# The Value of Bosses

Edward P. Lazear,
Kathryn L. Shaw,
and
Christopher T. Stanton

Stanford University,
Stanford University,
University of Utah

Preliminary
Do Note Quote

November 4, 2011

Abstract

As more productivity data become available, it is possible to examine the effects of people and practices on productivity. Arguably, the most important relationship in the firm is between worker and supervisor. The supervisor hires and fires, assigns work, instructs, motivates and rewards workers. Models of incentives and productivity build at least some subset of these functions in explicitly, but because of lack of data, little work exists that demonstrates the importance of bosses and the channels through which the productivity enhancing effects operate. Using a unique company based data set, supervisor effects are estimated and found to be large for technology-based services workers. The three most important findings are: The only "peer" that matters is the boss. In this environment peers have little or no effect on output, but bosses affect workers significantly. Second, there is substantial variation in boss quality, measured by their effect on worker productivity. Replacing a boss who is in the lower 10% of boss quality with one who is in the upper 10% of boss quality increases a team's total output by about the same amount as would adding one worker to a nine member team. Third, the marginal product of a boss is about 60% greater than that of a worker, commensurate with the ratio of their wages. Additionally, good bosses should be sorted to the best workers: although good bosses increase the productivity of both good and bad workers, they increase it by more for the firm's top performers.

Workers depend on their bosses in many ways. First, the hiring decision is generally made with input from a worker's superiors, sometimes direct, sometimes more removed. Second, the supervisor is likely to be important in motivating a worker, which in turn affects raises, promotions and other benefits. In extreme cases, supervisors discipline and terminate workers. Third, supervisors assign tasks to workers and tell them what they must do and may not do on the job. Fourth, the supervisor acts as mentor or coach, teaching his subordinates the techniques that will enhance their productivity.

Despite the clear and important role that supervisors play, the economics literature has been silent on the effects that bosses actually have on affecting worker productivity.[1] Even more to the point, the literature has not been able to speak to the importance of the various mechanisms through which boss effects might operate. Most of this is a data issue, but some of it reflects the fact that the literature has modeled the relationship between boss and worker at an abstract level and has not pushed beyond to ask about what is likely to be the most important relationship in the workplace.

The neglect is even more striking when contrasted with the interest in peer effects. There is a large literature, both theoretical and more recently empirical, that has focused on the effects of workers on their peers and team members.[2] Peer effects may be important, but except in a few

[1]Some early exceptions are Herbert Simon on firm size and compensation (1957) and Rosen on the span of managerial control (1982).

[2] For theory, see Kandel and Lazear (1992). For empirical examples, see Mas and Moretti (2009), and Falk and Ichino (2006). For work on teams and complementarities, see Ichniowski

industries, like academia, where the structure is very flat and workers have much authority over what they do, the relationship with one's boss is likely to be as or more important than that to any other worker. At a minimum, this remains an open question and one that should be investigated.

By using data from a large service oriented company, it is possible to examine the effects of bosses on their workers' productivity and to compare them to individual and peer effects. The primary findings are:

1. Bosses are important and vary in productivity. Replacing a boss who is in the lower 10% of boss quality with one who is in the upper 10% of boss quality increases a team's total output by about the same amount as would adding one worker to a nine member team.

2. The marginal product of a boss is about 60% greater than the marginal product of a typical worker. The ratio is consistent with differences in compensation levels.

3. Bosses are the only "peer" that matters. Peer effects are small or zero, whereas boss effects are substantial.

4. Good bosses increase the productivity of many different types of workers. Bosses who are good for old workers are also good for new workers.

5. The difference between the effect of good and bad bosses on high quality workers is greater than that on lower quality workers, which suggests that to the

---

and Shaw (2003).

extent that the same boss is good for both, the assignment of the good boss should be made to the higher quality worker. Comparative advantage is key. Allocating bosses appropriately can raise firm productivity.

## I. Theoretical Framework

Workers and bosses together produce output.

### A. The Output of Workers and Bosses

An individual's output, q, depends on human capital, H, which reflects both innate ability and previously learned skills, and on effort, E. A natural specification is multiplicative: harder work results in greater returns to human capital

(1) $q = H * E.$

For example, one measure of effort is time worked. H is normalized such that the average worker has H=1. H scales hours appropriately to turn effort, here hours, into units of output. A "unit" of output is then defined as the amount of output that an average worker produces in one period.

For now, let us focus on the motivating and teaching roles of supervisors and ignore task assignment, hiring, firing and other aspects of the supervisor job. It is necessary to define boss effects before they can be discussed. Because every worker has a boss, at least at some level, the boss effect cannot be the difference between having a boss and not having one. Instead, think of the boss effect as the importance of different quality bosses on the output of their subordinates. Without loss of generality, define an index $S_i$, i=t,m for teaching and motivation. The best boss is defined as the boss who has the largest positive effect on subordinates' output. The worst boss

is defined as the boss who has the smallest positive or possibly even most negative effect on worker output.

Then the boss effect is defined from (1) for teaching and motivating as

$$(2) \qquad \frac{\partial q}{\partial S_i} = E\frac{\partial H}{\partial S_i} + H\frac{\partial E}{\partial S_i} \qquad i = t, m$$

The first term reflects the effect of having a better boss in either the teaching (subscript i=t) or the motivating dimensions (subscript i=m), on a worker's output through the human capital (H) channel. Higher quality bosses impart more knowledge on their subordinates and the first term refers to that effect. The second term reflects the effect of having a better boss on a worker's output through the effort (E) channel. The indexes $S_t$ and $S_m$ refer to the amount of quality adjusted time that a supervisor places on H and on E, respectively. It seems reasonable that the effect of $S_t$, supervisor teaching time, would operate primarily through H and the effect of $S_m$,supervisor motivating time, would operate primarily through E, but nothing in the specification requires this.

One can think of $S_t$ and $S_m$ as partly reflecting endowments to supervisors of teaching and motivating skills and in partly reflecting choice over investment as time allocation between motivating and teaching skills or activity.

There are a number of different supervision processes that have economic interpretations. One possibility is that $S_t$ is exactly proportionate to $S_m$ and that nothing else matters. There is fundamentally only one kind of supervisory ability, S, and motivating ability and teaching ability are proportionate to it. In this case, the best motivators are also the best teachers. A regression of $S_t$ on $S_m$ would yield an intercept of zero and an $R^2$ of one.

Another possibility is that all bosses have an identical fixed amount of quality-adjusted supervisor time, and that those supervisors who, for whatever reason, spend more time motivating spend correspondingly less time teaching. A regression of $S_t$ on $S_m$ would yield a coefficient of negative one on $S_m$ , again with an $R^2$ of one.

Another view is that of $S_t$ and $S_m$ are endowed in bosses and not subject to choice at all. Whether $S_t$ and $S_m$ are observed to be positively or negatively correlated in the population of bosses would depend on the joint density of $S_m$ and $S_t$ that characterizes the population. One possibility is that nature endows skills in ways that result in positive observed correlations. Those who are best able to teach are also able to be efficient motivators. An alternative is that $S_t$ and $S_m$ are negatively correlated in the population. Drill sergeants may be good at getting subordinates to show up for work, but may not be great psychotherapists or nurturing teachers so that those who are endowed with high $S_m$ skills are also endowed with poor $S_t$ skills.

The question is an empirical one and cannot be settled a priori. If a positive correlation is observed, then it is necessarily the case that individuals differ in their overall endowments (or acquired levels) of quality adjusted time or of the specific teaching and motivating skills. It is still possible that bosses have the ability to choose how much of their time to spend on motivating versus teaching, but in order to observe positive correlations across people in $S_m$ and $S_t$, it is necessary that total quality-adjusted supervisory time varies.

Whether choice is involved remains crucial, but difficult to determine empirically. If there is room for choice, then the senior management can simply instruct supervisors to alter their time allocation in a direction that enhances supervisor productivity. If, on the other hand, little or no choice is involved and the observed $S_t$ and $S_m$ varies across people because of their

endowments, then senior management's only tools for affecting the allocation between teaching and motivating is the recruitment of the supervisors with the best combination of talents and the firing of those with the worst.

This framework suggests the following empirical questions:

**E1:** Do bosses matter? Do they raise workers' output? If so, by how much? Specifically, if bosses do matter, then some combinations of bosses' levels of $S_t$ and $S_m$ must differ across bosses.

**E2:** Do bosses matter because they teach or because they motivate? Which dominates? Given equations (1) and (2), some assumptions must be made to distinguish between teaching and motivation.

**E3:** Is a good boss good at both teaching and motivating because these are endowed traits; or, are teaching and motivating substitutes in the boss's allocation of time on the job?

### B. Sorting Bosses to Workers

Sorting is key. Should good bosses be matched with good workers? Suppose good workers are defined as those who have higher levels of human capital, H. From (2), the boss effect depends on the effect of H on output:

$$\frac{\partial^2 q}{\partial S_i \, \partial H} = \frac{\partial E}{\partial H}\frac{\partial H}{\partial S_i} + E\frac{\partial^2 H}{\partial S_i \, \partial H} + H\frac{\partial^2 E}{\partial S_i \, \partial H} + \frac{\partial E}{\partial S_i}.$$

The sign is ambiguous. Because E is a choice variable for the worker and because there may be a relation of H to E, $\partial E/\partial H$ cannot be assumed to equal zero. All other terms are positive with the exception of $\partial^2 H/\partial S \partial H$ and $\partial^2 E/\partial S \partial H$, which cannot be signed a priori. If a good boss were more valuable to less able workers than to more able ones, then $\partial^2 H/\partial S \partial H$ would be

negative. Similarly, if good motivating were more important for lazy workers than for energetic ones, $\partial^2 E/\partial S \partial H$ would be negative. Sufficiently strong effects of either or both types could mean that it is better to sort good bosses with low ability workers. This is an empirical question, but one that can be resolved by the data used in the empirical section.

Thus, additional questions are:

**E4:** Do boss effects differ for star workers and laggard workers? A laggard may have more room for improvement; a star may be more receptive to improvement.

**E5:** Are there complementarities between bosses and workers? Does a good bosses produce more output when matched with star workers than when matched with laggard workers?

### C. Workers are Additive; Bosses are Multiplicative

Why is a research scientist who has a great breakthrough so valuable to a firm? It is because the innovation enhances the productivity of a large number of workers. The effect is multiplied by the number of workers that it affects.

The same is true of bosses. A good claims processor can process a larger number of claims than a poorer one, but the effects are limited to the claims that the worker processes himself. The quality of supervisor affects output primarily through the work of subordinates and an increase in supervisor quality is multiplied by the number of individuals who are touched by that supervisor. Thus, the effects are multiplicative for supervisors and additive for workers.

Formally, the total output of the firm, $Q$, is the sum of the individual worker's (excluding supervisor) outputs, $q_i$,

(3) $$Q = \sum_i q_i$$
$$= \sum_i q\ (x_i) + \alpha_i + \sum_j \delta_j D_{ij}$$

where $x_i$ are observables that affect output (like tenure and time), $\alpha_i$ is worker i's fixed effect, $\delta_j$ is boss j's fixed effect and $D_{ij}$ is a dummy equal to one if worker i is a subordinate of boss j. It follows from (3) that

$$\frac{\partial Q}{\partial \alpha_i} = 1$$
$$\frac{\partial Q}{\partial \delta_j} = \sum D_{ij}$$

Thus, the effect of worker talent on output is just the effect itself, whereas the effect of the boss's talent on output is multiplied by the number of individuals that she supervises.

Peers could have the same multiplicative effect as bosses. If one peer influences all his team members, his effect is multiplicative in the same way as that of the boss.

**E6:** Are boss effects bigger than peer effects?

**II. Data**

The data contain four years of daily productivity transaction records between June 2006 and May 2010 from an extremely large services company. There are 23,878 unique workers and unique 1,940 bosses, for a total worker-day sample size of about 5.7 million observations. This company has multiple different service functions, but the data used come from one task classification where workers are involved in general customer transactions. This ensures that all workers in the sample perform approximately the same tasks. Because of confidentiality restrictions, most detail about the day-to-day tasks that workers perform must be suppressed.

The data come from many sites, but number of sites is also suppressed for confidentiality reasons.

The jobs are labeled, "technology-based service" jobs or "TBS jobs." Examples include insurance-claims processing, computer-based test grading, technical call centers, some retailing jobs such as cashiers, movie theater concession stand employees, in-house IT specialists, airline gate agents, technical repair workers, and a large number of other jobs.

Consider a detailed example of a TBS job: workers doing computer-based test grading. Most U.S. states expect students to take standardized tests, such as the "Star" tests in California. The students' handwritten essays (from science to English) are scanned into a computer, and then the graders of these tests sit in large rooms, where they grade each essay on a computer. Their work is timed and checked for quality. They must be at their desk a certain percent of the day (defined as 'uptime' below), which is recorded. They have modest amounts of incentive pay. They are often given daily feedback on their performance, and they can see some measures of the performance of other team members. Their bosses sit with them and teach and motivate the workers. While this may seem like an unusual example, we made a number of plant visits to companies like this, and all visits shared this typical scenario.

These are labeled TBS (technology-based service) jobs because the company uses some form of advanced IT system to record the beginning and ending time for each transaction, or to record the daily volume of transactions, for each worker. As described above, many production processes in services now fit this description. The technology that is used to measure performance may be a new computer-based monitoring system (as in the standardized test grading above), an ERP (Enterprise Resource Planning) system that records a worker's

productivity each day (such as the number of windshield repair visits done by each Safelite worker (Lazear, 1999; Shaw and Lazear, 20xx)), cash registers that record each transaction under an employee ID number, call centers, or computer-monitored data entry. These TBS jobs are likely to be widespread and represent a major IT-based shift in computerization and worker productivity. While some of these jobs are outsourced to firms outside the U.S., many remain in the U.S., particularly when the customer interaction is face-to-face or the work is idiosyncratic and skilled (as in test grading).

In our data, the TBS workers are doing reasonably technical work, with a computer interface. New products or processes are introduced over time, and thus there is constant learning on the job. Bosses are constantly teaching.

The workers are working in areas, which are labeled "teams" herein. In this firm, the average daily team size is 9.04 workers, and each team is managed by one boss. The team is identified through the worker's link to a boss identification number; all workers with the same boss that day are said to be part of the team. Workers switch bosses about four times a year.[3] It is these switches to different bosses that permits us to estimate the effect of bosses on worker productivity.

There are two measures of output. One is productivity, which is output-per-hour, and as shown in Table 1, in these data each worker handles about 10.3 transactions per hour. The second measure is uptime. In any hour at work, workers miss some of the time for breaks, etc., leaving their work areas and thereby slowing the entire system. The mean uptime is 96.3%, and

[3] The worker-boss pair is defined by the usual worker-boss pairing. If a boss were absent on any given day, the usual boss would be the one of record.

the standard deviation is a small 3.0%. See Table 1 footnotes for more details.

Most of the variation is in output-per-hour rather than in uptime. The standard deviation of output-per-hour is 30.8% of its mean; the standard deviation of uptime is 2.8% of its mean. Consequently, the initial discussion and results focus on variation in output-per-hour. Later, the analysis is done on uptime. Most of the workers' variation in performance operates through productivity rather than uptime. Temporal variation in the demand for a worker's services can be taken into account through the use of time dummies and by using the group mean output for other workers on any particular day (described below in footnote 8).

## III. Empirical Results: Boss Effects and Peer Effects

### A. Boss Effects

The most important finding is that bosses have large, varying and significant effects on worker productivity. The basic unit of observation is a worker-day so that each of the 5.7 million observations represents output for a given worker on a day on which he worked. The baseline productivity regression is

$$(4) \qquad q_{ijt} = X_{it}\beta + \alpha_i + \delta_j + t + \varepsilon_{ijt.}$$

The goal of estimating equation (4) is to recover estimates of individual worker and boss fixed effects as well as summary measures of the distributions of $\alpha_i$ and $\delta_j$. Taking the variance of the estimated boss effects and worker effects as an estimate of the true variances is problematic because the estimated boss and worker effects contain sampling error. Following Rockoff (2004), the estimated boss effects and worker effects are assumed to be drawn from a normal distribution with unknown variance $\sigma_\delta^2$ and $\sigma_\alpha^2$. The variance of the estimated boss and worker effects is then composed of the sum of sampling error and the true variance of the boss

and worker effects. The parameters $\sigma_\delta^2$ and $\sigma_\alpha^2$ can be recovered with maximum likelihood given a consistent estimate of the sampling variance of the fixed effects.

In some applications, the sampling error can be estimated using the variance-covariance matrix of the fixed effects. Because of the scope of the problem of estimating worker and boss effects jointly, the variance covariance matrix cannot be recovered because of computational constraints. For each boss fixed effect, we instead use $\sigma_{\varepsilon(j)}^2/(N_j - 1)$ as an estimate of the sampling variance, and for each worker fixed effect, we use $\sigma_{\varepsilon(i)}^2/(N_i - 1)$. This provides a slight upward bias in the estimated variance of the boss and worker fixed effects because the diagonal elements of $X'X^{-1}$ are larger than $1/(N_j-1)$ and $1/(N_i-1)$.[4] In each specification, the ML estimates are compared to weighted and unweighted estimates of the fixed effects.

Table 2 reports the basic regression results where output-per-hour is the dependent variable. Column 1 reports the R-squared from the most basic regression of output-per-hour on a fifth order polynomial of daily tenure, monthly time dummies, and day of the week dummies, while restricting $\alpha_i=\alpha$ for all i and $\delta_j=0$. Because tenure and monthly time dummies are highly correlated within worker, a "restricted tenure polynomial" is imposed to separately identify the month dummies from the tenure profile. Tenure enters the specification as $g(ten) = 1\{ten < 365\} * f(ten) + \varphi 1\{ten \geq 365\} + 1\{ten \geq 365\} * f(365)$, where f is a fifth order polynomial over the first year. After the first year, f is evaluated at 365, and $\varphi$ captures this censoring of the tenure profile. Not surprisingly, and consistent with prior work in other

[4] Using a subset of the data to compare this procedure with estimates produced using a heteroskedasticity robust variance covariance matrix clustered by boss suggests that the bias from ML is about 15-20%.

industries,[5] the output is increasing and concave in tenure (results are shown in section IV.A. below).

The rest of Table 2 is of primary interest. In column 2, worker fixed effects are added to the basic regression. Worker fixed effects are clearly important. The R-squared rises from 0.061 to 0.237 with an F(23877, 5705571) statistic of 55.6 (p-value = 0), rejecting the null hypothesis that the set of individual fixed effects are zero. The variation in fixed effects is large. A one-standard deviation increase in the worker fixed effects increases worker output-per-hour by 18%.[6] Column 3 includes only boss fixed effects. Boss fixed effects also matter. The R-squared increases from 0.061 to 0.092 with an F(1939, 5727509) statistic of 102.1 (p-value = 0), rejecting the null hypothesis that the boss fixed effects are jointly zero. Although the importance of the boss effect is striking, because of potential non-random assignment of workers to bosses, little can be inferred about the importance of bosses without taking worker effects into account.

The more important results are in column 4, which includes both worker and boss effects. Here, worker fixed effects and boss fixed effects are estimated jointly with dummy variables included so that $D_{ij} = 1$ if individual i has boss j as his supervisor on the day of the observation in question.[7] With both fixed effects, the R-squared rises to 0.243. Worker fixed effects and

[5]See Lazear (2000), Shaw and Lazear (2008) for examples with productivity data.

[6] It is well-known that there is significant variation in worker wages and that in panel data, the worker-specific fixed effects explain much of that variation. It is less well-known, primarily because of lack of data on individual worker productivity, that there is significant variation in worker productivity and that worker fixed effects explain much of the variation.

[7] Boss and worker effects are both estimated from a full set of dummy variables using a sparse matrix implementation. Because of the large size of the matrix of regressors, a conjugate gradient algorithm is used to minimize the sum of squared residuals. Credible standard error estimates are only available using complicated re-sampling techniques because of the two-way structure of the panel, so inference instead relies on comparisons of restricted and unrestricted

boss fixed effects are each significant. The F(23878, 5703630) statistic is 47.7 (p-value=0) on worker fixed effects and, for the boss effects, the F(1940, 5703630) statistic is 20.5 (p-value = 0). While the levels of workers' productivity can be affected by demand conditions for their services, careful analysis of the robustness of the results suggests that the magnitudes revealed here are little changed with a range of controls for varying demand conditions.[8]

One method for assessing the robustness of these results with respect to demand shocks is contained in the final column of Table 2. The mean of other team members' output is introduced by adding a variable that is the mean contemporaneous (daily) output for all other peer team members, excluding the worker. This mean value of daily team output will pick up changes in daily demand, and therefore, it has a positive coefficient. However, the standard deviation of the estimated worker effects and boss effects falls only very modestly when this stringent control for demand is introduced.

Bosses affect all workers that they oversee, and each boss effect must be multiplied by the number of workers supervised by the boss to get the effect of the boss on overall productivity. The last few lines of Table 2 report the boss effects while assuming that each boss

---

models: F-statistics are used to test for the introduction of alternative sets of fixed effects, which are the core theme of this paper. In addition, the fixed effects are identified within "connected groups" of workers and bosses. Over 99% of the workers and bosses in the data are connected in the same group. The estimated fixed effects are only identified up to normalizations, so the mean of the worker effects is set to zero and one boss effect is zero. For a precise statement of the identifying conditions, see Abowd, Creecy, and Kramarz (2002).

[8]Market conditions might affect output some, but this is really a question of how good the firm is at adjusting the number of employees so as to keep the transaction arrival rate, which will be referred to as demand, close to constant for any given worker. Because this firm adjusts the number of hours worked so as to ensure that workers have virtually no slack, there is relatively little variation in output that is caused by fluctuations in transaction arrival rates (demand shocks). Introducing daily time dummies as well as person and boss fixed effects leaves very little remaining variation; monthly time dummies are introduced and are significant.

is assigned to an average size team.

Even among the selected sample of those who are promoted to boss, there is large variation in the effect of bosses on worker output. The variation in the effect of bosses on output is about 2.5 times as great as the effect of individual variation in worker quality on total output. This is one of the most significant findings of the paper.[9]

It is not surprising that bosses differ in their quality and effect on output, but it is noteworthy that being assigned to one boss over another affects worker productivity by as much as individual worker variation does.

To get a sense of the magnitude of this effect, Figure 1 displays a histogram of the estimated boss effects multiplied by team size. In addition, the standard deviation of boss effects is between 2.3 and 4.1 times of the standard deviation of worker fixed effects, depending on the weighting. (The estimated boss effects can be weighted by worker-day or by boss, and the different results are described in the next sub-section B and the footnotes to Table 2.) Replacing a boss who is in the bottom 10% with one in the top 10% of quality is equivalent to gaining about 123% of a typical worker's output.[10] Put differently, changing from a boss in the bottom 10% to one in the top 10% would be about as important as adding a full worker to the team that

[9] There are three estimates of boss effects reported in the table: the weighted boss effects use the boss*agent day weights to give more weight to bosses who have greater presence in the data. The boss effects that are unweighted take an individual boss as the unit of observation. The unweighted boss effects likely reflect variation that mirrors the set of potential bosses. Because the set of bosses observed in the data is likely a selected sample, there is a plausible argument that the measure with more variation (unweighted) should be used. However, there is a tradeoff between precision of the estimated boss effects and sample selection—as a boss interacts with more workers for longer periods of time, the individual boss effect can be estimated more precisely. More will be said about this in section C below.

[10] This is calculated as 6.42-(-6.24) / mean output, where 6.42 and -6.24 correspond to the top and bottom quintile of the unweighted boss effects.

is supervised.

Finally, it is important to remember that the estimates of boss effects are lower bounds of the maximum boss effect because of the promotion rule. The worst conceivable boss is not likely to be in our sample of bosses.

**B. The boss effects are identified**

Holding constant the worker's quality, $\alpha_i$, the boss effect $\delta_j$, is identified by those workers who switch bosses. The boss effects in (4) are estimated off of "changers." In order to estimate the effect of a boss on worker productivity, the same boss must work with different workers, whose abilities are known through the worker fixed effects. Logically, if a given worker switches from boss A to boss B and his productivity rises, then the change in productivity is attributed to the change in bosses. For any given boss, the boss effect is therefore estimated as the average over all workers' changes to that boss. More precisely, the boss effects are estimated within "groups" of connected workers in the graph-theoretic sense.[11] If a separate group of bosses and workers is not connected, no worker or boss ever interacts with any other worker or boss in the non-connected group. Within each group, there must be one normalization of the boss effects and one normalization of the worker effects.

To get a sense of what this means, suppose that there are two bosses interacting with several workers in a single connected group. Because both bosses and all workers are connected, this means that some workers switch bosses. The relative difference in the boss effect is just $\delta_2 - \delta_1$. Constraining $\delta_1=0$, $\delta_2$ is estimated from the set of workers that switch bosses. Now suppose

[11] "When a group of [workers] and [bosses] is connected, the group contains all the workers who ever worked for any of the bosses in the group and all the [bosses with] which any of the workers were ever [assigned]" (Abowd, Kramerz, and Woodcock, 2006).

there is a third boss who is in the connected group. Being in the same connected group implies that at least some workers have been supervised by boss 3, and at least one of these workers has worked with another boss. In this case, $\delta_3$ is identified from switches between boss 3 and boss 2 or boss 3 and boss 1 because $\delta_2$ is already identified relative to $\delta_1$.

Stated more generally, within each connected group, $\delta_j = E(y \mid \text{Boss } j) - E(y \mid \text{Boss } 1)$ where $\delta_1$ is normalized to equal 0 for the first boss within each group. Each boss effect captures the bosses' change in productivity relative to the excluded boss. What conditions are necessary such that $\delta_j = E(y \mid \text{Boss } j) - E(y \mid \text{Boss } 1)$? To estimate this average treatment effect of changes in boss quality on worker productivity, either there is random sorting between bosses and workers after accounting for the worker's fixed effect and X, or the boss treatment effects must be homogeneous across workers. That is $\delta_j = E(y \mid \text{Boss } j, \text{worker } i) - E(y \mid \text{Boss } 1, \text{worker } i) = E(y \mid \text{Boss } j, \text{worker } k) - E(y \mid \text{Boss } 1, \text{worker } k)$ for all workers i and k. Sorting of workers to bosses and heterogeneity in the treatment effects are addressed in detail below in section VI.

How much data is there to estimate the boss effects within each connected group? The dataset is the population of workers in the firm from 2006 to 2010. For each worker, there is an average of 240 days of daily productivity data (or about a calendar year of data). Each worker changes bosses about 4 times during this interval. Therefore, when the boss is the unit of analysis, his team members have, on average, touched 4.7 other bosses. Given the average number of workers per boss, the number of worker changers per boss is 49 (or 80 if weighted by the number of observations per boss). These are sizable numbers. As a result, 99.99% of the daily data is in the largest connected group, with only 560 of the 5.7 million observations and 11 of the 1,940 bosses outside of the largest group.

### C. Serial correlation

Serial correlation that influences the assignment of bosses and workers may be a concern. If workers' output is serially correlated and assignment of bosses to workers depends on the past history of workers' output, then the variance of boss effects may be overstated. On the other hand, if workers' output is mean reverting and assignment is based on transitory shocks, then the boss effects may be understated. To investigate, equation (4) is re-estimated with lagged worker output on the right hand side. [12] Results are presented in the Appendix Table using two different lag specifications: a daily lag and a 14 day average of past worker output. In all cases, the estimated standard deviation of boss effects is substantially larger than the standard deviation of worker effects; additionally, the ratio of the variance of boss effects and worker effects appears stable.

### D. Are the results sensitive to the number of observations per boss?

Some bosses are long-lived in the data set; others are short-lived and thus have few workers per boss identifying their boss effect. The short-lived bosses introduce noise into the estimation of the boss effects, but the conclusion that bosses have large and varying effects is unchanged when we take this noise into account and attempt to correct for it.

The estimated standard deviation of the boss effect becomes smaller when focusing on the bosses that have many worker observations in the data. Figure 2 plots the estimated boss fixed effects as a function of the number of total observations (worker-days). The results look similar when plotting the estimated boss effect as a function of the total number of workers the

[12] This estimator is consistent, even with lagged output and worker fixed effects provided that T, the time dimension of the panel, goes to infinity. Otherwise, an instrument is necessary for a workers' lagged output.

boss supervises. Inspection of the figure reveals that the dispersion of the estimated boss fixed effects declines as the number of worker-days increases.

What causes this? The declining dispersion of the estimated boss effects with the frequency of worker observations per boss could be due to the sorting of bosses in and out of the firm, or due to measurement error in the estimated fixed effects. There is clearly an argument in favor of measurement error: as the number of workers per boss falls, the variance of the estimated boss fixed effects rises, because the extreme values of the boss fixed effects are estimated with very few workers per boss (this is clear in Figures 1 and 2). There is also an argument in favor of sorting; extremely good or bad bosses are more likely to be at the firm for short durations (due to firing or quits), implying that bosses with many worker observations represents a selected sample, but presents a conservative estimate of the variance in boss effects for core permanent bosses.

### E. Peer Effects

There is a growing literature on peer effects.[13] The basic specification with two-way fixed effects is run while adding a peer effect:

$$(5) \qquad q_{ijt} = X_{it}\beta + \alpha_i + \delta_j + \xi\, p_{ijt} + t + \varepsilon_{ijt}$$

where peer effect, $p_{ijt}$, is specified in several ways.

The naïve way to examine whether peers matter is to compute the average output of other

[13] Most current peer effects papers test whether workers learn from each other due to proximity, or adjust their effort in response to those who work around them (Falk and Ichnio, 2006) or who watch them (Mas and Moretti, 2009). Few papers test for the complementarity of skills within the teams that are formed among peers, because skills are unobserved and most data has come from production functions (like store clerks) that are largely individual output, not team output. That is true of this data.

workers in the team on a given day and see whether this affects worker productivity. In column 1 of Table 3, the peer effect variable is the average output of the team with which the worker works, excluding own output. The coefficient is .16 which suggests that a one standard deviation increase in a peer's ability increases own output by .063 units for a worker with average output of about 10 units per hour. This effect of any one peer on a given worker's output is calculated as

$$\frac{\partial OPH}{\partial Peer\ Output} = \frac{\partial OPH}{\partial Team\ Average\ Output} \frac{\partial Team\ Average\ Output}{\partial Peer\ Output}$$

where the change in the team average output is the change in an individual's output divided by the team size-1. A one standard deviation change in the quality of a peer is equal to about 3 units of output per hour, so the effect on a worker's output of working with a peer who is one standard deviation better is (.16) (3.16) / (9.04-1) = .063.[14]

This is a marked overestimate of the true peer effect. In addition to the standard concerns about the reflection problem (Manski, 1993), in these data, the calculated peer mean is a proxy for daily demand conditions: if daily demand falls, productivity falls for the team co-workers as it falls for the worker 'i' in the regression. Therefore, the estimated peer effect reflects demand variation, rather than the spillover of one worker's quality on another. The estimated peer effect has a strong upward bias. In these TBS jobs, demand can vary due to fluctuations in transactions from customers, from co-workers who are passing on work, or from technology mal-functions. The calculated peer team mean does an excellent job of controlling for all of these, and the

[14] The change in the peer can't be divided by the team size if you're using the boss*teamSize because the change in the peer affects all team members

estimated variance of the boss fixed effects is virtually unchanged (comparing to Table 2 boss effects).

One way to eliminate the temporal reflection problem and the time-based spurious effects is to use peers' fixed effects as measures of the peer output rather than the contemporaneous productivity of other team members. We use a two-step non-linear least squares routine to jointly recover estimates of peer effects, worker effects, and boss effects. The estimating equation for the joint model is

(6) $q_{ijt} = X_{it}\beta + \alpha_i + \delta_j + \xi\, p_{ijt} + t + \beta_{Peer}(TeamSize - 1)^{-1} \sum_{k \in jit\backslash\{i\}} a_k + \varepsilon_{ijt}$

where summation over $k \in jit\backslash\{i\}$ captures worker i's team on day t with boss j while excluding worker i. This specification allows the estimated peer effect to depend only on the permanent effect of co-workers on the team, $\alpha_k$, not on concurrent $q_{ijt}$. Estimation of the joint model is not feasible on the full set of data because of memory constraints; storage of the matrix of peer-indicators, even in sparse form, requires an order of magnitude more memory than storage of the data with only worker and boss indicators. Because workers and bosses rarely move establishments, the joint procedure can be applied using subsets of establishments. The estimation algorithm is a two-step procedure. The outer-loop "guesses" a value of $\hat{a}_{Peer}$ and then computes the remaining parameters via a linear conjugate gradient procedure in an inner-loop conditioning on the value of $\hat{a}_{Peer.}$ Search is then over $\hat{a}_{Peer.}$

The second column of Table 3 provides estimates of peer effects, worker effects, and boss effects recovered jointly using the two-step non-linear least squares routine. The regressions in column 2 uses a subset of the data corresponding to a typical region, because joint estimation of worker effects and unconstrained peer effects is only feasible on subsets of the

data. The estimated peer effects are zero in column 2. The peer effects are not economically significant relative to boss and worker effects.

Another method to estimate peer effects uses a peer's first few months of output as a proxy for the peer's current output. These results are provided in the last column. Again, the coefficient is negative but close to zero.

The conclusion is that peer effects are very small relative to boss effects.[15] Note that this production environment has relatively little teamwork because each worker primarily interacts with a customer, not with other workers.[16] While the workers can see each other and may learn from each other or compete with each other, the workers do not appear to be complements in production.

## IV. Why Do Bosses Matter: Teaching and Motivating

How much do workers learn, and how much do bosses matter for this learning? These questions are addressed by empirically examining the learning curve for workers and then examining the impact of bosses on learning.

In this firm, and in many other technology-based service jobs, workers are required to have product knowledge, and the products are constantly changing. Consequently, it would seem that learning would be an important component of the job. The issue is not whether learning is important, but rather whether the variation that we see across bosses reflects

[15] There is also possible sorting of workers into teams of correlated peers, because good workers will work together if given the choice of their preferred shift and there are similar preferred shifts for all workers. If this sorting is temporal, based on recent performance (as it is), introducing worker fixed effects for peer effects will reduce the bias. If the sorting is based on permanent performance, there will be an upward bias in the estimated peer effects. Given that the peer effects are zero or negative, this is not a concern.

[16] The same is true in Mas and Moretti ( 2009), who also find significant, but small peer effects.

difference in their teaching ability, or in something else that we refer to as motivational skills.

### A. Learning and Effort by Workers

Workers learn substantially in their first months on the job. For these jobs, a portion of the learning is firm-specific and a portion is occupation-specific, and the regressions do not hold constant the latter because the data contains only the start date with the current firm, not general occupational experience. Therefore, the tenure coefficients combine firm-specific learning with occupational learning for those who did not arrive with previous occupational experience, but estimate firm-specific learning for those who arrived with previous experience. For this type of job, there is negative self selection in exit: the best new hires are more likely to leave the firm than the worst new hires.

### B. Teaching and Motivating by Bosses

What drives the dispersion in boss effects? The most obvious factors, especially in the context studied, are teaching and motivating. Some effects of bosses are persistent and some are temporary. When a worker switches bosses, some of the productivity increase that comes from having been with the previous boss remains and some does not. Suppose that prior bosses affect the current period. Then call that which persists "teaching" and that which is only contemporaneous "motivation". Let λ be the persistent portion of the boss effect (teaching), and 1-λ of the boss effect is contemporaneous (motivation). If the persistent component of the boss effect depreciates, it is possible to identify the persistent contribution of past bosses because workers are often observed in many periods after a boss change. The following equation outlines the estimation approach:

$$(7) \qquad oph_{ijt} = \alpha_i + X_{it}\beta + \sum_j^J 1\{t \geq \tau_{ij}\}\gamma^{t-\tau_{ij}}\lambda\delta_j + D_{ijt}\delta_j + \varepsilon_{ijt},$$

where the term $\tau_{ij}$ captures the last calendar day that worker i works with boss j. Then the term t-$\tau_{ij}$ captures elapsed time since worker i was supervised by boss j. The term $\gamma$ is the monthly rate of geometric decay of the persistent portion of the boss effect. The estimated $\gamma$ is reported as a monthly rate of decay, using (t-$\tau_{ij}$)/30.

Table 4 contains the results.[17] The monthly rate of decay, $\gamma$, is estimated to be 0.87, and $\lambda$, the portion of the initial boss effect that is persistent, is estimated to be 0.78. It appears that bosses are mostly teaching. The weighted standard deviation of boss effects remains substantial, at 3.51, compared to a weighted standard deviation of worker fixed effects of 1.28. After 1 year, approximately 13% of a boss effect remains.

### C. Two Measures of Output: Productivity and Uptime

Table 5 provides results comparing output-per-hour and uptime. First note that the range of boss effects is larger, both in absolute and percentage terms, on output-per-hour than on uptime. The weighted standard deviation of boss effects is 3.31 for output-per-hour and 0.58 for uptime. This is in part a function of the fact that output per hour varies much more than uptime. The unconditional standard deviation of output per hour is 30.7% of the mean whereas the standard deviation of uptime is 2.8% of the mean. There is a limitation to how important bosses can be in motivating. But that is exactly the point. For service jobs of this type, where monitoring is easily performed by the information technology, the incremental effect on output of human monitoring or motivation is low. As a result, the variance in that effect is also low.

Most of the action then comes through changing the output per unit of worked time.

[17] The sample for this table comes from a subset of establishments. Estimation of (7) is conducted via non-linear least squares. The sample must be limited because storing the matrix of past boss histories for each worker-day is not possible for the full sample.

Little of this is likely to be due to human motivating, again because IT tracks worker output in terms of quantity and even quality because customers are surveyed on their experience. At least in the context of this type of service job, the supervisor's role appears to be one of coach, passing along information and tips on how to accomplish the work more efficiently.

Good bosses, however, appear good along both dimensions. The correlation between the boss effect on output-per-hour and the boss effect on uptime can be computed. The simple correlation of the two fixed effects (bosses on uptime and bosses on output-per-hour) is .16, which is significant at standard levels. Bosses who are better at increasing output-per-hour are better at increasing uptime in their workers as well.[18]

**Robustness**

The dataset is large. Therefore, it is possible to split the sample randomly into two separate groups partitioned by worker identifier to examine the extent of sampling error on the estimated boss effects. After splitting the sample, Sample A contains 2,862,270 person-days and Sample B contains 2,867,238 person-days.

Equation (4) is re-estimated separately for Sample A and Sample B. The correlation between the boss fixed effects in Sample A and Sample B is .36 (N=1854 for bosses in the same connected group in both Sample A and Sample B). Regressing the estimated boss effects on each other, $\delta_j^A = \alpha + \beta\delta_j^B$ for every boss j in the same connected group across samples, yields an estimated β coefficient of .36 with an R-squared of .13.

The importance of this analysis is that it provides a benchmark for what one would

[18] Worker fixed effects are already held constant so this is not a result of good workers sorting to good bosses.

expect when one regresses a boss fixed effect defined in one way on the same boss fixed effect defined in another. There are two reasons for the coefficient to be less than one. First, substantively the fixed effects are unrelated. Second, the fixed effects are estimated with error and errors-in-variables pushes the coefficient toward zero. The A, B sample approach above says that even when the coefficient should be 1 because of the random design for the subsamples, it is only .36, so all future regressions of one kind of fixed effect on another should be interpreted as deviations from some lower number, like .36, not 1.

## V. Heterogeneity in Boss Effects and in Worker Assignment to Bosses

The effect of bosses on a worker's productivity is likely to depend on the quality of the worker. Good bosses, especially those with teaching skills, may be most useful for those workers who have the most to learn (new workers who begin with low output). Alternatively, the ability of good teachers to raise the output of the best workers may be greater than that for low quality workers. The data provide ways to compare the quality of workers, which permits contrasting, for example, newly hired workers with experienced workers and star new hires with laggard new hires. Stars differ from laggards and old from new workers in a variety of ways that might suggest boss effects are different on the various groups.

Do the subgroups differ? Table 6 contains the summary statistics for group breakdowns. Old workers and stars have higher mean productivity than new workers and laggards. Old workers have more variation in output than new workers, and old workers have a slightly larger coefficient of variation than new workers (.32 compared to .29). Other interesting differences

come from comparing stars and laggards. Stars have higher variance in output than laggards, but the coefficient of variation is nearly identical for both groups (around .28).[19]

The next sections perform comparisons of estimated boss effects by group to determine for which groups bosses have the largest effects. The logic is as follows. If the boss effect is different for two groups, say, young and old, let $N_{it}$ be a dummy equal to 1 when individual i is young in period t, then

$$q_{ijt} = \alpha_i + \lambda^{New} N_{it} Q_{ijt} + \lambda^{Old} (1- N_{it}) Q_{ijt} + t + \varepsilon_{ijt} \quad (8)$$

where $Q_{ijt}$ is the single dimension of boss j quality, e.g., say, the boss's IQ, for the boss with whom worker i is matched in period t. Then $\lambda^{New}$ is the transformation of raw boss quality into worker productivity when the worker is a new worker and $\lambda^{Old}$ is the transformation of raw boss quality into worker productivity when the worker is an old worker.

Boss quality, $Q_{ijt}$, is unobservable, but boss fixed effects can be estimated for young groups and old groups separately. Thus, write (8) as

$$q_{ijt} = \alpha_i + \delta^{New}_j N_{it} + \delta^{Old}_j (1- N_{it}) + t + \varepsilon_{ijt} \quad (9)$$

where $\delta^{New}_j = \lambda^{New} Q_{ijt}$ and $\delta^{Old}_j = \lambda^{Old} Q_{ijt}$ and estimate (9).

Will the boss effects, $\delta_j^{Old}$ and $\delta_j^{New}$, be bigger for new or old workers (or star or laggard workers)? Theory provides guidance to examine differences in the new and old (as well

[19] Stars also appear more likely to leave than laggards: for the new hires sample, the mean maximum observed tenure for stars 664 days, compared to 820 days for laggards. But the distribution of maximum tenure is wider for laggards: while laggards are more likely to stay longer than stars, laggards are also likely to leave faster (perhaps due to firing).

as star and laggard) treatment effects. Recall that the effect of boss quality on worker output is

$$\frac{\partial q}{\partial S_i} = E\frac{\partial H}{\partial S_i} + H\frac{\partial E}{\partial S_i} \qquad i = t, m$$

from equation 2 . Good bosses, especially those with good teaching skills, may be most useful for those workers with low stocks of H because $(\partial H/\partial S_T)^{New} > (\partial H/\partial S_T)^{Old}$ and $(\partial H/\partial S_T)^{Laggard} > (\partial H/\partial S_T)^{Star}$ as predicted by most theories of human capital accumulation or learning by doing. A complication, though, is that each term is multiplied by the corresponding amount of effort or stock of human capital. Therefore, it is conceivable that better teaching bosses should be paired with star workers, not because stars have more to learn, but because each change in the quality of human capital is applied to a larger stock of human capital for stars. Similarly, if old workers worked harder than young workers (unlikely), it would be possible that the optimal pairing would have the best bosses with the old, not the young. Theory provides insights on why the effects may differ; data are needed to estimate the magnitudes of the offsetting effects.

**A. Does Non-random assignment of workers to bosses bias the estimated boss effects?**

There is not a random assignment rule; there would not be in any workplace. But it is often the case that actual assignment is nearly random, because worker turnover rates are high. High quality workers could be paired with high quality bosses because older workers and older bosses get their preferred work shifts. It is equally possible that there is no sorting bias, because many new workers and new bosses are higher quality than old workers and bosses. Recall that the highest quality workers are more likely to leave the firm than the lowest quality. However, a formal analysis of these assignment biases is required.

Stated more generally, to estimate this average boss quality treatment effect on worker

productivity, either there is random sorting between bosses and workers after accounting for the worker's fixed effect and X, or the boss treatment effects must be homogeneous across workers. As stated in section III.C above, this means $\delta_j$ = E(y | Boss j, worker i) – E(y | Boss 1, worker i) = E(y | Boss j, worker k) – E(y | Boss 1, worker k) for all workers i and k. We aim to let the treatment effects be heterogeneous across subgroups, so random assignment is needed to estimate unbiased treatment effects. To capture this, rewrite (9) as

(9') $$q_{ijt} = \alpha_i + \lambda^{New} N_{it} Q_{ijt}^{New} + \lambda^{Old} (1- N_{it}) Q_{ijt}^{Old} + t + \varepsilon_{ijt}$$

where the subgroups sort to different quality bosses, and quality differences enter the estimation of (9'). Alternatively, if every boss has at least some workers assigned to him in the new and old subgroups, then the estimated heterogeneous boss effects are unbiased—equation (9) prevails. Thus, even if boss 'j' has only 5 new workers and 25 old workers, his estimated boss effects by group will be unbiased, though the precision of the estimated effect for boss 'j' is lower for new workers given only 5 observations per boss. Therefore, the analysis of bias henceforth is an analysis of whether there are bosses excluded from one group or the other.

Using new and old workers as the comparison groups, the estimated distribution of boss effects is biased when the *excluded* group of bosses, namely those bosses who had either only new or only old workers, is different in their effects on worker productivity from the *included* group of bosses who had both new and old workers. If the assignment of workers to bosses is not random, then there is the potential that the effects of bosses on the two groups (new and old) is only relevant for those *included* bosses who have had both.

Is the proportion of bosses having both new and old workers relevant to this calculation of potential bias? Yes, but even were the proportion of omitted bosses small, it is still possible

that the excluded bosses could still be fundamentally different from the included bosses. It means, however, that we probably care less about the excluded group than the included group because it is a small part of the population. Put differently, the population distribution of boss effects, which is the weighted average of the included and excluded bosses' fixed effects, is likely to be close to the included group estimates alone when the included group is a large proportion of the total population.[20]

Let us examine the potential bias formally. The population parameter of interest is

$$\sigma^{New} = (\#Bosses\text{-}1)^{-1} * (\Sigma_{j \text{ with new}}(\delta_j - \delta)^2 + \Sigma_{j \text{ without new}}(\delta_j - \delta)^2)$$

where $\delta$ is the mean. Here $\Sigma_{j \text{ without new}}(\delta_j - \delta)^2$ is censored, and this omitted term is the source of bias (another source of bias may come from incorrectly estimating $\delta$, the population mean).

Notice that because of the summation, the weight on the term from the uncensored sample, $\Sigma_{j \text{ with new}}(\delta_j - \delta)^2$ , is the number of bosses with new workers over the total number of bosses. As the number of bosses with new workers relative to the total number of bosses increases, the estimated parameter approaches the population parameter.

There is a second issue. The analysis above focuses on whether the overall distribution of treatment effects is biased. However, much of the analysis of heterogeneity in estimated treatment effects uses the boss effects for only the included bosses, who have both new and old workers on their teams. For the set of bosses who work with both old and new workers, the regression $\delta^{New}_j = a + b\, \delta^{Old}_j$ is estimated, because it assesses whether a good boss is always a

[20] For this not to be the case, the parameter in the excluded group would have to be very different from that in the included group.

good boss for new and old workers.[21] But the regression is feasible only for the included group. If the excluded group is large, then there is sample selection bias introduced into the calculated correlation.

How big are the excluded groups in these data? Here is the breakdown.

New and Old Workers

| Excluded Bosses<br>New workers | Excluded Bosses<br>Old workers | Included Bosses<br>New and old workers | Total Bosses |
|---|---|---|---|
| 70 | 146 | 1720 | 1940 |
| | | | |

[21] The only way to compare the magnitude of boss j's effect on new workers with boss j's effect on old workers is if boss j works with new and old workers in the same connected group. Suppose a boss works with new and old workers in different connected groups; that is, boss k interacts with new workers in connected group 1 and boss k interacts with old workers in connected group 2. Then it is not possible to compare the magnitude of boss k's effect on new workers with boss k's effect on old workers because boss k's effect on new workers is $\delta_k$= E(y | boss k, new workers in group 1) – E(y | boss 1, new workers in group 1) whereas boss k's effect on old workers is $\delta_k$= E(y | boss k, old workers in group 2) – E(y | some boss ~=1, old workers in group 2). Notice that boss 1 is the excluded boss in connected group 1, but there is a different excluded boss in connected group 2.

Star and Laggard Newly Hired Workers

| Excluded Bosses<br>Laggard workers | Excluded Bosses<br>Star workers | Included Bosses<br>Laggard and star | Total Bosses<br>(for new hires) |
|---|---|---|---|
| 134 | 181 | 1711 | 1854 |
| | | | |

Thus, for the two alternative breakdowns, new/old and laggard/star, the proportion of the population that is excluded is very small: the included are 89 percent and 92 percent, respectively.[22]

This is suggestive that most worker assignment to bosses is close to random, and assignment bias is minor. No general statement can be made that there is no assignment bias for all possible partitioning of the data into subgroups. For example, if the assignment were to four different subgroups, rather than two, the possible bias could rise, and the proportion included falls. But this is unlikely; the groups above are the most likely breakdowns as described by the firm.

**B. Boss Effects for New and Old Workers**

The distribution of boss effects for old and new workers is given in the first column of Table 7. The standard deviation of boss effects for new workers (weighted by the boss's total

[22] For bosses who work with stars, star workers' mean (standard deviation) [Number of observations] daily output while matched with a boss who never workers with laggards is 11.5 (3.97) [2922], whereas output for star workers while matched with a boss who works with both stars and laggards 11.2 (3.1) [1713567]. Similarly, laggard workers' output while matched with a boss who never works with stars is 8.78 (3.0) [7073], whereas output for laggard workers while matched with a boss who works with both stars and laggards is 9.56 (2.67) [2005916].

frequency with new workers, but not adjusting for team size) is .39 compared to .44 for old workers (weighted by the boss's total frequency with old workers, but again not adjusting for team size). Bosses are slightly more important for old workers from this analysis. The difference between a good boss and a bad boss seems to have a larger effect on older workers than on younger ones, but only slightly so.

For the set of bosses who work with both old and new workers, the correlation in a boss's effect for old and new workers can be estimated. [23] The correlation in boss effects for old and new workers is positive and significant (0.29), suggesting that bosses who are good for new workers are also good for old workers. Note also that this correlation is very close to the simple A, B random group correlation of boss effects, which attempted to show what the effect would be were there mere errors in variables and were the true coefficient one. In this case, that would imply that b would equal 1 in

$$\delta^{New}_{j} = a + b\,\delta^{Old}_{j}$$

but the estimated coefficient would be .37 from the A, B comparison as long as the noise in the old/new comparison associated with estimate the boss effects was the same as that in the A,B comparisons above. This correlation is estimated within the largest connected group; 99.99

[23] The only way to compare the magnitude of boss j's effect on new workers with boss j's effect on old workers is if boss j works with new and old workers in the same connected group. Suppose a boss works with new and old workers in different connected groups; that is, boss k interacts with new workers in connected group 1 and boss k interacts with old workers in connected group 2. Then it is not possible to compare the magnitude of boss k's effect on new workers with boss k's effect on old workers because boss k's effect on new workers is $\delta_k$= E(y | boss k, new workers in group 1) – E(y | boss 1, new workers in group 1) whereas boss k's effect on old workers is $\delta_k$= E(y | boss k, old workers in group 2) – E(y | some boss ~=1, old workers in group 2). Notice that boss 1 is the excluded boss in connected group 1, but there is a different excluded boss in connected group 2.

percent of the sample falls within the largest group. The inference is that a one dimensional view of boss quality is a good description of what is going on for new and old workers.

Were the true coefficient, absent the errors-in-variable bias, one, there would be no clear advantage to pairing good bosses with new or old workers. Were the true coefficient greater than one, the best bosses should be paired with new workers. Were it less than one, best bosses should be paired with the old workers. The fact that the variance in boss effects for old and new workers is about the same is consistent with the true coefficient, corrected for errors-in-variable bias, being close to one.

### C. Boss Effects for Stars and Laggards

The second column of Table 7 provides the results for boss effects for stars and laggards. The weighted standard deviation of the boss effects for stars is 5.5 (0.61 if not multiplying by team size) and for laggards is 3.64 (0.4 without adjusting for team size). Whether the estimated standard deviation of boss effects reflects the population standard deviation of boss effects requires some care; some bosses only work with stars and some bosses only work with laggards. Recall from above that there are 1,806 bosses who ever work with laggards, 1,759 bosses who ever work with stars, and there are 1,711 bosses who work with both stars and laggards. The within boss correlation of boss effects for stars and laggards (for bosses with both stars and laggards in the largest connected group) is .18.

Again, the approach of regressing

$$\delta^{Star}_{j} = a + b\,\delta^{Laggard}_{j}$$

could be used to determine whether bosses are similar. The correlation of .18 from Table 7 suggests that those who are good for stars are also good for laggards, but that the relationship is

not perfect. The regression of estimated $\delta^{Star}$ on $\delta^{Laggard}$ produces an estimated coefficient on $\delta^{Laggard}$ of .29, with a constant of .31. The fact that .29 is less than one would, if it were bias-free, imply that the good bosses should be assigned to laggards rather than stars. However, the following fact alters this conclusion. First, the reverse regression of laggard fixed effects on star fixed effects yields an even lower coefficient. This is because the variation in boss effects is substantially greater when bosses are matched with star workers than when they are matched with laggard workers, the conclusion is that bosses affect the output of stars by more than they do laggards.

These results suggest that good bosses should be paired with the best workers. However, a "good boss" for one type of worker is not necessarily a good boss with another type of worker. Because the boss effects are not perfectly correlated for new and old workers or stars and laggards, there is room for assignment based on comparative advantage. The findings suggest that those bosses who are best at raising the productivity of new workers or laggards should be assigned to new workers or laggards, provided those bosses are not much better at raising the productivity of the complement group.

As an operational matter, a boss's quality, i.e., the boss's group specific fixed effect, can be determined at the level of a firm by using the approach above. It is then possible to decide to which boss a laggard should be assigned and to which a star should be assigned. This can be done by the firm to make more efficient assignments of workers to supervisors.

## VI. The Marginal Product of Bosses

Because the number of team members per boss varies, it is possible to identify the effect of additional boss time on worker productivity. This, in essence, estimates the marginal product

of boss time, or at least that component that works through enhanced worker productivity. The obvious problem is endogeneity. Better bosses may be assigned to supervise more workers, which would bias downward the observed effect of adding boss time on worker output. Additionally, the average tenure in the firm rose between 2006 and 2010, which would also affect productivity, but in the opposite direction were other things the same.

The firms' training policy provides an approach to dealing with potential endogeneity. When a worker leaves the firm, vacancies are filled by new workers. However, there is often a lag between a worker leaving and the availability of a replacement worker. Each new worker spends several weeks in training with other new workers, and vacancies are filled when a full training cohort is available upon the end of a training cycle.

Consequently, an instrumental variables approach is used. The key potentially endogenous variable is workers-per-boss (the team size). The first stage regression is

$$(10) \quad teamSize_{ijt}=\alpha_i+\delta_j+trainingCycle\delta_1+trainingCycle^2\delta_2+g(tenure_{it})\delta_3+t+\varepsilon_{ijt}$$

where team size is a function of the number of days since the last training class ended at each establishment. Monthly time dummies are used to control for a decrease in turnover and hiring during the recession. Column 1 of Table 8 provides details about the first stage. The residual sum of squares from regressing team size on a tenure polynomial, month dummies, worker fixed effects and boss fixed effects is 35,020,976 compared to a residual sum of squares of 34,885,127 when including a quadratic function of the number of elapsed days since the last training cohort entered the establishment. This yields a first stage partial F on the instruments of 11,105. The overall F statistic on the first stage is 109.4.

The estimates in the first stage imply that team size is .099 workers smaller after 19 days into the training cycle (the mean training cycle lag observed in the data) versus at the beginning. A one standard deviation increase in the training cycle, to a 52 day lag, decreases team size by an additional .184 workers. A one standard deviation change in the training cycle thus provides a roughly 2% change in team size.

The regression of interest is

(11) $q_{ijt}=\alpha_i+\delta_j+\widehat{teamSize}_{ijt}\beta_1+f(tenure_{it})\beta_2+t+\varepsilon_{ijt}$

where output-per-hour is regressed on team size, with controls for tenure, monthly time dummies, worker fixed effects, and boss fixed effects. The results are reported in Table 8 in the remaining columns.

From these numbers, it appears that reducing team size significantly increases each worker's output-per-hour. This is true in both OLS and IV versions. The coefficient is interpreted as the effect of adding one worker per boss on individual output. This implies that adding another worker to a team increases total output less than the impact of the additional worker's productivity. The coefficient on team size in the IV regression is greater than the OLS coefficient, consistent with the view that better bosses are assigned a larger number of workers, which would bias down the effect in OLS.[24]

[24] The IV estimates of the effect of team size are obtained using the projection from the first stage as a regressor in the second stage. Standard errors cannot be computed because of difficulty in inverting the matrix of regressors, but a comparison of the residual sum of squares with the model excluding team size and the corresponding F-test confirm that team size is statistically significantly related to output.

Using the IV estimates, total output is given by N*(oph-0.143*N/B) where N is the number of workers in the company and B is the number of bosses. Therefore, the marginal product of a boss is $0.143N^2 / B^2$. On a typical day, the marginal product of an additional boss is approximately 12.1 units of output. The marginal product of a worker can be calculated as oph – 0.286N / B. The marginal product of an additional worker is about 7.5 units of output. In terms of a boss's effect on output as it operates by increasing worker productivity, a boss is 60% more important than a worker. There may be things that a boss does as well that are not captured by the effect through workers alone, but these numbers are consistent with levels of compensation received by bosses and workers as well, where a boss could earn approximately 50% to 100% as much as a typical worker. [25]

## VII. Conclusion

Supervision and management are a fundamental concept in personnel economics and in the theory of the firm. Although we take as given that mangers matter, neither the mechanisms through which they affect productivity nor the actual size of the effects has been spelled out previously. By using a unique data set that gives very detailed daily output on workers and records the supervisors to which they are assigned on that day, it is possible to examine the effects of bosses on worker productivity.

Boss effects are large and significant. The value of a boss is about twice that of a worker. Further, bosses vary substantially. A very good boss increases the output of the supervised team over that supervised by a very bad boss by about as much as adding one member to the team.

[25] The company did not supply compensation data, but in conversations with managers about levels of compensation, a ratio of 2:1 boss-to-worker compensation is not out of line.

Additionally, peer effects are trivial. The only "peer" who matters in this work environment is the boss. Finally, good bosses increase the output of the better workers by more than they do for the poorer workers. Consequently, the assignment of supervisors to workers matters; productivity can be increased by sorting bosses appropriately to workers.

**References**


1. Abowd, John, Robert Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using LInked Longitudinal Employer-Employee Data." Working Paper.
2. Abowd, John, Francis Kramarz, and Simon Woodcock. 2006. "Econometric Analysis of Linked Employer-Employee Data." Working Paper.
3. Falk, Armin, and Andrea Ichino. 2006. "Clean Evidence on Peer Effects." *Journal of Labor Economics*, 24(1), 39-57.
4. Ichniowski, Casey, and Kathryn Shaw. 2003. "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices." *The Journal of Economic Perspectives,* 17(1), 155-180.
5. Kandel, Eugene, and Edward Lazear. 1992. "Peer Pressure and Partnerships." *Journal of Political Economy*, 100(4), 801-817.
6. Lazear, Edward. 2000. "Performance Pay and Productivity." *American Economic Review,* 90(5), 1346-1361.
7. Manski, Charles. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, (60), 531-542.
8. Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review,* 99(1), 112-145.
9. Rockoff, Jonah. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94(2), 247-252.
10. Rosen, Sherwin. 1982. "Authority, Control, and the Distribution of Earnings." *Bell Journal of Economics*, 13(2), 311-323.
11. Shaw, Kathryn, and Edward Lazear. 2008. "Tenure and Output." *Labour Economics,* 15, 710-724.
12. Simon, Herbert. 1957. "The Compensation of Executives." *Sociometry*, 20(1), 32-35.

Figure 1: Histogram of Estimated Boss Effects

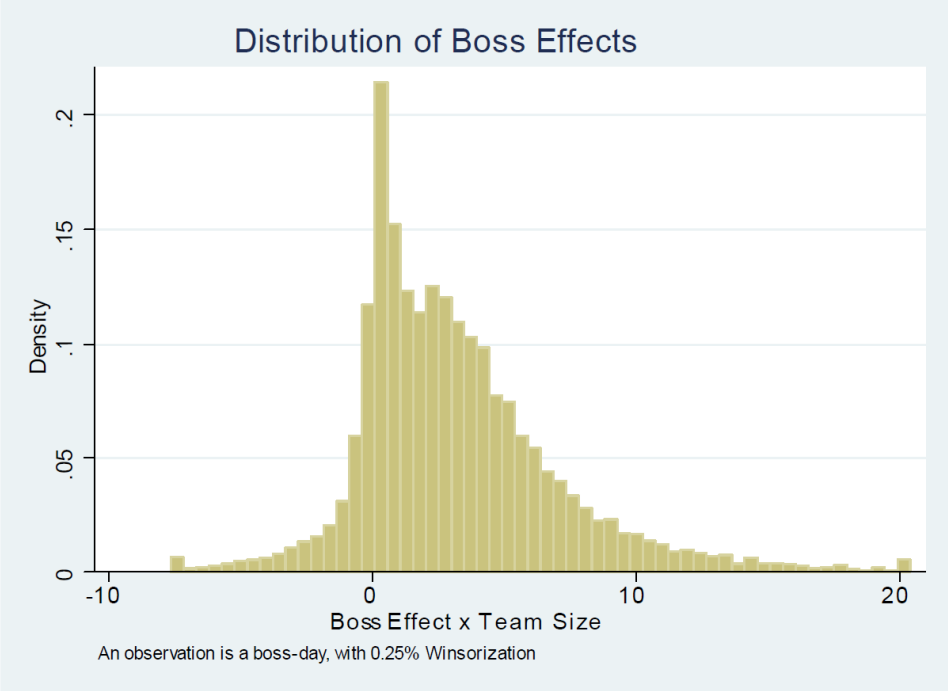


Figure 2: Estimated boss effects as a function of the number of worker-days per boss

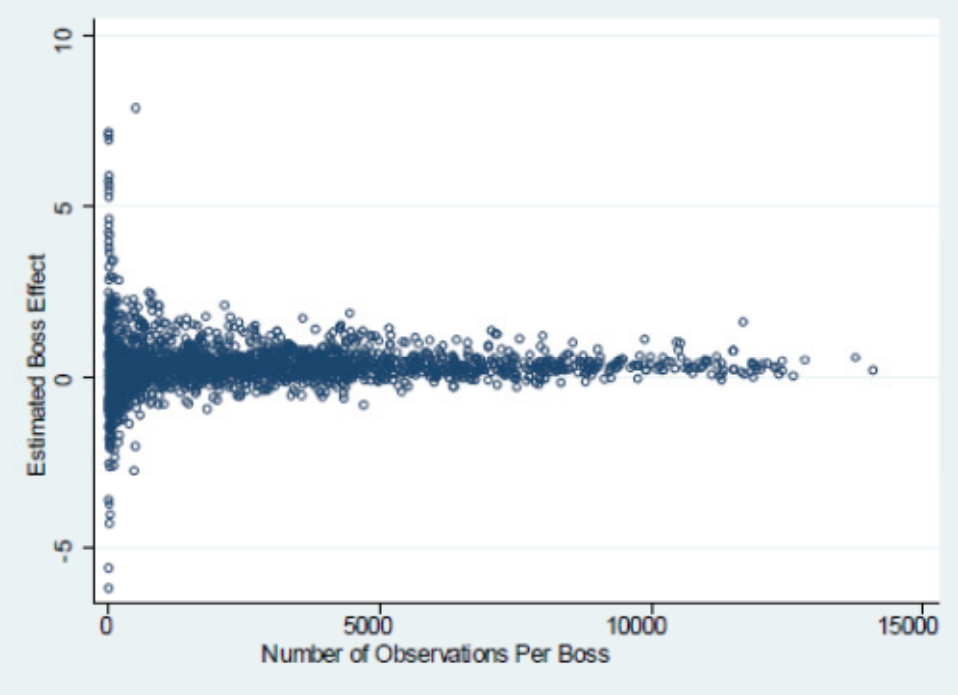

**Table 1: Summary Statistics**

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Output Per Hour | 5,729,508 | 10.26 | 3.16 | 0.1 | 40.0 |
| Uptime | 4,870,610 | 0.96 | 0.03 | 0.5 | 1.0 |
| Output Per Hour * Uptime | 4,870,610 | 10.01 | 3.00 | 0.4 | 40.0 |
| Tenure | 5,729,508 | 648.91 | 609.83 | 1.0 | 4235.0 |
| | | | | | |
| Number of Workers | 23,878 | | | | |
| Number of Unique Bosses Per Worker | 23,878 | 3.99 | 2.78 | 1.0 | 19.0 |
| Daily Team Size | 633,818 | 9.04 | 4.54 | 1.0 | 29.0 |
| | | | | | |
| Number of Bosses | 1,940 | | | | |
| Number of Unique Workers Per Boss | 1,940 | 49.15 | 35.41 | 1.0 | 250.0 |
| Mean Number of Other Bosses for Each Worker | 1,940 | 4.69 | 1.51 | 0.0 | 11.3 |

Notes:

The data contain daily worker productivity records from June 2006 to May 2010. Output per hour is calculated from records that contain the average daily transaction handling time for each worker. Uptime is calculated from an IT system that monitors the fraction of clock time that a worker is available to handle transactions. There is some missing data on uptime. The missing uptime data is concentrated toward the beginning fo the sample period. The mean of output per hour when restricting the sample to the 4,870,610 agent-days with non-missing uptime is 10.38 with standard deviation 3.08. Eliminating daily teams with 1 person removes 80,067 person-days and changes the mean daily team size to 10.20.

**ressions of Output-per-hour on combinations of fixed effects**

| | OLS | Worker Fixed Effects | Boss Fixed Effects | Worker and Boss Fixed Effects | Wo Ef |
|---|---|---|---|---|---|
| | 0.061 | 0.237 | 0.092 | 0.243 | |
| **tion of Worker Fixed Effects** | | | | | |
| observation per worker) | | 1.87 | | 1.85 | |
| rker-days (frequency) | | 1.34 | | 1.32 | |
| hood Estimates (1 observation per worker) | | 1.24 | | 1.22 | |
| | | 55.6*** | | 47.7*** | |
| **tion of Boss Effects** | | | | | |
| observation per boss) | | | 0.98 | 0.83 | |
| rker*boss days (frequency) | | | 0.58 | 0.38 | |
| hood Estimates (1 observation per boss) | | | 0.62 | 0.39 | |
| | | | 102.1*** | 20.5*** | |
| **tion of Boss Effects Multiplied by Average Team Size (9.04)** | | | | | |
| observation per boss) | | | 8.86 | 7.5 | |
| rker*boss days (frequency) | | | 5.24 | 3.44 | |
| hood Estimates (1 observation per boss) | | | 5.6 | 3.53 | |
| | | | | | |
| ner team members' mean output per hour | | | | | |
| | | | | | |
| Effects | | | | 53.3*** | |
| vations | 5,729,508 | 5,729,508 | 5,729,508 | 5,729,508 | |
| ers | 23,878 | 23,878 | 23,878 | 23,878 | |
| es | 1,940 | 1,940 | 1,940 | 1,940 | |
| e in largest connected group | | | | 99.99 | |

contain a fifth order polynomial function of tenure (with a 365 day cutoff and cutoff indicator), monthly time dummies, an
n each connected group of workers and bosses, worker fixed effects are mean zero and one boss fixed effect is restricte
by the number of observations correspond to the sample frequency, whereas unweighted or "1 per-person" measures pe
(either bosses or workers) as the unit of observation. Because boss effects are not estimated precisely if the boss has v
e maximum likelihood estimates assume the true boss and worker fixed effects are normally distributed, and the estimate
normally distributed with variance equal to the sum of sampling error and the true variance. Maximum likelihood of estim
orted. The sampling variance is computed as the residual variance for each worker or boss, which slightly understates th
e there is no accounting for off-diagonal elements in the matrix of regressors.

## Table 3: The Effect of Peer Quality on Output-per-hour

| **Estimation method:** | OLS | Joint NLS | Peer Proxies |
|---|---|---|---|
| R-Squared | 0.2504 | 0.2356 | 0.243 |
| Coefficient on Peers' Mean Ability | 0.16 | 0.001 | -0.022 |
| Standard Deviation of Peer Effects | 0.063 | 0.022 | 0.009 |
| Standard Deviation of Boss Effects (Weighted by frequency) | 0.33 | 0.31 | 0.38 |
| Standard Deviation of Worker Effects (Weighted by frequency) | 1.31 | | 1.32 |
| Number of Workers | 23,878 | 1679 | 23,878 |
| Number of Bosses | 1,940 | 155 | 1,940 |
| Number of Observations | 5,729,508 | 391,730 | 5,729,508 |

Notes:

Specifications in columns 1 and 3 contain a 5th order polynomial in tenure (with a 365 day cutoff and cutoff indicator), month, day of week, boss, and worker fixed effects. Column 2 does not contain day of week dummies. Standard errors cannot be computed. The first column contains the mean contemporaneous output for all other team members on a given day. The joint estimation column use non-linear least squares, taking the mean of the team members' individual fixed effects as a measure of peer quality. The joint estimation procedure is computationally demanding; an "outer" loop is used to search over the peer effect coefficient, while an inner loop conditions on the outer loop value and solves for the parameters using a conjugant gradient procedure. The joint procedure is not possible on the full data because of memory issues in Matlab; storage of the matrix of peer fixed effects requires an order of magnitude more memory than using a single-dimensional index of peer quality. The peer proxies uses mean output on the first three months on the job as the value of peer quality. If a worker's first three months are not observed, then the mean value of all observed workers' first three months is used.

To calculate the standard deviation of peer effects, it is assumed that one peers' output increases by a standard deviation change in output per hour, or 3.16 units. This is then multiplied by the Coefficient on Peer's Mean Abiliity and divided by (9.04-1), the mean number of other team members.

## Table 4: Teaching and the Fadeout of Boss Effects

| | |
|---|---|
| Teaching (λ) | 0.78 |
| Monthly Rate of Decay (γ) | 0.87 |
| Amount of Boss Effect Remaining After 1 Year (γ^12)*λ | 0.13 |
| **Standard Deviation of Worker Fixed Effects** | |
| Weighted by frequency | 1.28 |
| **Standard Deviation of Boss Effects Multiplied by Average Team Size (9.04)** | |
| Weighted by frequency | 3.51 |
| Number of Workers | 1679 |
| Number of Bosses | 155 |
| Number of Observations | 391,730 |

Notes: The results of estimating equation (7). The specification contains a 5th order polynomial in tenure (with a 365 day cutoff and cutoff indicator), month, boss, and worker fixed effects. Estimation is conducted via nonlinear least squares, where search over the reported parameters involves an "outer" loop, while an inner loop conditions on the outer loop values. Estimation is not possible on the full sample, so a subset of establishments are included in the analysis.

**Table 5: Comparison of Output-per-hour and Uptime**

| | Output-per-hour | Uptime |
|---|---|---|
| Mean of Dependent Variable | 10.26 | 0.96 |
| R-squared | 0.243 | 0.0994 |
| **Standard Deviation of Worker Fixed Effects on Total Output (oph*uptime)** | | |
| Unweighted (1 observation per worker) | 1.78 | 0.21 |
| Weighted by worker-days (frequency) | 1.27 | 0.1 |
| Maximum Likelihood Estimates (1 observation per worker) | 1.18 | 0.12 |
| F statistic | 47.5*** | 18.6*** |
| **Standard Deviation of Boss Effects on Total Output (oph*uptime) Multiplied by Average Team Size (9.04)** | | |
| Unweighted (1 observation per boss) | 7.23 | 2.58 |
| Weighted by worker*boss days (frequency) | 3.31 | 0.58 |
| Maximum Likelihood Estimates (1 observation per boss) | 3.4 | 1.61 |
| F statistic | 20.3*** | 16.9*** |
| F on Joint Fixed Effects | 53.5*** | 20.7**** |
| Number of Observations | 5,729,508 | 4,870,610 |
| Number of Bosses | 1,940 | 1,726 |
| Percent of sample in largest connected group | 99.99 | 99.99 |
| Correlation of Worker Oph and Uptime Fixed Effects | 0.15 | |
| Correlation of Boss Oph and Uptime Fixed Effects | 0.16 | |

Notes:

All specifications contain a fifth order polynomial function of tenure, monthly time dummies, and day of week dummies. Worker fixed effects are mean zero, and one boss fixed effect is restricted to be zero within each connected group. Some data on uptime is missing toward the beginning of the sample period, giving differences in the number of observations. The calculation of the standard deviation of worker fixed effects on total output with respect to oph multiplies the standard deviation of worker fixed effects with respect to oph by the mean uptime from Table 1. The standard deviation of worker fixed effects with respect to uptime multiplies the standard deviation of worker fixed effects on uptime by the mean oph from Table 1. The calculation of the standard deviation of boss fixed effects is similar.

**Table 6: Summary Statistics for Heterogeneous Worker Groups**

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| | **Old and New Workers Sample Split** | | | | |
| **New Workers** | | | | | |
| Output Per Hour | 2,435,999 | 9.97 | 2.91 | 0.1 | 40.0 |
| Tenure | 2,435,999 | 166.19 | 104.91 | 1.0 | 365.0 |
| Number of Workers | 19,676 | | | | |
| **Old Workers** | | | | | |
| Output Per Hour | 3,293,509 | 10.48 | 3.32 | 0.1 | 40.0 |
| Tenure | 3,293,509 | 1005.94 | 582.24 | 366.0 | 4235.0 |
| Number of Workers | 14,167 | | | | |
| | **Stars and Laggards Sample Split** | | | | |
| **Stars** | | | | | |
| Output Per Hour | 1,716,489 | 11.20 | 3.11 | 0.2 | 40.0 |
| Tenure | 1,716,489 | 324.62 | 278.61 | 1.0 | 1542.0 |
| Maximum Observed Tenure | 1,716,489 | 663.57 | 357.18 | 1.0 | 1542.0 |
| Maximum Observed Tenure (1 observation per worker) | 8,374 | 406.88 | 332.59 | 1.0 | 1542.0 |
| Number of Workers | 8,374 | | | | |
| **Laggards** | | | | | |
| Output Per Hour | 2,012,989 | 9.56 | 2.67 | 0.1 | 40.0 |
| Tenure | 2,012,989 | 397.01 | 327.78 | 1.0 | 1542.0 |
| Maximum Observed Tenure | 2,012,989 | 820.45 | 404.40 | 1.0 | 1542.0 |
| Maximum Observed Tenure (1 observation per worker) | 8,579 | 449.70 | 412.57 | 1.0 | 1542.0 |
| Number of Workers | 8,579 | | | | |

**Table 7: Heterogeneous Boss Effects**

| | Output-per-hour New and Old Workers | Output-per-hour Stars and Laggards |
|---|---|---|
| R-squared | 0.2448 | 0.2445 |
| **Standard Deviation of Worker Fixed Effects** | | |
| Maximum Likelihood Estimates for Old Workers / Stars | 1.18 | 1.06 |
| Maximum Likelihood Estimates for New Workers / Laggards | 1.2 | 0.85 |
| **Standard Deviation of Boss Effects * Team Size** | | |
| For Old Workers / Stars (Weighted by frequency with old workers) | 4 | 5.5 |
| For Old Workers / Stars (ML) | 3.9 | 5.85 |
| For New Workers / Laggards (Weighted by frequency with new workers) | 3.54 | 3.64 |
| For New Workers / Laggards (ML) | 3.22 | 3.95 |
| P Value of Heterogeneous versus Homogeneous Boss Effects | 0 | 0 |
| Number of Observations | 5,729,508 | 5,729,508 |
| Percentage of Observations in the Largest Connected Group | 99.99 | 97.51 |
| Number of Observations in the Second Largest Connected Group | 0.01 | 2.46 [See Note A] |
| Number of Bosses with Old Workers (Star Workers) | 1,794 | 1,759 |
| Number of Bosses with New Workers (Laggard Workers) | 1,870 | 1,806 |
| Number of Bosses with Both Worker Types in the Largest Connected Group | 1,709 | 1,532 |
| Correlation of New (Laggard) and Old (Star) Boss Effects [See Note B] | 0.29 | 0.18 |

Notes:

All specifications contain a fifth order polynomial function of tenure (with a 365 day cutoff), monthly time dummies, and day of week dummies. Worker fixed effects are mean zero, and one boss fixed effect is restricted to be zero for each sub-group. In the star/laggard specification all workers, including those who cannot be classified as stars or laggards because their beginning productivity is not observed, are used to estimate the tenure profile, month and day of week dummies. Amongst stars and laggards only, the number of observations is 3,729,478. Standard deviations of boss fixed effects are weighted by the number of observations or number of bosses within each connected group, taking deviations from the mean of the connected group. Correlations of boss effects are restricted to bosses whose new/old and star/laggard effects are estimated within the same connected group. For the new/old specification, 1720 bosses work with both new and old workers, with 1709 bosses having new and old workers within the same connected group. For the star/laggard specification, 1854 bosses work with either stars or laggards; 1711 bosses work with both stars and laggards, with 1532 bosses having both stars and laggards within the same connected group.

[A]: The second largest connected group contains all laggard workers and boss*laggard vectors for bosses working with laggards who never work with stars or other laggards who are connected to other bosses. There are 91,658 observations and 170 unique boss*laggard fixed effects in this group; the standard deviation of boss fixed effects for laggards in this group (weighted by frequency) is .64. However, only 7 unique bosses in the group never work with stars. The third largest connected group has only 501 observations and contains only laggards and boss*laggard vectors. The largest group containing only stars and boss*star vectors has 333 observations. There are 26 overall connected groups.
[B]: The correlation is computed for bosses who have old/new or stars/laggards within the same connected group.

## Table 8: Analysis of Team Size and the Marginal Product of Workers and Bosses

| | First Stage OLS | OLS | IV |
|---|---|---|---|
| | Team Size | Output-per-hour | |
| **First Stage Results** | | | |
| Mean elapsed days since last training cohort entered (Training cycle): | 19 | | |
| Standard deviation: | 33 | | |
| Implied change in team size at 19 days | -0.099 | | |
| Implied change in team size at 19+33=52 days | -0.283 | | |
| First stage partial F(2, 5703636) on the instruments | 11,105 | | |
| First stage F(25878, 5703629) overall on workers per team | 109.4 | | |
| **Main Results** | | | |
| Workers per team | | -0.043 | -0.143 |
| Marginal Product of a Boss | | 3.6 | 12.1 |
| Marginal Product of a Worker | | 9.3 | 7.5 |
| Standard Deviation of Worker Effects (Weighted) | | 1.32 | 1.32 |
| Standard Deviation of Boss Effects* Team Size (Weighted) | | 3.44 | 3.34 |
| R-squared | 0.324 | 0.243 | 0.242 |
| Number of observations | 5,729,508 | 5,729,508 | 5,729,508 |

Notes:

All specifications contain a 5th order polynomial in tenure (capped at 365), month dummies, day of week dummies, boss fixed effects and worker fixed effects. Standard errors cannot be computed, but F tests of each model against the restricted version show that the results are highly significant at conventional levels. Total firm output is $N*(oph-\beta*N/B)$ where N is the number of workers, B is the number of bosses, and $\beta$ is the estimated coefficient on Workers Per Team. Differentiating this expression with respect to B gives the marginal product of a boss. Differentiating with respect to N gives the marginal product of a worker.

**Appendix Table: Regressions of Output-per-hour with lagged dependent variables**

| Lag Type | 14 Day Avg. | 14 Day Avg. | 1 Day | 1 Day |
|---|---|---|---|---|
| Number of Lags | 1 | 1 | 1 | 2 |
| Worker Tenure | No | Yes | Yes | Yes |
| R-squared | 0.2599 | 0.2617 | 0.2516 | 0.2558 |
| Coefficient on the first lag | 0.402 | 0.375 | 0.105 | 0.0968 |
| Coefficient on the second lag | | | | 0.0749 |
| **Standard Deviation of Worker Fixed Effects** | | | | |
| Weighted by frequency | 0.79 | 0.83 | 1.19 | 1.09 |
| **Standard Deviation of Boss Effects Multiplied by Average Team Size (9.04)** | | | | |
| Weighted by frequency | 2.26 | 2.12 | 3.07 | 2.85 |
| **NPV of a Standard Deviation of Boss Effects for an Average Team** | | | | |
| | 3.78 | 3.39 | 3.43 | 3.16 |