Bonuses versus Commissions: A Field Study

Sunil Kishore
(kisho013@umn.edu)

Raghunath Singh Rao
(Raghunath.Rao@mccombs.utexas.edu)

Om Narasimhan
(o.narasimhan@lse.ac.uk)

George John
(johnx001@umn.edu)

January 2013

Forthcoming Journal of Marketing Research

Sunil Kishore works for McKinsey & Company. Raghunath Rao is Assistant Professor of Marketing, McCombs School of Business, University of Texas at Austin. Om Narasimhan is Professor, London School of Economics and George John is Professor, Carlson School of Management, University of Minnesota, and Distinguished Visiting Professor, KAU University, Jeddah, Saudi Arabia. We are grateful to the executives of the focal firm (who wish to remain anonymous) for providing us with the data and for very helpful discussions. The paper is based on the first author's doctoral dissertation at Univ. of Minnesota. The first two authors contributed equally and are listed alphabetically.

Bonuses versus Commissions: A Field Study

Abstract

Quota-based bonuses and quota-commissions are the two most common incentive compensation plans. We uncover differential effects of these plans from a natural field-based experiment featuring 14,000 monthly observations over 3 years from 458 sales territories of a pharmaceutical firm that was persuaded to switch from a bonus to an *equivalent* commission plan. Our intervention led to significant sales productivity improvement; this effect was heterogeneous across ability deciles with much larger increases occurring at lower ability deciles. We found stark differences across these plans on i) effort against non-incentivized tasks and ii) output fluctuations induced through "timing games". At this firm, the bonus plan was strictly inferior to the implemented commission plan with respect to short-term revenues and timing games. On the other hand, the commission plan induced greater neglect of non-incentivized tasks (not directly affecting observable output). We build a simple theoretical model in the personnel economics tradition to organize our findings. Our novel finding that the multi-tasking concerns are small under bonuses once the quota has been met provides a nuanced rationale for the widespread existence of lump-sum bonus plans.

(*Keywords*: Sales Compensation, Field Experiments, Timing Games, Multi-tasking)

INTRODUCTION

Firms have long recognized the importance of designing appropriate incentive plans to induce their salespeople to exert effort. An extensive theoretical literature in marketing and economics, centered on principal-agent models, attempts to understand how different incentive compensation schemes affect effort choices, with the aim of characterizing optimal incentive contracts (see Prendergast (1999) for a review of studies in economics and Albers and Mantrala (2008) for studies in marketing). Although the theoretical work yields a number of important insights into the design of incentive plans, there is surprisingly little empirical work on the efficacy of various incentive schemes actually employed by firms in practice.

In the real world, most incentive plans are invariably nonlinear and the variable component begins when salespeople meet a threshold that is set above the minimum output level. This threshold is commonly known as a “quota” in the sales management literature. A survey by Joseph and Kalwani (1998) of sales compensation practices at Fortune 500 companies found that 95% of the firms employed plans that featured a quota. Quotas have been rationalized as devices to account for salesperson territory heterogeneity (e.g., Zoltners et al. 2008). Additionally, psychological goal theory (see Locke and Latham (1991) for a comprehensive review) suggests that quotas provide salespeople with challenging objectives whose attainment offer psychological rewards (Jain 2010).

Two forms of variable pay are widely used in conjunction with quotas: i) lump-sums (bonuses) paid upon reaching quota, and ii) per unit payouts (commissions) paid on sales beyond the quota (see Figure 1). For ease of exposition, in what follows we refer to these as *bonus* and *commission* plans respectively, with the understanding that both involve (non-negative) quotas. Intuitively, a bonus plan provides a strong incentive for a salesperson to reach the quota, but it also encourages one to minimize one's efforts thereafter, whereas a commission plan provides an incentive to keep working hard even after the quota has been reached. Oyer (1998) and Jensen (2003) formalize these insights with models that show bonuses tempt salespeople to manipulate the timing of orders (e.g., delayed selling, forward selling, etc.). As such, it appears that bonus plans are perhaps strictly inferior to linear commissions if the output is the sole criterion to judge the efficacy of a plan (see Oyer 2000 for an exception). Notwithstanding theoretical conjectures about the relative merits of these plans, the empirical literature is sparse and inconclusive. The empirical issue that has received some systematic attention in this area is the issue of deliberate postponement or advancement of sales by salespersons (referred to as 'timing games') (Oyer 1998; Steenburgh 2008; Misra and Nair 2012).

The "multi-tasking" distortion associated with high-powered incentives (Holmstrom and Milgrom 1991) has largely been ignored by this empirical literature. We remarked earlier that commission plans appear to provide stronger incentives than bonus plans on measurable tasks used in the calculation of incentive pay. As such, if sales people are assigned a portfolio of tasks, some of which are not explicitly included in the computation of incentive pay (either because of the unavailability of good outcome measures or due to ease of "shading" of these measures by agents), commission plans should induce agents to neglect these other tasks to a greater degree than would bonus plans.

We employ data from a large firm-level field experiment to study the impact on productivity as well as the impact upon output fluctuations and multi-tasking concerns of the focal firm. The experiment involved switching the firm's sales compensation from bonuses to linear commissions. From a methodological viewpoint, there is increasing attention to field experiments to study contracting (e.g., Paarsch and Shearer 2000; Bandiera et al. 2005) as they alleviate endogeneity concerns. A small literature has also recently tried to estimate the structural parameters of sales contracts from observational data (Chung et al. 2010; Misra and Nair 2012). None of these studies contrast bonuses with commissions, which is the objective of our paper. More specifically, we designed a quasi-experiment to address the following issues:

1. *What are the productivity consequences of bonus plans compared to commissions?* We are interested in the overall effect as well as differences in impact across agents.
2. *How do timing games manifest themselves under bonuses and commissions?* Sales fluctuations induced by incentive plans can be inefficient for organizations (Jensen 2003) and our study sheds light upon sales timing behavior.
3. *Under which plan (bonuses or commissions) are multi-tasking concerns more severe?* While the neglect of non-incentivized tasks under high-powered incentives is well-known conceptually, we know little about them empirically across bonus and commission plans.

Preview of our findings: (1) The commission plan increases sales force productivity by around 24% over the bonus plan. Specifically, the switch has *heterogeneous impact*: It leads to higher improvements at lower ability deciles. (2) Salespeople "push sales into the future" if they are unlikely to meet quota and "pull sales in from the future" if they are near quota. These patterns are more pronounced under the bonus plan. (3) Once a quota has been met, the bonus scheme is much better than commissions in terms of engaging the salespeople to undertake tasks that are not directly compensated, but are important to the firm.

An immediate issue that arises from single-firm field studies such as the one proposed here relates to generalizability. Consider each of the three questions above in turn. The generalizability of our productivity predictions (question 1) turns on defining plan *equivalence.*

Suppose the quota set by a firm is very hard to reach under bonuses. Now, if the firm were to switch to a commission plan where the quota was set at a relatively low level of sales, it is intuitively obvious that the commission plan would likely improve productivity, but this would not generalize to other commission (or bonus) plans. As such, we compare bonus and commission schemes that are “equivalent” (a term we define precisely very shortly). It is important to note that our comparison does not require either plan to be “optimal” - all we require are data on two equivalent plans. The case for generalizability is even stronger for our timing games predictions (question 2) and multi-tasking distortion predictions (question 3). In neither case do we require plans to be either optimal or equivalent. To give an example, salespeople can distort the sale in the final month of a quarter through “pulling in sales” from the future or “pushing out sales” into the future, and we predict that once the quota is met, the former (latter) is more likely to occur under commissions (bonuses). This result is quite general and does not rely upon the process that went into the firm’s incentive design or whether the contracts are equivalent. Taken together, these theoretical elements enable generalizability even with data from a single firm design. That said, we reiterate that a single firm experiment permits detailed data collection and controlled implementation, albeit at the cost of context generalizability.

Turning to the implementation of the study, we successfully urged a division of a large pharmaceutical firm in an emerging economy to change their extant bonus plan to a commission plan. As a result, we have a before-after quasi-experimental design; that is, the change in scheme was exogenous. We exploit this exogenous change to identify the unique effects of across the two plans.

The rest of the paper is organized as follows. In the next section we describe our empirical context followed by a stylized theoretical model inspired by our empirical setting. Next, we report empirical findings and conclude the paper with a discussion of implications of our findings. Details of the theory model are contained in the self-contained Web Appendix A. Web Appendix B is a detailed discussion of data and identification issues and robustness checks.

EMPIRICAL SETTING

We obtained detailed territory-level data from a large pharmaceutical firm in an emerging market. This firm sells drugs in multiple therapeutic classes and employs a dedicated sales force for each class. We worked with the sales force for one of these therapeutic classes. Our data consist of observations about sales groups of varying sizes assigned to territories, with unique quarterly quotas/targets for each territory[1]. This firm was actively examining changes to their extant incentive pay plan for salespeople, which consisted of lump sum bonuses paid when the territory sales reached the quota assigned at the beginning of that quarter. We worked with the firm in setting up a commission scheme taking into consideration our theoretical framework.

Institutional Details

There are three sets of actors in this market: salespersons, pharmacies and distributors. Consider each in turn. The salespeople comprising a sales group are responsible for a specific territory. The number of salespeople in a sales group is determined by the market potential of the territory, which is roughly a function of the number of doctors and pharmacies operating in the territory. Each salesperson is assigned a specific list of doctors and pharmacies. The salesperson calls on doctors to encourage them to prescribe the firm's drugs (referred to as

[1] To preserve the anonymity of the company that provided us with the personnel records, we have scaled all the time series data by a *single common positive number*. All monetary values in the main body of paper and the tables are in '00,000s in local currency except for the industry sales, which are in '00,00,000s in local currency.

“doctor visits”), and pharmacies to encourage them to buy and stock the firm’s drugs (“pharmacy visits”). However, sales cannot be parsed unambiguously; when a customer purchases a drug from a pharmacy, the firm cannot trace it back to a particular doctor (and thus credit a specific salesperson). Notice that this is not an issue in single salesperson territories because all sales in the territory are credited to that salesperson. Each salesperson reports the number of doctor and pharmacy visits made daily in a call report system. Managers conduct random post-report checks by calling or visiting doctors and pharmacies mentioned in the salesperson’s call report.

Pharmacies are typically single-owner establishments with a few employees. The owner (or an employee) is required to have a pharmacy certification. There are multiple pharmacies within each territory, and they are quite competitive with each other. They stock a limited volume of drugs and rely on frequent deliveries from distributors. Distributors are mid-sized trading firms who buy the drugs from the firm and are responsible for ensuring that drugs are delivered to the pharmacies. There are generally multiple distributors within each territory.

Incentive compensation is computed by crediting all sales in a territory equally. All salespeople in the territory get equal credit; thus, the incentive compensation is split equally. Salespersons in a territory routinely meet up and exchange information on detailing and sales achievements. It is crucial at this point to emphasize that even with multiple salespeople working in a single territory, there is no team selling as such. While salespeople do meet and compare notes routinely, they do not make joint calls, or otherwise complement each other. We expected concerns about free-riding in multi-person territories given the firm’s equal credit allocation for sales, but the managers we interviewed told us that free riding was not a major problem. Our interviews with salespeople suggest a possible reason. The salespeople in a territory meet routinely for office conferences. Presumably, the peer monitoring that occurs in these meetings dampens free riding as has been shown previously (e.g., Knez and Simester 2001) - we discuss this issue further when discussing our empirical results.

Doctor visits: Each salesperson visits the doctors assigned to him. These visits are a classic example of a "pull" marketing tactic. During a typical visit (that lasts anywhere from 5 to 20 minutes), the salesperson highlights the effectiveness of the drugs, distributes technical literature and provides free samples. Another way to think about these doctor visits is to consider them akin to *informative* selling, wherein provision of information is used as a tool to encourage doctors to prescribe the drugs. While these visits undoubtedly build preference and loyalty for the company's products, they do not translate into immediate sales because of two crucial institutional features. First, doctors in this country may legally write just the drug's active ingredient on the prescription. Typically, there are multiple equivalent brands because of limited patent protection[2]. Second, e*ven when a doctor prescribes a particular brand, a pharmacist may legally substitute a different brand, as long as it is chemically equivalent*. Despite these factors, the importance of the doctor remains; patients are rarely assertive and generally defer to doctors – the reasons are partly cultural, in that doctors are high status individuals, and partly institutional in that direct-to-customer drug advertising is not permitted by law.

Pharmacists tend to stock a limited number of brands from the plethora of options available and hence in practice, when a patient walks into a pharmacy, she is likely to buy from among the brands that are in stock at the pharmacy. Also, while most instructions on the package are

[2] In this market, the legal system allows only process patents and not product patents. As soon as a new drug is launched, a number of equivalent products manufactured using non-infringing processes get launched fairly quickly.

in English, the large majority of patients are not fluent in English, giving the pharmacist considerable influence.

If a doctor has a high preference for a particular brand (as a result of salesperson provided information, among other things) and conveys that to patients, patients might insist on getting that particular brand at the retail level. This also results in pharmacies being more likely to stock these brands. In short, while marketing efforts made at the doctor level may not translate into immediate sales at the retail level, they do have important consequences for the firm. Given this importance of doctor visits, but absent a clear output measure, the firm enforces a minimum level of self-reported doctor visits per quarter and imposes a penalty if this level is not met.

Pharmacy visits: Pharmacy visits are a “push” marketing tool for this firm. A typical pharmacy visit is longer than a doctor visit (25 to 45 minutes). During such a visit, the salesperson and the pharmacy owner discuss sales, stock levels, future orders, credit and other trade terms. Salespeople often also obtain competitive intelligence and pricing information at the pharmacy as pharmacists interact with salespeople from multiple manufacturers. A pharmacy visit is less formal than a doctor visit; among other things, the timing and volume of a booked sale is often influenced by a salesperson’s ties with a pharmacy owner. One can also think of these efforts as *persuasive* selling because convincing a pharmacy to stock more of the firm’s products results in booking a sale that immediately gets credited to the salesperson and impacts his compensation. Salespeople have some pricing authority within a band and can negotiate terms of trade, such as the credit period, billing cycle, samples, etc. This has important implications for the “timing games” discussed later.

Market Conditions

The overall prescription drug market has grown at double digit rates in the past decade as access to health care in the country has grown rapidly. While primary demand is influenced by many factors, the medical reps are the principal marketing resource deployed by drug firms. Our focal category of products treats a specific non-chronic ailment whose demand at the patient level is subject to seasonal variation. Patients do not stockpile these drugs since they are prescribed for a particular period (typically 2-3 weeks). Stockpiling by pharmacies is limited because prices are relatively stable and products come with an expiration date. We do not observe wholesale prices, but our understanding is that they were fairly stable during the period of our analysis. Two additional facts about prices are important: i) firms do not seem to compete on price – prices across competing brands are fairly similar; and ii) no firm is large enough to influence prices at the market level – within this category, there were at least 15 firms offering products, and market shares rarely exceeded 10% for any firm.

Given the facts above, it appears that a crucial driver of demand at the individual firm level is sales force management - both the intensity of selling efforts and the manner in which salespeople engage in *informative* (i.e., doctor visits) and *persuasive* (i.e., pharmacy visits) aspects of the selling process. As for overall market conditions, the number of pharmacies that salespersons visit in each territory stayed stable over the observation period. Finally, even though the overall market was growing, the size of sales groups within the period of our analysis remained largely stable (See Figure B1 in Web Appendix).

The Experiment Design

The focal firm employed an incentive plan wherein each salesperson received a *monthly* salary and a *quarterly* lump-sum bonus upon achievement of quota. After discussing the pros and cons of different incentive plans, we persuaded the firm to shift to a commission plan where each person received a fixed monthly salary and earned commissions at a fixed rate on any sales exceeding commission quota. This new plan was devised following a two-step procedure.

First, we asked the firm to first set quotas for their extant (bonus) plan based on their existing quota-setting process (see details of quota setting process in Web Appendix B) for the quarter at the start of the intervention. Next, using these quotas as a benchmark, we asked the firm to revise the quotas for a commission plan such that a sales group *achieving the quota from the first step would earn the same amount of incentive pay under the new plan*. This two-step procedure ensured "equivalence" of the old and new plans, a concept we define more precisely in the next section. The revised quotas under the commission plan were *lower* than the quotas under the bonus plan set in the first step (as is required by our theoretical setup). Parenthetically, there are many possible combinations of quota and commission rate that would generate an "equivalent" treatment plan (see Fig. 1 for illustration of several possible commission plans both equivalent and non-equivalent) and we left it to the firm to come up with their preferred combination of quota and commission rate[3,4].

In addition, we designed the experiment so as to minimize experimental artifacts. In particular, we were concerned about Hawthorne effects and differential attrition problems that are common in field settings. To manage these issues, we (and encouraged by the firm), eschewed a treatment-control groups design because such designs highlight the contrast

---

[3] The final choice of the commission scheme from the many possibilities was dictated by the existing quota setting process within the firm and the expected incentive payout, among other things.

[4] An "equivalent" scheme was operationalized as follows: In the pre-intervention period, the quarterly lump sum bonus paid in each territory was about 2.5% of each bonus sales quota. For the first quarter of the post regime, the firm calculated the new bonus sales quota (Q_{iBN}). At this stage, we provided several arithmetic examples of equivalent commission schemes, and quizzed the managers to ensure that they understood equivalence unambiguously. For the post regime, the firm's managers first settled on a commission rate of 8.33% (for sales above the commission quota) for every territory. Then, to maintain equivalence with the pre regime, they set the new quotas as: Q_{iBN}*(2.5%) = (8.33%)(Q_{iBN} -Q_{icN}), which yields Q_{icN}=0.7(Q_{icB}). In subsequent quarters, these quotas were updated following its established procedures (see Web Appendix B).

between the treated and control groups. To manage attrition artifacts, we persuaded the firm to launch the new plan in exactly the same fashion as the firm normally does in making changes to compensation. As we document later, our efforts appeared to be successful, with no measurable changes in quit rates[5]. The availability of repeated observations of sales pre and post, along with other covariates, enables us to uncover the mechanics of the two schemes.

THEORETICAL FRAMEWORK

In this section, we develop a theoretical framework that allows us to gain a better understanding of the subsequent empirical results. The models are deliberately kept simple with an eye on our empirical context described earlier. Importantly, the aim here is not *to derive an optimal sales compensation scheme*. The goal is to illuminate how the use of bonuses and commissions might involve trade-offs between productivity, multi-tasking concerns, and distortions induced by timing games.

Incentive Pay and Productivity

To illustrate the impact of incentive schemes (bonus vs. commissions) on productivity, consider the following stylized model of rational salespeople with differential abilities. For salesperson *i* exerting effort e_i , sales are given by: $y_i=\psi_i e_i$, where $\psi_i>0$ and $e_i \geq 0$, where ψ_i represents the innate "ability" of salesperson *i* that results in higher output for the same level of effort for a rep with higher ability compared to one with lower ability. We assume it to be distributed continuously with M~ $[\psi_{min}, \psi_{max}]$ with a higher ψ indicating a more productive salesperson. The firm knows the ability distribution but not the exact ability of each salesperson. We assume throughout that M is strictly increasing with a density m. Our approach is in line with recent

[5] We do not wish to suggest that before-after design is superior to randomized control/treatment approach (RCT). However unlike in a classic RCT, in our context it was not possible to isolate the treatment and control groups.

work in “Personnel Economics” (e.g., Lazear 1995, 2000) that has de-emphasized the effort-insurance trade-off central to agency models so as to focus more sharply upon the role of incentives and salesperson heterogeneity.

The utility of a salesperson with ability ψ_i when she earns a wage F and puts in effort e is given by: $u(\psi_i;e_i) = F-(\Theta e_i^2)/2$, where , $\Theta>0$, represents the cost (of effort) parameter assumed to be common across salespersons[6]. A quota-based bonus scheme is given as:

$$F_B = \begin{cases} W & \text{if } 0 \le y_i < Q_B \\ W + B & \text{if } \; y_i \ge Q_B \end{cases} , \tag{1}$$

where $W > 0$ is the fixed salary irrespective of the output and the salesperson receives a lump sum bonus $B > 0$ upon reaching or exceeding quota threshold Q_B. Plainly, an agent will either pick effort 0 (or some positive level of minimum enforced effort) or earn W, or pick just enough effort to earn W+B. Because of the heterogeneity in abilities, salespersons would require different levels of efforts to reach the bonus quota threshold given by:

$$e_{iB}=Q_B/\psi_i. \tag{2}$$

As can be seen from (2), salespeople with higher ability require a smaller level of effort to reach the bonus quota. Hence it can be readily verified that that the utility associated with earning a bonus is decreasing with the decrease in ability. Further, with sufficient heterogeneity in salesperson types, there exists a critical ability threshold above which the salesperson will work more than the minimum required effort and earn the bonus while the rest will simply earn the fixed wage. This threshold level of ability, denoted by ψ_B, can be derived as:

[6] The heterogeneity in salesperson abilities could be equivalently modeled using a heterogeneous cost parameter, with higher ability represented by a lower cost of effort (Θ).

$$W + B - \frac{\theta(Q_B/\psi_B)^2}{2} = W \text{. This yields: } \psi_B^{\ 2} = \frac{\theta Q_B^{\ 2}}{2B}. \quad (3)$$

Assuming that parameters B and Q_B are chosen in such a way that ψ_B lies in the interior of the ability parameter distribution (that is, at least some people get the bonus), salespeople with abilities $(\psi_B, \psi_{max}]$ respond with more than minimum effort and earn a bonus while salespeople with abilities $[\psi_{min}, \psi_B]$ put in the minimum effort and get paid a fixed wage.

Similarly, a quota-based commission scheme with commission rate $0<\alpha<1$ is given by[7]:

$$F_C = \begin{cases} W & \text{if } 0 \leq y_i < Q_C \\ W + \alpha(y_i - Q_C) & \text{if } \ y_i \geq Q_C \end{cases}. \quad (4)$$

As before, we can derive the threshold ability of the salesperson who is indifferent between earning a fixed wage and a positive level of commission as: $\psi_c = (\theta e_c^{2} + 2\alpha Q_c)/(2\alpha e_c)$. (5)

A salesperson who (optimally) exerts more than minimum effort solves:

$$\underset{e_i}{\arg\max}\ \alpha(y_i - Q_C) - (\frac{\theta e_i^{\ 2}}{2}) \text{, which yields the optimal effort level: } e_i^{\ *} = \frac{\alpha \psi_i}{\theta}. \quad (6)$$

Equation (6) shows that the equilibrium effort of a salesperson who decides put more than minimum effort level is increasing in ability. Substituting (6) into (5) yields:

$\psi_c^{\ 2} = (2\theta Q_c)/\alpha$ (7)

As detailed before, we define a bonus scheme as "equivalent" to a commission scheme if, for an output Q_B, both schemes result in the same pay. Hence, the condition for the equivalency of two schemes is given by: $W + B = W + \alpha\ (Q_B - Q_c)$, or $B = \alpha\ (Q_B - Q_c)$. (8)

Using (8) in conjunction with the other expressions, we obtain our first key result:

Result 1: *Under equivalent bonus and commission schemes, the salesperson who is indifferent between earning a positive level of commission and a fixed salary is of lower ability*

[7] There could be instances of real incentive contracts where α is more than 1.

compared to the salesperson who is indifferent between earning a bonus and a fixed salary. In other words, $\psi_c < \psi_B$.

Proof: See Web Appendix A.

This simple result highlights one source of productivity differences across two equivalent schemes. A positive mass of salespeople who would otherwise engage in minimum effort and earn a fixed salary under the bonus scheme would end up earning commissions and putting in higher effort. Also, at the higher end of the spectrum, the highest ability salespeople are no longer constrained by a fixed bonus and tend to invest in higher effort and make more money than they would otherwise make under an equivalent bonus scheme. To formally see this, we can write the effort as a function of ability under the two schemes as below:

$$e_{iB} = \begin{cases} 0 & if\ \psi_i < \psi_B \\ Q_B / \psi_i & if\ \ \psi_i \geq \psi_B \end{cases} \qquad e_{iC} = \begin{cases} 0 & if\ \psi_i < \psi_C \\ \dfrac{\alpha \psi_i}{\theta} & if\ \ \psi_i \geq \psi_C \end{cases} \qquad \text{(9 a \&b)}$$

Two features are worth noting in the equation (9): First, the participation of agents who respond to incentives has gone up under the commissions regime, as evidenced by $\psi_c < \psi_c$ (Result 1). Second, of the agents who respond to incentives under bonuses, the effort is *inversely* related to ability, i.e., higher types *need to incur less effort* than lower types to earn a bonus. The opposite holds under commissions: The level of efforts of agents who earn commissions is *positively* related to ability; higher types *put in more effort* and earn higher commissions than lower types.

Comparing across regimes, the effect on the middle ability types who were already earning a bonus is somewhat ambiguous - for example, see Fig. 2a (left panel) where there exists a middle segment that puts in less effort than under commissions while in Fig. 2b (left panel), all the ability types put (weakly) more effort under commissions than under bonuses. But what is not ambiguous is that the increase in effort level of people who were not previously earning bonuses but who now earn commissions is significant.

While the firm cannot directly observe the effort levels of the salespersons, the incentive plan it employs impact the efforts and the effort can be *inferred* from the contract and the distribution of abilities. Thus the firm can impact the output through the appropriate design of incentive pay. The rational salesperson response to the equivalent bonus and commissions contracts can be written as the following output equation:

$$y_{iB} = \begin{cases} 0 & if\ \psi_i < \psi_B \\ Q_B & if\ \psi_i \geq \psi_B \end{cases} \qquad y_{iC} = \begin{cases} 0 & if\ \psi_i < \psi_C \\ \dfrac{\alpha \psi_i^2}{\theta} & if\ \psi_i \geq \psi_C \end{cases} \qquad 10(a\ \&\ b)$$

This output function is plotted in Figure 2 ((a) and (b), right panels). It shows that while output remains constant under the bonuses at higher ability levels, it increases at an increasing rate under commissions. Again, the most significant change is for the segment at the lower end wherein some salespersons who were just putting minimal effort (under bonuses) now respond to incentives and exhibit significant productivity improvement. The middle types' productivity change is somewhat ambiguous, while productivity goes up for the higher types. Summing up, *we would expect to see the biggest change in productivity at the lower ability levels.*

We wish to re-emphasize that this result is *not* meant to convey that commissions are strictly superior to bonuses in terms of productivity. We highlight two sources of possible productivity gains under commissions when two schemes are equivalent (even if sub-optimal): low ability agents putting in more than minimum effort for a chance to earn non-zero commission and high ability agents putting in higher effort as there is no earning ceiling.

Multi-tasking and Timing Games

The model above included one task and a deterministic relationship between effort and output. However, as we described earlier, this firm's salespeople mainly undertake two tasks that are not equally influential in determining immediate sales (which is the basis for determining incentive pay in both of the pay plans). Furthermore, they could pull in sales and/or push out sales. We seek to understand these issues with a combination of a standard multi-task model (Holmstrom and Milgrom 1991) and a simple timing game model (Oyer 1995).

Consider a salesperson engaged in two activities over two quarters, each consisting of two periods[8]. To match up this model to our empirical setting, assume that the firm asks a salesperson to engage in two activities, denoted "doctor visits" and "pharmacy visits" respectively. Furthermore, in the second period of the first quarter, the salesperson can possibly "game" the sales below or above the "natural" level of sales; this has consequences for sales in the third-period (the first month of the second quarter). The salesperson is paid on total observable output in a quarter. In any given period t (month), observed output (sans gaming) is given by: $y_t=h_dT_{dt}+h_pT_{pt}+\varphi$, where T_{dt} and T_{pt} denote the time spent on doctor visits and pharmacy visits respectively in period t and h_d and h_p are the positive marginal products of these activities on observed output. The time spent proxies effort substitution across these activities and φ represents random noise distributed symmetrically around mean zero. The value added to the firm arising from the time spent on these activities is given by: $v_t = f_dT_{dt} + f_pT_{pt} + \epsilon$, where f_d and f_p are the positive marginal products of these activities on firm value and ϵ represents mean zero random noise with properties similar to φ as described above.

[8] Plainly, real-world quarters consist of three months, but we assume two months per quarter to simplify the exposition of our model while still capturing the sales variation that is present within the months of a quarter.

Furthermore, $f_d > f_p$ and $h_d > h_p$, setting up a classic multi-tasking problem wherein the activities that enhance the firm's welfare are valued less by the salesperson because of their smaller impact on observable output. The salesperson cannot be compensated directly on v_t because it is virtually impossible to quantify an individual agent's contribution to the value of the firm. The total time spent on the two activities, T, is fixed in a given period t (salespeople are assumed to work a certain fixed number of hours in a month), and $T=T_{dt}+T_{pt}$ with the minimum number of doctor visits in any period fixed at $T_{d\,min}$. In the bonus plan, in quarter 1, the salesperson is compensated $W+B$ if $y_1+y_2 \geq Q_B$, otherwise he gets W. Similarly, in quarter 2, if $y_3+y_4 \geq Q_B$, he makes $W+B$, else W[9]. In the commissions plan, the salesperson makes $W+\alpha\,(y_1+y_2-Q_c)$ if $y_1+y_2 \geq Q_c$ with $0<\alpha<1$, otherwise he gets W. The same scheme applies in quarter 2. Equivalence requires $\alpha\,(Q_B-Q_C)=B$ and trivially, $Q_B \geq Q_C$.

A well-established literature has shown that salespersons often game output in response to high-powered incentives (Jensen 2003; Steenburgh 2008; Misra and Nair 2011). Following Oyer (1995), we model this by allowing that in period 2 of quarter 1, the salesperson could potentially *game* the system by either pulling in sales from period 3 or pushing out sales to period 3. Both these activities are costly (to the firm). We model this gaming through a variable $0<\lambda<2$, where $\lambda=1$ implies a "natural" level of sales while $\lambda>1$ indicates pull-in and $\lambda<1$ indicates push-out. Finally, we assume that when a salesperson is indifferent between carrying out either doctor or pharmacy visits, he focuses upon doctor visits. A summary of the timeline of this game is given in Figure 3. The critical variable that impacts a salesperson's gaming

[9] We are abstracting away from so-called "ratcheting" effects by assuming a constant quota over time (Gibbons 1987). Admittedly, this is not an innocuous assumption but we have both institutional and empirical evidence to suggest that ratcheting is unlikely to be an important strategic consideration among salespersons in our setting. Please see the Discussion section and Web Appendix B for details.

strategy and activity substitution is the realized sales y_1 that translates into the state variable S_{Q1B} (distance to quota).

Given this framework, the observed output and the contribution to firm value by the salesperson in each of the four periods can be written mathematically. These derivations have been relegated to Web Appendix A and key mathematical expressions are given in Table 2. It is worth emphasizing that in this formulation, we are seeking to understand salesperson behavior under a *given* bonus scheme and an equivalent commission scheme while abstracting away from the firm's problem.

We use the salesperson's problem laid out in Table 2 to gain insights into the choices made in the last month of a quarter, and the impact of these choices in the first month of the subsequent quarter, under bonus and commissions regimes respectively. We collect two key empirically testable results below –Web Appendix A details an example of the algebra involved in deriving these results. These two results show that when the quota is too far (and hence unlikely to be reached), salespeople exhibit similar behavior across bonus and commission regimes; by contrast, when the quota has been met, salesperson behavior is dramatically different across the two regimes in terms of timing games and multi-tasking distortions.

Result 2: *If at the end of period 1, the quota for bonus is too far (we label this condition "FAR" in our empirical analysis), then:*

- *a) The salesperson will push sales out to period 3.*
- *b) The focus will shift (weakly) to doctor visits resulting in attenuation of multi-tasking concerns.*

If at the end of period 1, the quota for commissions is too far, then:

- *c) The salesperson will push sales out to period 3.*
- *d) The focus will shift (weakly) to doctor visits resulting in attenuation of multi-tasking concerns.*

Result 3: *If at the end of period 1, the quota for bonus has been achieved (we label this "EXCEEDED" in the empirical analysis) then*

- *a) The salesperson will push sales out to period 3.*

b) The salesperson will focus largely on doctor visits resulting in the attenuation of multi-tasking concerns.

If the quota for commissions has been achieved then

c) The salesperson will pull sales in from period 3.

d) The salesperson will focus largely on pharmacy visits while keeping doctor visits at the minimum level, resulting in the amplification of multi-tasking concerns[10].

Intuition: The intuition behind result 2 is simple - there is very little chance that the salesperson will reach the quota in either regime, so the marginal return to pull-in almost zero. Hence, he will push sales out to improve his chances of reaching/exceeding the quota in the next quarter. Since doctor visits dampen the current output but returns to the current output decrease are non-negative, we expect salespersons to weakly focus on doctor visits. In result 3, one can see the clearest distinction between the effects of the two schemes. Under bonuses, the salesperson has no incentive to improve output once he reaches the quota, and in fact, has every incentive to push sales out to period 3 to maximize the probability of reaching quota in the next period as well. On the other hand, the salesperson will readily focus on doctor visits even though pharmacy visits are more impactful in the realization of observable output, because the marginal return to increased output is zero after reaching the quota. Under the commission scheme, the marginal return to increased output is positive and hence the salesperson will largely focus upon pharmacy visits to maximize current sales. He will also pull sales in from period 3 because he has uncertainty regarding reaching the quota in the next period and having reached the quota in this period, he wants to sell as much as possible. To recap*, a realistically achievable bonus plan tends to amplify distortions induced by timing games whereas an equivalent commissions plan exacerbates multi-tasking concerns*.

[10] The terms “near”, “too far” and “far” (stretch) are defined more precisely in the Technical Appendix with reference to the model. The corresponding empirical measures are detailed in the section titled Empirical Analysis.

EMPIRICAL ANALYSIS

Data Description

Our data include an almost complete history of incentive plans and payments, monthly output and quarterly quotas for 458 territories for a period of 3 years, from April 2007 (the start of the financial year) to March 2010. We also have average daily doctor and pharmacy visits at the monthly level. The intervention in October 2008 marks the beginning of the third quarter of the 2008–2009 financial year. The pre-intervention period of 18 months (under the bonus plan) starts in April 2007 and ends September 2008, and the post-intervention period of 18 months for these same territories (under the commission plan) starts in October 2008 and ends March 2010. In all, the data include more than 14,000 monthly observations on both output and visits.

The descriptive statistics reported in Tables 3 and 4 show large differences in quotas across territories; plainly there are significant differences in sales potential across territories. Figure 4 plots total monthly sales. Visually, revenues appear to have increased post intervention. Substantial seasonal variation is also evident. Table 3 presents summary statistics for revenues, productivity (defined as revenues per salesperson per quarter), and incentive pay earned by the compensation plan. Consistent with Figure 4, the numbers in Table 4 indicate that revenues and productivity are higher in the intervention period. Figure 5 presents a plot of industry sales over time during the observation period, while Figure 6 shows total monthly sales of the firm including the 5^{th} and 95^{th} percentile territories.

Linking Theory to Empirical Application

The suitability of our empirical context and design to the task of assessing the validity of our theory results is elaborated upon here. Consider the results in turn.

Result 1 on productivity relies upon two key elements of the model: i) Heterogeneity in the innate ability of the salespeople, and ii) equivalence of the two schemes. Since we have already described the operationalization of the equivalent scheme, we focus on (i) here. It is heterogeneity combined with chronic/persistent ability that yields the outcome wherein the lower types respond to incentives under commissions ($\psi_c < \psi_B$). Does our context yield data with sufficient heterogeneity in ability but also persistence in this ability over time, specifically across bonus and equivalent commissions? Given that we observe the productivity of the each sales group across multiple quarters under bonuses and equivalent commissions, we can readily test this identification condition. Using the productivity equation (11), we estimate the fixed effects associated with each sales group separately for bonuses and commissions (by running different regressions for each regime). Figure 7 plots these fixed effects; plainly, we have heterogeneity in ability estimates, but ability persistence across the regimes as well (the correlation coefficient between the fixed effects is 0.742).

Furthermore, while Result 1's prediction of salespersons responding more to incentives under commissions is supported by the quota attainment numbers (under the bonus regime, only 14% of salespeople met their quarterly quota, while 60% did so for the commission regime), our design also generates additional evidence for the mechanism. While we do not have direct measures of effort and ability, the efforts can be linked to output through equation (10) and as we just showed, we can *infer* the ability of salespeople through repeated observations. Thus, the direction of the productivity change at different ability levels across bonuses and commissions in our experimental set-up allows us to assess the change in effort, which is accomplished via quantile regressions in the next section.

Our multi-tasking results hinge upon the institutional fact that pharmacy (doctor) visits do (not) translate into immediate sales gain. We arrived at this crucial working assumption through extensive conversations with the management as well as our perusal of field visits. We are also able to test the empirical validity of this crucial assumption by running a regression of monthly sales against pharmacy and doctor visits and other controls (Table B4 in Web Appendix B). Indeed, we find the coefficient of pharmacy visits to be positive (0.0603, p<0.01) while the coefficient of doctor visits is negative (-0.007, p=0.659) but statistically not significant[11]. This accords with our two task multi-tasking model outlined earlier, and discloses the appropriateness of our empirical setting to the theoretical insights laid out in Results 2 and 3.

Intervention Effect on Productivity

Figure 8 plots the kernel density of quarterly productivity (defined as unit sales per salesperson). From this figure, it is clear that the commission distribution lies to the right of the bonus distribution. Furthermore, the peak value of the density function in the bonus plan is lower than that of the commission plan. There is also more concentration of productivity around the modal value under commissions than under bonuses. Table 4 describes mean pre and post intervention productivity levels as 26.23 and 28.87 respectively, which provides directional evidence that the intervention increased productivity.

We estimate the following model to analyze the intervention effect more formally:

$$\log(y_{it}) = \alpha_0 + \sum_i \alpha_i S_i + \sum_t \gamma_t YM_t + \beta NewPlan_t + \delta_1 \log(QtrTarget_{it}) + \delta_2 GroupSize_{it} + \varepsilon_{it} \quad (11)$$

where, y_{it}: Revenue per salesperson generated in territory *i* in period *t* (e.g., July 2008).

[11] If the total number of visits are fixed (as in the theory model), then one might question how the effect of doctor and pharmacy visits are separately identified. The identification comes from the fact that we do observe some variation in total visits across groups as well as over time.

S_i : Dummy for territory *i* that controls for the unobserved "ability"/territory effect of the sales group.

YM_t : Year-month dummy for period *t* (e.g., Jan 2008). This allows for month dummies to be estimated off the within-month variation across the 458 territories and provides a stringent control for unobserved differences in demand/competition/business-environment between the before and after periods.

$NewPlan_t$: Intervention dummy set to 0 for pre-intervention periods.

$QtrTarget_{it}$: Quarterly quota for territory *i* for period *t* that provides observed control for the differences across territory potential over time.

$Groupsize_{it}$: Number of salespeople in territory *i* in period *t* provides another observed control for the territory effects over time. The disturbance term ε_{it} is clustered at the territory level.

Identification of productivity estimate. The main identification assumption for the productivity estimate by plan is the exogeneity of the intervention (the compensation switch was designed as a quasi-experiment). Since we do not have a control group, the use of month-year and territory dummies along with the measures of quarterly quota and sales group size is useful in controlling for unobserved and observed territory, time and other market level heterogeneity. A few issues are worth noting. Quarterly quota could potentially be endogenous and while the covariates and unobservable heterogeneity control alleviate this concern somewhat, we also estimate our productivity specification without the inclusion of the quota variable (see Table 5). This also alleviates the concern that the quota systematically varies with scheme change and its inclusion might create biased estimates of the new plan[12].

[12] The impact of the quota might be different across the two schemes and an estimation of a specification of (11) that includes an interaction term of new plan dummy and quarterly quota gives very similar results.

Selection and quits and group-size variation. We have interpreted the territory dummy as being a control for a sales group's underlying ability and territory potential. This interpretation is problematic if the team composition is changing through quits/ hires (see further details in Web Appendix B). We account for turnover and re-ran the entire analysis, by constructing fixed effects at the territory-team level rather than just at the territory level. For example, if a sales group had 12 members during the first 14 months, 10 members during the next 6 months (due to quits) and 13 members during the last 16 months (due to 3 new members joining), then each of these three territory-team pairs were given three distinct IDs and were estimated as distinct fixed effects in the analysis. In Table 5 we present the results of all the specifications of (11) using these new territory-team controls as well. The results are fairly similar across the specification with the territory fixed effects (Columns 1, 2, 3) and team-territory fixed effects (Columns 4, 5, 6). We also re-ran the analyses by dropping all the territories that experienced any turnover - the results are similar, and are presented in Web Appendix B along with other robustness results.

Column 3 of Table 5 shows the estimate of our intervention effect on productivity. Evaluated at the mean, an average territory is 23.9% more productive after the intervention. The new plan productivity estimate, when accounting for the turnover, shows a very similar productivity improvement of 24.9% (see Column 6 of Table 5). Since these are the net estimated effects after controlling for calendar month-year and unobserved territory differences, it is reasonable to attribute these to the increased effort by salespeople. (The coefficients of all but one period dummy are significant, indicating a presence of seasonality as well as a time trend). The non-inclusion of the quarterly target results in a higher value of this estimate (see columns 2 and 5), while non-inclusion of the month-year dummies results in a lower value this estimate (see columns 1 and 4). Overall, based on Table 5 with different specifications, the most conservative estimate is an approximately 22% productivity improvement as a result of the switch.

Heterogeneous Intervention Effect

Recall that we predicted differential responses to the plan change across agents of different abilities from our model. We use the following quantile regression specification to estimate these differences:

$$Quant_{\theta}(\log(y_{it})\,|\,.) = \alpha_{0\theta} + \sum_{t}\gamma_{t\theta}YM_{t} + \beta_{\theta}\mathrm{NewPlan}_{t} + \delta_{1\theta}\log(QtrTarget_{it}) + \delta_{2\theta}\mathrm{GroupSize}_{i} + \varepsilon_{it} \qquad (12)$$

where all variables are as previously defined. Columns (a) to (e) in Table 6 report the simultaneous estimates of (12) at different quantiles indexed by *Θ*. The results illustrate that the average intervention effect reported above masks substantially different changes throughout the conditional distribution of log productivity.

Our intervention increases the productivity of the 10^{th} quantile territory by 41.5%, the 25^{th} quantile by 37.0 %, the 75^{th} quantile by 15.1% and the 90^{th} quantile by 20.0%. Overall, the quantile regression estimates confirm that the switch to the commission plan has differential effects on the productivity of the best and worst sales groups, with the greatest impact on the least productive groups. Also, the variance in output goes down post intervention because lower productivity groups experience a significant increase in productivity and move closer to the productivity of higher performing groups.

Timing Games

Table 1 summarizes the predicted effects on timing games. Parenthetically, we wish to clarify that our timing games effects do not imply that a salesperson at a certain distance from the quota will *not increase* effort to reach the quota. We argue only that a part of any observed acceleration in sales as one approaches the quota could be *explained by the salesperson pulling in sales from the future*. As such, the more rigorous test of our prediction comes from the distinctive predictions based on how far a salesperson is from the quota in the specific regime in question. Evidence of pull-in (push-out) from the future in the final month of a quarter is not simply an observed increase (decrease) in sales but whether, in addition, there is a statistically significant decrease (increase) in sales in the first month of the following quarter. Following Steenburgh (2008), we define categorical variables that capture the impact of past performance on an individual's current output. Variables are constructed using performance to date against quota immediately prior to the final month of the quarter. Performance to date is defined as the ratio of the cumulative revenue produced in the first two months of the quarter to the quota that needs to be met in that quarter. This performance to date is then used to create the categorical variables "*EXCEEDED*", "*NEAR*", "*STRETCH*", and "*FAR*" for the last month of the quarter. We also create variables "*POST EXCEEDED*", "*POST NEAR*", "*POST STRETCH*", and "*POST FAR*" in the month following the quarter-ending month. Detailed definitions are provided in Table 3.

As mentioned before, if salespeople are only varying efforts over months within quarters, then spikes or dips in revenue production could occur in the last month of the quarter but not in the month immediately following the quarter. However, if salespeople are playing timing games, then spikes and dips in revenue in the last month of a quarter will be followed by dips and spikes respectively in the first month of the next quarter. The coefficients on the *POST* variables can thus be used to infer whether timing games are present. We estimate the following model of revenue in territory group i in period t as follows:

$$\log(r_{it}) = \alpha_0 + \sum_i \alpha_i S_i + \sum_m \upsilon^m \Gamma_{it}^{\ m} + \sum_{post-m} \upsilon^{post-m} \Gamma_{it}^{\ post-m} + \delta_1 \log(QtrTarget_t) + \delta_2 GroupSize_i + \varepsilon_{it} \quad (13)$$

where r_{it} is the average revenue produced in territory i in period t, S_i, $QtrTarget_t$, $Groupsize_i$ are all defined previously, $\boldsymbol{\Gamma}_{it}^{m}$ consists of the *EXCEEDED, NEAR, STRETCH* and *FAR* dummy variables describing sales to date in the last month of a quarter in territory *i* in period *t*, $\boldsymbol{\Gamma}_{it}^{Post\text{-}m}$ consists of the *POST EXCEEDED, POST NEAR, POST STRETCH* and *POST FAR* dummy variables describing the state in the first month of the subsequent quarter.

We estimate the above regression separately on data from the bonus and commission plans. Column 1 in Table 7 shows evidence of timing games under the bonus regime. Specifically, both the *EXCEEDED* and the *FAR* coefficient are negative and significant (-0.417, p<0.01 and -0.316, p<0.01). This suggests that sales groups might be pushing sales out: if i) they have exceeded quota, and/or ii) they expect that they cannot make the quota. Looking at the POST variables, we observe that both of the coefficients on the *POST EXCEEDED* and *POST FAR* variables are positive and significant (0.308, p<0.01 and 0.163, p<0.01). Overall this provides confirmation for results 3(a) and 2(a) respectively in Table 1 that sales groups play timing games under the bonus plan; both sales groups that have exceeded quota and sales groups that are far from achieving quota reduce effort in the final month of the bonus period and push out sales to the next period. Also, notice that the *POST EXCEEDED* and *POST FAR* estimates are smaller in absolute value than the *EXCEEDED* and *FAR* estimates, suggesting that in the final month effort (reduction) and push-outs are being used in conjunction. The estimates for *NEAR* and *STRETCH* coefficients are positive and significant, suggesting that sales groups that are close might be pulling sales in (and increasing effort) to achieve quota and earn incentives. The *POST STRETCH* estimate is negative and significant, suggesting that sales groups are pulling in sales from the next period in order to meet quota and earn incentives in line with result 3(a). On the other hand, the *POST NEAR* estimate is not statistically significant.

Steenburgh (2008) reports an absence of timing games in his analysis of the durable goods office equipment sector, and suggests that the finding is very likely a function of his particular institutional context. Plainly, our context differs greatly from Steenburgh's. For instance, in his context order-booking involves, among other things, long negotiation periods, management involvement from both sides, bank credit lines, and working capital positions. Coupled with the sheer dollar size of the orders, these factors make it very difficult to move the order booking date forward or backward. By contrast, our selling context is much more fragmented, with a much smaller monetary value for each transaction, rendering it much more amenable to manipulation.

Column 2 in Table 7 reports results under the commission regime. In the FAR condition, we expect that even under the commission regime the salesperson would push sales out to the next quarter. The estimates of the *FAR* and *POST FAR* variables are -0.192 and 0.073 respectively, both directionally consistent with our result 2c in Table 1 – however, the latter is not statistically significant. The estimate of *EXCEEDED* is positive and statistically significant (0.279, $p<0.01$) but the analogous POST estimate is only marginally significant. So while we predict a weak pull-in, the evidence seems to point towards *limited amount* of timing games. In other words, under the commission regime we find mostly directionally consistent but statistically weak evidence for results 2c-3c in Table 1. The near absence of timing distortions in the commission regime is intriguing and calls for deeper theoretical and empirical investigation.

Multitasking distortions

Recall that our salespeople were encouraged to visit doctors and pharmacies. These doctor visits were considered to have little short-term impact on sales, but to have long-term consequences for the firm. The firm requires each territory to make at least 80% of their doctor visits in a quarter. If sales groups fall below this level, the firm imposes a penalty, which is usually a fraction of the quarterly incentives earned, if any. In contrast, the firm does not place any corresponding requirements for pharmacy visits. As explained previously, since pharmacy visits are closely linked to losses in short-term sales from insufficient stocking, salespeople will undertake necessary effort to ensure that pharmacies stock and push products even without any defined minimum call levels.

Our call report data show that most salespersons make between 14-16 visits per day (doctor and pharmacy visits) with the 80% enforcement limit falling close to 7.5 doctor visits per day. Table 8a reports average daily doctor and pharmacy visits by regime. Under commissions, average daily doctor visits *decrease* from 10.08 to 7.93, but average daily pharmacy visits *increase* from 4.92 to 6.74. These averages are statistically different at the 5% level. Apparently, attention shifted to tasks that are directly related to compensation, away from tasks that are nevertheless important to the firm. But these raw averages hide a much more nuanced story, which is illustrated in Table 8b. Looking at the upper panel of Table 8b, it is obvious that salespeople that managed to reach quota were just fulfilling the minimum level of doctor visits under both the bonus and commission regimes (8.02 and 8.50 respectively) before reaching the quota. This is in line with Corollary 1 (Web Appendix A). But once the quota has been achieved, the salesperson under the bonus regime shifts his attention largely to doctor visits (12.98 > 8.02, $p < 0.01$), whereas under the commission regime attention shifts away from doctor visits (6.86 < 8.50, $p < 0.01$). These findings are exactly in line with results 3b and 3d in Table 1[13].

To probe these effects formally, we run the following regression(s) separately for bonus and commission plans:

$$\log(A_{it}) = \alpha_0 + \sum_i \alpha_i S_i + \sum_t \gamma_t YM_t + \sum_m \upsilon^m \Gamma_{it}^{\ m} + \sum_m \kappa^m (NewPlan)_t * \Gamma_{it}^{\ m} + \delta_1 log(QtrTarget)_t + \delta_2 GroupSize_i + \varepsilon_{it} \quad (14)$$

where A_{it} is the average of doctor (pharmacy) visits reported in territory *i* in period *t*, and the remaining variables have all defined previously. Note that while we include the vector Γ_{it}^{m} consisting of the four dummy variables describing sales history in the last month of the quarter, we exclude the POST counterparts of these variables as we are not interested in timing effects here.

Table 9 reports the results from this analysis. We focus on doctor visits, since the results for pharmacy visits are complementary. Recall that our predictions across the two plans are sharpest for salespersons who have achieved their quota. Specifically, Results 3b and 3d say that in the bonus plan, salespersons who have achieved/exceeded their quota will emphasize doctor visits while those in the commission plan will emphasize pharmacy visits.

Our results support precisely this conclusion; the *NewPlan* coefficient is -0.256, the *EXCEEDED* coefficient is +0.288 and *New Plan* EXCEEDED* is -0.450, all significant ($p < 0.01$). Intuitively, in the bonus plan, the salesperson is willing to focus on doctor visits because there is no immediate opportunity cost. In the commission plan, however, salespeople continue to get

[13] Salespeople who never made the quota exceed the minimum (10.10) under the bonus regime while they just meet the minimum under commissions (7.91). One way to rationalize this is that under bonuses, the quota is relatively harder to reach and salespersons far from the quota are willing to substitute into doctor visits as they do not care for the output during the current quarter.

commissions from any sales they book, so it makes sense for them to try to increase sales even after realizing their quota. None of the other results are statistically significant.

We feel that the result that multi-tasking concerns get minimized once the bonus-quota is reached is a novel finding and makes an important theoretical contribution to the literature. It also provides a nuanced rationale for the existence of lump-sum bonuses, in that while the presence of a bonus provides an incentive for agents to increase effort, the ceiling imposed through a bonus allows the principal to engage agents in those activities that benefit the principal even though they are not directly compensated.

Overall, our results are largely supportive of our theoretical framework. Nevertheless, many alternative explanations remain, especially with respect to the overall identification in a before-after framework, and issues related to turnover and ratcheting. We discuss these issues in the next section and present the details of some robustness checks in Web Appendix B.

DISCUSSION

Detailed longitudinal data from a pre-post experimental design implemented at a single firm enabled us to uncover the relative merits of quota- bonus compensation plans versus quota-commissions. Our empirical analysis can be summarized by four key findings.

First, the switch to an equivalent commission plan improved sales productivity by 24% on average, after accounting for industry, territory, and period effects. *Second*, the switch has heterogeneous effects at different ability levels. Specifically, the switch increases productivity at the lower deciles of ability much more so than at the higher deciles of ability. *Third*, sales groups engage in timing games under the bonus plan. Specifically, sales groups that have either achieved quota or are far from quota tend to reduce effort and push sales out to the next period; sales groups that can reach quota if they stretch tend to pull-in sales from future periods. These effects are attenuated under the commission plan. *Fourth*, the bonus plan is much better than equivalent commissions in engaging salespeople to focus effort on tasks that are not directly related to short-term output (and hence incentive pay).

Assessing Threats to Validity

Our before-after design with multiple observations per territory over time rules out a number of threats arising from unobserved differences across territories, by including effects for territory, time and targets in our specifications. In Web Appendix B, we describe a number of additional potential threats. Here, we describe briefly the supplemental analyses we undertook.

Individual vs. multi-person territories. Recall that the firm only knows sales at the territory level, which includes both single person and multi-person territories. We re-ran our regression models using only observations from single-person territories; the results were largely unchanged, suggesting no artifacts arising from multi-person territories.

Ratcheting. Ratcheting refers to quotas being raised in follow-on quarters when a salesperson has a good sales realization in a quarter. Realizing this, salespeople may strategically adjust their effort downwards. In our fieldwork, we brought up this issue, and discovered that management was keenly aware of this potential problem, and assiduously sought to avoid it by excluding the immediate past quarter's sales realization in the formula for the next quarter's quota. Nevertheless, since past performance enters the quota setting process along with company sales, industry sales and territory features, we assess ratcheting concerns directly using the Misra and Nair (2011)'s specification of a reduced form regression that predicts quotas in period *t* from prior period sales and quotas. We find no empirical evidence of ratcheting. (See Table B1)

Differential attrition. All experiments, including randomized trials are threatened when subjects quit on account of the treatment. To this end, we compared quit rates before and after the intervention; the rates are virtually identical (See Table B2). We also re-ran our regression models with additional measures accounting for turnover (see Table B3).

Serially correlated outputs. Given the 36 month period of our observations, we re-ran our regression models to allow for serially correlated errors. No qualitative changes were found, ruling out unobserved time effects as a validity threat.

Despite these supplemental analyses, there are other limitations of the present experiment that warrant scrutiny in future research. First, we have not modeled the quota setting process of our firm – a complete model would enable us to explicitly account for ratcheting concerns and other dynamic considerations. Second, our single firm experiment raises generalizability issues; it would be interesting to study other institutional contexts.

Managerial Implications

Our results offer sharp managerial implications for two common quota based plans. For small sales task portfolios (which rule out large multi-tasking problems), commissions are preferable to lump sum bonuses. Commission plans are also more desirable when a company desires a smoother selling pattern (i.e., when timing games carry significant costs, commissions trump bonuses). On the other hand, when firms want to incentivize their salespeople to produce more but the task portfolio is more diverse, bonus schemes work better in getting salespeople to engage those tasks in the portfolio that do not have an immediate short-term impact on pay.

Our study is located within the recent literature in marketing that has focused on understanding issues related to sales force productivity. Misra and Nair (2011) focus their energies upon the issue of quota dynamics in general and the ratcheting effect in particular.

Their prescription calls for implementing incentives without caps and using a shorter horizon for incentive payment. Interestingly, as discussed above, our firm seems to have overcome ratcheting concerns through the use of aggregate performance measures in updating quotas. As such, removal of caps is likely to lead to higher output, but our study suggests caution towards this prescription since this could potentially be counterproductive when significant multi-tasking concerns are present.

Table 1: Results from Analytical Model

Result	Productivity	Heterogeneity in Productivity Increase	
1	Bonus < Commissions	↑Low ability > ↑High ability Medium ability ↔	
	Timing Games	**Too Far from Quota (*FAR*)**	**Exceeds Quota (*EXCEEDED*)**
2a-3a	Bonus	Push sales out	Push sales out
2c-3c	Commissions	Push sales out	Pull sales in
	Multi-tasking	**Too Far from Quota (*FAR*)**	**Exceeds Quota (EXCEEDED)**
2b-3b	Bonus	Doctor↑ Pharmacy↓ (small effect)	Doctor↑ Pharmacy↓
2d-3d	Commissions	Doctor↑ Pharmacy↓ (small effect)	Doctor↓ Pharmacy↑

Table 2: Salesperson Optimization Problem

	Probability of reaching quota in Qtr 1	Expected Incentive Qtr1	(Expected) Probability of reaching quota in Qtr2	Expected Incentive Qtr2	Cost of gaming
Bonus	$(1-G(\frac{S_{Q_{1B}}}{\lambda}-h_dT-(\Delta h)T_{p2}))$	B	$(1-G(S_{Q_{2B}}-h_dT-(\Delta h)T_{p4})$	B	$\mu(\lambda-1)^2 D$
Commissions	$(1-G(\frac{S_{Q_{1C}}}{\lambda}-h_dT-(\Delta h)T_{p2}))$	$\alpha(\lambda(h_dT+(\Delta h)T_{p2})-S_{Q_{1C}}$	$(1-G(S_{Q_{2C}}-h_dT-(\Delta h)T_{p4})$	$\alpha(h_dT+(\Delta h)T_{p4}-S_{Q_{2C}})$	$\mu(\lambda-1)^2 D$

Salesperson Optimization Problem:

arg max$_{(\lambda,\ Tp2)}$ *[(Probability of reaching quota in Qtr 1)x(Expected Incentive Qtr1)]+β[E(Probability of reaching quota in Qtr2)x(Expected Incentive Qtr2)]- [Cost of gaming]*

Table 3: Data Description

Variable	Definition	Mean	Standard Deviation
New Plan	A dummy variable equal to 1 if sales group is on the Commission plan during that quarter/month	0.51	
Group Size	Total members in the sales group	2.17	3.05
Qtr Productivity	Quarterly revenues generated per salesperson	27.60	15.63
Monthly Productivity	Monthly revenues generated per salesperson	9.26	5.63
Incentives	Quarterly incentives earned by a salesperson	0.24254	0.42
Net Revenues	Quarterly revenues minus total incentives paid out to sales group	57.15	72.56
Industry Sales	Total quarterly industry sales	210.66	40.96
Target	Quarterly sales quota	62.37	76.49
Sales History Variables			
EXCEEDED	If quarterly performance to date is >=1, then this variable takes a value 1 during the last month of the quarter, 0 otherwise	0.02	
NEAR	If quarterly performance to date >=2/3 but less < 1, then this variable takes a value 1 during the last month of the quarter, 0 otherwise	0.10	
STRETCH	If quarterly performance to date >=1/3 but less < 2/3, then this variable takes a value 1 during the last month of the quarter, 0 otherwise	0.17	
FAR	If quarterly performance to date >=0 but less < 1/3, then this variable takes a value 1 during the last month of the quarter, 0 otherwise	0.02	
POST EXCEEDED	Takes value 1 if Exceeded=1, 0 otherwise	0.02	
POST NEAR	Takes value 1 if Near=1, 0 otherwise	0.08	
POST STRETCH	Takes value 1 if Stretch=1, 0 otherwise	0.14	
POST FAR	Takes value 1 if Far=1, 0 otherwise	0.02	

Note: Productivity, Net Revenues, Targets and Incentives have been disguised by multiplying actual values by a single common positive number. Further, the currency is suppressed for confidentiality.

Table 4: Key Variables by Incentive Plan

	Bonus Plan		Commissions Plan	
Variable	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**
Monthly Productivity	8.99	5.63	9.51	5.61
Qtr Productivity	26.23	15.97	28.87	15.20
Qtr Net Revenues	51.87	64.42	62.07	79.08
Qtr Incentives	0.138	0.391	0.339	0.427
Group Size	2.06	2.75	2.23	3.21
Quarterly Target	70.44	82.13	57.95	75.12
10^{th} Percentile Qtr Target	18.47		13.14	
90^{th} Percentile Qtr Target	132.94		114.89	

Table 5: Monthly Productivity Results

Dependent Variable = log (Monthly Productivity)

	1	2	3	4	5	6
	Turnover not accounted			**Turnover accounted**		
New Plan	0.222***	0.558***	0.239***	0.238***	0.576***	0.249***
	(0.008)	(0.0287)	(0.0300)	(0.008)	(0.0287)	(0.0301)
Group Size	-0.091***	-0.0804***	-0.0863***	-0.048***	-0.0120	-0.0136*
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
	0.420***		0.178***	0.432***		0.192***
log(Qtr Target)	(0.012)	-	(0.0172)	(0.012)	-	(0.0175)
Constant	0.534***	1.73***	1.39***	0.391***	1.52***	1.19***
	(0.0492)	(0.028)	(0.061)	(0.047)	(0.028)	(0.061)
Group Effects	Included	Included	Included	Included	Included	Included
Month-Year Effects	Not Included	Included	Included	Not Included	Included	Included
Observations	14,499	14,499	14,499	14,499	14,499	14,499
R-squared	0.597	0.656	0.659	0.603	0.662	0.665
Sales Groups (Clusters)	458	458	458	528	528	528

Notes: ***p<0.01, **p<0.05,*p<0.1, Standard errors in parentheses. (1), (2) and (3) use territory fixed effects as the group effects while (4),(5) & (6) report the results where team composition (turnover) is accounted for and each new team in a territory is assigned a unique dummy to create the "team-territory fixed effects".

Table 6: Effect of Commission Plan on Productivity (Quantile SUR Regressions)

Dependent Variable = log (Monthly Productivity)

Variables	**(a)** **10th**	**(b)** **25th**	**(c)** **50th**	**(d)** **75th**	**(e)** **90th**
New Plan	0.415*** (0.098)	0.370*** (0.063)	0.219*** (0.047)	0.151*** (0.040)	0.200*** (0.053)
Group Size	-0.169*** (0.007)	-0.147*** (0.005)	-0.115*** (0.004)	-0.092*** (0.002)	-.094*** (0.001)
log(Qtr Target)	0.857*** (0.016)	0.696*** (0.012)	0.576*** (0.011)	0.524*** (0.010)	0.567*** (0.012)
Constant	-1.85*** (0.095)	-0.892*** (0.067)	-0.117** (0.049)	0.426*** (0.035)	0.529*** (0.064)
Month-Year Effects	Included	Included	Included	Included	Included
Observations	14,499	14,499	14,499	14,499	14,499
(Pseudo) R-Squared	0.289	0.2607	0.2459	0.2454	0.2550

Note: ***p<0.01, **p<0.05,*p<0.1, Standard errors (bootstrapped) in parentheses.

Table 7: Sales Variation across Time

Dependent Variable: log (Monthly Revenues)

VARIABLES	A **Bonus Plan**	B **Commissions Plan**
EXCEEDED	-0.417*** (0.085)	0.279*** (0.027)
NEAR	0.147*** (0.023)	0.096*** (0.013)
STRETCH	0.051*** (0.012)	0.053*** (0.015)
FAR	-0.316*** (0.041)	-0.192* (0.100)
POST EXCEEDED	0.308*** (0.070)	0.046* (0.025)
POST NEAR	0.026 (0.038)	-0.001 (0.012)
POST STRETCH	-0.121*** (0.014)	-0.005 (0.014)
POST FAR	0.163*** (0.041)	0.073 (0.078)
Group Size	0.0256 (0.031)	0.018*** (0.005)
log(Qtr Target)	0.379*** (0.018)	0.508*** (0.017)
Constant	0.899*** (0.089)	0.647*** (0.062)
Fixed Effects	Included	Included
Observations	7107	7392
Clusters	444	438
R-squared (Overall)	0.8061	0.8530

Note: ***p<0.01, **p<0.05,*p<0.1, Standard errors in parentheses, clustered by territory

Table 8a: Multitasking Effects

	Bonus Plan	Commissions Plan
Doctor Visits	10.08	7.93***
Pharmacy Visits	4.92	6.74***

Note: ***Denotes significant difference at 0.01 level.

Table 8b: Multitasking Effects (By Target Achievement)

	Panel 1 Bonus Regime		Panel 2 Commissions Regime	
Doctor Visits	**Before reaching Quota (A)**	**After reaching Quota (B)**	**Before reaching Quota (C)**	**After reaching Quota (D)**
Reached Quota	8.02	12.98***	8.50	6.86***
Did not reach Quota	10.10		7.91***	
Pharmacy Visits	**Before reaching Quota (E)**	**After reaching Quota (F)**	**Before reaching Quota (G)**	**After reaching Quota (H)**
Reached Quota	6.3	3.37***	6.27	8.45***
Did not reach Quota	4.90		6.71***	

Note: ***Denotes significant difference at 0.01 level.

Table 9: Doctor and Pharmacy Visits

Variables	Pharmacy Visits	Doctor Visits
New Plan	0.309 (0.005)***	-0.256 (0.004)***
EXCEEDED	-0.383 (0.022)***	0.288 (0.010)***
NEAR	-0.015 (0.014)	-0.009 (0.008)
STRETCH	0.007 (0.008)	0.005 (0.003)
FAR	0.002 (0.013)	0.0006 (0.008)
New Plan * EXCEEDED	0.645 (0.024)***	-0.450 (0.021)***
New Plan* NEAR	0.0386 (0.016)**	0.0007 (0.011)
New Plan* STRETCH	0.003 (0.011)	-0.0167 (0.105)
New Plan* FAR	0.012 (0.032)	-0.0174 (0.053)
Group Size	0.005 (0.005)	0.0026 (0.003)
log(Qtr Target)	0.011 (0.006)*	-0.008 (0.006)
Constant	1.51 (0.030)***	2.329 (0.023)***
Month-Year Effects	Included	Included
Fixed Effects	Included	Included
Observations	7107	7,392
Clusters	444	438
R-squared (overall)	0.4245	0.3192

Note: ***p<0.01,**p<0.05,*p<0.1, Standard errors in parentheses, clustered by territory

Figure 1: Plots of Incentive Compensation Plans

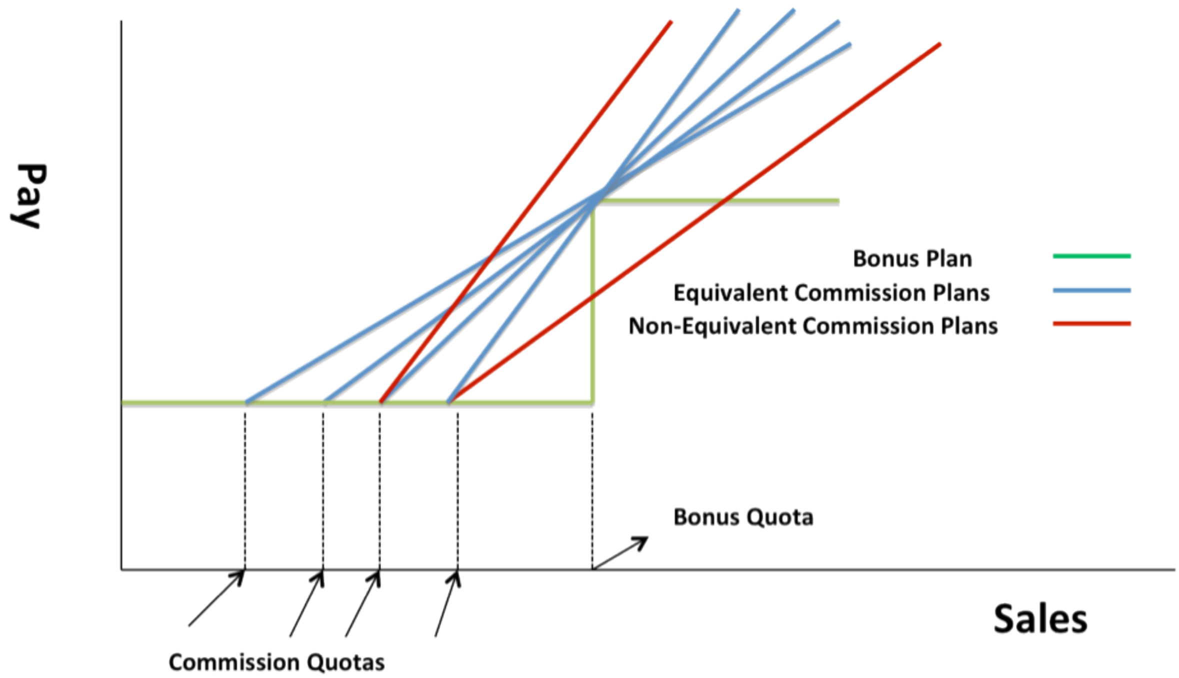


Figure 2: Effort & Output under Equivalent Bonus and Commission Plans[14]

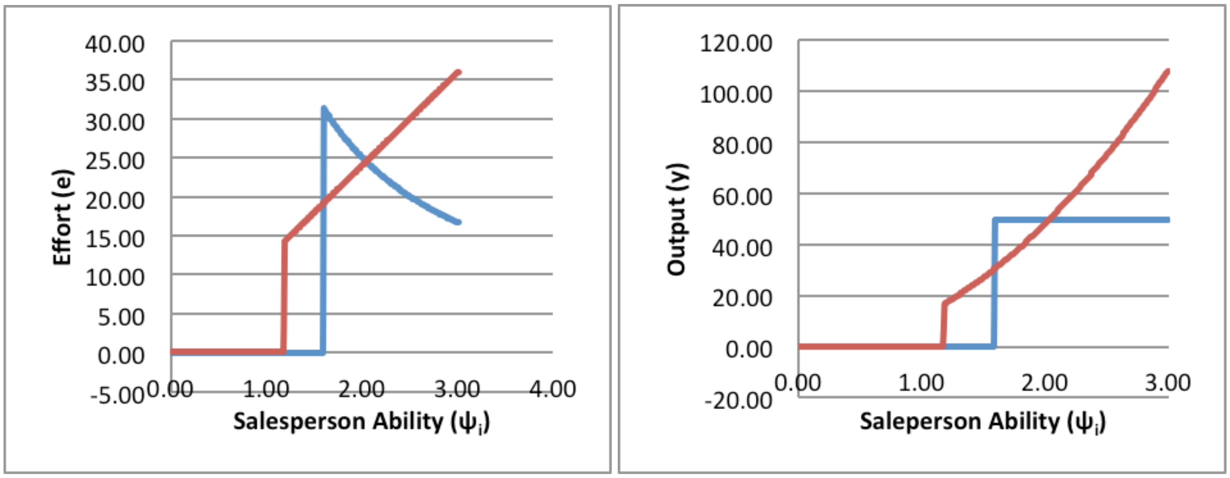


Figure 2(a)

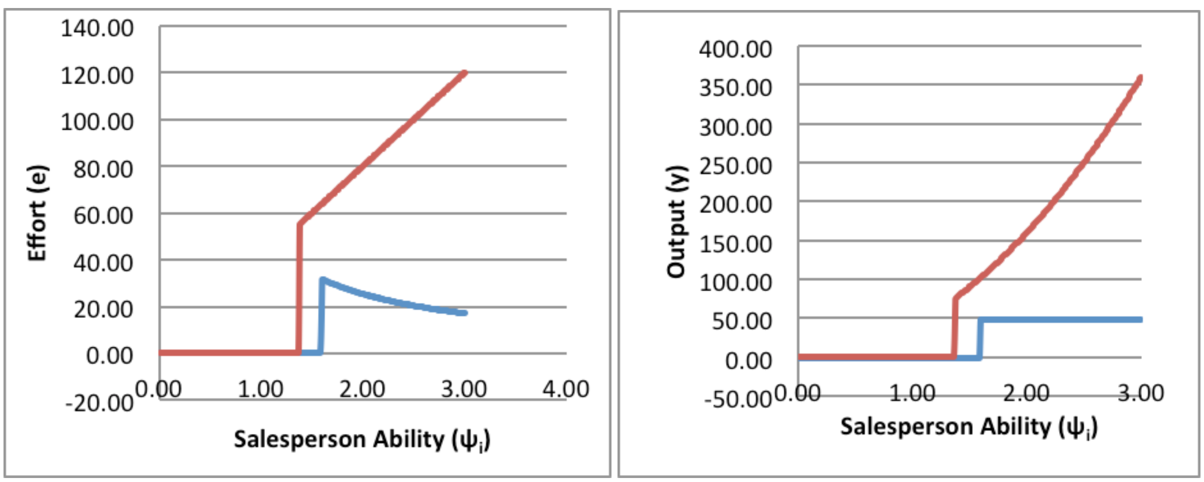
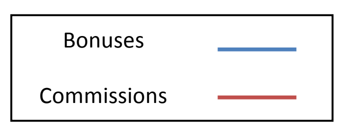


Figure 2(b)



[14] The parameter values used in constructing these figures are B=5, Q_B=50, ϴ=0.01, ψ~ u [0, 3]. The parameter which is different across 2a and 2b is the commission rate (α=0.12 and 0.40 respectively).

Figure 3: Sequence of Events in Multi-tasking and Timing Game Model

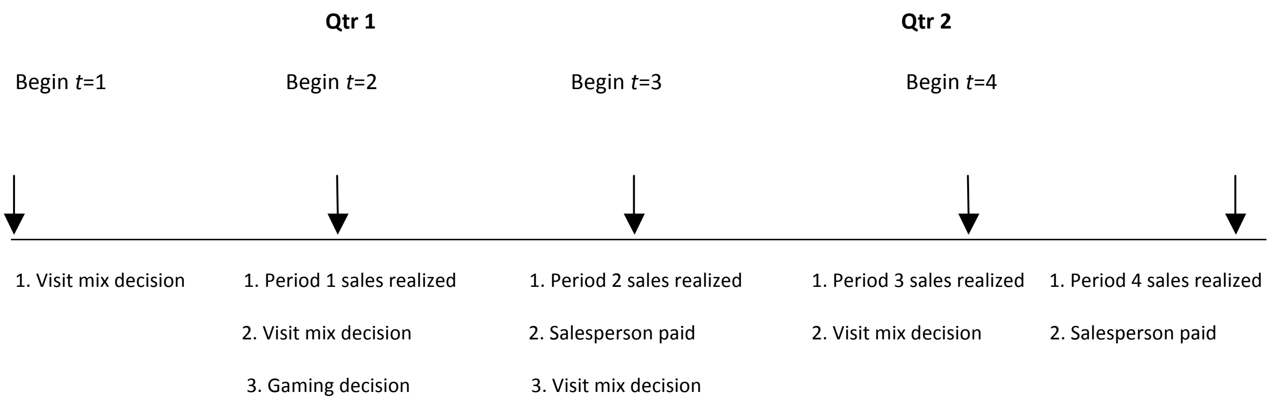


Figure 4: Total Monthly Sales over time

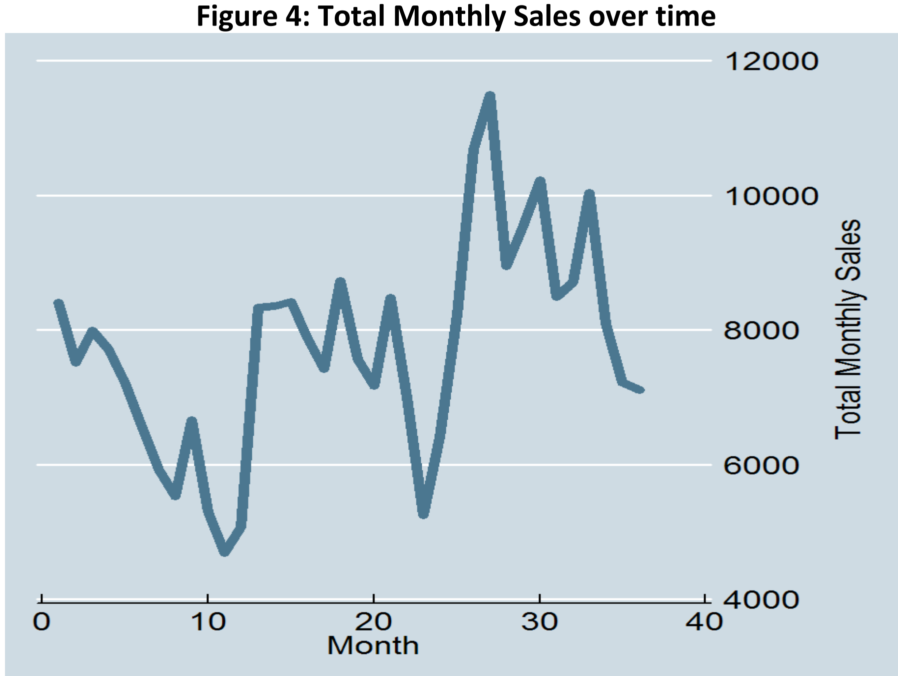


Figure 5: Monthly Industry Sales over Time

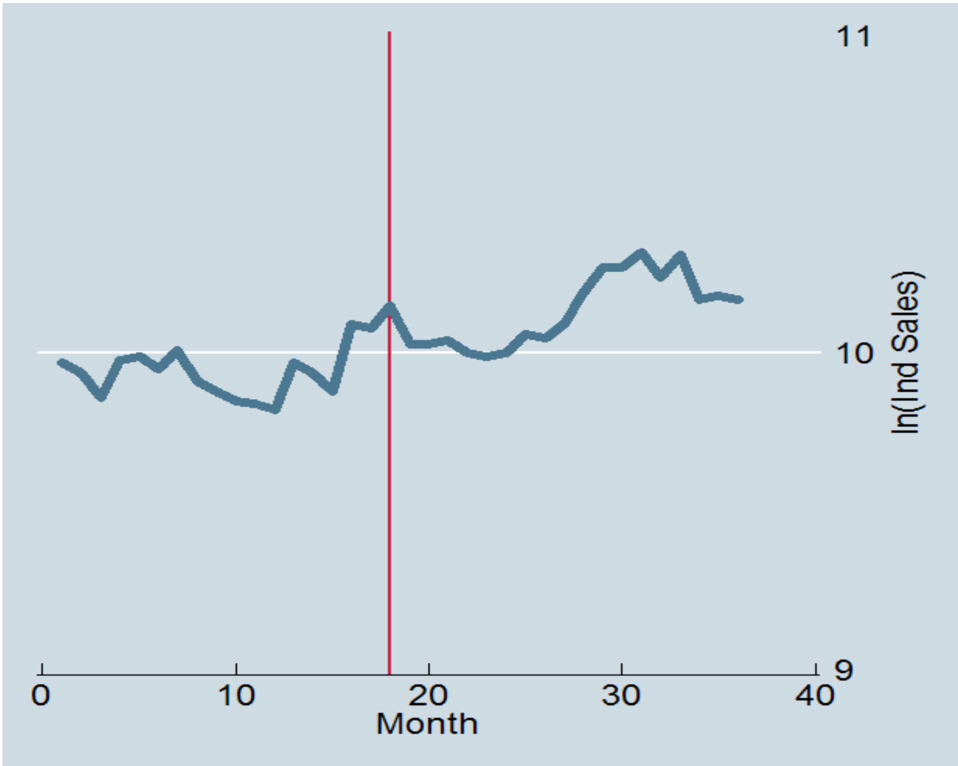


Figure 6: Total Monthly Sales over Time

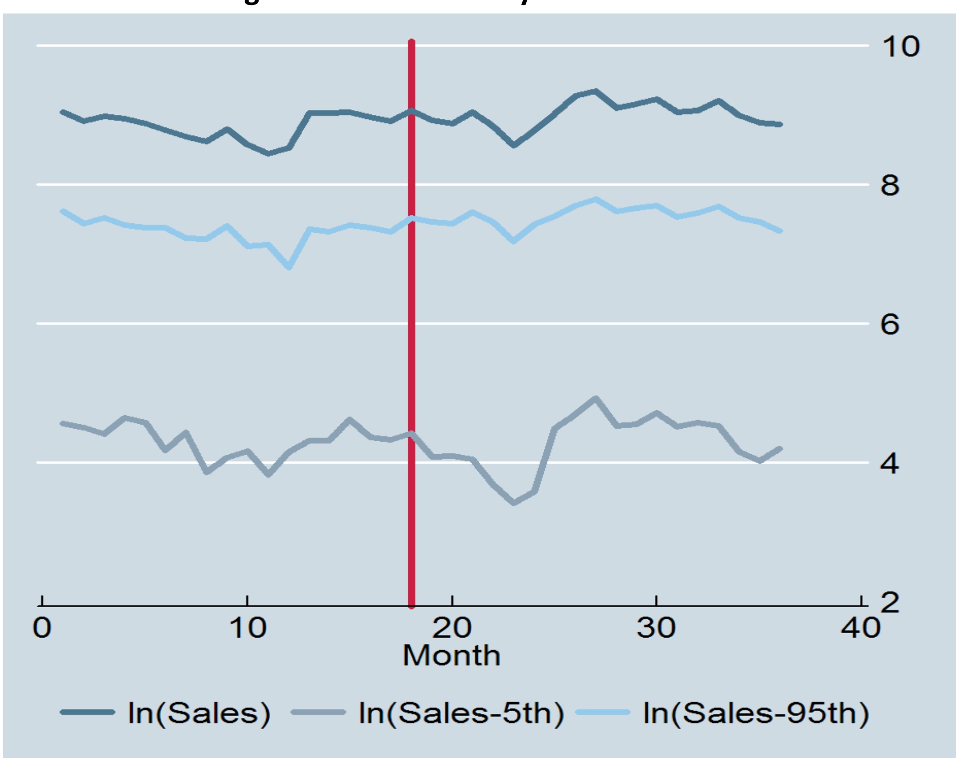
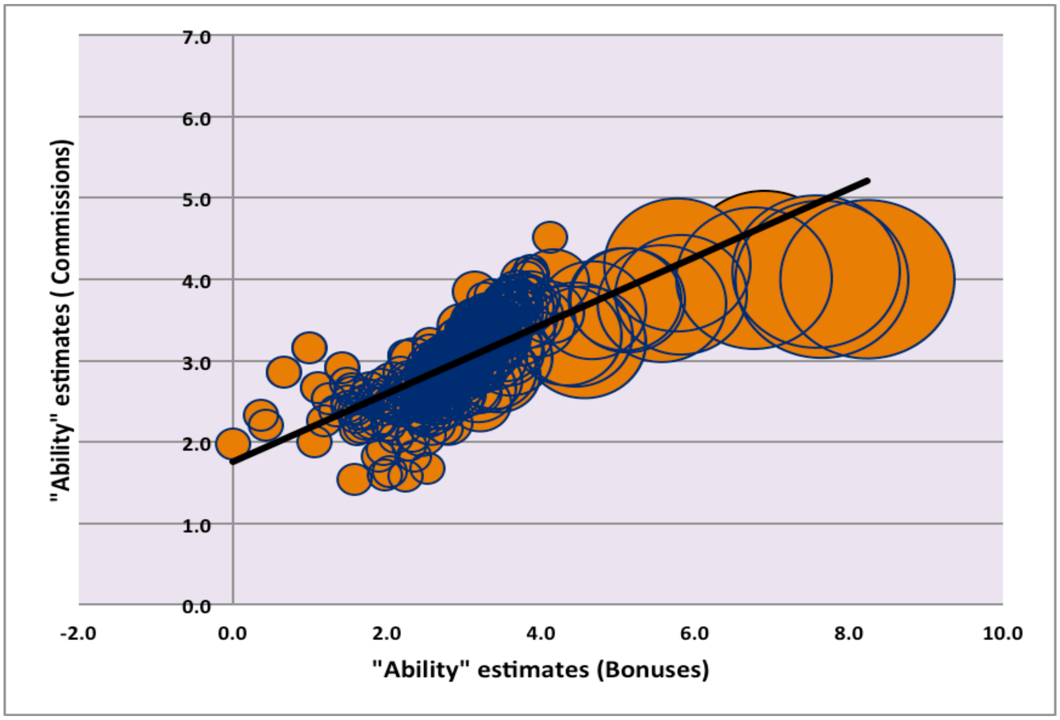
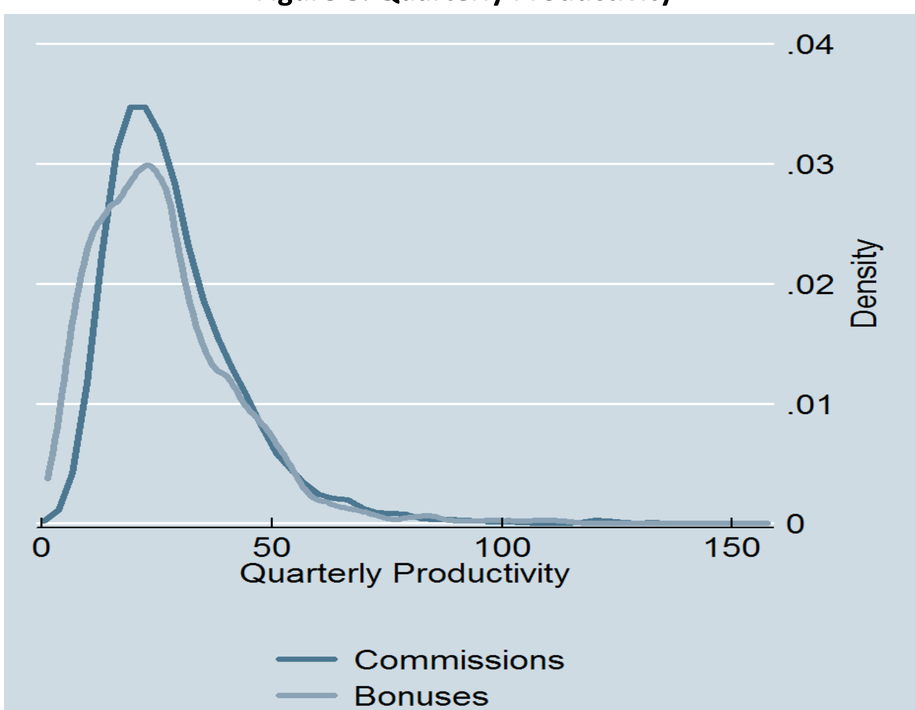


Figure 7: Ability Estimates across Bonuses and Commissions



Note: The size of each circle represents the average size of the sales group during the period of observation and the center of the circle is the associated fixed effect estimate. The black solid straight line represents the linear trend line. Each territory fixed effects pair is estimated by running equation (11) for each regime. The smallest fixed effect has been normalized to zero.

Figure 8: Quarterly Productivity



REFERENCES

Albers, Sonke and Murali K. Mantrala (2008), "Models for Sales Management Decisions", Handbook *of Marketing Decision Models*, Wierenga, Bireng (Ed.), Springer.

Alchian, A. A. and H. Demsetz (1972). "Production, Information Costs, and Economic Organization," *American Economic Review, 62,* pp. 777-795

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2005), "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *Quarterly Journal of Economics*, 120(3), 917-962

Chung, Doug, Thomas J. Steenburgh, and K. Sudhir (2010), "Do Bonuses Enhance Productivity? A Dynamic Structural Analysis Bonus-based Compensation Plans," *Harvard Business School Working Paper* (1491283)

Gibbons, Robert (1987), ""Piece-Rate Incentive Schemes" *Journal of Labor Economics*, 5 (October): 413-29

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan (2003), "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 111(3), pp. 465-497.

Hardin, Garrett (1968), "The Tragedy of the Commons," *Science*, 162:1243-48

Holmstrom, Bengt (1982), "Moral Hazard in Teams", *The Bell Journal of Economics* 13 (2), pp. 324—340.

________ and P. Milgrom (1991), "Multitask Principal-agent analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization,* 7 (Special Issue), pp. 24-51.

Jain, Sanjay (2010), "Self-control and Optimal Goals: A Theoretical Analysis", *Marketing Science*, 28(6), 1027-1045

Jensen, M. C. (2003), "Paying People to Lie: The Truth about the Budgeting Process*", European Financial Management,* IX, pp. 379–406

Joseph, K. and M. Kalwani (1998),"The Role of Bonus Pay in Sales force Compensation Plans", *Industrial Marketing Management*, 27, pp. 147-159.

Knez, Marc, and Simester, Duncan (2001) "Firm-wide Incentives and Mutual Monitoring at Continental Airlines", *Journal of Labor Economics*, 19, pp. 743-72

Latham, G. P., & Locke, E. A. (1991), "Self-regulation through goal setting", *Organizational Behavior and Human Decision Processes*, L, pp. 212–247.

Lazear, Edward P. (1995), "Personnel Economics", Cambridge and London: MIT Press

______ (2000), "Performance Pay and Productivity," *American Economic Review,* 90(5), pp. 1346-1361

Misra, Sanjog and Harikesh Nair (2011), "A Structural Model of Sales-force Compensation Dynamics: Estimation and Field Implementation," *Quantitative Marketing and Economics*, 9, 211-257

Oyer, Paul (1995), "The Effect of Sales Incentives on Business Seasonality," *Working Paper #354*, Industrial Relations Section, Princeton University

________ (1998), "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality", *Quarterly Journal of Economics*, 113, pp. 149-185.

Paarsch, Harry J. and Bruce Shearer (2000), "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records," *International Economic Review*, 41(1), 59-92

Prendergast, Candice (1999), "The Provision of Incentives in Firms", Journal *of Economic Literature,* 37 (1), 763

Steenburgh, Thomas (2008), "Effort or Timing: The Effect of Lump-Sum Bonuses", *Quantitative Marketing and Economics*, 6, pp. 235-256

Zoltners, A., Prabhakant Sinha and Sally E. Lorimer (2008), "Sales Force Effectiveness: A Framework for Researchers and Practitioners", *Journal of Personal Selling and Sales Management*, 28 (2), pp. 115-131