon_\beta + b^2\mu_s & \quad = \alpha + h\left(\mu_s^E\right) + b\theta_\beta\beta_s + b^2\mu_s + b\nu_\beta \end{aligned} \tag{24}$$

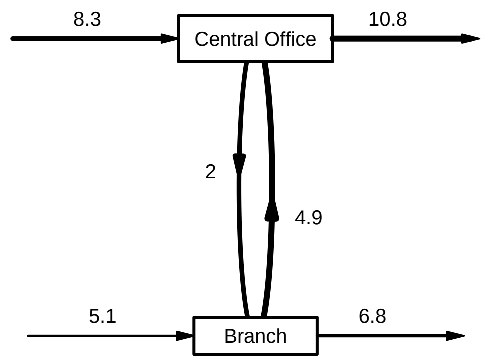
where we substitute in the linear projection of $\varepsilon_\beta = \frac{cov(\varepsilon_\beta, \beta_s)}{var(\beta_s)}\beta_s + \nu_\beta = \frac{cov\left(\varepsilon_\beta, \beta_s^E + \varepsilon_\beta\right)}{var(\beta_s)}\beta_s + \nu_\beta = \theta_\beta\beta_s + \nu_\beta$.

By the same logic as before, we cannot sign how expected employee compensation relates to μ_s^E. 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. Finally, since output increases in μ_s, compensation also increases. This establishes point 3.

For point 4, supervisor compensation, we note that, as before, expected output of a worker-supervisor pair net of worker compensation does not vary with β_s^E or β_s, and increases in μ_s^E and μ_s. Thus, earnings of the supervisor are independent of β_s^E and increase in μ_s^E.

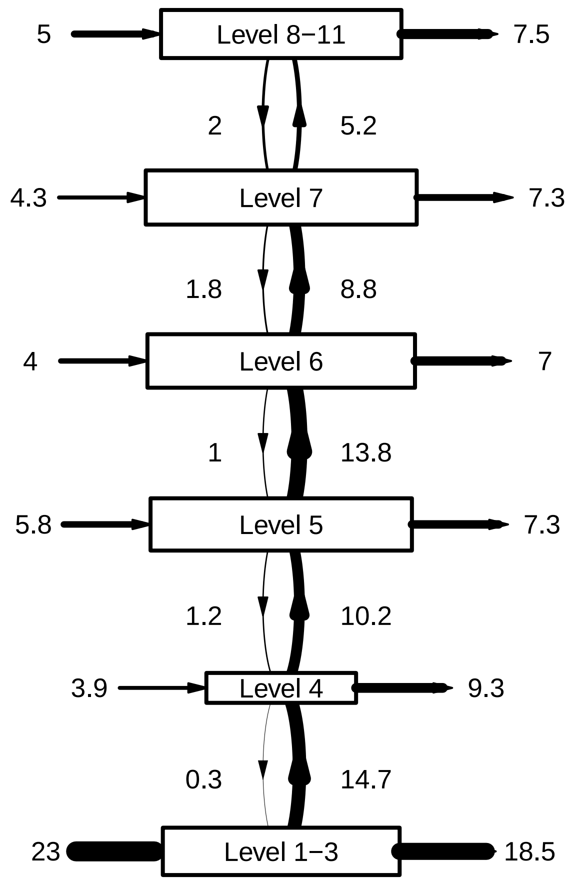
C Appendix Figures and Tables

Figure A1: Flows Across Central Office and Branch Network



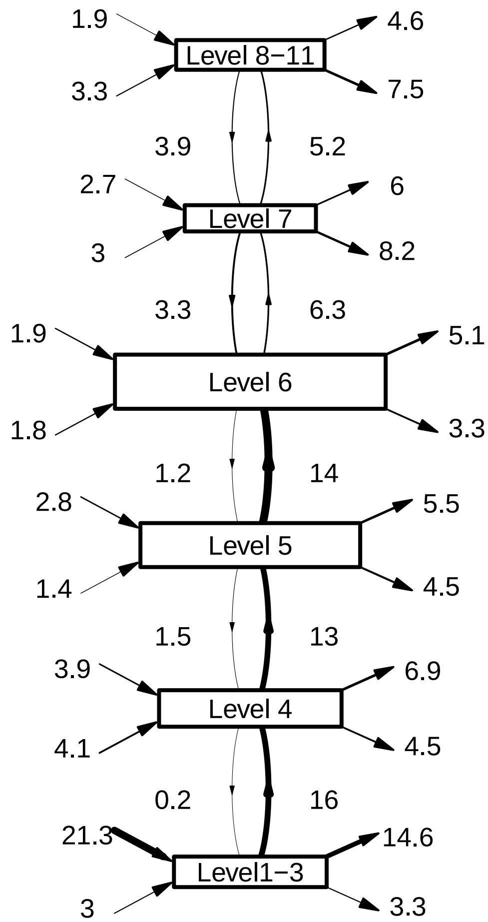
The diagram shows flows across different parts of the firm – the central office and the branch network. The rectangles are proportional to the number of employees in each part. On the left are the percentages of the employees at an indicated part entering the firm. On the right are those leaving. For example, 9.8 percent of employees working at the central office entered in the preceding year; 10.6 percent left the firm. Flows between branches and central are expressed as a percent of the origin part. For example, 2.4 percent of the employees in the central office switch to the branches. The numbers are averages over our sample period.

Figure A2: Flows Across Job Level



The diagram shows flows across different job levels for the firm as a whole. The rectangles are proportional to the number of employees at each level. On the left are the percentages of the employees at an indicated level entering the firm. On the right are those leaving. For example, 5.2 percent of employees working at level 4 entered the firm into that level in the preceding year. 8.7 percent quit or were laid off from that level in a given year. The flows between job levels are expressed as a percent of the origin level. For example, 9.6 percent of workers in level 4 are promoted to level 5 in a given year; 0.2 percent are demoted. The numbers are averages over our sample period. Flows across job levels not indicated in the graph (for example a move from level 4 to level 6) are rare.

Figure A3: Flows within the Branch Network



The diagram is restrict to the branch network of the firm. It is similar to Figure A2, except we also include flows between the branch system and the central office. The rectangles are proportional to the number of employees at each level. On the left are the percentages of the employees at an indicated level entering the branches, either from outside the firm (top arrow), or from the central office (bottom arrow). The arrows to the right analogously show exits from the firm (top) or to the central office (bottom). For example, 2.7 percent of employees working at level 4 in the branch network entered the firm into that level in the preceding year; 3 percent of employees entered level 4 in the branch network from the central office; 6 percent of workers in level 4 of the branch network leave the firm in a given year and 8.2 percent move to the central office. The flows between job levels are expressed as a percent of the origin level, and restrict to moves within the branch network. For example, 11 percent of workers in level 4 are promoted to level 5 in a given year; 0.2 percent are demoted; the level receives 9.2 percent of those working in levels 1-3 in a given year. The numbers are averages over our sample period. Flows across job levels not indicated in the graph (for example a move from level 4 to level 6) are rare.

Table A1: Summary Statistics, Branches and Corporate

	Central Function			Branch Network		
	Mean	Std. Dev.	N	Mean	Std. Dev	N
Outcomes:						
Pass	0.49	0.50	48,075	0.43	0.49	37,194
Earnings[1]	1.99	1.33	43,275	1.67	0.46	34,407
Received bonus	0.32	0.47	43,275	0.29	0.46	34,407
Bonuses (including zeros)[1]	0.16	0.91	43,275	0.03	0.09	34,407
Stay in sample[2]	0.83	0.38	43,275	0.84	0.37	34,407
Stay in business unit[2]	0.78	0.41	43,275	0.65	0.48	34,407
Stay with supervisor[2]	0.54	0.50	43,275	0.51	0.50	34,407
Promotion[3]	0.11	0.31	41,839	0.11	0.31	33,358
Demotion[3]	0.01	0.09	41,839	0.02	0.14	33,358
Two-year layoff rate[2]	0.01	0.10	38,316	0.01	0.10	31,211
KPI Rating	na	na	na	0.53	0.28	7,871
Financial performance	na	na	na	-0.074	0.126	2,502
Bottom-Up Evaluation	4.68	1.06	41,795	4.78	0.92	33,198
Controls:						
Full-time	1.00	0.00	48,075	1.00	0.00	37,194
In Branches	0.00	0.00	48,075	1.00	0.00	37,194
Age	44.16	10.20	48,075	43.85	11.25	37,194
Tenure	16.12	12.84	48,075	20.38	13.47	37,194
Female	0.40	0.49	48,075	0.49	0.50	37,194
Supervisor Age	45.00	7.87	48,075	45.46	8.06	37,194
Supervisor tenure	16.97	11.37	48,075	22.99	10.92	37,194
Supervisor Female	0.32	0.47	48,075	0.22	0.41	37,194

Note: The summary statistics are reported for the sample used to estimate the fixed effects in the ratings equation (see section 2). Not all variables are available for all observations in this sample. "Pass" is our constructed performance measure that equals 1 if the subjective performance evaluation was 4 or 5, and equals 0 if it was 1, 2, o Stay in firm, in business unit, with supervisor, promotion and demotion refer to any change in the worker's status over the next year. Business unit is the branch or functio the central office. KPI rating is the branch-level ranking divided by the number of peer branches in the comparison set. Financial performance is the year-over-year growth of the individual's financial portfolio. Bottom-up evaluation is the average of seven questions workers answer regarding their satisfication with their supervisors. Response range from 1 to 10; we average answers on all responses and norm the variable to have a standard deviation of 1. "In Branches" equals 1 if the worker was in the branch network and 0 if in the central corporate office.

1) Divided by average earnings in the country. Income variables not available in last year of data, 2014. 2) Restricted to not right-censored obs, excluding the last year of d (last two years for layoff rate). "Stay in sample" denotes the probability of being retained in the estimation sample in the following year. By far the most common reason f leaving the sample is to leave the firm within 2 years. 3) Restricted to not right-censored obs that did not quit or get laid off in respective year.

Table A2: Worker Mobility and Supervisor Changes

Type of employee transition		**All present in t and t+1**		**Conditional on Supervisor Switch**		
		Distribution	**Supervisor Change**	**Distribution**	**Change in ϕ**	**Change in ε**
Changed Business Unit	Promoted	0.03	0.85	0.06	-0.004	-0.128
	Demoted	0.005	0.92	0.01	0.009	0.010
	Same Level	0.10	0.66	0.18	-0.003	-0.027
Same Business Unit	Promoted	0.09	0.38	0.09	-0.014	-0.067
	Demoted	0.01	0.50	0.01	0.003	0.008
	Same Level	0.77	0.30	0.64	-0.006	0.014
	All	**1.00**	**0.37**	**1.00**	**-0.006**	**-0.01**
Observations		64,691		23,666		

Notes: Table is restricted to workers present in the firm in t and t+1. Column (1) shows the distribution of moves by business units and job levels (shares sum to 1); column (2) shows the share changing supervisor, conditional on the transition type. Columns 3-5 restrict to observations where a supervisor switch did occur (as well as the indicated type of worker transition); column (3) shows the distribution across transition types, conditional on superisor switch; column (4) shows the change in φ (the supervisor fixed effect in ratings), conditional on switching; column (5) shows the change in ε (the transitory ratings component), conditional on switching. For row 1, for example, column (1) shows that 3% of all observations present in the firm in adjacent years are promoted and switch to a new business unit. Among this group (promoted and changed business units), 85% change supervisors (column 2). These supervisor changes make up 6% of all supervisor switches (column 3). The supervisor change results in an average change in phi (the supervisor ratings effect) of -0.004 (column 5) and an average change in ε (the transitory ratings component) of -0.128 (column 6).

Table A3: Supervisor Mobility for Workers who remain in the Same Position but Switch Supervisors

Type of supervisor transition		**(1) Distribution across categories**	**(2) Change in ϕ following switch**	**(3) Change in ε following switch**
Changed Business Unit	Promoted	0.04	0.038	0.020
	Demoted	0.03	0.002	0.010
	Same Level	0.12	0.003	0.014
Same Business Unit	Promoted	0.08	0.002	0.006
	Demoted	0.03	-0.037	0.063
	Same Level	0.48	-0.006	0.015
Left Firm/Sample		0.23	-0.006	0.005

Notes: The table shows where supervisors of workers in the prior period moved conditional on a worker experiencing a supervisor switch. The table is restricted to workers present in the firm in t and t+1 who remain in the same job level and business unit but switch supervisors (11,639 worker-year observations). Column (1) shows the distribution of moves of the supervisor, by business units and job levels (shares sum to 1); column (2) shows the change in φ (the supervisor fixed effect in ratings), conditional on the type of supervisor move; column (3) shows the change in ε (the transient component of ratings), conditional on the type of supervisor move. For row 1, for example, column (1) shows that for 4% of all employees who switch supervisors between t and t+1 (while remaining in the same job level and business unit), their supervisors are promoted and move business units. At the same time, these workers experience an increase in φ of 0.038 (column 2) and an increase in ε of 0.02 (column 3) upon switching supervisors.

Table A4: Log(Earnings) and Ratings Components, Restricted to Certain Supervisor Moves

Dependent Variable: Log(Earnings)	(1)	(2)	(3)	(4)
Supervisor ratings effect (φ)	0.095***	0.084***	0.090***	0.144***
	(0.013)	(0.014)	(0.021)	(0.041)
Worker ratings effect (α)	0.098***	0.091***	0.101***	0.106***
	(0.003)	(0.004)	(0.005)	(0.010)
Pass residual (ε)	0.021***	0.009***	0.010**	0.010
	(0.001)	(0.003)	(0.004)	(0.008)
Observations	77,682	8,918	4,503	1,791
R-squared	0.818	0.844	0.843	0.825
	Split-Sample IV			
Supervisor ratings effect (φ)	0.117***	0.115***	0.140***	0.203**
	(0.023)	(0.025)	(0.037)	(0.083)
Worker ratings effect (α)	0.117***	0.118***	0.127***	0.135***
	(0.004)	(0.007)	(0.009)	(0.018)
Observations	74,641	8,537	4,334	1,697
R-squared	0.814	0.836	0.834	0.816
Worker Stays, Supervisor Switches		X	X	X
Supervisor Moves Levels, Units or Exits			X	X
Supervisor Exits				X

Notes: Column 1 presents OLS regressions of log earnings on ratings components for the full sample (see table 5). Column 2 restricts to observations where the worker remained in the same job level and business unit between t and t-1 but experienced a change in supervisor. Column 3 further restricts to observations where the t-1 supervisor moved job levels, business units, or left the sample. Column 4 further restricts to observations where the t-1 supervisor left the sample. The split-sample IV estimates supervisor and worker effects in even and odd years, separately, and use estimates in even years as instruments for estimates in odd years and vice versa. All regressions also include controls listed in Table 3. Standard errors are clustered by supervisor. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Full Earnings Dynamics and Supervisor Heterogeneity

Dependent variable	Log earnings										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Supervisor FE (ϕ):											
Contemporaneous ϕ	0.095***	0.096***	0.058***	0.088***	0.048***	0.076***	0.037***	0.068***	0.030**	0.068***	0.043***
	(0.013)	(0.014)	(0.011)	(0.014)	(0.012)	(0.015)	(0.012)	(0.017)	(0.014)	(0.019)	(0.016)
Lag 1 ϕ			0.050***		0.026***		0.019**		0.017*		0.002
			(0.008)		(0.007)		(0.009)		(0.009)		(0.012)
Lag 2 ϕ					0.032***		0.016**		0.013		0.018
					(0.008)		(0.007)		(0.009)		(0.012)
Lag 3 ϕ							0.028***		0.016*		0.002
							(0.009)		(0.009)		(0.010)
Lag 4 ϕ									0.020**		0.018
									(0.010)		(0.012)
Lag 5 ϕ											0.010
											(0.009)
Restricted		X		X		X		X		X	
Observations	77,682	57,828	57,828	42,642	42,642	31,418	31,418	22,609	22,609	15,430	15,430
Partial R-squared	0.818	0.813	0.813	0.814	0.815	0.820	0.821	0.821	0.822	0.824	0.824

Notes: See table 11. "Restricted" samples include observations with non-missing values for the number of lags shown in the next column. All regressions contain the same number of lags in (ε) as in (ϕ) and control for the same set of controls as in the main specification reported (Table 5). Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.