Do Agents Game Their Agents’ Behavior? Evidence from Sales Managers

By Alan Benson*

Job Market Paper

Draft: January 22, 2013

This paper explores intermediary agency problems, specifically the use and misuse of authority and incentives in organizational hierarchies. I offer a principal-agent-supervisor model inspired by sales settings. Through it, I propose organizations delegate authority over front-line salespeople to front-line sales managers (even though workers’ performance is known to the firm) because managers’ private information allows them to distinguish ability from luck. The model yields the result that managers on the cusp of a quota have a unique personal incentive to retain and provide downward quota adjustments to poor performing subordinates, permitting me to distinguish managers’ interests from those of the firm. I parametrically estimate the model using detailed person-transaction-level microdata from 244 firms that subscribe to a “cloud”-based service for automating transaction processing and compensation. I estimate 13-15% of quota adjustments and retentions among poor performers are explained by the managers’ unique personal interest in meeting a quota. I discuss how firms attempt to mitigate the costs of managerial gaming, and agency theory’s implications for when various mitigation practices are most effective.

The constraints on principals’ ability to induce efficient behavior through their economic agents are the defining determinants of economic organization. In the classic principal-agent model, a principal (e.g. the firm, which provides employment and a compensation plan) contracts directly

* PhD Candidate, MIT Sloan School of Management. This is a draft of the third chapter of my dissertation and job market paper (exp. graduation, May 2013). I thank David Autor, Martin Conyon, Bob Gibbons, Ian Larkin, Danielle Li, Paul Oyer, Mike Powell, and seminar participants at MIT Sloan, MIT Economics, Rutgers SMLR, Maryland Smith, Chicago Booth, Illinois SLER, and the 2012 People and Organizations Conference at the Wharton School of the University of Pennsylvania for comments and advice. I also thank the individuals at the disclosing company and practitioner interviewees involved in this project. The usual disclaimer applies. Correspondence: alanmb@mit.edu.
with its agents (e.g. the worker, who provides productive effort). In practice, profit-maximizing principals are far-removed from rank-and-file agents. For example, shareholders of publicly traded firms rely on a long chain of intermediary executives and managers to set and monitor workers’ employment practices on their behalf. As such, models invoking a profit-maximizing firm implicitly assume the interests of their intermediary agents, even if they are not identical to those of shareholders, are sufficiently aligned that their ultimate employment practices also approximate profit-maximizing behavior.

Although it is well-known that agents’ incentive plans may encourage activities that are inconsistent with the interests of principals, evidence of such “gaming” draws almost exclusively from the top and bottom of organizations (e.g. CEOs and rank-and-file workers; see Murphy 1999 and Lazear and Oyer 2009 for reviews). However, little empirical work examines how misaligned managerial incentives propagate to subordinates. While this issue is difficult to examine, it is also important, since managers are the intermediary agents responsible for making decisions on behalf of the “Firm.” Indeed, early organizational researchers dismissed profit maximization as the chief motive governing managerial decision making. Based on their observations, they concluded managers are imperfect and self-interested coordinators of economic activity, that firms should not be treated as monolithic, and that the inability of organizational hierarchies to coordinate activities efficiently determines firm structure, governance, and scope (classic studies include Baumol 1959; Chandler 1977; Coase 1937; Crozier 1964; Cyert and March 1963; Penrose 1959; Simon 1957, 1964; and Williamson 1963, 1967).

In this paper, I theorize why firms delegate authority to intermediary managers, identify the misuse of authority and incentives, and describe how sales organizations attempt to control the costs of managerial gaming affecting salespeople. I propose firms delegate authority over staffing and incentive decisions to immediate managers (even though sales are observed by the firm) because sales managers accumulate private information allowing them to distinguish salespeoples’ ability from exogenous circumstances affecting performance, such as market conditions, product quality, or territory quality. Managers’ private information allows them to screen and incentivize salespeople more efficiently than would a firm that conditions these decisions on sales figures alone. Although managerial incentives generally align managers’ decisions with profit maximization, authority and

---

1 This emphasis might be attributed to the three common strategies for acquiring data—using publicly-available accounting data, company-researcher data use agreements for single-firm studies, or sports statistics for athletes.

2 For brevity, I use “sales” to refer to performance measures. In the data, performance measures also include presale, support, and renewal activities.
quotas provide an opportunity and a motive for managers to shift sales to their desired measurement periods through staffing and incentive decisions that affect subordinates.

The model yields the hypothesis that quotas distort managerial incentives to make decisions that are consistent with the interest of the firm. Intuitively, the model captures the institutional features that allows decisions motivated by the manager’s personal interests to be identified: (i) that sales managers’ have a unique interest in the marginal sales that meet their quotas, (ii) their quota attainments are determined by the cumulative credited business of their subordinates, and (iii) that managers can affect the timing of sales through staffing and incentive decisions affecting subordinates. The model yields the hypotheses that managers will be more likely to forgo terminating poor performing subordinates (Hypothesis 1) and will provide downward quota adjustments (Hypothesis 2) when the managers are on the cusp of meeting a quota, compared to when they are not. The model captures the institutional account— that managers’ decisions effectively “pull in” sales into the current fiscal year at the expense of future sales and contrary to the interests of the firm.

I test this hypothesis using a novel and uniquely well-suited data featuring salespeople working at firms that subscribe to an on-demand (over “the cloud”) sales performance management service. The data include longitudinal detail on the hierarchical positions, incentive plans, performance, and pay of 7,492 sales managers and their 61,092 immediate subordinates in 244 firms. I parametrically estimate the formal model, distinguishing the turnover and quota adjustments of salespeople whose sales are critical for the manager to meet a quota (the quasi-experimental “treatment”) with salespeople working under managers who would or would not meet a quota anyway (the controls). As such, the identification strategy uses sales of a subordinates’ peers as an exogenous source of variation affecting whether a subordinate’s sales will be crucial for the manager to meet a quota, and the sales of a subordinate as a “treatment bubble.” This allows distortions in subordinates’ staffing and incentives to be causally attributed to managers’ personal interests, thereby addressing a key challenge for empirical agency research. I estimate 13-15% of quota adjustments and retentions among poor performers are explained by the managers’ unique personal interest in meeting a quota.

[FIGURE 1]

To illustrate a puzzling implication of gaming by intermediary agents, Figure 1 shows that the cumulative sales of the manager’s subordinates often just reach the manager’s quota. Indeed, for both rank-and-file salespeople and also their managers, there are nearly four times as many quotas surpassed within 5% as there are quotas missed within 5%. This figure excludes the 7% of instances...
that managers’ quotas are the sum of subordinates’ quotas, and so this pattern cannot be explained by cumulatively-aligned quotas alone.

The model offers an explanation for the delegation of authority in organizational hierarchies, with results providing evidence that intermediary agents extract rents on their private information on personnel quality and market conditions. This enables managers to pull in sales to reach 100% quota through decisions affecting subordinates. I discuss how sales organizations use a variety of rules and practices to prevent such gaming from becoming too costly.

I. Sales Management, Weak Monitoring, and Gaming in Sales

A. Background

Like rank-and-file salespeople, sales managers receive variable pay that can vary widely depending on measured performance. Unlike rank-and-file salespeople, sales managers’ performance is measured largely by the cumulative sales of their subordinates. In the data, mean annual variable pay is about one-half of base pay for both managers and non-managers.

Variable pay includes commissions and bonuses. Their rates depend on quota attainment. Quotas are specific thresholds at which workers typically receive a discrete bonus (in an archetypal quota-bonus plan) and/or begin to earn commissions on marginal sales (in an archetypal commission plan). Quotas are generally set in advance of a measurement period. However, organizations typically allow managers to adjust subordinates’ quotas for reasons outside of the salespersons’ control.

Salespeople who exceed their quotas typically become eligible progressive bonuses or accelerators that increase the rate at which commissions are paid. Salespeople who consistently exceed their quotas may also be considered for promotions, transfers, superior leads, or superior accounts. Making quotas and other discrete benchmarks also confer prestige, influence, and symbolic rewards (Larkin 2009). Salespeople who do not meet quotas typically earn a base pay, which reduces risk borne by salespeople and provides income to new recruits. Guaranteed income is often temporary and may be phased-out or drawn from future variable pay (in the case of “draws”).

B. Why Organizations Use Quotas

Based primarily on interviews, I find three main reasons sales compensation plan designers use quotas. First, quotas focus variable pay around marginal effort. Some sales positions, such as account managers with an account renewal quota, can achieve some large share of their quota with relatively little effort. In the classical agency model, this is similar to the property that a firm can
capture maximum rents by paying an agent’s participation constraint for the first-best effort. The firm may have a better understanding of this level of effort than it does the marginal cost of this effort, which is needed under a linear incentive plan (i.e. a piecerate or pure commission plan).

Second, quotas communicate minimum acceptable performance. Consistently missing a quota is generally understood to be a ground for dismissal. Terminating a salesperson for performance reasons when that salesperson consistently makes quota may be interpreted as symptomatic of poor communication by management. In most sales settings, sales people consume territories, sales leads, and support resources that are inherently valuable. As such, firms are generally not willing to retain a salesperson indefinitely at low pay for low performance. Interviewees explicitly referred to retaining a poor performing sales representative in terms of the opportunity cost of replacing them with a productive new recruit. In a classical agency framework, this is analogous to a firm communicating that it is incentive compatible for a firm to replace an agent for performance beneath a threshold.

Third, managers and plan designers widely believe that quotas have a behavioral effect on workers that tends to boost performance. Managers explain the importance of communicating how much sales people should expect to make if they make their quota. I interpret these reports to signify that managers use quotas to invoke loss aversion, thereby eliciting greater effort for less pay when below a quota threshold.

C. How Incentives Affect the Timing of Sales Activities

Outside of executives and sales, variable pay is often a small or negligible component of compensation. One reason is that incentive contracts encourage workers to “game” plans by engaging in activities that are correlated with performance measures but contrary to the interests of the firm. Examples abound of how misaligned incentives prompt undesired behaviors. At the top of organizations, executives adapt accounting procedures, accrual procedures, and voluntary disclosures to maximize bonus rewards (Aboody and Kasznik 2000, Healy 1985, Yermack 1997). At the bottom of organizations, several studies show how seemingly innocuous pay-for-performance

---

3Heath, Larrick, and Wu (1999) discuss how performance goals are analogous to reference points in Prospect Theory. Psychologists typically find explicit targets can improve measured performance when they are challenging, specific, attainable, and supported by coaching and other practices (for reviews, see Shinkle 2012, Steel and König 2006). Although this literature is not conclusive, sales plan designers widely report their belief that quotas have psychological importance.

4Interviewees often view “gaming” as pejorative, noting that it is the responsibility of the firm to design incentives and for salespeople to pursue them. I use this term as shorthand.
schemes have backfired.\footnote{Job Training Partnership Act training agencies manipulate the timing of students’ graduation dates to boost the share of graduates with jobs (Courty and Marschke 1997, 2003). In lending, the desire to avoid appearing to have poorly assessed borrowers’ risks led bank loan officers to fail to disclose bad news to their supervisors (Hertzberg, Liberti, and Paravisini 2007). After implementing a per-passenger commission, bus drivers in Chile had a higher incidence of traffic accidents than prior to the implementation or compared with a competing bus company that did not pay by commission (Johnson, Reiley, and Muñoz 2011). Baker (1992), Ethiraj and Levinthal (2009), Feltham and Xie (1994), Holmstrom and Milgrom (1991), and Kerr (1974) provide other examples.}

Despite their ubiquity in sales, empirical work shows misaligned and nonlinear incentives distort sales activities as well. Using data from an enterprise software vendor, Larkin (2007) finds that accelerating quarterly commissions lead salespeople to use discretionary discounts to concentrate transactions into fewer quarters, costing the employer an estimated 6-8% of revenues. Using Compustat data, Oyer (1998) exploits variation in fiscal years by company and within industry to show that manufacturing firms’ sales rise in the fourth quarter. He interprets this finding as consistent with the incentive effects of annual quotas, although the firm-level public disclosure data do not permit a direct test or analysis of gaming at any level of the organization. Incentives to boost the size of subscriptions led account managers at Dun & Bradstreet to overstate their clients’ historical usage, spurring millions of dollars in lawsuits (Roberts 1989).

Salespeople describe several additional practices used to shift credit across measurement periods. Salespeople may boost sales figures by enticing distribution channels to place large orders to keep as inventory, a practice referred to as “channel stuffing.” Salespeople may delay closing deals until future measurement periods, a practice referred to as “sandbagging.” Salespeople and managers may exchange credit for sales across measurement periods. Salespeople and managers may misrepresent the quality of their territory to affect the sales forecasts used as the bases of their quotas. Sales managers can provide incentives, called “SPIFs,” directly to salespeople at downstream firms who sell their products.

Employer “ratcheting,” or the practice of moving quota thresholds based on past performance, also provides an incentive to manipulate the size and timing of sales. Organizational researchers conducting fieldwork have long recognized the pervasiveness of restricting output to avoid quota increases.\footnote{For example, this practice has been labeled “soldiering” by Frederick Winslow Taylor (1912), “targeting a bogey” by Elton Mayo (Roethlisberger and Dickson 1939), and “quota restriction” by Donald Roy (1953).} Murphy (2000) shows that firms that set managerial quotas according to internal standards (such as a budget or past performance) have less-variable bonuses and smoother earnings than those that use external standards that cannot be gamed. Leone, Misra, and Zimmerman (2006)
find evidence of dynamic sales quota ratcheting in a Fortune 500 firm. They note that quotas rise with over-performance more readily than they fall with under-performance. Asymmetric ratcheting further compresses incentives around meeting quota by weakening the benefits of missing quotas (because quotas are unlikely to fall) and weakening incentives to exceed quotas (because quotas are likely to rise).

D. How and Why Firms Monitor Sales Activities

Sales hierarchies devote considerable resources to identifying and retaining salespeople who exceed quotas. In the data, turnover is 47% per year, and sales performance is highly skewed. However, in many cases it is difficult for the firm to attribute sales numbers to the skill of a salesperson, rather than exogenous factors such as the quality of the product, territory, or market conditions. As such, managers play a large role in identifying and retaining high performers.

Early organizational research emphasizes how managerial behaviors depart from profit maximization. Coase (1937), Penrose (1959), and Williamson (1967) invoke diminishing returns to management and the alienation of managerial interests to explain the limited growth of firms. Simon (1957) argues managers “satisfice,” adopting decisions that meet some non-maximizing acceptability threshold. Cyert and March (1963) argue managers possess neither the motives nor the cognitive means to make profit-maximizing decisions, and managers’ private information allows them to pursue tangential objectives. Crozier (1964) argues that hierarchies use impartiality as a pretense for the centralization and consolidation of organizational power. Chandler (1977) argues that managerial hierarchies are independent sources of power, permanence, and continued growth. Baumol (1959), Gordon (1961), and Williamson (1963) interpret profit as a constraint to which manager’s other goals—such as job security, influence, prestige, and advancement—may be pursued.

Recent advances in agency theory incorporate the role of supervisors in reducing gaming (see Gibbons 2005 or Miller 2005 for a review). Monitoring allows firms to condition employment and payment on agents’ inputs (such as effort) and discourages opportunism. To prevent opportunism and politicking among managers, firms may use bureaucratic rules and internal auditing. When performance measures are not contractible, firms may commit to subjective awards by relying on its reputation or by delegating subjective awards to an impartial supervisor. Subjective bonuses have

---

7The sales industry often cites the “80-20 rule,” the rule-of-thumb that 80% of sales are made by 20% of the salesforce. In the data, this slightly exaggerates the variation in sales performance at most firms. Prior to controlling for tenure, about 25% of salespeople are responsible for 75% of sales at the median firm. The variation is greatest for sales representatives at enterprise software companies and narrowest for engineers and support occupations.
other challenges, however; managers use evaluations to distribute performance rewards as they see fit, potentially eliciting cognitive biases, influence activities, and perceptions of unfairness. Conyon and He (2004), using evidence from CEO compensation committees, find three-tier agency models are better able to explain decision-making among executive compensation committees, compared to managerial power and collusion models.\textsuperscript{8}

Social psychologists have examined why workers and managers pursue an organization’s interests, even when such activities are unrewarded, and the circumstances under which individuals are willing to harm the organizations or other individuals to pursue personal goals. Bennett and Robinson (2000) and Berry et al. (2007) find that individuals who commit deviant behaviors harming individuals are also more likely to commit deviant behaviors harming the organization. Organizational researchers also offer explanations why the solutions offered by the standard agency theory may not work in practice. Neihoff and Moorman (1993) find that monitoring reduces organizational commitment behavior, and Gneezy and Rustichini (2000) find pecuniary penalties reduces guilt for breaking norms. Nickerson and Zenger (2008) hypothesize that perceptions of unfairness, envy, and dissatisfaction with relative pay restrict efficiency in settings such as sales, where individuals are highly motivated by pecuniary rewards. Studies by Benford and Snow (2000), Kaplan (2008), and Obloj and Sengul (2012) suggest managers will learn to frame their activities within new organizational initiatives, and learn to game their plans.

Salespeople are stereotypically driven by incentives rather than loyalty to their firms, as reflected by high turnover rates and high variation in compensation that typifies sales settings. Perhaps due to these reasons, information is rarely observed or communicated perfectly in sales hierarchies. For example, sales managers and their subordinates learn how difficult it is to sell a given product in a given territory, while other functions (potentially sales operations, the CFO, or marketing) use past performance, subjective reports, or other sources of information to produce forecasts. One purpose of these forecasts are to set quotas, making subordinates’ reports potentially unreliable.

While sales managers are responsible for monitoring gaming behavior, nonlinearities in their incentive plans encourage activities not perfectly aligned with profit maximization. Moreover,\textsuperscript{*}

managers typically focus their effort on negotiating and closing deals—the point at which there is the greatest opportunity to affect the timing of measured performance.

As such, sales organizations also rely on a variety of reporting practices to reduce information asymmetries, maintain incentive alignment, and discourage gaming. Customer relationship management (CRM) tools allow salespeople to report progress on their sales pipeline and share information regarding how clients’ purchasing decisions are made. These too may be gamed; interviewees report that subordinates may misrepresent the status of intermediary sales activities to avoid interventions by managers, whose desired closing date for sales may conflict with their own, or to avoid others from expropriating their client relationships and accounts. Moon and Mentzer (1999), in a study of a sales organization, found salespeople grossly misrepresented forecasts and the state of their sales pipeline, which they believed were used to adjust quotas.

Some firms use subjective bonuses or promotions to reward managers perceived to be acting primarily on the interests of the firm (for evidence outside sales, see Cappelli and Conyon 2011). Some firms restrict managers’ staffing, incentive, and pricing decisions, requiring large decisions to be approved or reviewed by sales operations or superiors. Indeed, the effort and expense firms dedicate to designing plans, monitoring activities, and improving coordination suggest that hierarchical coordination is indeed costly.

II. Managerial Quotas’ Effect on Staffing and Incentives

This study focuses on the manager’s interest to achieve the discrete pay, recognition, and job security associated with quota attainment. While firms also desire that their managers meet and surpass their quotas, this study examines managerial behaviors that (i) affect the staffing and incentives of subordinates, (ii) are encouraged by annual quotas, and (iii) are consistent with the interests of the manager, and not the interests of the firm or subordinates. Specifically, this study examines the timing of termination decisions and subordinate quota adjustments.9

The model connects the relevant institutional details of sales to an agency framework, and serves three purposes. First, it explains why firms delegate authority over staffing and incentive decisions

9 Perhaps the best-identified example of the effect of manager’s incentive plans on their subordinates’ employment and performance is Bandiera, Barankay, and Rasul’s (2007) field experiment on supervisors of fruit-pickers. Introducing a piece rate improved efficiency by leading supervisors to be more selective in whom they recruit and leads them to focus effort on assisting the most productive workers. Because workers consist of migrants who lived on the farm and their work consists of picking fruit, recruitment costs are negligible, the difficulty of the task is relatively well-known and homogenous, workers require minimal training, and there are no quotas. These characteristics distinguish fruit-pickers from the complex sales settings that distort incentives among sales managers.
to managers, even though the performance measure (sales) is observable to firms. Second, it illustrates how sales quotas distort the timing of staffing and incentive decisions. Third, the model yields the strategy for identifying decisions motivated by the unique interests of the manager.

A. Managerial Quotas and Subordinates’ Staffing

Developing a sales team is among a sales manager’s chief responsibilities. This involves recruiting, training, assisting, and disciplining subordinates. Because managers’ performance is measured primarily through the cumulative sales of their subordinates, their plans incentivize them to build productive teams. However, the incentives provided by quotas affect the timing of staffing decisions.

The intuition follows. Hiring and training new salespeople consumes time. In sales settings involving complex products or services, the typical applicant for a sales position is intensively screened, recruits receive training and/or a shadowing period, and newly-trained salespeople are given several months to develop skills and establish a “sales pipeline” beginning with initial leads and ending with a purchase order (and potentially installation, renewal, and support). Salespeople refer to this as the “ramp up” period. The sales industry often uses twice the length of the sales cycle as a heuristic for the ramp up time. For business-to-business sales, which constitute the majority of sales in these data, interviewees suggest a typical ramp up time would be six to twelve months. This estimate is consistent with results presented in the next section.

For this reason, replacing a poor-performing but experienced salesperson with a new recruit is an investment involving the substitution of present sales with greater expected future sales as the new recruit is hired and ramped up; in Oyer’s (1998) framework, retaining a poor performer is a way for managers to “pull in” sales from future measurement periods. Therefore, annual quotas create incentives for managers to retain poor-performing subordinates who would otherwise be terminated.

A natural question arising from agency theory is: Why do firms delegate termination decisions to managers, rather than specifying termination criteria in contracts? Based on interviews, I propose that it is very difficult for firms to translate sales figures into a claim about the quality of a salesperson, particularly when sales people are covering different products, territories, or functions. As such, manager’s chief responsibilities include: (i) selecting, mentoring, and screening a sales force, and (ii) developing an awareness of the qualities of sales territories and products. I interpret these responsibilities to signify that managers are responsible for screening for high ability salespeople, which cannot be distinguished from sales figures alone due to luck. Here, I use the term “ability” to include the personal characteristics that contribute to a salespersons’ sustained

10
performance. I use the term “luck” to include exogenous factors beyond a salesperson’s control, including the quality of the product, the quality of the territory, the quality of the leads, and so on. The model then captures the following insight: Firms delegate authority for termination decisions to managers to use their private knowledge of salespeoples’ ability and luck, allowing them to accelerate screening for new hires when exogenous factors affecting performance make it difficult for the firm to do so using sales figures alone. 10

To analyze the decision to employ a supervisor and the effects of a sales quota on a supervisor’s staffing decisions, first consider the following firm-worker model where the firm observes production (sales) but not ability or luck directly. For now, I abstract from the wage and effort decisions, and the only choice is the firm’s decision to retain or replace a worker, implicitly at the worker’s reservation rate. The employment relationship in the model may be thought of as a firm filling a unique job tied to a valuable asset; I abstract from the firm’s cost of acquiring that asset and the external competition that would lead the firm to adopt a reservation level of profitability for that asset and otherwise replace the worker. For example, for a medical device manufacturer, the asset may be the exclusive ability to sell a proprietary product to hospitals in California, which is tied to a specific sales job; for a newspaper, the asset may be an existing set of advertisers, who are assigned to an account manager. 11

Firms are risk-neutral, there are infinite periods, and firms discount future periods at δ. In period t = 1, the firm hires a worker, and then production occurs with the firm observing output. In periods t > 1, the firm chooses an action \( A_t \) ∈ \{retain, replace\}, and then production occurs with the firm observing output. The production of worker i in period t is

\[
y_{it} = r_{it} + \alpha_i + \varepsilon_{it}
\]

where \( r_{it} \in \{0, 1\} \) denotes whether the worker is “ramped up,” \( \alpha_i \in \{0, 1\} \) denotes the worker’s period-invariant ability, and \( \varepsilon_{it} \in \{0, 1\} \) denotes the worker’s period-specific luck. Let \( \Pr(\alpha_i = 1) = \)

---

10Analytically, the value of the supervisor in this setup most-closely resembles Harris and Raviv’s (1978) model in which a firm is willing to pay to contract on a risk-neutral worker’s effort rather than output. The strategic manipulation of information to affect decision-making has long traditions in organizational theory and decision theory (see especially Barnard 1938; Crozier 1964; Cyert and March 1963), with agency theory giving increasing attention to incorporating bureaucratic rules and politicking behavior into (see Gibbons, Matouschek, and Roberts 2012; or Tirole 1986, 1992; for a review).

11Here, assets might be used interchangeably with ‘resources’ in the resource-based view parlance; see Barney (1991), Rumelt (1984), or Wernerfelt (1984). That is, these assets are inherently valuable, firm-specific, and not easily imitated by competitors.
Pr(ε_{it} = 1) = 0.5, and r_{it} = 0 in the worker’s first period of employment and r_{it} = 1 thereafter if that worker is ever retained. A ramped up worker may be thought of as a worker with accumulated firm- and client-specific human capital and a mature sales pipeline. Crucially, suppose the firm observes y_{it} and r_{it}, but does not observe α_i or ε_{it}; the firm observes performance but not the worker’s ability or luck directly. It can be shown that, for δ ∈ (0.5, 1), the net present value (NPV) of a new recruit exceeds the NPV of a revealed low ability worker and the firm replaces a worker if and only if production is y_{it} = 0 for a new recruit or y_{it} = 1 for a ramped-up worker. 12 These are the two conditions in which the worker is revealed to be low ability. 13

For δ < 0.5, the NPV of a ramped-up low ability worker is greater than the NPV of a new recruit of unknown ability such that the firm retains all incumbents regardless of ability. Therefore, for the rest of this model I impose the parameter restriction δ ∈ (0.5, 1). In this case, the NPV of workers are:

\[ v_N = 1 + δ(0.5v_H + 0.5v_L) \]  \hspace{1cm} (2)
\[ v_H = 2.5(1 - δ)^{-1} \]  \hspace{1cm} (3)
\[ v_L = 0.5v_N + 0.5(1.5 + δv_L) \]  \hspace{1cm} (4)

where \( v_N \) denotes the NPV of a new recruit, \( v_H \) denotes the NPV of an experienced high ability worker, and \( v_L \) denotes the NPV of a low ability worker. For δ ∈ (0.5, 1), \( v_N > v_L \) and the firm replaces revealed low ability workers with new recruits. By substitution, the NPV of a new worker is \( v_N = (1 - 0.25δ - 0.125δ^2)(1 - 1.75δ + 0.75δ^2)^{-1} \).

Now, suppose instead that the firm may choose to delegate the replacement decision to a supervisor. I assume the supervisor is risk-neutral and may be paid a reservation rate from any surplus the supervisor generates (thereby allowing us to abstract from supervisor turnover). The crucial assumption is that the supervisor observes (but cannot verify) productive inputs α_i and ε_{it}. 14

12 This decision rule, in which a firm conditions terminations on both performance experience, is also consistent with the tendency for firms to raise quotas as a new recruit is ramped up.

13 Note that, by construction, the firm’s posterior belief regarding the worker’s ability is equal to the anterior belief for “medium” performance, Pr(α_i = 1|r_{it} = 0, y_{it} = 1) = Pr(α_i = 1|r_{it} = 1, y_{it} = 2) = 0.5. This simplifies the exposition.

14 In practice, firms’ performance evaluation procedures and customer relationship management tools may be an example of how the firm may try to verify α_i and ε_{it}. Such evaluations are notoriously unreliable and subject to bias, with meaningful subjective assessments of subordinates’ abilities eschewed by managers in the interest of avoiding conflict (Bretz, Milkovich, and Read 1989). Baliga (1999) shows that firms may choose to hire a supervisor even in the presence of
Formally, if the firm does not hire a supervisor, play proceeds as above. If the firm hires a supervisor, in period $t = 1$, the supervisor hires a worker, production occurs, and then the supervisor observes $\alpha_i$, $\varepsilon_{it}$, and $r$. In periods $t > 1$, the supervisor chooses an action $A_s \in \{\text{retain}, \text{replace}\}$, production occurs, and then the supervisor observes $\alpha_i$, $\varepsilon_{it}$, and $r$. Let $v^s$ denote the NPV of a subordinate with a supervisor.

The supervisor’s contract specifies a piecerate and bonus for production exceeding a quota, $Q_i$, normalized to zero. Define the supervisor’s total quota attainment in the event the subordinate is retained as

$$Q^H_i \equiv Q^L_i + y_{it}$$  \hspace{1cm} (5)

where $Q^L_i \in [-5, 5]$ is a random exogenous variable denoting what the quota attainment would be without subordinate $i$’s contribution, and $y_{it}$ is the subordinate’s contribution. Let the supervisor’s contract take the form $w_{it} = cS_{it} + b_{it}$, where $c \geq 0$ is the rent-sharing (commission) rate, $S(\delta) \equiv v^s_N - v^s_H$ is the surplus created when the firm hires the supervisor, and $b_{it}$ is a bonus, where $b_{it} = B \geq 0$ if $Q^L_i + y_{it} \geq Q_i$ and zero otherwise.

The firm can make terminating low ability subordinates strictly incentive compatible for all values $Q^L_i$ by choosing an arbitrarily-small $c$ and by setting $B = 0$, i.e., by eliminating the incentives for meeting quota and paying a linear piecerate. To see this, note that when the manager replaces low ability workers, the firm’s NPV of new, high ability, and low ability workers are:

$$v^s_N = 1 + \delta(0.5v^s_H + 0.5v^s_L)$$  \hspace{1cm} (6)

$$v^s_H = 2.5(1 - \delta)^{-1}$$  \hspace{1cm} (7)

$$v^s_L = v^s_N$$  \hspace{1cm} (8)

By substitution, the value of the supervised new worker is $v^s_N = (1 - 0.25\delta)(1 - 1.5\delta + 0.5\delta^2)^{-1}$, which is greater than the value of the unsupervised worker, $v_N$, yielding $S(\delta) > 0$ for $\delta \in (0.5, 1)$. Intuitively, the rent-sharing rule makes it incentive compatible for the supervisor to use the private information of $\alpha_i$ and $\varepsilon_{it}$ to accelerate the screening process for new workers.\textsuperscript{15}

---

\textsuperscript{15}The model features firms that learn the ability of the worker over time. This is because $\varepsilon_{it}$ is an independent and random draw in each period. For firms, the more troubling (but plausible) scenario is that $\varepsilon_{it}$ is serially-correlated within a position; since individuals are employed at positions over time, firms may misattribute sustained high performance to the individual rather than the position (e.g. a good territory). If managers enjoy private information regarding the value of the firm’s assets, this could allow managers to extract sustained rents on this information by pursuing non-verifiable information and the potential for collusion.
Although the model predicts firms may induce supervisors to terminate low ability subordinates by providing simple linear incentives, such plans are exceedingly rare in practice. Rather, incentive plans routinely feature quotas and other target-based goals. Unfortunately, there is no standard, empirically-established modeling technique yielding quotas and discrete incentives in incomplete contracts (for a theoretical discussion, see Frankel 2011, Kim 1997, Levin 2003, Oyer 2000, Park 1995, Steel and Konig 2006). Rather, consider the manager’s decision whether to replace an experienced \((r_{it} = 1)\), low ability \((\alpha_i)\) worker when a bonus for meeting a quota \((B > 0)\) is treated as exogenous.

For \(B > 4cS(\delta)\), contrary to the desires of the firm, managers retain low ability workers when \(Q^L_i \in [-2, -1]\). That is, when the bonus is sufficiently high, rent-sharing is sufficiently small, and the future is sufficiently discounted, quotas lead managers on the cusp of meeting a quota to pull in sales by retaining experienced, low ability subordinates.\(^{16}\) Intuitively, because the experienced worker enjoys \(r_{it} = 1\) with certainty whereas the new recruit enjoys \(\alpha_i = 1\) with probability 0.5, the immediate production for an experienced, low ability worker first-order stochastically dominates production for a new recruit.

Figure 2 illustrates how quotas distort incentives to screen low ability workers.

![FIGURE 2](image)

The model’s predictions are driven by managerial incentives to meet a quota. For identification purposes, this prediction is particularly powerful because the treatment concerns an interior range of quota attainment values. However, perhaps the larger distortion emerges from commission accelerators, which make pay convex in sales and provide incentives to concentrate sales into fewer measurement periods. Although the prediction that managers push out sales by replacing poor performers when they are far below quota may be unintuitive, interviews suggest this is common when the manager’s job is secure and marginal incentives are weak; interviewed sales managers referred to the general practice of concentrating losses in a single measurement period (including making staffing changes) as “taking a bath.”

Drawing from interviews and from the model they inform, I hypothesize that annual quotas lead managers to refrain from terminating poor-performing subordinates late in measurement periods personal objectives (such as pulling in sales, as above, or also collusion with subordinates, nepotism, or so on).

\(^{16}\)Note that an implication of the model is that gaming around quotas could be reduced by boosting local linear incentives around the threshold and reducing the discrete benefits of achieving it.
if the poor performer’s sales are needed for the manager to meet quota, and to terminate poor performers late in the measurement period if doing so is unlikely to affect whether the manager will meet quota.

**HYPOTHESIS 1:** *Managers will delay terminating poor performing subordinates until future measurement periods if and only if those subordinates are essential to the manager’s quota achievement (i.e., subordinates are “quota-critical”).*

This hypothesis requires four specific hypothesis tests. For months in the fourth fiscal quarter, I hypothesize that turnover of quota-critical poor performers will be lower than (H1a) turnover among poor performers whose manager *would not have* met quota anyway, and (H1b) turnover among poor performers whose manager *would have* met quota anyway. Furthermore, I examine whether these foregone turnovers are delayed until the following fiscal year. Specifically, in the month following the annual measurement period (the “thirteenth” fiscal month), I hypothesize the turnover of quota-critical poor performers will be higher than (H1c) turnover among subordinates whose managers did not meet quota, and (H1d) turnover among managers who did meet quota, but would not have without the subordinate’s credit.

**B. Managerial Quotas and Subordinates’ Incentives**

The second hypothesis concerns the alignment of subordinates’ incentives with managers’ interests. In particular, I examine whether subordinates are more likely to receive quota adjustments in the fourth quarter when the manager is in the neighborhood of reaching a quota.

Quota adjustments allow firms to adapt incentives to unforeseen circumstances. Otherwise, salespeople who are far below their quota for reasons unrelated to their prior effort may suffer from weak marginal incentives late in measurement periods, prompting them to quit or hoard sales until the next measurement period. To discourage salespeople from hoarding nearly-closed deals (commonly referred to as “sandbagging”), managers are typically instructed to reduce quotas only for reasons outside a worker’s control, to communicate how circumstances meet predetermined criteria for quota adjustments with subordinates, and be mindful of adjustments’ reputational effects.

In context of the model, I interpret this as evidence that managers observe $\alpha_i$ and $\varepsilon_{it}$ prior to production, are delegated authority to choose $A^s_q \in \{\text{keep quota, reduce quota}\}$, and that firms want managers to choose to keep quota for $\{\varepsilon_{it} = 1, \alpha_i = 1\}$, to reduce quota for $\{\varepsilon_{it} = 0, \alpha_i = 1\}$, and to fire the worker (as before) for $\alpha_i = 0$. The quota adjustment is a decision that creates surplus if the quota is kept for the output corresponding to $\{\varepsilon_{it} = 0, \alpha_i = 1\}$ and is reduced for the
output corresponding to \( \{ \varepsilon_{it} = 1, \alpha_i = 0 \} \). Intuitively, quotas specify an output level, but the key assumption is that firms cannot contract quota adjustments on \( \alpha_i \) or \( \varepsilon_{it} \), just as they could not for terminations. However, the immediate productivity boosts only depend on \( y_{ij} \); a downward quota adjustment may boost marginal incentives when production is low either due to bad luck or lower prior production due to low ability.

Although firms may immediately boost marginal incentives by adjusting quotas for all subordinates far below their quotas late in measurement periods, in practice, downward adjustments are not routine. This is because downward adjustments are also implicitly costly, as they distort incentives and harm morale if adjustments are anticipated or not viewed as exogenous. Otherwise, salespeople may ratchet effort, sandbag sales, and misrepresent forecasts early in measurement cycles in anticipation of adjustments. Subordinates may also interpret downward adjustments as “selective intervention,” and evidence of favoritism, entitlements, or managerial opportunism. These beliefs may impair the firm’s or the manager’s ability to commit to future quota adjustments only for exogenous and pre-defined circumstances (Foss 2003).\(^{17}\)

I interpret this to signify that a profit-maximizing firm would like to condition a quota on \( \varepsilon_{it} \) and credibly commit to it in advance, thereby avoiding moral hazard and sandbagging if downward adjustments are expected even if \( \varepsilon_{it} = 1 \). However, because firms observe only \( y_{ij} \) and not \( \varepsilon_{it} \), they delegate quota adjustment decisions to a supervisor. Because contracts specify \( y_{ij} \) and the quota’s effect on workers’ effort effective for both \( \{ \varepsilon_{it} = 0, \alpha_i = 1 \} \) and \( \{ \varepsilon_{it} = 1, \alpha_i = 0 \} \), supervisors on the cusp of making a quota retain poor performers and provide a downward quota adjustment. As such, providing a downward quota adjustment for a \( \{ \varepsilon_{it} = 0, \alpha_i = 1 \} \) worker is also tantamount to pulling in sales, as it boosts immediate incentives but invites future moral hazard as subordinates sandbag effort in anticipation of quota reductions.

The second hypothesis examines whether managers on the cusp of meeting their quotas are more likely to reduce subordinates’ quotas.

**HYPOTHESIS 2:** *Subordinates are more likely to have their quotas adjusted when their manager’s ultimate quota attainment will be near the quota threshold.*

While a quota adjustment is a relatively interpretable and standardized outcome variable, quota adjustments are arguably not the main way managers shift subordinates’ incentives. Managers

\(^{17}\)Indeed, firms recognize that quotas can become out of reach for subordinates, but rather than adjusting quotas downward, may tolerate the sandbagging and hoarding to avoid the reputational costs of downward quota adjustments when the exogenous circumstances cannot be verified.
may also shift incentives by changing subordinates’ implicit commission rates. For example, discretionary bonus pools may be distributed through targeted commissions and bonuses, through incentives for channels, through rewards for selling a certain product or to a certain client (“bounties”), through tournaments, or through other means. For salespeople greatly exceeding their quotas, a downward adjustment can boost marginal incentives by moving the salesperson to a higher commission rate tier.

The identification strategy requires distinguishing subordinates who managers believe necessary to meet their quota. As such, the model relies on calculating each parameter in \( Q_i^H \equiv Q_i^L + y_i \), and examining the turnover and quota adjustments among poor performing subordinates for whom \( Q_i^H > 100\% \) and \( Q_i^L < 100\% \).

### III. Data

#### A. Description of the Data

Data come from a firm that offers on-demand (over “the cloud”) sales performance management (SPM) software. Data include how 22 million transactions are credited to 7,492 sales managers and their 61,092 immediate subordinates in 244 client firms. Client firms upload their compensation plans to a server, and then the clients’ salespeople log in to the server (e.g. through a computer or smart phone) to report credited transactions and track their progress toward quotas or other benchmarks. The service is designed to make incentives and real-time performance transparent to salespeople and their managers, to calculate and automate compensation, to enable monitoring, to produce an audit-trail, and to promote flexibility in adapting compensation plans. Data begin in January 2008 and end in October 2011, although not all firms are represented throughout this period. No one firm represents more than 13% of workers or worker-months. Table 1 provides descriptive statistics.

[TABLE 1]

For each worker within a position, relevant data include a unique worker identifier (linkable if a person changes positions), a job title, a position title, the parent position, and the compensation plan. Position and parent position identifiers allow the construction of longitudinal organizational hierarchies, which also determine performance monitoring and other privileges within the SPM software. The most common job titles among workers with one level of subordinates are ‘territory manager,’ ‘sales director,’ ‘regional director,’ ‘regional manager,’ ‘sales engineer manager,’ and ‘regional vice president.’ Each transaction includes a timestamp, the share of credit for each worker
credited on the transaction, the amount and currency of the order, and the incentive plan to which the credit pertains. Each credited transaction further specifies associated quotas, commissions, and bonuses. Pay calculations from credited business are checked against a summary of terminal payments for each worker’s payroll period. Likewise, annual quota attainment calculations made from quota thresholds and credited business are checked against the SPM software’s transaction-level rolling measure for percent quota attainment, which was updated in real-time and made visible to workers and managers via the client software’s virtual dashboard. Because calculated pay may be automatically linked to payroll, forecasts, and audit reports, it is unusually incentive compatible to enter plans and transactions accurately.

The data are unusually rich in that they allow a large number of workers’ pay and performance to be tracked longitudinally and in a fashion that is reliable and standardized across firms. Because data come from an on-demand SPM software service, they largely avoid the selection dilemmas presented by data from single firms that opt-in as research sites. Likewise, this also helps address external validity concerns inherent to single-firm studies. However, the data also feature limitations. Analytically, the chief limitation is that employment is subject to truncation; for example, some salespeople enter the data when the company subscribes to the service or are still employed when the data end. Because it is optional for system administrators to input incumbents’ hire dates, tenure cannot always be determined for workers employed prior to the subscription to the SPM service. Although not essential to the analysis, the data offer two descriptive limitations. First, non-monetary rewards are not reported in the system; these may include prizes with great pecuniary or psychic value. Second, the data do not include education, demographics, or other descriptives sometimes available in empirical personnel research.

Compared to single-firm data, it is also difficult to identify and describe the institutional detail underlying data collected by on-demand SPM software. Single-firm studies typically complement quantitative results by interviewing workers whose experience is largely-similar to those in the data (Ichniowski and Shaw 2009 refer to this process as “insider econometrics”).

B. Bringing the Data to the Model

The model yields the prediction that the manager will be less likely to terminate poor performing subordinates and more likely to provide downward quota adjustments when the subordinate’s sales are needed for the manager to meet a quota. To do so, the data must distinguish subordinates within the “treatment bubble,” whose sales are needed for the manager to make quota. In the model, subordinates inside the treatment bubble are those for whom $Q_H^i > 100\%$ and $Q_L^i < 100\%$;
those below the treatment bubble are those for whom $Q_i^H < 100\%$ and $Q_i^L < 100\%$ (the manager would miss quota anyway), and those above the treatment bubble are those for whom $Q_i^H > 100\%$ and $Q_i^L > 100\%$ (the manager would make the quota anyway).

First, I estimate Equation 1, $y_{it} = r_{it} + \alpha_i + \varepsilon_{it}$, the production given the productive inputs of being “ramped up,” ability, and temporarily lucky. Second, I estimate terms in Equation 5, $Q_i^L + y_{it}$, the quota attainment absent the subordinate’s contribution plus the subordinates contribution. By definition, this yields $Q_i^H$.

Subordinate’s Production $y_{it}$. To estimate the subordinates’s contribution $y_{it} = r_{it} + \alpha_i + \varepsilon_{it}$, I estimate an OLS spline regression for the total monthly business credited as a function of tenure (the “ramp up,” $r_{it}$), a worker fixed effect (the worker ability, $\alpha_i$), and the noise term (the worker’s period-specific luck, $\varepsilon_{it}$). I use a spline regression with quarterly knots because the functional relationship between sales and tenure as a compromise between the weak assumptions of month fixed effects and power of linear approximations. I do this for each standard job classification within a firm. The regression takes the form

$$\ln(y_{it}) = \beta_0 + \beta_1 M_{3it} + \beta_2 M_{6it} + \beta_3 M_{9it} + \beta_4 M_{12it} + \alpha_i + \varepsilon_{it}$$

where $i$ is the individual salesperson, $t$ is the month, $\ln(y_{it})$ is the natural logarithm of the salesperson $i$’s total credited business in month $j$. $M_3$, $M_6$, $M_9$, and $M_{12}$ respectively denote months into the first, second, third, and fourth quarters, are zero for months prior to their respective quarters, and are three for months following their respective quarters. $M_{13}$ is zero for months prior to one year tenure, and otherwise denotes months greater than twelve. $\alpha_i$ is an individual fixed effect. I perform this regression separately for 834 standard job category-firm combinations that collectively feature 71,001 employment spells and 679,523 employee-months with credited transactions. Table 2 presents these regressions at the industry (rather than firm-job) level of aggregation. The worker fixed effect $\alpha_i$ estimates worker-invariant characteristics, while the residual $\varepsilon_{it}$ represents the idiosyncratic noise affecting measured performance in a given month. The empirical distributions of both $\alpha_i$ and $\varepsilon_{it}$ are approximately normal.

[TABLE 2]

\[^{18}\text{Specifically, } \ln y_{it} \text{ is the logged sum of the salesperson’s split order credit—the credit value of transactions multiplied by the share credited to the individual salesperson. The split order credit is the elemental unit of measured performance.}\]
First, consistent with the model’s assumption regarding the ramp-up period, Table 2 shows that workers become more productive with tenure. Credited business generally rises most rapidly for new recruits and then decelerates, presumably as salespeople accumulate basic skills and develop a sales pipeline. Recall that this is a necessary condition to yield the model’s predictions; intuitively, managers may expect an experienced, poor-performing incumbent with a mature pipeline to outperform a new recruit.

Second, Table 2 reports variation in workers’ performance $\alpha_i$ controlling for tenure. The hypotheses concern workers who are revealed to be low ability by sustained low performance; for the remainder of the paper, I denote workers in the bottom quartile as “poor performers.” Intuitively, these are the workers who have the bottom-quartile credited sales for workers of their job classification and tenure. Even if this method misclassifies the ability of some workers due to unobserved factors, it may offer a better-approximation of the supervisor’s beliefs. Because poor performers have higher turnover, workers with bottom quartile $\alpha_i$ values represent 25% of workers but only 18% of worker-months.

Lastly, Table 2 reports variation in a workers’ period-specific “luck” $\epsilon_{it}$. Depending on the industry, the worker fixed effect explains 70%-88% of the variation in measured performance after controlling for tenure. Because this table aggregates to the industry level, it overstates the variation explained by worker effects when the regression model is run for job classifications within firms.

Conditional managerial quota attainments, $Q^H_i$ and $Q^L_i$. Following Equation 5, define $Q^H_i \equiv Q^L_i + y_i$, where $Q^L_i$ is the expected quota attainment if the poor performing subordinate turns over in the fourth fiscal quarter (months 10 - 12), $y_i$ is the subordinate’s contribution in the fourth quarter, and their sum is the manager’s expected quota attainment if the subordinate is retained in the fourth quarter.

To estimate $Q^L_i$, I recalculate the manager’s quota attainment under the counterfactual scenario that subordinate $i$’s sales in the fourth quarter are set to zero. This includes actual sales for the full fiscal year of all of the subordinate’s peers under the same manager, plus the actual sales for fiscal months 1 - 9 for subordinate $i$.

To estimate $Q^H_i$, first I re-estimate the firm-job regressions in equation 8 with fiscal month fixed effects. Exponentiating the predicted values yields a prediction for sales credit in months of the fourth quarter that adjusts for fiscal month effects and subordinate $i$’s prior performance. These values are then summed to estimate predicted sales in the fourth quarter, $y_i = \hat{y}_{i10} + \hat{Y}_{i11} + \hat{Y}_{i12}$.

Note that $y_i$ is the sum of the three $y_{it}$’s corresponding to months of the fourth quarter.
This prediction is strictly positive (following from exponentiating the logged dependent variable), and the difference between the logged predicted sales and the logged actual sales credit for those who stay is approximately normal with mean zero. Then \( Q_i^H = Q_i^L + \hat{y}_{i10} + \hat{y}_{i11} + \hat{Y}_{i12} \), where \( \hat{y}_{im} \) is the exponentiated predicted values from the regression in month \( m \). Note that I use the anterior predicted sales rather than actual sales even when actual sales are revealed in the data. I do this because the independent variable of interest is the manager’s expectations, and because actual sales aren’t observed for workers who turn over.\(^{20}\)

Peer effects, the manager’s private information, and other factors introduce measurement error into \( Q_i^L \) and \( Q_i^H \). However, the power of the test is that it occurs at an interval, and so measurement error would lead to attenuation bias, reducing the likelihood of finding a significant result.

**Defining Turnover.** The data treat turnover events as any severance in employment for the purpose of recordkeeping within the SPM software. As such, turnover events include quits, layoffs, and fires. Turnover events also include internal job transfers to a position not covered by the SPM software, but do not include transfers to a position covered by or added to the SPM software. For example, a salesperson who transfers within sales or who is promoted to a sales manager is likely to remain in the data (and therefore not be counted as a termination). However, a salesperson who transfers out of sales may drop out of the data and be counted as a turnover event. Data include 38,159 turnover events and 15,695 internal transfers. Turnover is highly periodic, with peaks at the end of measurement cycles.

The reasons for turnover (terminations or quits) are immaterial in this setting, since it distinguishes the likelihood of turnover inside the “treatment bubble” where the manager has a personal interest in retaining the worker. In doing so, the identification strategy also solves a problem for research on turnover, since the distinction between terminations and quits is often blurred; workers who would be fired are often counselled to quit, and workers who quit may have instead been persuaded by enticements to stay.

**IV. Results**

**A. Results for Managerial Quotas and Subordinates’ Staffing**

To test Hypothesis 1, I test whether the turnover of “quota-critical” poor-performing subordinates in the fourth quarter (i.e. \( Q_i^L < 100\% < Q_i^H \)) is lower than the turnover for poor performers for \(^{20}\)As a check, I compare predicted sales against actual sales among retained subordinates, and confirm that actual log-sales are approximately normally distributed with mean zero about predicted log-sales.
managers whom the model predicts would or would not meet a quota anyway (i.e. $100\% < Q^L_i < Q^H_i$ or $Q^L_i < Q^H_i < 100\%$). To test the specific hypothesis that turnover is delayed for a month, rather than foregone, I also test whether the turnover of quota-critical poor performers in Month 13 is higher than non-quota critical poor performers.

Consistent with Hypothesis 1, Table 3 shows that turnover is 5.6% when the manager reaches the quota threshold if and only if the subordinate’s credited transactions are counted, substantially lower than the 18.6% turnover rate when the manager does not meet quota (H1a) and the 22.2% turnover rate when the manager would meet quota anyway (H1b). Both differences are significant with $p < 0.01$. These results suggest managers forego terminating poor performers late in the measurement period when these subordinates are needed for the manager to meet quota. Table 3 also shows that “thirteenth month” turnover among poor-performing subordinates is substantially higher when the manager met quota but would not have without the subordinates’ fourth quarter credited transactions, compared to managers who did not meet quota (H1c) or would have met quota anyway (H1d). Taken together, results suggest managers delay terminating poor performers until the following fiscal year when doing so would reduce their likelihood of making an annual quota.

Although the result does not hold for non-poor performers, this result should be treated with caution. Most non-poor performers are making their quotas, suggesting that their performance is sufficient to avoid termination. The highest performers are likely to delay quitting until the end of the year to collect their bonus pay.\footnote{The data are generally consistent with this account. Only 31% of quotas among poor performers are met, and the median variable pay of poor performers (including team incentives) is only 4% of fourth quarter salary. In contrast, 60% of quotas among non-poor performers are met, 80% of non-poor performers receive variable pay in the fourth quarter, and the median variable pay among non-poor performers is 60% of base salary. For this reason, the primary incentive for improving performance among non-poor performers may be increasing variable pay, while the primary incentive among poor performers may be avoiding dismissal.} To provide an estimate for how much turnover among poor performing salespeople is foregone because the manager is near a quota threshold, I restrict the sample to the poor performing salespeople whose sales are not necessary for their managers to meet quota and perform a logistic regression predicting the likelihood a subordinate will turn over as a linear function of the manager’s and subordinate’s quota attainment. I compare this predicted likelihood to the actual turnover rates among salespeople whose sales were necessary for their managers to meet quota, and estimate that
total actual fourth quarter turnover among poor performers is 15% (s.e.: 1%) less than it would be if turnover among these “quota-critical” salespeople followed the same linear trend.

[FIGURE 3]

To examine turnover rates in greater detail, Figure 3 presents fourth quarter turnover rates within 5% bins of the expected quota attainments when the subordinate turns over, $Q_i^L$, rather than retained, $Q_i^H$. Fourth quarter turnover of poor performers drops sharply when the model predicts the manager will barely make quota. The low turnover among poor-performing subordinates whose managers who (barely) didn’t need the subordinate to meet quota suggests managers may be conservative in terminations.

Next, I perform a falsification test against effects and common shocks occurring at the level of the industry, the level of the firm, and the level of the manager’s peers. To do so, I compare turnover rates of subordinates’ against the quota attainment of a manager who shares the same upper-manager as their own direct manager. Intuitively, if managers are thought of as the “parent” position in the organizational hierarchy, subordinates may be compared against the quota attainment of the manager filling their “uncle” position.

[FIGURE 4]

Figure 4 presents two series of data. The hollow markers show the turnover rates at the uncle’s quota attainment levels for the universe of their “niece/nephew” subordinates. The solid markers restrict the sample to subordinates outside the treatment bubble of the parent position. This is important, since the parent’s and uncle’s quota attainments are not independent. The hollow series shows that the turnover among poor performers in the neighborhood of the uncle’s quota exhibits a small dip, which is not as pronounced as it is around the parent’s quota. Turnover within 10% of the uncle’s quota is 2.4% (with a standard error of 0.3%) lower than it is within 50% of the uncle’s quota. The restriction introduced in the solid markers eliminates statistically significant effects in the neighborhood of the uncle’s quota threshold, suggesting the dip among the hollow markers is caused by correlation in the uncle and parent’s quota attainment.

Results show turnover among poor performers is lower late in annual measurement periods when poor performers’ credited transactions would affect whether the manager would meet quota, and greater in the following month. More broadly, results suggest managers delay terminating poor performers until the following fiscal year when doing so is likely to affect their perceived likelihood of making quota. By showing that the parent position’s quota attainment predicts turnover net of
the “uncle” position’s quota attainment, I rule out common shocks at the level of industry, firm, or division within the firm (at the next level of the organizational hierarchy).

B. Results for Managerial Quotas and Subordinates’ Incentives

To test the hypothesis that managers’ quotas affect their subordinates’ incentives, I test whether subordinates are more likely to have their quotas adjusted just prior to the end of the fiscal year when the manager’s sales are projected be near quota.

I run six logistic regressions to evaluate the likelihood of quota adjustments over months of the fourth quarter, as a function of the manager’s and subordinate’s expected quota attainment. The manager’s expected quota attainment is calculated as $Q^H_i$ above; it is the actual quota attainment through the third quarter plus the sum of the predicted attainment of the subordinates. The subordinate’s expected quota attainment is sales as a percent of quota, including actual quota attainment through the third quarter plus the expected fourth quarter quota attainment. These regressions take the form

$$\ln \left( \frac{\hat{p}}{1 - \hat{p}} \right) = \hat{\beta}_0 + \mathbf{X}_m \hat{\beta}_m + \mathbf{X}_s \hat{\beta}_s$$

where $\hat{p}$ is the logit-estimated probability for the outcome of interest, $\mathbf{X}_m$ is a dummy vector indicating the 5% quota attainment bin of the manager and $\mathbf{X}_s$ is a dummy vector indicating the 5% quota attainment bin of the subordinate. This regression is performed for the six outcomes of interest corresponding to $\{\text{Upward Adjustment, Downward Adjustment}\} \times \{\text{Month 10, Month 11, Month 12}\}$.

[TABLE 4]

[FIGURE 5]

Table 4 shows the logistic regression results for downward adjustments in Months 10, 11, and 12. Figure 5 illustrates the regression’s estimated probabilities at the manager’s and subordinate’s expected quota attainments, holding the other’s constant at their mean, for both upward and downward adjustments for each month.

Before discussing the main results, some trends deserve note. First, downward adjustments are more common than upward adjustments. Upward adjustments, while potentially bad for morale, may adjust for an exogenous circumstance that boosts performance measures (e.g. a product launch) or may be implemented as a penalty (too many clients canceled sales). Second,
the probability that quotas are adjusted downward in months ten and eleven declines slightly as
the manager’s expected quota attainment increases. This slight downward trend may be because
when a manager has a high quota attainment, it suggests the subordinate’s peers are doing well,
and that the subordinate is responsible for poor sales and should not be awarded a downward
quota adjustment. Third, the probability that quotas are adjusted downward in months ten and
eleven are more-sensitive to the quota attainment of the subordinate. One explanation is that
managers interpret greatly-surpassed quotas as a signal the quota is already too low (consistent
with ratcheting models). Fourth, quota adjustments in months ten and eleventh are more likely to
place subordinates in the neighborhood of ultimately reaching their quota, rather than far above
or far below. Fifth, quota adjustments are both more-rare and less-sensitive to the manager’s and
subordinate’s quota attainment in the twelfth fiscal month than they are in the tenth and eleventh
fiscal months. Lastly, in addition to these general trends, subordinates throughout the distribution
receive both downward and upward quota adjustments, including downward adjustments for
salespeople expected to surpass their quotas and upward adjustments for salespeople expected
to miss their quotas. This may be because adjusting one subordinate’s quota may also require
adjusting comparable peers’ quotas as well, in the interest of fairness.

Table 4 and Figure 5 lend support for the main prediction. The regression estimates that the
likelihood a subordinate receives a quota adjustment is significantly lower \( p < 0.05 \) in each of the
six 5% intervals of the manager’s quotas between 75-90% and 110-125%, compared to the interval
100-105%, for both fiscal months 10 and 11. The jump in the probability of an adjustment in the
neighborhood of quota attainment does not appear in the twelfth fiscal month. I do not find strong
evidence that upward adjustments are more likely in the neighborhood of either a subordinate’s or
manager’s quota threshold.

To provide an intuition of the magnitude, I restrict the sample to workers whose managers are not
within 10% of making quota, re-estimate the likelihood of receiving a downward quota adjustment as
a linear function of the manager’s ultimate quota attainment, and compare the estimated likelihood
of receiving a quota adjustment against the actual likelihood of receiving a quota adjustment among
workers whose managers were within 10% of the quota threshold. These estimates suggest 13% 
(s.e.: 1%) of downward quota adjustments in months 10 and 11 are explained by the jump in quota
adjustments among managers within 10% of making quota.

The likelihood of quota adjustments in the tenth and eleventh months of the fiscal year rise in
the neighborhood of making quota, but not more pronounced for those who barely-met a quota
compared to those who barely-miss it. As such, subordinates are more-likely to receive downward
adjustments when their managers will not meet a quota anyway. One interpretation is that quota adjustment is a blunt instrument for boosting sales. Channel stuffing, discounting, credit-trading, sales hoarding, and reallocating effort (both in quantity and toward deals near completion) may offer more-immediate and precise ways of making a quota.

V. Solutions to Managerial Gaming?

Results suggest sales functions’ turnover and quota adjustments are different than what they would be in the absence of managers’ discrete interest in making quotas. Presumably, there are profit-maximizing criteria for terminating poor performers and adjusting quotas, and these are distinct from manager’s interest in the marginal sale associated with his or her quota attainment. Moreover, this study tests for a narrow, relatively well-identified gaming behavior. Managers’ decisions routinely shift the timing of sales, including the timing and implementation of product pricing, product promotions, and sales campaigns. Managers also govern the allocation of resources across presale, final negotiation, and support activities. Interviews suggest that pricing, channel stuffing, and gaming discretionary bonuses may be more costly to the firm. As such, the gaming behavior described in this paper should be construed as a part of broader behaviors elicited when incentives are strong and information is poor.

Agency and organizational theory provide some guidance as to how managerial gaming may be mitigated. Any potential solution to gaming around quota thresholds should be weighted against the three chief benefits described in the review section: incentivizing marginal effort, communicating minimum acceptable performance for retention, and invoking loss aversion.

With those in mind, the model calls attention to specific levers for reducing managerial gaming. First, there are levers for smoothing incentives around quota thresholds:

- **Locally-linear incentives.** Firms may use locally-linear incentives around quota thresholds and eliminate bonuses for 100% quota attainment (setting $B = 0$). This implicitly deemphasizes the quota itself. However, further research is required to ascertain whether these practices also diminish quotas’ benefits.

- **Quota obfuscation.** In practice, distance to quota ($Q_i - Q_i^L$) is often obfuscated by complex crediting procedures, cancelled sales, or credit clawbacks. In principle, this uncertainty would smooth expected pay around expected sales; Ederer, Holden, and Meyer (2012) explore a theoretical model in which obfuscating performance measures reduces gaming. However, due to morale and incentive concerns, plan designers generally eschew complex and nontransparent
plans; Englmaier, Roider, and Sunde (2012) provide evidence that obfuscating piece rates dilutes incentive effects.

Second, there are levers around the firm’s discovery of salespeople’s luck:

- **Bureaucratic rules and monitoring.** If exogenous circumstances affecting a sales performance $x_t$ could be codified and verified to the firm *ex post*, then firms can deter managerial gaming through bureaucratic decision rules and monitoring. In the model, this is not possible because the firm does not observe $x_t$. In practice, organizations do try to deter gaming through formal procedures, including performance appraisals, bureaucratic rules, complaints, audits, and monitoring (*e.g.* through CRM tools). Decisions that could be motivated by opportunism, such as large discounts, may require approval or review by a senior manager. Firms may also require managers to justify termination and quota adjustment decisions; these criteria may be unverifiable (almost got a big client, but was unlucky) or verifiable (poor sales were due to unfavorable exchange rates). However, such rules can reduce flexibility and raise bureaucratic costs. Neihoff and Moorman (1993) and Gneezy and Rustichini (2000) provide evidence that monitoring and pecuniary penalties reduce organizational citizenship behaviors.

- **Improving forecasts.** If exogenous circumstances affecting sales performance $x_t$ could be contracted upon in advance, then the firm would not need to rely on a manager’s discretion. Indeed, forecasts are used to set sales quotas for this purpose. However, accurately forecasting sales is notoriously difficult, particularly when selling new products. Some organizations condition forecasts or quotas on easily-observed future circumstances, such as exchange rates in commodities. Forecasting using past performance can also distort incentives through the ratchet effect, and salespeople can game the customer relationship management (CRM) tools used to generate forecasts (Moon and Mentzer 1999).

- **Relative performance measures.** If salespeople are performing the same tasks and are subject to common shocks, firms may use plans that reward “relative” performance. Sales tournaments are a common application. However, these plans are also challenging to implement: exogenous shocks may be specific to an individual salesperson, relative performance measures discourage teamwork and diffusion of best selling practices, and anticipated tournaments encourage sandbagging.

- **Equalizing opportunities.** If there is little accumulation of position-specific human capital, firms can reorganize work to create comparable jobs. For example, rather than assigning
a salesperson all leads from a narrow product or territory, an organization could distribute shares of leads from a wide range of products or territories, or rotate jobs, making sales performance comparable across salespeople.

Third, there are levers around aligning managerial incentives directly:

- **Conditioning managerial quotas.** Firms may adjust managerial quotas downward when they make new hires (reducing $Q_t$ by $r$ when the manager employs a new hire). While this would eliminate a manager’s incentive to retain a poor-performing subordinate, it would also reduce incentives for retaining capable salespeople.

- **Promoting altruism.** Firms may attempt to make managers interested directly in the well-being of the firm, aligning managers’ interests with those of the firm (making the manager’s utility that of the firm, like that in Akerlof and Kranton’s (1986) model of identity). In industries where high variable pay is common, salespeople are notoriously driven by commissions rather than by company loyalty. Behavioral research may offer insight for promoting organizational citizenship behavior and mitigating gaming (see, for example, Podsakoff et al. 2009 for a review).\(^{22}\)

Indeed, each theoretical solution has real world analogues. Together, they provide insight as to the conditions in which specific levers are effective.

The model also provides guidance as to how the characteristics of the selling environment determine the structure of sales hierarchies. The model predicts that firms would not hire a supervisor if it could observe workers’ person-specific luck, $x_t$. Indeed, in settings where the solutions above are easy to implement (such as inside sales representatives within call centers, where performance is easy to measure and may be compared against workers subject to the same exogenous factors affecting performance), there are a higher ratio of first-line salespeople to first-line managers, and these managers are given less discretion to retain a poor performer. As such, cheaper and more-effective monitoring technology may not only improve incentives, but it may also flatten bureaucracies.

\section*{VI. Conclusion}

This study examines the imperfect ability of organizational hierarchies to motivate intermediary managers to act on the organization’s behalf. It does so by exploiting the institutional features that

\(^{22}\)SAS Corporation and Apple serve as famous examples of companies that have enjoyed high sales performance and low turnover without commissions.
(i) that sales managers’ have a unique interest in the marginal sales that meet their quotas, (ii) their quota attainments are determined by the cumulative credited business of their subordinates, and (iii) that managers can affect the timing of sales through staffing and incentive decisions affecting subordinates. Because an incumbent poor performer with accumulated job-specific skills and an established sales pipeline is likely to be more productive in the short term than a new recruit, sales managers may pull in sales by retaining poor performers until the following fiscal year. Because incentives for salespeople far below quota are weak, adjustments that make quotas “within reach” allow managers to boost immediate sales, while risking future moral hazard and disillusionment as adjustments become expected and viewed as entitlements. Following the model, I propose firms delegate authority for these decisions to managers due to their private knowledge of salespeoples’ abilities and the sales environment, which is difficult for the firm to observe and verify. Although piecerates encourage managers to use this information to improve the screening and motivation of the salesforce, discrete quotas and sales goals provide managers an incentive to game the staffing and incentives of subordinates. This study makes three main contributions.

First, this study makes a methodological contribution to empirical organizational research. I connect an agency model to predictions within a “treatment bubble,” yielding a novel strategy for identifying the interests of individuals within organizations. I execute this strategy within a novel class of data, generated by on-demand sales performance management software. This service creates standardized, reliable, longitudinal, cross-firm data, enabling the study of compensation plans and performance for a large number of workers while mitigating the habitual selection issues and external validity concerns of studies in which participating firms “opt-in.”

Second, this study contributes a theory for the delegation of authority in organizational hierarchies, inspired by the institutional details of managing sales. Using a firm-supervisor-worker model, I hypothesize that firms use managerial incentives to make terminating low ability workers incentive compatible, since a firm cannot fully contract on a worker’s ability or the difficulty of the selling environment. I show how firms attempt to use the many theoretical solutions to managing gaming, including the practices of monitoring, reporting, incentives, bureaucratic rules, job rotations, and attempts to get workers to identify with the firm. Moreover, the model helps explain why organizations in which relative performance appraisal is easy (such as call centers), hierarchies tend to be flatter and immediate managers have less discretion. I show how organizational and agency theory can inform under what conditions these practices may be effective.

Third, by highlighting the case of sales managers, this study contributes well-identified evidence for a specific failure of agency intermediation. While both classic and recent theoretical work
emphasizes misalignment of managerial incentives in determining the structure of the firm, there has been scant evidence misaligned incentives propagate within organizational hierarchies. As such, this study corroborates the classic hypothesis that the interests of managers, who are the intermediary agents responsible for performing the activities of the firm, are inconsistent with profit maximization.

REFERENCES


31


### Table 1—Descriptive Statistics by Industry

<table>
<thead>
<tr>
<th></th>
<th>Information &amp; Enterprise Software</th>
<th>Scientific, Professional, &amp; Technical Services</th>
<th>Manufacturing</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>a. Basic Descriptives</td>
<td>Firms</td>
<td>56</td>
<td>38</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Firm-Months</td>
<td>1,032</td>
<td>679</td>
<td>1,165</td>
</tr>
<tr>
<td></td>
<td>Persons</td>
<td>18,662</td>
<td>14,448</td>
<td>22,297</td>
</tr>
<tr>
<td></td>
<td>Person-Months</td>
<td>286k</td>
<td>237k</td>
<td>330k</td>
</tr>
<tr>
<td>b. Standardized Jobs</td>
<td>Sales Managers</td>
<td>1,960</td>
<td>1,658</td>
<td>2,139</td>
</tr>
<tr>
<td></td>
<td>Sales Reps</td>
<td>4,028</td>
<td>5,255</td>
<td>6,829</td>
</tr>
<tr>
<td></td>
<td>Account Managers</td>
<td>2,227</td>
<td>2,239</td>
<td>2,174</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>12,159</td>
<td>8,207</td>
<td>12,238</td>
</tr>
<tr>
<td>c. Orders &amp; Payments</td>
<td>Transactions</td>
<td>23m</td>
<td>22m</td>
<td>74m</td>
</tr>
<tr>
<td></td>
<td>Commissions</td>
<td>15m</td>
<td>14m</td>
<td>39m</td>
</tr>
<tr>
<td></td>
<td>Bonus Payments</td>
<td>49k</td>
<td>47k</td>
<td>176k</td>
</tr>
<tr>
<td></td>
<td>Other V. Payments</td>
<td>14k</td>
<td>6k</td>
<td>54k</td>
</tr>
<tr>
<td></td>
<td>Var. Pay (USD)</td>
<td>2,277m</td>
<td>1,462m</td>
<td>3,729m</td>
</tr>
<tr>
<td>d. Key Variables</td>
<td>Turnover Events</td>
<td>8,968</td>
<td>3,170</td>
<td>9,346</td>
</tr>
<tr>
<td></td>
<td>Comp Plans</td>
<td>1,877</td>
<td>1,052</td>
<td>1,842</td>
</tr>
<tr>
<td></td>
<td>Quotas</td>
<td>28,702</td>
<td>21,087</td>
<td>32,006</td>
</tr>
<tr>
<td></td>
<td>EE-Fiscal Years</td>
<td>35,703</td>
<td>30,472</td>
<td>44,461</td>
</tr>
</tbody>
</table>

**Note** — Firms include all those for which data are complete. “Other” industries include retail trade (14 firms), wholesale trade (13), administrative support (12), and finance and insurance (10). The suffix “m” denotes the count is in millions. Data include 89 private and 133 public companies and subsidiaries thereof. Sales managers include persons with reporting subordinates. Credited transactions include only unique order-worker combinations, and do not count, for example, annuity tails or redundant individual/team credits. Other variable payments include, for example, draws that are arguably part of base pay.
Table 2— OLS Spline Regression Predicting Log-Credited Business of Subordinate Employees with Employee Fixed Effects, by Industry

<table>
<thead>
<tr>
<th></th>
<th>Information &amp; Enterprise Software</th>
<th>Scientific, &amp; Technical Services</th>
<th>Manufacturing</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Months 1 - 3 ($M_3$)</td>
<td>0.281*</td>
<td>0.132*</td>
<td>0.268*</td>
<td>0.217*</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.015)</td>
<td>(0.011)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Months 4 - 6 ($M_6$)</td>
<td>0.157*</td>
<td>0.125*</td>
<td>0.162*</td>
<td>0.166*</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Months 7 - 9 ($M_9$)</td>
<td>0.136*</td>
<td>0.020*</td>
<td>0.070*</td>
<td>0.102*</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Months 10-12 ($M_{12}$)</td>
<td>0.109*</td>
<td>0.018*</td>
<td>0.063*</td>
<td>0.0747*</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Months 13+ ($M_{13}$)</td>
<td>-0.099*</td>
<td>0.002</td>
<td>0.009*</td>
<td>0.048*</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.023)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>SD of $\alpha_i$</td>
<td>4.794</td>
<td>6.359</td>
<td>3.999</td>
<td>5.503</td>
</tr>
<tr>
<td>SD of $\varepsilon_{it}$</td>
<td>3.066</td>
<td>2.415</td>
<td>2.522</td>
<td>2.429</td>
</tr>
<tr>
<td>Share of var(ln $y_{it}$) explained by $\alpha_i$</td>
<td>0.710</td>
<td>0.874</td>
<td>0.715</td>
<td>0.837</td>
</tr>
<tr>
<td>Fixed Effects (EEs)</td>
<td>11,397</td>
<td>10,961</td>
<td>8,314</td>
<td>7,864</td>
</tr>
<tr>
<td>Obs (EE-Months)</td>
<td>141,493</td>
<td>91,373</td>
<td>168,101</td>
<td>69,383</td>
</tr>
</tbody>
</table>

Note — * : $p < 0.01$. These regressions are performed at the level of four industries, while distinguish poor performers use regressions at the level of 834 job-firm combinations, where employee fixed-effects are normalized such that employee fixed effects for jobs-in-firms have a mean of zero. Regression results are for workers without subordinates who ultimately had greater than six months tenure and who had at least four peers with their firm-job category.
Table 3—Monthly Turnover by Estimated Quota Attainment With and Without Subordinate’s Credits, for Poor and Non-Poor Performing Subordinates

<table>
<thead>
<tr>
<th>Parent’s Q Attainment</th>
<th>$Q^H_i &lt; 100%$</th>
<th>$Q^H_i \geq 100%$</th>
<th>$Q^H_i \geq 100%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discounted Q Attainment</td>
<td>$Q^L_i &lt; 100%$</td>
<td>$Q^L_i &lt; 100%$</td>
<td>$Q^L_i \geq 100%$</td>
</tr>
<tr>
<td>Outcome of Parent’s Q Attainment wrt. Sub.:</td>
<td>“Misses Quota Anyway”</td>
<td>“Makes Quota Conditionally”</td>
<td>“Makes Quota Anyway”</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
</tr>
</tbody>
</table>

a. Poor Performing Subordinates

<table>
<thead>
<tr>
<th>Fourth Quarter</th>
<th>mean</th>
<th>18.6%</th>
<th>5.6%</th>
<th>22.2%</th>
</tr>
</thead>
<tbody>
<tr>
<td>s.e.</td>
<td></td>
<td>(0.69%)</td>
<td>(0.36%)</td>
<td>(0.35%)</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td>3,114</td>
<td>3,972</td>
<td>13,853</td>
</tr>
</tbody>
</table>

“Month 13”

<table>
<thead>
<tr>
<th>mean</th>
<th>7.4%</th>
<th>14.4%</th>
<th>4.1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>s.e.</td>
<td>(0.53%)</td>
<td>(0.61%)</td>
<td>(0.19%)</td>
</tr>
<tr>
<td>n</td>
<td>2,427</td>
<td>3,306</td>
<td>10,137</td>
</tr>
</tbody>
</table>

b. All Other Subordinates

<table>
<thead>
<tr>
<th>Fourth Quarter</th>
<th>mean</th>
<th>10.4%</th>
<th>11.1%</th>
<th>18.8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>s.e.</td>
<td></td>
<td>(0.30%)</td>
<td>(0.32%)</td>
<td>(0.16%)</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td>9,857</td>
<td>9,133</td>
<td>56,887</td>
</tr>
</tbody>
</table>

“Month 13”

<table>
<thead>
<tr>
<th>mean</th>
<th>3.7%</th>
<th>4.7%</th>
<th>1.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>s.e.</td>
<td>(0.22%)</td>
<td>(0.25%)</td>
<td>(0.05%)</td>
</tr>
<tr>
<td>n</td>
<td>6,946</td>
<td>6,862</td>
<td>45,358</td>
</tr>
</tbody>
</table>

Note – Turnover rates are for subordinates in the fourth quarter and the first month of the year following the fiscal year for which the manager’s quota attainments and the subordinate-adjusted quota attainments are reported. “Month 13” results include only cases where the manager-worker pair matches across fiscal years (i.e. neither the manager nor subordinate were transferred, the manager did not turnover, and the subordinate did not turn over in the previous quarter).
Table 4—Logit Predicting Subordinate’s Quota Adjustment by Manager’s and Subordinate’s Realized Quota Attainment, for Fiscal Months 10 - 12

<table>
<thead>
<tr>
<th>Quota Range</th>
<th>Month 10</th>
<th>Month 11</th>
<th>Month 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta^m$</td>
<td>$\beta^s$</td>
<td>$\beta^m$</td>
</tr>
<tr>
<td>75-80%</td>
<td>-0.775**</td>
<td>-0.318</td>
<td>-0.530*</td>
</tr>
<tr>
<td></td>
<td>(0.210)</td>
<td>(0.178)</td>
<td>(0.209)</td>
</tr>
<tr>
<td>80-85%</td>
<td>-0.379**</td>
<td>-0.525**</td>
<td>-0.152</td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td>(0.157)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>85-90%</td>
<td>-0.615**</td>
<td>-0.180</td>
<td>-0.543**</td>
</tr>
<tr>
<td></td>
<td>(0.147)</td>
<td>(0.133)</td>
<td>(0.160)</td>
</tr>
<tr>
<td>90-95%</td>
<td>-0.325**</td>
<td>-0.425**</td>
<td>-0.472**</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.128)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>95-100%</td>
<td>-0.233</td>
<td>0.212*</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.108)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>100-105%</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td>105-110%</td>
<td>0.009</td>
<td>0.184*</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>(0.0821)</td>
<td>(0.088)</td>
<td>(0.091)</td>
</tr>
<tr>
<td>110-115%</td>
<td>-0.360**</td>
<td>-0.190</td>
<td>-0.410**</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.097)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>115-120%</td>
<td>-0.440**</td>
<td>-0.217*</td>
<td>-0.327**</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.091)</td>
<td>(0.092)</td>
</tr>
<tr>
<td>120-125%</td>
<td>-0.523**</td>
<td>-0.274**</td>
<td>-0.522**</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(0.104)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>Exterior</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5% F.E.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-2.096*</td>
<td>-2.213*</td>
<td>-2.923*</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.075)</td>
<td>(0.158)</td>
</tr>
<tr>
<td>Obs.</td>
<td>41,665</td>
<td>41,090</td>
<td>38,228</td>
</tr>
</tbody>
</table>

Note – ** : $p < 0.01$, * : $p < 0.05$. Each logit regression spans two columns. Includes manager-subordinate pairs in which both ultimate quota attainments are within 50% of quota. Exterior controls include 5% quota attainment bins (shown in the accompanying figure). Standard errors clustered by the manager-worker pair.
Figure 1. : Histogram of Quota Attainment, Salespeople and their Immediate Managers

a. Managers  

b. Subordinates

Note – Histograms represent counts of realized total quota attainment (at 5% intervals) at the end of annual measurement periods among salespeople and sales managers employed over the entirety of the measurement period. All panels corresponding to managers include only plans with orders that are credited to at least four subordinates, who themselves do not have subordinates, and whose subordinates’ sum of quotas are not within 5% of the manager’s quota. The quartiles of the sum of subordinates’ quotas are 50%, 100%, and 200% of the manager’s quota, and only 7% of the sum of the subordinates quotas are exactly the manager’s quota.
Figure 2: Expected Payoffs: Retaining and Terminating Low Ability Workers

Supervisor's Expected Payoff (Bonus B = 1)

- Replacing \( \alpha_r = 0, r_n = 1 \) workers is incentive-compatible.
- Replacing \( \alpha_r = 0, r_n = 1 \) workers is not incentive-compatible.
- Replacing \( \alpha_r = 0, r_n = 1 \) workers is incentive-compatible.

QL, Attainment Credited to Other Subordinates

Replace
- Retain
Figure 3: Turnover of Poor-Performing Subordinates at Managers’ Quota Thresholds

a. Turnover by Expected Attainment if Subordinate $i$ is Retained, $Q_i^H$

b. Turnover by Expected Attainment if Subordinate $i$ Turns Over, $Q_i^L$

Note – Quota attainment is discretized into 5%-wide bins. Whiskers represent 95% confidence intervals.
Figure 4. : Turnover of Poor-Performing Subordinates at Uncle-Position’s Quota Thresholds

Note – Among rank-and-file subordinates, I use “uncle” to denote a randomly-selected manager who reports to the subordinate’s parent’s parent (“grandparent”). Fewer than 1% of all subordinates have no “uncles.” Solid markers omit workers whose sales were predicted to be critical to meeting the parent position’s quota. Whiskers represent 95% intervals.
FIGURE 5. Predicted Probabilities a Subordinate’s Quota is Adjusted, by Manager’s and Subordinate’s Own Ultimate Quota Attainment

Managers

Pr(Sub Quota Adjusted),
Month 10

Subordinates

Pr(Sub Quota Adjusted),
Month 10

Pr(Sub Quota Adjusted),
Month 11

Pr(Sub Quota Adjusted),
Month 12

Parent's Expected Quota Attainment

Sub's Expected Quota Attainment

Note – Predicted probabilities by manager’s quota are at mean values for subordinate quota attainment bins, and predicted probabilities by the subordinate’s quota threshold are at the mean values for manager quota attainment bins. Quota attainment bins are post-adjustment. Whiskers represent standard errors.