Supervisors and Performance Management Systems

Anders Frederiksen, Lisa B. Kahn, and Fabian Lange

Aarhus University, Yale School of Management, and McGill University[*†]

April 11, 2017

Abstract

Supervisors occupy central roles in production and performance monitoring. 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 observed supervisor heterogeneity in ratings to worker, supervisor, and firm outcomes. We find that subordinates are paid significantly more, and their pay is more closely aligned with performance, when they are matched to a high-rating supervisor. 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 7-14% of an annual salary.

[*]We are grateful for helpful comments from seminar participants at the GAPE conference at Aarhus Univeristy, 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 Illinoiss, Queens, NBER Summer Institute, Zurich University, and Royal Holloway, London.

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

1 INTRODUCTION

Subjective performance evaluations are 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, supervisor 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 drivers of ratings heterogeneity across supervisors, the degree to which firms are informed about any heterogeneity, and the impact of such heterogeneity on workers' careers inside the firm.

If supervisors give different ratings for the same underlying performance, then the performance management system will be ineffective. Workers will dislike bearing unnecessary risk and firms will be limited in their 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 how much 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. Firms would very much like to be informed about these differences so they can reward higher effort. Indeed, the quality and type of information available on worker performance is central to the theory and practice of optimal incentive design and ownership assignment.

In this paper, we strive to understand the degree and 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 find that there is substantial heterogeneity in ratings across supervisors: we estimate that a worker receives a 32 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 7 to 14 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. Because we find that both subordinates and supervisors fare better, in terms of pay, when the supervisor is a higher rater, and teams perform better on objective output metrics when managed by a high rater, we conclude that heterogeneity in ratings is driven primarily by real differences in managerial ability that firms are at least partially informed about.

In our model, 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 choose to exert. Supervisors observe worker output and report on this output to the firm. In our model, supervisors differ along two dimensions. First, they differ in how much weight they place on reporting ratings truthfully as opposed to favorably, which we refer to as "leniency bias."[1] Second, they differ in their managerial ability, which affects the marginal costs of effort on the part of their subordinates. 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 show that there is substantial ratings heterogeneity across supervisors. In a regression of performance ratings on supervisor and worker fixed effects, as well as controls, we find that a one-standard-deviation increase in the supervisor effect is associated with a 32 percent increase in reported performance. This estimate adjusts for the well-known correlated measurement error problem inherent in double fixed effects models following Card, Heining, and Kline (2013, CHK hereafter).[2] Second, we 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 CHK correction.[3]

[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] CHK show how to estimate the variance-covariance matrix of the unobserved effects in the well-known double fixed effects equation used in the study of wage dynamics across workers and firms (Abowd, Kramarz, and Margolis, 1999).

[3] In our analysis we adapt the CHK methodology to a stacked system of equations with a double fixed effect structure to obtain unbiased estimates of the variance-covariance matrix of the unobserved components in performance ratings and earnings, and we show how to adjust coefficients in a regression of the unobserved components in performance ratings on earnings.

We find that subordinates of a high rater are paid more than subordinates matched to lower raters. This finding could be explained, in the context of our model, as being driven by heterogeneity in managerial ability, or by heterogeneity in leniency about which the firm is uninformed. However, we also find that teams managed by higher raters tend to outperform those managed by lower raters on a set of objective criteria available at the branch level. This outcome can only be accounted for in our model by differences in managerial ability to elicit output. Two further findings corroborate this interpretation. First, we find that higher raters tend to 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 ratings heterogeneity across supervisors stems from their ability to impact the marginal cost of effort. Finally, from self-reports, we also know that workers matched to higher raters are more satisfied with their immediate supervisors, and we find they are less likely to quit or change supervisors. 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 can explain 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 and optimally set stronger pay-for-performance components to the subordinates whose marginal cost of effort is reduced by better managers. However, the evidence from quitting behavior and worker satisfaction surveys also suggests that employees do earn economic rents from working for higher 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 from these employees.

We go beyond the confines of our static model to quantify the effects of rater heterogeneity for workers' careers inside the firm. We allow supervisors to dynamically impact worker pay, both directly and through promotion probabilities. As mentioned above, we find that assignment to higher rating supervisors has substantial effects on individual compensation. These arise because the effects on pay persist for some time and because being matched with a high raters increases the odds of a promotion. We thus conclude that better managers have large and real impacts on the careers of their subordinates.

Our work relates to various literatures in personnel and labor economics. The literature on the productivity effects of managers predominantely studies upper management and CEOs (Bennedsen, Perrez-Gonzales, Wolfenzon 2007; Bertrand and Schoar, 2003; Kaplan, Klebanov, and Sorensen 2012). To our knowledge, Lazear, Shaw, and Stanton (2015) is the only paper in this literature who, as we do, explore how productivity varies across supervisors lower in the firm hierachy. They exploit the daily rotation of line managers to estimate variation among subordinates in a low-skilled service task (transactions per hour) associated with these managers. 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 perfomance metrics are difficult to craft and firms instead rely on subjective ratings. Our analysis sheds light on the wide-spread and growing practice on the crucial role managers play in performance management systems that rely on subjective ratings of performance.

Previous researchers and practitioners have worried about the value of performance ratings given their subjective nature (Medoff and Abraham 1980, 1981). 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 far improves our understanding of performance management inside the firm.

Even though we allow for and find evidence of 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 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 perfor-

[4]An overview of this literature can be found in Frederiksen, Lange and Kriegel 2017. They summarize empirical patterns in the data on subjective performance evaluations that has been exploited in 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).

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

mance evaluations even when, as is 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.[6] On a practical level, this suggests 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.[7]

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. 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 from a large Scandinavian service sector firm. The firm consists of a central corporate office and an extensive branch network. The branches comprise 44 percent of workers. From Figure 1, churn is fairly low at this firm, with roughly 10 percent of employment at the central office and 6 percent of employment in the branch network entering and exiting in each year. There is also some movement between the branch network and the central office. Across branches, jobs are comparable and involve close client contact. In 2013, there were several hundred branches.[8] The median branch had 17 employees.

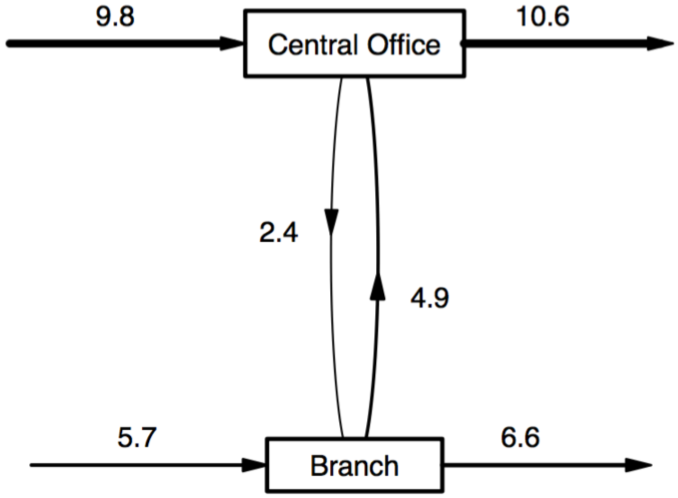
The firm has 11 identifiable job levels (see appendix figures A1 and A2). Workers in the central corporate office have a variety of functions and there are more high-level jobs. In contrast, the typical branch has a branch manager (levels 9–11), a deputy branch manager (levels 8–9), 6–9 senior workers in client-facing roles (level 6), 5–7 junior workers in client-facing roles (levels 3–5), and sometimes a trainee (level 1). We pay

[6]Bloom and Van Reenen (2007) show substantial heterogeneity in management practices across firms.

[7]For theoretical work on the trade-off between rules and discretion or power inside firms, see Bolton and Dewatripont (2012), Dessein (2002), Alonso and Matouschek (2008), Aghion and Tirole (1997), and Li, Matouschek, and Powell (2017).

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

Figure 1: Flows Across Central Office and Branch Network



The flow diagram shows flows across different sectors of the firm – the central office and the branch. The rectangles are proportional to the number of employees in each sector. On the left are the percentages of the employees at an indicated sector entering the firm. On the right are those leaving the firm. The percentages referring to flows between sectors are percentages of the origin sector. For example, 9.8 percent of employees working at the central office entered the sector in the preceding year. 10.6 percent are quitting or are being laid off from the firm. 2.4 percent of the employees switch to the branch. The numbers are averages over our sample period.

particular attention to the branch network because we have access to objective (financial-, and customer-satisfaction-based) performance measures of the branches. We have no such measures for the central office functions.

The bulk of our data stems from the performance management system of this firm and our sample comprises all employees engaged in domestic activities between 2004 and 2014.[9] The performance management system was introduced just prior to 2004, where our data begins, and the system was rolled out in the following years. In 2004, 42 percent of the employees received performance ratings but the system spread rapidly so that by 2008 almost 82 percent of the employees were covered. The coverage stayed at that level or slightly above throughout the remainder of the sample period.

As part of the performance management system, each worker receives a performance rating that is meant to describe their aggregate performance in a given year. In the branch network, ratings are typically given by the branch manager, but we also observe that deputy branch managers rate employees. Across the firm we find that the typical manager has a span of control of 10.47 (s.d. of 6.88).

The performance ratings range from 1 (unsatisfactory) to 5 (outstanding). The distribution of the

[9]The firm is a market leader within the domestic market with some international activities. We focus our analysis on the domestic workforce.

Table 1: Performance Distribution

	Fail			Pass	
Rating	1	2	3	4	5
frequency	0.11%	3.00%	49.33%	41.28%	6.27%
	52.44%			47.56%	

Note: This table is based on the estimation sample, which consists of those 77,077 individuals with 2 or more ratings for whom we can estimate fixed effects (see table 2).

performance score is shown in Table 1. As is common among companies using performance management systems (Frederiksen, Lange, and Kriechel, 2017), the tails of the distribution are rarely used and ratings are concentrated in a small subset of the support: 90 percent of ratings are either a 3 or a 4. For this reason, our empirical investigation is built around a "pass-fail" 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."

In our data we have a total of 136,286 employee-year observations.[10] We drop small fractions of these observations that are part-time workers or low-level staff such as cleaners and apprentices.[11] Most importantly, we drop 32,462 observations because they lack a performance measure. Many of these come from the years prior to 2008 when the system was not fully implemented across the whole firm, or from individuals in their first or last year at the firm who are less likely to be present during ratings collection.[12] Finally, we drop an additional 3,037 observations because they come from supervisors with fewer than two subordinates or from a small number of additional supervisors who do not yield independent variation once worker fixed effects and controls are included in the empirical specification. This results in our estimation sample of 77,077 worker-year observations, corresponding to 56.5 percent of the full data set and 67.5 percent of the full-time sample.

Table 2 provides summary statistics for the full sample as well as for the estimation sample. We report earnings (and its components) relative to average per capita earnings in this country. The workforce at this firm is highly educated and the firm is known to be an attractive employer. Thus earnings at this firm average 185 percent of the national mean. In the data it is possible to distinguish between base pay and annual bonus. Roughly 30 percent of the workers receive a bonus, and the bonus pool has historically been close to 5 percent of the wage pool.

[10]This reflects roughly 20,000 unique workers, though the firm requests we do not reveal the exact number. This sample includes all workers for whom we have pay information and basic data (demographics, tenure, job level etc.). This includes all but 4 percent of the raw data we received from the firm.

[11]We drop 22,344 observations that are part-time and 1,466 of the remaining observations because the subordinate is rated by a supervisor below job level 3, revealing themselves to be cleaning staff or apprentices and therefore outside the scope of our analysis.

[12]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 account for only a small fraction of the variation in missing ratings.

Of those employees present in consecutive years, about 10 percent are promoted and less than 1.5 percent are demoted annually. Supervisor relationships are somewhat sticky; about half of the workers present at the firm across adjacent periods and neither promoted nor demoted keep the same supervisor. Finally, in our regression sample, 5 percent of workers quit in the next two years while 1 percent are laid off.[13]

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 1.6 more years tenure in the firm (19.7 vs. 18.1 years).

Our data contains two measures of objective performance. For the period 2007–2010, we have branch-level rankings that are based on a set of Key Performance Indicators (KPIs) within a group of peer branches, defined by the firm. The KPIs include measures of financial performance of the branches, as well as other metrics (for example, customer satisfaction). The set of KPIs changes from year to year as the firm's focus evolves. Branches are placed into peer groups based on size and customer base, and these peer groups vary from year to year. The average peer group has 17 branches. We call these branch rankings "KPI rankings" hereafter.

Our second measure of objective performance reflects financial performance and is available at the individual level. We cannot reveal the precise content of these financial measures, 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 percentage increase in the value of the portfolio across two consecutive years. We refer to these measures as "financial performance" hereafter.

We also have access to job satisfaction surveys. These surveys include questions about the employees' perceptions of supervisor performance.[14] 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. On average employees rate their supervisors at 8.24 with a standard deviation of 1.50.[15]

[13] We take a two-year perspective because workers are less likely to receive a performance review in their last year at the firm. These rates are higher in the sample as a whole, 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] 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.

[15] It is unusual to have employee satisfaction data merged with personnel files (Frederiksen, 2017). The reason is that 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 (GSOP), 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.

Table 2: Summary Statistics

	Full Sample		Estimation Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Outcomes:				
With performance	0.714	0.452	1	0
Pass	0.456	0.498	0.476	0.499
Earnings[1]	1.73	1.17	1.85	1.05
Bonuses (including zeros)[1]	0.10	0.70	0.10	0.69
Received bonus	0.307	0.461	0.310	0.462
Promotion[2]	0.108	0.310	0.111	0.314
Demotion[2]	0.014	0.116	0.012	0.111
Same supervisor[3]	0.551	0.500	0.506	0.500
Quit[4]	0.123	0.328	0.052	0.222
Layoff[4]	0.033	0.178	0.010	0.097
Controls:				
Full-time	0.836	0.370	1	0
Tenure	17.5	13.5	18.1	13.3
Age	43.5	11.3	44.0	10.7
In branch	0.445	0.497	0.443	0.497
Female	0.519	0.500	0.439	0.496
Supervisor female	0.278	0.448	0.270	0.444
Supervisor age	44.2	10.4	45.2	8.0
Supervisor tenure	19.6	11.8	19.7	11.6
Observations	136,286		77,077	

Note: The full sample consists of all observations between 2004 and 2014 with a wage measure and basic job characteristics. The estimation sample consists of observations that additionally have performance measures, are working full-time, and are not a cleaner or apprentice or other low-level employee outside our scope. We restrict to workers and supervisors for whom we can estimate double fixed effects specifications.

1) Reported relative to average earnings in the country.

2) One-year promotion/demotion probabilities are summarized for the sample of workers present at the firm across adjacent periods, 124,805 observations in the full sample and 74,602 in the estimation sample.

3) We report the probability that a worker remains with the same supervisor in the following year for workers present at the firm across adjacent years and not promoted or demoted, 109,622 in the full sample and 65,415 in the estimation sample.

4) We report two-year quit/layoff probabilies since workers are less likely to receive performance evaluations in their last year, and thus exclude observations in the last year of data, keeping 123,860 for the full sample and 69,350 in the estimation sample.

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.

2.2 Variation in Performance Measures

We now demonstrate that supervisors differ substantially in how they rate their subordinates. 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 worker controls (X_{it}) and supervisor controls ($Y_{s(i,t),t}$):[16]

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

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 3.1 different supervisors. Employees who were with the firm throughout the entire period had on average 3.9 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 10.47 (s.d. of 6.88) employees in a given year and 25.9 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, ideal for estimating the fixed effects that we require.

We would like to estimate the variation in the unobserved effects $\{\alpha_i, \phi_s\}$ using the variation in the estimated fixed effects $\left\{\hat{\alpha}_i, \hat{\phi}_s\right\}$. However, we cannot do so directly since we run into an 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. These fixed effects are thus 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 and the estimation error can be expected to correlate across worker and supervisor effects.[17]

[16] The controls include for the worker (X_{it}) indicators for five-year age and tenure groups, gender, and job level. For the supervisor (Y_{st}) the controls include indicators for five-year age groups, gender and job level. We also control for business unit indicators (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.

[17] 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

Table 3: Variances of Ratings Components

	(1) Unadjusted	(2) CHK Correction
Var(worker effects) (α)	0.136	0.074
Var(supervisor effects) (ϕ)	0.033	0.023
Var(residual) (ε)	0.110	0.124
Covariance(α,ϕ)	-0.016	-0.011

Notes: See Section 2.2 and Appendix A.1. Column 1 reports unadjusted estimates from a regression of worker fixed effects (α), supervisor fixed effects (ϕ), and controls; ε are the residuals. CHK correction adjusts the column 1 estimates based on the variance-covariance matrix of the estimation error of the fixed effects. Controls include five-year age and tenure groups, gender, and job level of the worker, five-year age groups, gender, and job level of the supervisor, business unit indicators (whether or not the worker is in a branch network), and year fixed effects.

To account for this problem, we adapt the approach of Card, Heining, and Kline (2013) and adjust 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 and their correlation compared to a naive estimator. See Appendix A.1 for more details on this procedure, which we hereafter term the CHK correction.

Table 3 shows both unadjusted (column 1) and CHK-adjusted (column 2) estimates of the second moments of α, ϕ, and ϵ^p. The adjustment for sampling error has a modest effect on the moments, reducing their magnitude by a third to a half. Either way, we find that ϕ_s varies substantially across supervisors. Using the CHK adjusted moments in column 2, a one-standard-deviation increase in ϕ_s amounts to a 15.3 percentage point (32%) increase in the probability of receiving a passing grade. Thus, a move from the 10th to the 90th percentile in the distribution of ϕ_s, assuming that ϕ_s is normally distributed, is associated with a 39 percentage point increase in the probability of receiving a passing grade. The heterogeneity at the worker level is even larger — a standard deviation increase in α_i amounts to a 27.2 percentage point increase (57%) in the probability of receiving a passing grade.

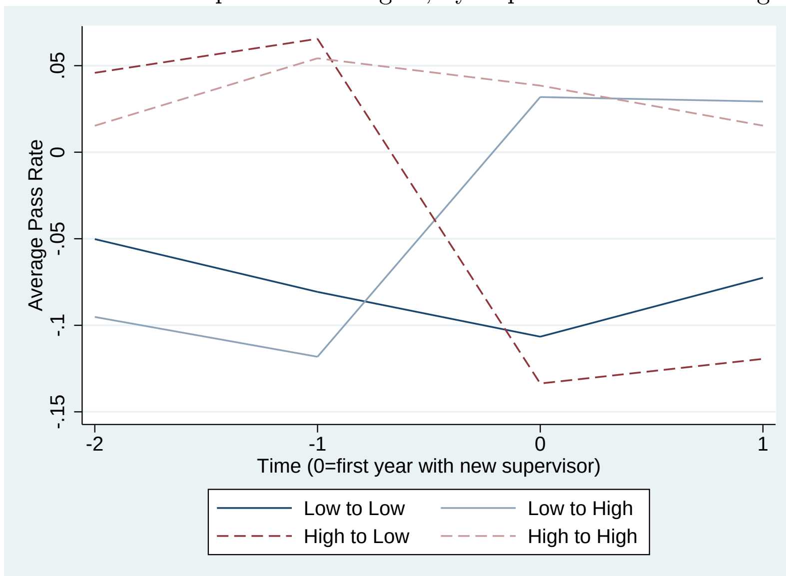
We also find that there is substantial idiosyncratic variation in ratings, holding constant these fixed effects and a rich set of time-varying controls. This residual variance is an input into the CHK correction and must also be adjusted. We use a within-transformation of the error term, demeaning by 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 (-0.011 using the CHK adjustment). Thus any systematic worker sorting across supervisors based on fixed performance differences is likely very small.[18] However, with double fixed effects specifications, such as that in equation

to bring the predicted values for observations associated with that worker back through the mean.

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

Figure 2: Mean Performance of Supervisor Changers, by Supervisor Effect at Origin and Destination



1, we are naturally worried that sorting based on time-varying performance might bias our estimates. Figure 2 presents an event-study of performance for workers who change supervisors. Following Card, Heining, and Kline (2013), this figure can help generate intuition for whether the basic additive specification is apt.

We split the set of supervisors into low and high raters based on whether their fixed effect, i.e., their average propensity to pass subordinates, was above or below the median. We plot demeaned 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. Figure 2 contains several important points.

First, in the two years before moving, workers matched to a high rater have similar (high) performance, regardless of whether they are about to move to a different high rater or to a low rater. Similarly, workers matched to a low rater have similar (low) performance regardless of their destination. This lack of pre-trend in performance is comforting and shows that worker sorting does not play an important role in our data.

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 but stay within the same supervisor type experience little change in performance.

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. The magnitudes are also similar to the average difference in pass probability across high and low raters, 0.26. This suggests that the basic specification with additively separable worker and supervisor effects is a good characterization of the data.

3 MODEL

In this section, we develop a model with testable predictions that allow us to distinguish between the sources of ratings heterogeneity and to determine how informed the firm is about such ratings variation across supervisors. 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).[19]

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

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.

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.

[19] 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).

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."[20]

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:

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

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:

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

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.

[20]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 marignal 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).

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

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

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.

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.[21] 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.[22]

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 $w_i = a_{is} + b_{is} r_i$.[23] The parameters (a_{is}, b_{is}) of these wage contracts vary with the common information on worker and supervisor types available at the contracting stage. We also 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.

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:

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

[21] 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.

[22] 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. Furthermore, the Waldman (1984) promotion-as-signal hypothesis posits that the market has better information about workers at higher job levels.

[23] 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.

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, 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 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 a high effort is exerted. In that case, better managers reduce the effort cost born by workers significantly and wages decline with μ_s.

The intuition behind those parts of the proposition related to variation in β_s 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.[24] We parameterize the share of total variation in β and μ unknown to agents as θ_β and θ_μ so that

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

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.

[24] 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.

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.

First, 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.

A second notable difference is that despite the firm's interest in neutralizing all known variation in β_s^E when compensating workers to remove any risk that is not of use in setting incentives the firm is only able to neutralize the component about which it is informed. This implies that expected compensation increases in β_s. It also implies that 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

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.*

that panel data on performance ratings alone does not allow separate identification of the 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.

At this point, we find it useful to consider extreme cases in order 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. First, that worker wages may vary with supervisor heterogeneity if either supervisor ability is important or leniency is important and the firm is uninformed. In the former case, productivity increases and hence so will ratings and compensation. In the latter case, the uninformed firm will have difficulty undoing increases in ratings that are driven by lenient managers.

Second, piece rates are declining in effort costs that the firm can observe. Thus, when supervisor heterogeneity is driven by ability and the firm is informed about this heterogeneity, the model predicts a positive relationship between supervisor heterogeneity and the incentive intensity measured by the loading on piece rates. Otherwise we should see no relationship.

Third, productivity is increasing in managerial ability since the worker always observes the supervisor's impact on his or her effort cost. In contrast leniency does not impact output. This implication is a key insight in the test between leniency and ability as it is true regardless of the information structure.

Fourth, supervisor wages are increasing in their effectiveness only if the firm (and the market) are informed. Supervisors will be kept at their outside option.

Finally, we note that in our model workers are made just indifferent between staying at this firm and moving elsewhere. Thus the fully informed firm will extract all rents a worker gains from being assigned to a more effective supervisor. Any rents the worker might earn by having a reduced effort cost will be undone through the intercept of the wage function. However, an imperfectly informed firm will be limited in its ability to do so and the worker's surplus therefore increases in both managerial ability and leniency bias.

4 TESTING THE MODEL

The model's predictions contingent on the nature of supervisor heterogeneity (ability and leniency) and the information structure are made explicit in Table 4. In this section, we empirically evaluate these predictions using detailed personnel data.

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:

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

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. First we apply a naive estimator that regresses 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 2.2. This estimator is biased for the same 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 sample 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 error across the two sets of estimates will be 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. Splitting the sample into even and odd years maximizes the overlap of workers and supervisors across the two samples.[25] Because of the low turnover in our sample, we can use 95 percent of our worker-year observations; i.e., 95 percent of our observations consist of workers and supervisors observed in both even and odd years. We hereafter term this the split-sample IV estimation.[26] It should be noted, however, that while this method allows for estimation of β_1 and β_2 it does not allow us to estimate β_3. The reason is that the error term from one subsample is uncorrelated with the unobserved effects (α and ϕ) in that subsample as 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.

Our third strategy is to use the methodology developed by Card, Heining, and Kline (2013), which we discussed and applied in Section 2.2. To estimate the three parameters $(\beta_1, \beta_2, \beta_3)$ the CHK procedure has to be expanded to a joint system of two double-fixed effects regressions (one for ratings and one for earnings). Once the second moment matrices of the unobserved effects are obtained using the CHK methodology they can be transformed into implied regression coefficients (see Appendix A.2 for details). These estimates are only shown for worker earnings because they are excessively computationally intensive. 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 the entire estimation sample, not only the one consisting of workers and supervisors present in both even and odd years. It also allows for specifying a variety of correlation structures for the error terms across time and space. Finally, this approach provides estimates of β_3, the coefficient on the error term in performance ($\hat{\epsilon^p_{it}}$), while our split-sample IV approach does not.

Table 5 presents results using the three methods. Column 1 presents results using the naive OLS estimator based on the estimated fixed effects. Column 2 shows the results from the split-sample IV, and column 3 contains results based on the CHK approach. In columns 1 and 2, we cluster standard errors by supervisor, the level of variation underlying our dependent variable.

[25]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.

[26]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 145 for predicting ϕ and 2,935 for predicting α.

Table 5: Log(Earnings) and Ratings Components

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

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

Across all specifications, we find that working for a high-rating supervisor is associated with substantially higher earnings. The estimate from column (1) implies that moving from a supervisor who never passes subordinates to one who passes all of them increases earnings by just over 10 percent. In Section 2.2, we found the standard deviation of ϕ_s to be 0.153. Thus, a one-standard-deviation increase in ϕ_s is associated with an increase in earnings of about 1.6 percent. We find fairly similar effects across specifications, though the unadjusted estimates (column 1) are 10-20 percent smaller in magnitude, as we would expect if the estimation error is interpreted as "measurement error." Using the split sample IV, our preferred method, we find that a one-standard-deviation higher rater increases earnings by 2.1 percent.

We also find that worker effects correlate positively with earnings. A one-standard-deviation higher α is associated with earnings increases of 2.5 to 3 percent and having an idiosyncratically high rating period gives workers a positive but modest earnings boost.

Overall, we find strong evidence that having a high-rating supervisor has positive, significant effects on earnings. Recall from table 4 that this result is 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

Table 6: Pay-for-Performance and Ratings Components

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

Notes: OLS results. See Table 5.

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 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 coefficient on this interaction measures whether individual earnings are more responsive to performance measures 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 the split-sample IV and the CHK correction cannot be adapted 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. The earnings increase associated with a one-standard-deviation higher rater is 0.9 percent for a worker who does not pass his or her own performance review. However, the worker assigned to a higher rater who passes receives an additional 1.3 percent earnings boost. Conditional on receiving a bonus, that bonus is about 7 percent higher for a worker assigned to a one-standard-deviation higher rater who passes the performance review, compared to one who does not. This roughly doubles the overall effect of being assigned to a high rater on bonus. Though, we find no effect on the probability of receiving a bonus.

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 for workers, which increases productivity. In contrast, lenient supervisors do not alter productivity. These associations are true irrespective of the firm's level of information about supervisor heterogeneity. Hence, the correlation between supervisor heterogeneity and productivity is a crucial component to disentangling supervisor ability from leniency. Note that the performance ratings themselves cannot be used to establish such a relationship as they may be contaminated by leniency bias. Instead, objective performance/productivity measures should be used and we rely on two such measures.

We have access to two sets of 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.[27] For the years 2014 and 2015 we have information on individual financial performance. This performance 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 performance measures are available. In relation to 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 observed in the last year of our personnel data, and, as usual, cluster standard errors by supervisor.[28]

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.[29] 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.[30] In these aggregated regressions, we control for a limited set of variables, either averaged to the branch-year level or at the individual level.[31] For the financial performance regressions (column 5), the fixed effects pertain to the individual workers and their supervisors

[27] 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.

[28] 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 $corr(\phi_t, \phi_{t+2}) = 0.766$.

[29] 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.

[30] 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.

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

Table 7: Branch KPI Performance and Ratings Components

	(1)	(2)	(3)	(4)	(5)
	KPI rankings:				Individual
Dependent variable:	Rank score	Pr(Top)	Pr(Top 5)	Pr(Top half)	financials
(mean)	(0.53)	(0.06)	(0.30)	(0.46)	(-0.07)
Panel A: OLS					
	Branch-year-level averages:				
Supervisor FE (φ)	0.165**	0.038	0.193	0.259*	0.038**
	(0.080)	(0.065)	(0.127)	(0.140)	(0.019)
Worker FE (α)	0.014	0.042	0.088	0.039	0.011
	(0.070)	(0.057)	(0.111)	(0.122)	(0.008)
Panel B: Split-Sample IV					
	Branch-year-level averages:				
Supervisor FE (φ)	0.267*	0.037	0.264	0.367	0.064*
	(0.156)	(0.127)	(0.248)	(0.271)	(0.039)
Worker FE (α)	0.005	0.034	0.066	0.050	-0.003
	(0.101)	(0.081)	(0.160)	(0.175)	(0.013)
Observations	781	781	781	781	2,202

Notes: Observations in columns 1-4 are at the branch-year level for 2007-2010 and at the worker level for years 2014-15 in column 5. Inverse rank score is -1 times the branch's KPI ranking in that year divided by the number of branches it is ranked against. Regressions in columns 1-4 also include year effects, the branch-year averages of worker age, tenure, share female, and job-level dummies; column 5 includes typical controls (see Table 3) and cluster standard errors by supervisor. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

and we include our typical individual-level controls (see Table 3).

For both types of measures and when we use OLS or our split-sample IV, we find that branches and individuals with higher rating supervisors perform better. Using our IV estimates, we find that a branch with a one-standard-deviation higher ϕ has a 0.04 higher (7.7%) inverse rank score (-1 times the branch's ranking divided by the number of branches in the peer group), is 0.6 percentage points (10%) more likely to be the top-ranked branch, is 4.0 percentage points (13%) more likely to be ranked among the top 5 branches in the peer group, and 5.6 percentage points (12%) more likely to be ranked in the top half. These magnitudes are economically large, though often statistically insignificant. Compared to the standard deviation of the financial performance metric, 0.126, our split sample IV estimate implies that a one-standard-deviation higher rating supervisor is associated with one tenth of a standard deviation in the distribution of financial performance of subordinates.

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 the results are therefore based on relatively small samples, which at times challenges the statistical significance of the point estimates.

The evidence we do provide, however, points in the same direction of a positive relationship. This conclusion is made stronger by the fact that we have two measures on performance and that our two sets of regressions are separate in time. Hence, our results support the hypothesis that manager ability (μ_s), rather than leniency bias (β_s), drives supervisor heterogeneity. [32]

4.4 Supervisor Wages

The fourth comparative static relates supervisor heterogeneity in ratings to the supervisors' own pay. Supervisor compensation would not correlate with supervisor heterogeneity if firms are uninformed about ϕ_s. Nor would firms compensate supervisors for being more lenient. Only in the case where supervisor heterogeneity reflects managerial ability that the firm is informed about would we observe a positive correlation between ϕ_s and supervisor compensation.

To investigate this relationship we regress components of supervisor pay on their own ratings fixed effect, as well 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.[33] These regressions control for the characteristics of the supervisor and the average characteristics of the group of workers being supervised, and standard errors are clustered by supervisor.

Results are reported in Table 8. Supervisor earnings are strongly positively influenced by their ratings style (as well as the quality of the team they are supervising). 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 5 percent for each standard deviation in ϕ_s. Much of this increase comes through an increase in the bonus received.

Consistent with our earlier findings, the positive relationship between supervisor compensation and supervisor heterogeneity provides support for the hypothesis that supervisor heterogeneity reflects ability differences that the firm is informed about rather than differences in leniency.[34]

4.5 Worker Surplus

The last comparative static that 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

[32] The estimated impacts of α on objective performance in Table 7 are statistically insignificant. The point estimates are, however, typically positive and the two-standard error bounds include large positive effects. The results are thus consistent with worker quality that correlates positively with branch performance.

[33] As in the branch-year regressions, 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.

[34] Despite the fact that our model cannot rationalize why supervisors are paid more for supervising workers of fixed higher quality (α) it is an intriguing result. One possibility is that perhaps the firm cannot perfectly separate the ability of supervisors from the ability of workers.

Table 8: Supervisor Outcomes and Ratings Components

Dependent variable:	(1) Log (earnings)	(2) Pr(bonus)	(3) Log (bonus)
Panel A: OLS			
Supervisor FE (φ)	0.217***	0.0407	0.511***
	(0.0331)	(0.0340)	(0.117)
Average worker FE (α)	0.196***	0.0906***	0.470***
	(0.0231)	(0.0335)	(0.0877)
Supervisor-year observations	8,436	8,436	4,982
Panel B: Split sample IV			
Supervisor FE (φ)	0.311***	0.007	0.812***
	(0.048)	(0.061)	(0.171)
Average worker FE (α)	0.232***	0.071	0.482***
	(0.038)	(0.056)	(0.138)
Supervisor-year observations	8,131	8,131	4,820

Notes: Observations are at the supervisor-year level. Outcomes are supervisor pay variables in the given year. In Panel B, we estimate supervisor and worker fixed effects on odd and even years separately. We instrument for the supervisor effect and supervisor-year-level average worker effects in odd years with those obtained in even years and vice versa. Controls are listed in Table 3; worker controls are the average for characteristics. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

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 quit behavior as well as the worker satisfaction surveys to look for evidence regarding rents due to ϕ_s. Workers can, to some extent, influence their assignments across units and branches within the firm, and they have direct control over their decision to quit the firm. In addition to this, worker satisfaction surveys provide direct evidence on how workers perceive their supervisors.

Table 9 presents results on mobility and worker satisfaction. Panel A shows OLS estimates and Panel B shows split-sample IV estimates. Column 1 shows that workers are less likely to quit in the next two years if working for a higher-rating supervisor.[35] Column 2 shows that workers are more likely to remain with their current supervisor when assigned to a high rater.[36] However, both results are statistically insignificant. The results in column 3 are based on data from the employee job satisfaction survey. The dependent variable is the average across the seven 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

[35] As noted above, we allow for a quit over the next two years because often workers do not receive a performance rating in their last year at the firm and our sample is restricted to observations with both pay and performance measures.

[36] In order to isolate moves that the worker might have possibly influenced, we restrict this sample to workers present at the firm in adjacent years at the same job level (i.e., we exclude workers who exited, or were promoted or demoted in $t+1$).

Table 9: Do Workers Value High Raters?

Dependent variables:	(1) Quit	(2) Same supervisor	(3) Bottom-up evaluation
	Panel A: Unadjusted		
Supervisor FE (ϕ)	-0.011	0.032	0.158***
	(0.007)	(0.027)	(0.048)
Worker FE (α)	-0.029***	0.064***	0.165***
	(0.004)	(0.008)	(0.018)
Pass residual (ε)	-0.005**	0.004	0.067***
	(0.002)	(0.006)	(0.011)
Observations	69,350	65,415	67,832
	Panel B: Split sample IV		
Supervisor FE (ϕ)	-0.009	0.039	0.258***
	(0.013)	(0.052)	(0.088)
Worker FE (α)	-0.042***	0.090***	0.187***
	(0.005)	(0.013)	(0.027)
Observations	66,463	62,339	64,897

Notes: Column 1 estimates the probability that the worker quit by t+2 for all workers observed in t, excluding the last year of data where t+2 outcomes are not observed. Column 2 estimates the probability that the worker is with the same supervisor in the next year among those observed in adjacent years in the firm at the same level of the hierachy (i.e., not promoted or demoted). Column 3 reports the worker's self-reported satisfaction of their supervisor. All regressions include time-varying worker and supervisor controls (see Table 3). Standard errors clustered by supervisor. Significance levels are represented using stars: *** p<0.01, ** p<0.05, * p<0.1.

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 0.04 increase in the bottom-up rating a worker ascribes to his or her supervisor (the mean of that variable is 5.5).

Together, these findings are indicative that workers earn rents when assigned to high rating supervisors, but that such rents are of moderate magnitude. 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.[37]

4.6 Discussion

In this section, we tested the predictions of our model to determine the nature of supervisor ratings heterogeneity and whether the firm is informed about this heterogeneity. We found that (1) individual earnings increase with ϕ_s, (2) piece rates increase in ϕ_s, (3) team productivity as measure by the KPI ranking or

[37] 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.

financial performance increase in the average ϕ_s within a branch, (4) supervisor pay increases in ϕ_s, and (5) workers appear to earn (moderate) rents from being matched to higher raters, since they are less likely to quit or change supervisors, and their self-reported work satisfaction is higher when ϕ_s is higher. 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 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 high-rating supervisors, 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 can be explained 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 terms of the compensation contracts of both workers and 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. 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 talent.

Of course, outside our model there are other reasons why firms might share rents associated 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, the presented results have a clear interpretation within the context of our model. However, one could write down other models of compensation and bonuses that might drive these same results. Regardless of our model, we have shown that there is substantial heterogeneity in performance ratings across supervisors and that this heterogeneity indeed is 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?

Above (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 $\phi_{s(i,t)}$.[38] These lags allow ϕ_s to influence earnings for up to k periods. Estimates from 8 do not represent the full dynamic effects of being assigned a higher rater (ϕ_s) because we control for job levels and because we control for the ratings type of supervisors in other periods. We control for job-level effects (l_{it}) to control for any variation in ratings style across job levels. Part of the effect of ratings heterogeneity on future earnings arises through promotions and we will explore that effect below. By controlling for ratings types in other periods, we remove 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 holding removing any effects attributable to persistence in supervisor ratings styles.

Results are summarized in Table 10. 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 contemporaneous earnings effect estimated on a restricted sample of workers present for at least five periods in the firm, that is, with at least four lags in supervisor effects.[39] The coefficient is a bit smaller in magnitude for the sample of more stable workers, 0.074 compared to 0.103, but still qualitatively similar.

Column 3 presents results including all four lags of supervisor effects. The coefficient on the contemporaneous earnings effect drops to 0.032. This is because part of the supervisor effect on earnings comes

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

[39] Since our panel is only 11 years, we lose too much data when we restrict to workers who have been at the firm for more than the 5 consecutive years we estimate here.

Table 10: Earnings Dynamics and Supervisor Heterogeneity

Dependent variable	Log earnings		
	(1)	(2)	(3)
Supervisor FE (ϕ):			
Contemporaneous ϕ	0.103***	0.074***	0.032**
	(0.014)	(0.016)	(0.013)
Lag 1 ϕ			0.017*
			(0.009)
Lag 2 ϕ			0.015*
			(0.009)
Lag 3 ϕ			0.017**
			(0.009)
Lag 4 ϕ			0.023**
			(0.010)
Non-missing lags		X	X
Observations	77,077	22,569	22,569
Partial R-squared	0.818	0.821	0.822

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

through persistence of supervisors across periods (as shown in Table 9) and associated cumulative impacts of pay. Indeed, we find that impacts of supervisor on pay are quite persistent. The coefficients on the lagged ϕ range between 0.015 and 0.023, while the contemporaneous effect is 0.032. This suggests that roughly half to two-thirds of the contemporaneous pay increase associated with having a high-rating supervisor persists one to four periods later. The decline in the size of the coefficients does imply that effects of being assigned to higher raters are somewhat but not completely transitory. The full contemporaneous effects do not persist. However, the large fraction that remains could indicate direct salary impacts that benefit workers in subsequent years (as opposed to benefits driven solely by transient bonuses), or perhaps because of lasting effects on human capital.[40]

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 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 11 shows that a higher

[40] Appendix Table A1 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 11: Worker Outcomes and Ratings Components

Dependent variables:	(1) Promotion	(2) Demotion	(3) Layoff
	Panel A: Unadjusted		
Supervisor FE (ϕ)	0.044***	-0.006**	-0.006***
	(0.009)	(0.002)	(0.002)
Worker FE (α)	0.095***	-0.018***	-0.005***
	(0.004)	(0.001)	(0.001)
Pass residual (ε)	0.055***	-0.006***	-0.001
	(0.004)	(0.001)	(0.001)
Observations	74,602	74,602	67.32
	Panel B: Split sample IV		
Supervisor FE (ϕ)	0.034*	-0.010*	-0.010***
	(0.019)	(0.005)	(0.004)
Worker FE (α)	0.104***	-0.023***	-0.008***
	(0.006)	(0.002)	(0.002)
Observations	73,327	71.024	64.552

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 year 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.

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.[41] We find that a one-standard-deviation higher rater increases the probability of promotion by 6 percent (0.7 percentage points). It decreases the probability of a demotion by 7.7 percent and a layoff by 15 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 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

[41] 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.

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.

The percent impact of a given supervisor effect in period t ($\phi_{s(i,t)}$) on earnings in period $t+k$ is thus given in equation 9. It equals the persistent component of the within job-level pay effect, β_k, 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 assume that careers on average last another 20 years in our data. We thus obtain $\frac{1}{W_{t+k}}\frac{dW_{t+k}}{d\phi_{s(i,t)}}$ for $k\leq 20$ 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 10. 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 10 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 from Table 11. 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 workers across job levels. In our data, the average earnings increase associated with moving up adjacent job levels is 14.8 percent.[42] 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. 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 estimate $\sigma_\phi=0.153$ (Table 3), we determine that a one-period, one-standard-standard deviation increase in ϕ_s is associated with an increase of the PDV of earnings of 2.7 to 5.3 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 1.5 to 4.1 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 0.8 percentage points smaller.

Overall, we find these effects to be quite large. When comparing workers assigned to supervisors at the 90th and 10th percentiles of the ratings distribution, the former can expect a 7 to 14 percent higher PDV

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

of earnings over the next 20 years at the firm. This is true for a one-time assignment, holding constant the stream of supervisors over the remaining years.

6 CONCLUSION

In this paper we provide evidence that supervisors differ widely in their ratings behavior. A worker matched to an especially high rater (at the 90th percentile) is, on average, 39 percentage points (82%) more likely to receive a passing score (a performance score in the upper half of the performance scale) than the same worker matched to an especially low rater (10th percentile). 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.[43] This conclusion is based on the empirical finding that worker pay, pay for performance, supervisor pay, 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 supervisors' ratings behavior.

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 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 (1) to highlight the surprising and sizable variation in ratings across supervisors, and, (2) disciplining ourselves to one plausible model, 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 how they manage their employees interact in important ways.

[43] 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.

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.[44] 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. Although workers may worry that the subjective nature of the performance ratings makes them bear unnecessary risk, firms should exercise care when they consider introducing forced curves or other guidelines restricting supervisors in their ratings behavior.

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.

[6] 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.

[7] 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.

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

[44] During the hiring process, there is evidence that reducing discretion among hiring managers can improve decision making. Autor and Scarborough (2008) show that job testing can improve hiring decisions and Hoffman, Kahn, and Li (2016) show that firms can do better by removing some discretion of hiring managers and relying more on a high-quality job test.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

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

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

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

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

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

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

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

[22] Frederiksen, Anders, Fabian Lange, and Ben Kriechel (2017), "Subjective Performance Evaluations and Employee Careers," *Journal of Economic Behavior and Organization*, forthcoming.

[23] 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.

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

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

[26] 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.

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

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

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

[30] Hoffman, Mitchell, Lisa B. Kahn, and Danielle Li (2016), “Discretion in Hiring,” NBER Working Paper #21709.

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

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

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

[34] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A Estimating the Covariance Matrix of Unobserved Effects

In this section we describe how we adapt the Card, Heining, and Kline (2013) method to (1) obtain unbiased estimates of the variance-covariance matrix of worker and supervisor effects in performance ratings and (2) adjust coefficients in a regression of these unobserved components of performance ratings on earnings.

A.1 Components of Performance Ratings

The worker and supervisor fixed effects obtained from estimating equation (1) are estimated with error. This error is correlated across workers and supervisors. The variance-covariance matrix of estimated fixed effects is therefore a biased, and inconsistent estimator. It is, however, possible to adjust the empirical estimate using the variance-covariance matrix of the estimation error of the fixed effects. We do so using the CHK correction, which generates unbiased estimates, following the procedure used by Card, Heining, and Kline (2013).

Given N worker-year observations, with N^* unique workers and S unique supervisors, our regression is as follows:

$$r = D'\alpha_r + F'\phi_r + \epsilon_r \tag{10}$$

The vector α are N^* worker fixed effects; ϕ are S supervisor effects; the NxN^* design matrix D identifies employees; the NxS design matrix F identifies supervisors; ϵ_r is an N^*-vector error term. r is a vector of performance measures that have been residualized on control variables.

Let $Z = \begin{pmatrix} D \\ F \end{pmatrix}$ and $\xi = \begin{pmatrix} \alpha_r \\ \phi_r \end{pmatrix}$. Then

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

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

$$V_{\hat{\xi}} = (Z'Z)^{-1} Z'\Omega Z (Z'Z)^{-1} \tag{11}$$

The matrix $V_{\hat{\xi}}$ is central to the correction procedure and depends on correctly specifying the variance structure of the unobserved ϵ_r. We assume that errors are iid across (i, t). This assumption is necessarily violated by our data since the ratings are limited dependent variables and their variances thus depend on the unobserved effects of the individual specific probability of receiving a passing grade. This assumption, however, greatly facilitates the analysis, which is why we invoke it here. Note that it is possible to allow for more general error structures than the iid assumption we make as long as it is possible to identify Ω.

The iid assumption implies that $\Omega = \sigma_r^2 I_N$, where σ_r^2 is the variance of ϵ_r . To obtain an unbiased and consistent estimate of this variance, we use a within transformation. We demean the residuals by match (worker-supervisor pairs), and obtain the variance that corrects for degrees of freedom (number of observations minus number of matches). This differences out unobserved effects that are not consistently estimated.

Estimating equation 10 gives fixed effect estimates $\hat{\xi}$ that are unbiased, but inconsistent estimators of ξ. For an unbiased estimate $\hat{\xi}$ of ξ and any matrix A there is a simple expression for the expectation of the quadratic form $E\left[\hat{\xi}' A \hat{\xi}\right]$, where tr is the trace of the matrix:

$$E\left[\hat{\xi}' A \hat{\xi}\right] = \xi' A \xi + tr\left(A V_{\hat{\xi}}\right) \tag{12}$$

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

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

where Q is the demeaning matrix.[45] Defining $A_{D\alpha_r} = \begin{bmatrix} D'QD & 0 \\ 0 & 0 \end{bmatrix}$ conformable with ξ, we get $\hat{\sigma}^2_{D\alpha_r} - \frac{1}{N^*-1} tr(A_{D\alpha_r} * V_{\hat{\xi}}) \rightarrow \sigma^2_{D\alpha_r}$. We proceed in the same manner for the other variances and covariances required. Exact details are shown in Card, Heining, and Kline (2013).

A.2 Regressing Performance Components on Earnings

Our estimates of equation 7, repeated here for convenience, will be biased since the independent variables $(\hat{\phi}_{s(i,t)}, \hat{\alpha}_i)$ are estimated with error and furthermore since the estimation error is potentially correlated:

$$log(w_{it}) = \beta_0 + \beta_1 \hat{\phi}_{s(i,t)} + \beta_2 \hat{\alpha}_i + \beta_3 \hat{\varepsilon^p_{it}} + \beta' X_{it} + \gamma' Y_{s(i,t)t} + \nu_{it} \tag{13}$$

We adapt the CHK correction to adjust estimates $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$. The basic approach is to estimate double fixed effects models for both performance and wages and then use adjusted estimates of the variance-covariance of the full set of fixed effects to calculate coefficients in equation 13.

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

Consider the stacked equation, where w is log earnings after having residualized on controls, ϵ_w is a residual, and the other terms are defined above:

$$\begin{pmatrix} r \\ w \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix}' \begin{pmatrix} \alpha_r \\ \alpha_w \end{pmatrix} + \begin{pmatrix} F & 0 \\ 0 & F \end{pmatrix}' \begin{pmatrix} \phi_r \\ \phi_w \end{pmatrix} + \begin{pmatrix} \epsilon_r \\ \epsilon_w \end{pmatrix} \tag{14}$$

The basic approach is to estimate fixed regressions on the system (14), and then adjust the variance-covariance matrix of $\left(\hat{\alpha_r}, \hat{\alpha_w}, \hat{\phi_r}, \hat{\phi_w}\right)$ in the exact same manner as in the single equation. We require the additional assumption that $(\varepsilon_{r,i,t}, \varepsilon_{w,i,t})$ are iid across(i,t). The variance-covariance matrix of the residuals, Ω, is now $2Nx2N$ block diagonal, with 4 NxN blocks. The top left block has the within-transformation variance of ϵ_r on the diagonal. The bottom right block has the within-transformation variance of ϵ_w on the diagonal. The top right and bottom left blocks have the within-transformation covariance of (ϵ_r, ϵ_w) on the off diagonals.

Once we have the adjusted variance-covariance matrix of the unobserved effects, we can estimate equation 13. First note that we can write the wages as a function of the unobserved components of the wage equation above (worker and supervisor effects and residuals), which are, in turn, functions of the unobserved components of performance.

$$E\left[w_{it}|\alpha_r, \phi_r, \varepsilon_{r,t}\right] = E\left[\alpha_w|\alpha_r, \phi_r, \varepsilon_{r,t}\right] + E\left[\phi_w|\alpha_r, \phi_r, \varepsilon_{r,t}\right] + E\left[\varepsilon_{w,t}|\alpha_r, \phi_r, \varepsilon_{r,t}\right] \tag{15}$$

This can be simplified because we have defined $E[\varepsilon_{rt}|i,s] = 0$ so that $cov\left(\varepsilon_{rt}, \alpha_w\right) = E\left[\varepsilon_{rt}\alpha_i^w\right] = E\left[E\left[\varepsilon_{rt}\alpha_i^w|i,s\right]\right] = E\left[\alpha_i^w E\left[\varepsilon_{rt}|i,s\right]\right] = 0$. The same argument holds for $cov\left(\varepsilon_{rt}, \phi_s^w\right)$. It also ensures that $E\left[\varepsilon_{w,t}|\alpha_r, \phi_r, \varepsilon_{r,t}\right] = E\left[\varepsilon_{w,t}|\varepsilon_{r,t}\right]$:

$$E\left[w_{it}|\alpha_r, \phi_r, \varepsilon_{r,t}\right] = E\left[\alpha_w|\alpha_r, \phi_r\right] + E\left[\phi_w|\alpha_r, \phi_r\right] + E\left[\varepsilon_{w,t}|\varepsilon_{r,t}\right] \tag{16}$$

All of these components can be estimated. Using the best linear approximation to each of the components we get:

$$E\left[w_{it}|\alpha_r, \phi_r, \varepsilon_{r,t}\right] = \left(b_{\alpha_w|\alpha_r} + b_{\phi_w|\alpha_r}\right)\alpha_r + \left(b_{\alpha_w|\phi_r} + b_{\phi_w|\phi_r}\right)\phi_r + \frac{cov(\varepsilon_{w,t}, \varepsilon_{rt})}{var(\varepsilon_{rt})} * \varepsilon_{rt} \tag{17}$$

where $b_{\alpha_w|\alpha_r}$ denotes the population regression coefficient from a regression of α_w on (α_r, ϕ_r) which can be easily obtained from the variance-covariance matrix of the unobserved effects that we estimated using the

CHK method. The other b's are analogously defined.

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 18, 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).

$$q = e + \alpha + \varepsilon^q \tag{18}$$

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 $c(e)$, repeated here in equation 19.

$$c(e) = -\frac{1}{2\mu_s}e^2 \tag{19}$$

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 20:

$$u(w_s, q, r) = w_s + \tilde{\beta}_s\left(r - q\right) - \frac{\tilde{\gamma_s}}{2}\left(r - q\right)^2 \tag{20}$$

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

$$r = q + \tfrac{\tilde{\beta}_s}{\tilde{\gamma_s}} = q + \beta_s. \tag{21}$$

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:

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

Substituting (18) in (21) and denoting by e_s the equilibrium effort that subordinates of supervisor s exert, we get:

$$r = \alpha + (e_s + \beta_s) + \varepsilon^q = \alpha + \phi_s + \varepsilon^q \tag{22}$$

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.

$$E[w - \frac{1}{2\mu_s} e^{*2} | I_C] - \frac{1}{2}\psi var\left(w - \frac{1}{2\mu_s} e^{*2} | I_C\right) \geq \underline{u\left(\alpha\right)} \tag{23}$$

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

$$e^* = b_s \mu_s \tag{24}$$

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. 24 into the certainty equivalent (23) and simplifying, we obtain the participation constraint:

$$a_{is} + b_s(\alpha + \beta_s) + \frac{1}{2}b^2\mu_s - \frac{\psi}{2}b_s^2\sigma_q^2 \geq \underline{u}(\alpha) \tag{25}$$

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. 24). Simplifying yields the following maximization problem for the firm's choice of b_s:[46]

$$b_s^* = \underset{\{b\}}{argmax}\left\{\alpha + b_s\mu_s - \frac{b_s^2}{2}\left(\mu_s + \psi\sigma_q^2\right)\right\} \tag{26}$$

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

$$b_s^* = \frac{\mu_s}{\mu_s + \psi\sigma_q^2} \tag{27}$$

Substituting the optimal effort (equation 24) and piece rate (equation 27) into the output equation 18 results in $E[q|\alpha,\mu_s,\beta_s] = \alpha + E[e|\mu_s,\beta_s] = \alpha + b_s\mu_s = \alpha + \frac{\mu_s}{\mu_s + \psi\sigma_q^2}\mu_s$. This establishes point 2: expected output increases one–for-one with α, does not vary with β_s and increases with μ_s.

[46] 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.

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

$$\alpha + b\mu_s - a_{is} - b_s\left(\alpha + \beta_s + b\mu_s\right) - w_s\left(\mu_s, \beta_s\right) = 0 \tag{28}$$

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

For expected compensation of workers (point 3), note that solving equation 28 implies that the firm will set worker pay so that their certainty equivalent exactly equals the outside option: $E[w|I_C] = \underline{u\left(\alpha\right)} + \frac{1}{2\mu_s}e^{*2} + \frac{1}{2}\psi var\left(w|I_C\right)$. From the equation 24, 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. 27).[47] Thus, we obtain

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

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 28) 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

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

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: $\frac{\partial w_s(\mu_s)}{\partial \mu_s} > 0$.

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 (β_s^E, μ_s^E) 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

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

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 (β_s^E, μ_s^E) 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:

$$a + b\left(\alpha_i + \beta_s^E\right) + b^2\frac{\mu_s^E}{2} - \frac{\psi}{2}\left(b^2\left(\theta_\beta\sigma_\beta^2 + \sigma_q^2\right) + \frac{b^4}{4}\theta_\mu\sigma_\mu^2\right) \geq \underline{u}\left(\alpha\right) \tag{29}$$

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 (29) 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:

$$\Pi\left(\mu_s^E, \beta_s^E, \alpha\right) = \underset{\{a,b\}}{Max}\left\{\alpha + b\mu_s^E - a_i - b_s\left(\alpha + \beta_s^E + b_s\mu_s^E\right) - w_s\left(\beta_s^E, \mu_s^E\right)\right\} \tag{30}$$

s.t. the participation constraint (29).

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. 30:

$$\mu_s^E = b_s\left(\mu_s^E + \psi\left(\theta_\beta \sigma_\beta^2 + \sigma_q^2 + b_s^2 \frac{\theta_\mu \sigma_\mu^2}{2}\right)\right) \tag{31}$$

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 31, 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. 29) 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 32, 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).

$$\begin{aligned} E\left[w|\alpha, \beta_s^E, \mu_s^E, \beta_s, \mu_s\right] = & \quad \alpha + h\left(\mu_s^E\right) + b\left(\left(\beta_s - \beta_s^E\right) + b\mu_s\right) \\ = \alpha + h\left(\mu_s^E\right) + b\varepsilon_\beta + b^2\mu_s & \quad = \alpha + h\left(\mu_s^E\right) + b\theta_\beta\beta_s + b^2\mu_s + b\nu_\beta \end{aligned} \tag{32}$$

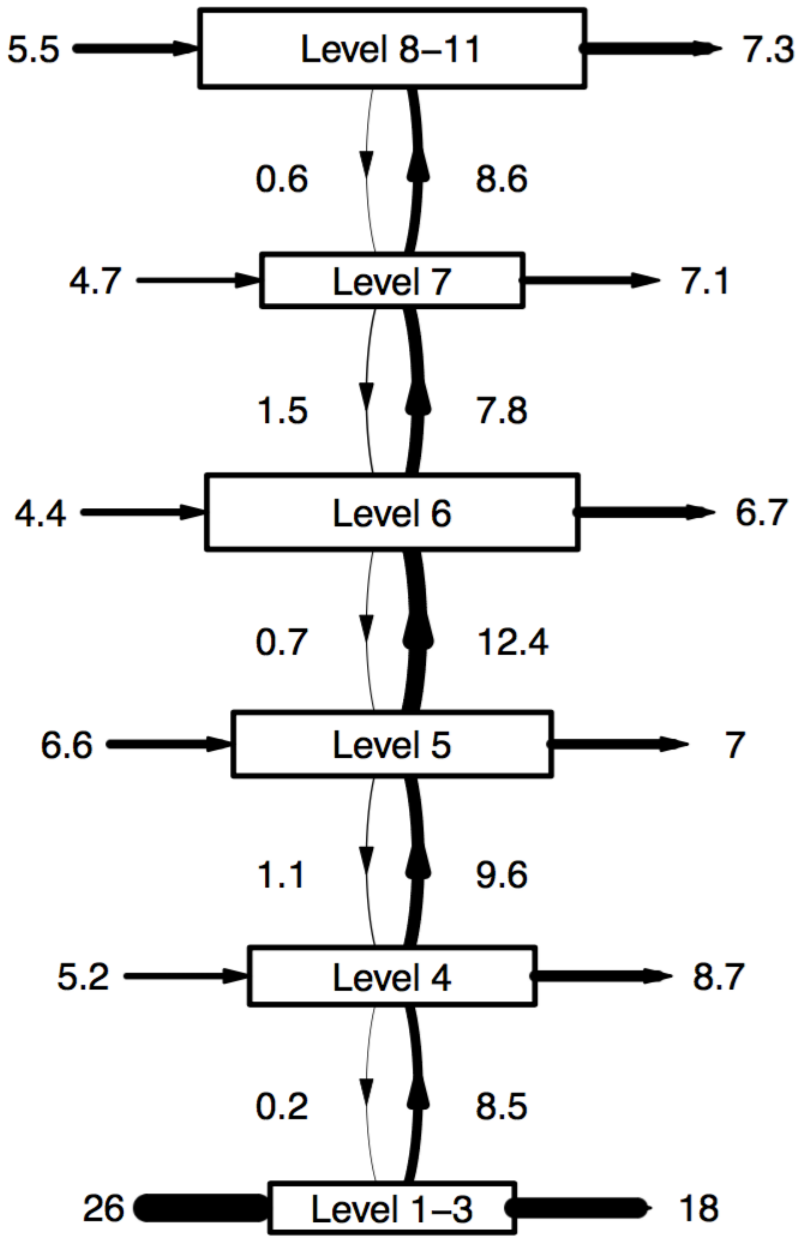
where we substitute in the linear projection of $\varepsilon_\beta = \frac{cov(\varepsilon_\beta, \beta_s)}{var(\beta_s)}\beta_s + \nu_\beta = \frac{cov\left(\varepsilon_\beta, \beta_s^E + \varepsilon_\beta\right)}{var(\beta_s)}\beta_s + \nu_\beta = \theta_\beta\beta_s + \nu_\beta$.

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.

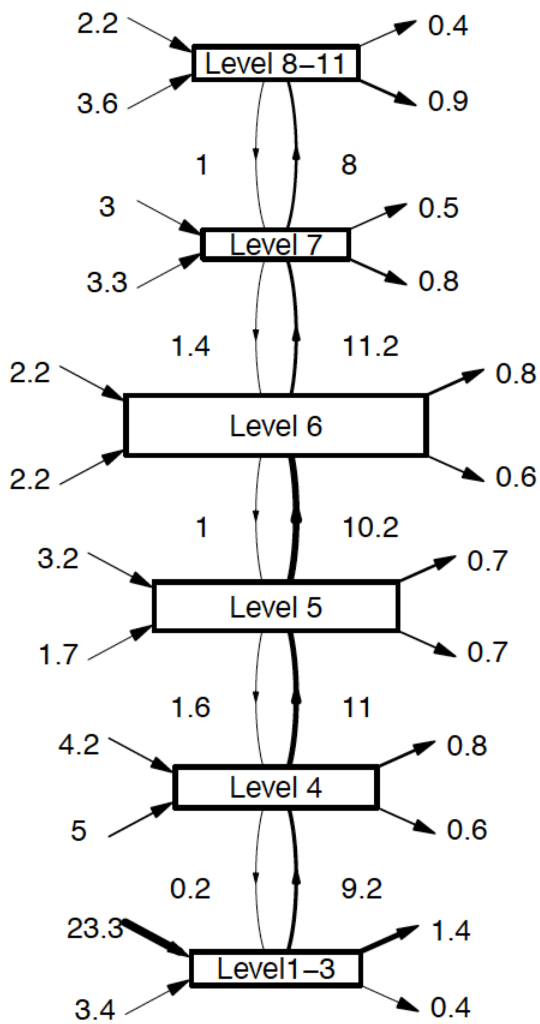
C Additional Figures and Tables

Figure A1: Flows Across Job Level



The flow diagram shows the flows across the different job levels in the firm. 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 the firm. The percentages referring to flows between job levels are percentages of the origin job level. For example, 5.2 percent of employees working at level 4 entered into that job level in the preceding year. 8.7 percent are quitting or are being laid off from level 4. 9.6 percent of the level workforce are promoted in a given year and 0.2 percent are demoted to levels 1-3. The numbers are averages over our sample period. Flows between other than the indicated job levels are rare and not shown in the graph.

Figure A2: Flows within the Branch Network



The flow diagram shows the flows across the different job levels within the branch of the firm. 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 (the top arrow), or entering the branch from central office (the bottom arrow). On the right are those leaving the firm (the top arrow) or moving to central office (the bottom arrow). The percentages referring to flows between job levels are percentages of the origin job level. For example, 4.2 percent of employees working at level 4 entered into that job level in the preceding year, and 5 percent of them were coming from the central office. 6.7 percent are quitting or are being laid off from level 4 and 4.6 percent are are switching to the central office. 12.4 percent of the level workforce are promoted in a given year and 0.2 percent are demoted to levels 1-3. The numbers are averages over our sample period. Flows between other than the indicated job levels are rare and not shown in the graph.

Table A1: 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	0.103***	0.102***	0.062***	0.093***	0.050***	0.082***	0.040***	0.074***	0.032**	0.072***	0.042***
	(0.014)	(0.015)	(0.011)	(0.014)	(0.012)	(0.015)	(0.012)	(0.016)	(0.013)	(0.017)	(0.015)
Lag 1 ϕ			0.054***		0.027***		0.019**		0.017*		0.005
			(0.008)		(0.007)		(0.008)		(0.009)		(0.012)
Lag 2 ϕ					0.036***		0.017**		0.015*		0.020*
					(0.008)		(0.007)		(0.009)		(0.011)
Lag 3 ϕ							0.032***		0.017**		0.003
							(0.009)		(0.009)		(0.010)
Lag 4 ϕ									0.023**		0.019*
									(0.010)		(0.011)
Lag 5 ϕ											0.010
											(0.009)
Restricted		X		X		X		X		X	
Observations	77,077	57,744	57,744	42,575	42,575	31,366	31,366	22,569	22,569	15,401	15,401
Partial R-squared	0.818	0.813	0.813	0.814	0.815	0.820	0.820	0.821	0.822	0.823	0.824

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