

Do Bonuses Enhance Sales Productivity?
A Dynamic Structural Analysis of Bonus-Based Compensation Plans*

Doug Chung, Yale School of Management
Thomas Steenburgh, Harvard Business School
K. Sudhir, Yale School of Management

May 2009

* We thank the participants at the UTD-FORMS conference, the Yale Marketing Seminar and the Yale Doctoral Workshop for helpful comments and suggestions.

Do Bonuses Enhance Sales Productivity?

A Dynamic Structural Analysis of Bonus-Based Compensation Plans

Abstract

Using data on individual level sales force performance over a 3 year period at a Fortune 500 firm, we propose and estimate a dynamic structural model of sales force response to a bonus-based compensation plan. We combine Arcidiacono and Miller's new EM algorithm approach within a two step conditional choice probability (CCP) estimator to allow for sales force response heterogeneity, while preserving the computational advantage of the CCP estimators. Our model accommodates hyperbolic discounting on the part of the sales force.

We find that bonuses enhance sales productivity. The current bonus-based plan that includes quarterly and annual bonuses and an overachievement commission rates produces 18.5% higher revenues relative to a simple linear commission-only plan. Even adjusting for commission rates to allow for the same total compensation cost as in the current plan, the current bonus-based plan obtains 4.6% higher revenues. Eliminating over-achievement commission rates for sales beyond quotas reduces revenues (and profits) by roughly 6%.

The elimination of quarterly bonuses (holding the annual bonus and overachievement rates constant) reduces total revenues by 13.6% and December (year-end) revenues by 7%. Even if the eliminated quarterly bonus amounts are added to the annual bonus, revenues fall by 4%. Thus annual quotas or over-achievement commission rates have smaller impact on performance *without the quarterly quotas*.

I. Introduction

Personal selling is one of the most important elements of the marketing mix, especially in the context of B2B firms. An estimated 20 million people work as salespersons in the United States (Albers et al. 2008). Sales force costs average about 10% of sales revenues and as much as 40% of sales revenues for certain industries (Mantrala et al. forthcoming). In the aggregate, U.S. firms spent over \$800 billion on sales forces in 2006, a number that is three times larger than advertising spending (Zoltners, Sinha and Lorimer 2008).

Marketing researchers routinely create response models for marketing mix instruments such as price, sales promotion and advertising. Meta-analysis of various research studies estimate that the sales force expenditure elasticity is about 0.35 (Albers et al., 2008), relative to about 0.22 for advertising (Assmus, Farley and Lehmann 1983) and about -2.62 for price (Bijmolt et al. 2005). While relative sales force expenditure elasticity is useful in determining the relative effectiveness of different instruments in the marketing mix, they give us little insight on how to design a sales force compensation plan, which is widely understood to be the primary tool by which firms can induce a sales force to exert the optimal levels of effort and thus to optimize the use of sales force expenditures.

A compensation plan can consist of many components: salary, commissions, and bonuses on achieving a certain threshold of performance called quotas. Figure 1a shows the distribution of which components are used by firms based on a survey by Joseph and Kalwani (1998). Figure 1b shows a variety of compensation plans that include combinations of these components.

Which of these different types of contracts should a particular firm offer to its sales force to maximize profits? What combination of salary, commission or quota-based bonuses should one use? Having chosen the compensation components, what should be the specific parameters for commission rate, quotas and bonus levels? Further, what is the right frequency for quota targets? For example, should there be quarterly or annual quotas? A firm needs to understand how the sales force will respond to different dimensions of the compensation plan to develop an appropriate plan.

The goal of this paper is to therefore build and estimate a structural model of how the sales force responds to alternative compensation instruments and their levels in order to help

design a compensation plan that optimizes sales force performance. Specifically, we seek to answer two substantive questions: First, do bonuses improve sales productivity, relative to straight commissions and by how much? Second, what should be the frequency of bonuses? Should one use a quarterly bonus in addition to an annual bonus?

Quotas and bonuses are widely used by firms. Figure 1a shows the frequency distribution of the types of compensation plans reported in Joseph and Kalwani (1998). Only about 25% of firms use a pure commission-based plan; the rest used some form of quotas. According to the Incentive Practices Research Study (2008) by ZS Associates, quota based compensation is used by 73%, 85% and 89% in the pharma/biotech, medical devices and high tech industries respectively. Yet, despite the ubiquity of quota-based compensation, there is considerable controversy in both the theoretical and empirical literature about the effectiveness of quotas and bonuses relative to straight linear commission plans.

In the theoretical literature, Basu et al. (1985) apply the principal agent framework of Holmstrom (1979) and demonstrate that a combination of salary plus commission (usually *nonlinear with respect to sales*) will be optimal. Rao (1990) also shows a similar result on the optimality of nonlinear compensation plans. However, the need for nonlinear compensation plans is questioned in Holmstrom and Milgrom (1987) and Lal and Srinivasan (1993). Using the specific assumptions of linear exponential utility and normal errors (LEN) they show that a *linear* commission incentive scheme can achieve the best possible outcomes for the firm.

Yet, why do most firms have quota based compensation plans? Why are compensation plans nonlinear? Raju and Srinivasan (1996) suggest that even though a commission over quota plan may not be theoretically optimal, they provide the best compromise between efficiency and ease of implementation. Others argue that quota based plans offer high powered incentives that can motivate salespeople to work harder (e.g., Darmon 1997). Park (1995) and Kim (1997) demonstrate that a quota-bonus plan may lead to the first-best outcome, but in their framework, quota-bonus plan is just one of many possible plans that lead to first best outcomes. Oyer (2000) shows that when participation constraints are not binding, a quota-bonus plan with linear commissions beyond quotas can be uniquely optimal because it can concentrate the compensation in the region of effort where the marginal revenue from effort minus the cost of compensation is maximized.

There is limited empirical work addressing the issue. Based on an analysis of aggregate sales across different industries in different quarters, Oyer (1998) concludes that the negative effects of quota based plans encouraging sales people to maneuver the timing of orders is greater than the benefits obtained from more effort. Steenburgh (2008) questions whether aggregate data can be used to reach this conclusion. Using individual sales performance data from the same firm used in this study (utilizes compensation plan F in figure 1b), he finds that the net improvement in revenues from effort dominates the inefficiencies induced by inter-temporal dynamic considerations. Using a dynamic structural model, Misra and Nair (2008) model data from a firm using compensation plan C in Figure 1b and conclude that a nonlinear compensation plan using quotas lead to substantial inefficiency. But their compensation plan does not have a bonus; hence the plan does not provide incentives for sales people to stretch themselves in order to meet quotas.

In this paper, we seek to address the question of whether bonuses can improve performance by stretching the sales force to work harder. By estimating a dynamic structural model, we are able to also assess how various dimensions of the compensation plan affect sales force performance. To the best of our knowledge, the substantive question about frequency of bonuses has not been addressed at all in the literature. Can frequent bonuses serve to enhance performance? And if so, why? In the education literature, researchers have argued that frequent testing leads to better performance outcomes (Bangert-Drowns et al. 1991). Can quarterly quotas serve a similar role to improve outcomes? Like in the education literature, where frequent exams keep students prepared for the comprehensive final exam; frequent quota-bonus plans may serve as a mechanism to keep the sales force motivated to perform in the short-run well enough to be in striking distance of the overall annual performance quota.

There are three major modeling and estimation issues that make the problem of structural estimation of response to compensation plans, especially those with quotas and bonuses difficult. First, in a typical structural model, one observes the agent's action in response to the firm's action. For example in a consumer response model, one observes consumers' choices in response to the firm's choice of marketing mix such as price, advertising or sales promotion. In contrast, for a sales force response model, one does not observe the actions of the sales force, i.e., the exerted effort. One only observes the outcome of the agent's effort, i.e., sales, which is correlated (but not a one-to-one mapping) with effort. Hence one has to make an inference about the agent's action (effort) that leads to sales from

the observed realized sales. This requires some modeling assumptions on the link between sales and effort.¹

A second challenge is that unlike marketing mix variables that change over time, the compensation plan remains stationary over at least a year. How can one identify how the sales force will respond to a compensation plan, when there is no variability in the plan? Here we draw on an empirical insight from Steenburgh (2008) that can help identify the sales force response, when the compensation plan involves payments for reaching quotas. The insight is that in any given period, a sales agent's optimal effort depends on her state: how close the person is to achieving her quota. A sales agent may find it optimal to reduce effort when she is close or very far from achieving quota, but may stretch herself to reach the quota, when she has a moderate chance of achieving this quota. This implies that the optimal level of effort (and therefore sales) would vary from period to period as a function of the agent's state (distance to quota).

A third issue follows from the discussion of the second. While quotas enable identification of sales force response, it also induces inter-temporal dynamics in optimal sales force response behavior. An agent has to be concerned not just with the current payoff by expending effort, but the future payoff that she can obtain by being in a more favorable state that can facilitate in obtaining a bonus. This implies that the estimated structural model needs to account for forward-looking behavior on the part of sales agents. Hence we need to estimate a dynamic structural model; since our model includes both quarterly and annual bonuses, we need to model forward looking behavior in response to both distances to quarterly and annual quotas.

We address the modeling challenges above to develop a dynamic structural model of sales force response to various features of the compensation plan. We estimate the proposed dynamic structural model using sales force output and compensation data from a Fortune 500 firm, selling office durable goods. This firm used Plan F in Figure 1b; the richness of the compensation plan allows us to address the sensitivity of sales force response to different dimensions of compensation. Importantly, the plan also includes quarterly and annual bonuses, which allows us to address the issue of frequency of quotas and bonuses. Armed with structural parameters of the model, we are able to perform counterfactual policy

¹ The issue has parallels in empirical channel response models. For example, Sudhir (2001) makes an inference about manufacturer actions (wholesale prices) from the observed retail price and sales to infer competition between manufacturers.

simulations that help us assess the two main substantive questions that we seek to answer: effectiveness of quota-bonuses and their frequency.

In terms of estimation methodology, we use the recently developed two stage estimation approaches using conditional choice probabilities to estimate the dynamic structural models (Hotz and Miller 1993; Bajari, Benkard and Levin 2007). The advantage of these estimation approaches is the computational simplicity relative to traditional nested fixed-point estimation approaches (Rust 1987). However, a primary limitation of the two-step CCP approach has been that it is unable to handle unobserved heterogeneity among agents. We use the innovations in Arcidiacono and Miller (2008) to incorporate unobserved heterogeneity in sales force response within a two step CCP based estimation framework. This enables to maintain much of the computational simplicity of two step estimation approaches, but still allowing for unobserved heterogeneity in sales force response, which we find is empirically critical in our data. We believe this approach will make dynamic structural estimation more computationally feasible without sacrificing the modeling of unobserved heterogeneity so critical in marketing applications. Finger (2008) has applied the approach to allow for unobserved heterogeneity across chemical plant productivity in an analysis of the chemical industry. To the best of our knowledge, this is the first paper in marketing applying the Arcidiacono and Miller approach to account for unobserved heterogeneity in the two-step dynamic structural estimation framework.

Discount factors are an important element of dynamic models. Since it is impossible to identify the discount factor non-parametrically from only choice data without functional form assumptions (Rust 1994; Magnac and Thesmar 2002) or additional exclusion restrictions, it is typically the norm to assume a discount factor that is consistent with an interest rate. However, the psychology literature finds that when experimental subjects make choices, they tend to discount the future much more than can be explained by interest rates.² This implies that the future (and therefore dynamics) may be seemingly more important for the sales force when a discount rate consistent with interest rates are used. This might make it appear that bonuses are important, when in reality, they may be less important because the sales force discounts the future much more than the interest rate warrants. Further, in contrast to the constant discount rate (Samuleson 1937), researchers have shown evidence of ‘hyperbolic discounting’ where individuals discount the immediate future from the present

² Frederick and Loewenstein (2002) provide a detailed comprehensive review and summary.

more than do they discount the same interval in future dates and hence a declining discount rate – Thaler (1981). To capture this anomaly economists have used a relatively simple functional form for the discount factor $D(k)=\beta\delta^k$ and refer to the functional form as *quasi-hyperbolic discounting*. – Phelps and Pollak (1968), Elster (1979), Laibson (1997, 1998) etc. These hyperbolic discounting effects might be particularly important in the context of work, which is a form of "pain" that is encountered by the sales person. Hence to be conservative in assessing the value of bonuses, we perform a grid search for the discount rates that minimizes errors in predicted sales both under a constant exponential discounting and hyperbolic discounting settings. Fortunately, in our application we do have appropriate exclusion restrictions that can help us non-parametrically identify the discount factor. We find that the best fitting discount factor is considerably lower than is consistent with an annual interest rate, and find that hyperbolic discounting fits better than exponential discounting. Despite reducing the value of the bonuses that arrive in the future substantially (through higher discounting), we find that bonuses still have a significant impact in improving sales force performance.

Our work is related to several static structural models of worker behavior such as Ferrall and Shearer (1999) and Paarsch and Shearer (2000), who endogenize the optimal contract choice of the firm. In contrast to these papers, we are interested in the dynamic response of sales agents, but do not model the contract choice, because we do not have data to model the selection effects of contract choice. In that sense, our paper is similar to Copeland and Monnet (forthcoming) and Misra and Nair (2008). Copeland and Monnet model agent's productivity in sorting checks as a function of nonlinear incentives; unlike sales force productivity, there is limited unobserved uncertainty in check sorting productivity. Much of the variation can be explained by observed characteristics such as machine breakdowns etc. In terms of substantive context and estimation methodology, Misra and Nair (2008) is the closest. Their estimation methodology relies on accommodating heterogeneity through estimating each sales person's individual utility function separately, requiring the need for a large number of client level data. This approach is not feasible in many sales force settings including ours, where the sales force performance is best assessed as an aggregate of all client sales, not at each client level. Hence, our approach to handle unobserved heterogeneity uses the more traditional latent class modeling framework, and makes the

approach more relevant to general contexts.³ Given the differences in the compensation plans, the substantive conclusions about the effectiveness of quotas differ. Finally, they address the issue of ratcheting of quotas based on past performance, an issue we abstract away in this paper. First, the updates in quotas for the focal firm are not as frequent as in the data of Misra and Nair (2008) where the quotas are updated after every bonus cycle. Furthermore, quota updating is not taken at an individual level but at a regional level and hence makes the incentive to ratchet smaller.

The rest of the paper is organized as follows. Section 2 summarizes the institutional details of the firm and data used within the empirical analysis. We present the model and the estimation methodology in sections 3 and 4. Section 5 discusses the estimation results and the counterfactual analysis. Section 6 concludes.

2. Institutional Details and the Data

The focal firm under study is a highly regarded multinational Fortune 500 company that sells durable office products. Products range from simple machines targeted for local small businesses with a price tag of less than a thousand dollars to highly sophisticated systems of machinery for multinational companies and government agencies costing several hundreds of thousands of dollars. In addition to its line of products, the firm also provides services such as equipment maintenance and system consulting.

The sales force is directly employed by the focal firm and is mainly divided into two groups: account managers and product specialists. The account managers mostly focus on selling basic products while the specialists sell more advanced and sophisticated products. There are several different types of product specialists each having different line of products. While evaluating performance, only one sales agent receives credit for a unit of sales and the firm traditionally does not encourage group work or team cooperation among the sales force.

The data consists of three years of monthly transactions from 1999 to 2001 where we observe the monthly revenues generated by each salesperson and the quarterly and annual quotas associated with each region.

The firm's compensation structure follows the pattern in Plan F of Figure 1. Every month, sales people receive a fixed monthly salary and commissions per volume of sales

³ Misra and Nair's (2008) approach to accommodate heterogeneity is similar to the estimation of individual level utility functions in conjoint analysis, by using a large number of observations related to a particular individual.

generated in that month. There are no caps on revenues for which an agent could obtain commissions. In the first three quarters, a quarterly lump-sum bonus is given when each of the quarterly quotas are met. And at the end of the year (or end of the fourth quarter) an annual lump-sum bonus is paid if the annual quota is met and an overachievement commission is given for cumulative revenues beyond the annual quota.

The firm utilizes the second most common form of compensation structure in regards to the survey of Joseph and Kalwani (1998) with three different components, the fixed component (the month salary), the linear component (commissions), and the non-linear component (lump-sum bonuses and overachievement commissions). Fortunately, the compensation structure of the focal firm features almost all of the dimensions of a compensation plan available; hence we can evaluate the sales force's response to many dimensions of the plan. The details of the focal firm's compensation schedule are described in Table 1.

We focus our study on the account managers because the product set that they carry is most comparable to the overall product set carried by the agents. We also restrict our analysis to 280 account managers who were present for the entire three year period of the data. This allows us to have panel information for at least three years in terms of each agent's response to annual quotas. The descriptive statistics of the data is in Table 2.

Figure 2a shows the average revenue generated for each calendar month by the sales force in the study. We can see the relatively high sales at the end of first three quarters (March, June, September) and a steep increase in the last quarter (December) relative to other months. At a first glance, this suggests that sales agents are highly responsive to quotas and bonuses.

Could there be other reasons other than quotas and bonuses that could contribute to higher sales in these quarters? For example, could it be possible that demand also happens to be greater during the period of high sales? Other than the direct sales force of which is the main focus of this study, the focal firm also utilizes an indirect sales force whose compensation structure is mainly that of pure commissions.⁴ Looking at their sales levels

⁴ The firm utilizes the indirect sales channel to perform sales activity in rural areas. The indirect sales channel is composed of approximately eight hundred small representative firms that resell the focal firm's products. The focal firm does not directly compensate the sales force of these rep firms. The representative firms simply receive commissions from the focal firm, and they in turn give a percentage of that commission to the sales agents.

over the different months, should help delineate the effects of quotas and bonuses as opposed to pure demand effects. Figure 2b shows the monthly revenue of the indirect sales force. One notices that the shape of average monthly revenue for the indirect sales force also mirrors that of the direct sales. We use this as a control for the monthly demand shifter in the model. Without this control, we may exaggerate the impact of quota-bonuses on sales.

Why should we expect sales to be so seasonal and coincide by quarter? Given that the focal firm's products are a B2B business selling discretionary good whose timing of purchase are also flexible, one possibility is that spending may expand during the month in which the fiscal year ends for its customers due to accounting procedures employed by clients. Figure 2c shows the distribution of fiscal year-ends across months for the year 2000 (the mid-point of our sample). Indeed, we see that over 66% of the firms have a December fiscal year-end, which might explain a significant boost in sales during that month. End of quarter months also have peaks relative to other months, but the skew is not as large. Hence fiscal year-ends may not explain all of the seasonality in sales experienced by the firm, but could be an important ingredient. In our data, the presence of the pure commission sales force turns out to be an attractive control for demand-side seasonality, allowing us to isolate the effects of quotas and bonuses on sales.

Figure 3 shows the average quarterly quotas over the 36 months. The quotas reflect the demand seasonality seen in the sales data, suggesting that the firm is clearly aware of the higher sales potential in certain periods.

3. Model

Before developing the model, we first illustrate how quotas and lump-sum bonuses can serve as a stretch goal to induce more effort in a simple static setting. Assume that there is a direct one-to-one mapping of sales and effort (effort equals sales) and that the disutility associated with effort is convex and utility from wealth is linear in proportion to the commission rate such that the utility function can be represented as

$$U^c(e) = -de^2 + re$$

where $-d$ is the disutility parameter and r is the commission rate. Assume further that $d=1$ and $r=10$. Then, with just a pure linear commission plan, a sales agent would exert effort up to the level that would maximize the above utility function, which would be at $e^{c*}=5$. For graphical illustration, see Figure 4.

Now, suppose that there is a lump-sum bonus B associated with meeting a quota Q then the utility function can be written as

$$U^b(e) = -de^2 + re + B \cdot 1_{\{e \geq Q\}}$$

Further, assume that $Q=10$, and $B=30$. Then the agents would maximize utility at an optimal effort level of $e^{b*}=10$ since they would obtain higher level of utility than at $e^{c*}=5$. So quotas serve to stretch agents and induce greater effort.

From the firm's perspective, in order to induce agents to work at e^{b*} by a pure linear compensation plan the commission rate r has to be increased to 20 and would cost 200. In contrast the cost for the quota-based plan to induce the same effort would only be 130. As a result, the firm would be more profitable by utilizing quota bonus plans to induce sales agents to work at an effort level of $e^{b*}=10$. In a dynamic model, the effort levels and the need to stretch to obtain a bonus would depend on the distance to quota. Hence a sales person needs to look forward dynamically when exerting effort in an early period in order to be in a "good" state to reach the quota and receive a bonus.

We now build a dynamic model of sales force response to a quota-based compensation scheme. We begin by discussing the timing of the decisions in the model. First, the firm chooses the compensation plan for individual i at the beginning of time t_0 . Then, conditional on the compensation plan, and their level of past sales which are sales agent specific states of the world, the sales agents exert effort. The effort plus an idiosyncratic sales shock that is realized after the agent's effort decision lead to observed levels of sales. The realized level of sales determines the states in the next period. In summary, the timing of the agents is as follows:

1. At the beginning of each year, firm chooses the annual compensation plan.
2. Each month, agents observe their current state and exert effort in a dynamically optimal manner.
3. An idiosyncratic sales shock is realized; the shock plus agent's effort determines the agent's realized sales for the period. Agent receives compensation.
4. The realized sales of the current period affect the agent's state of the next period. Steps 2-3 are repeated each month until the end of the year and Steps 1-3 are repeated over the years.

We next describe the model in four parts: (i) the compensation plan (ii) the sales agent's utility function (iii) the state transitions and (iv) the optimal effort choice by the sales agent.

3. 1. Compensation Plan

The compensation plan has three components. They are: (1) the monthly salary w_{it} , (ii) end-of quarter bonus, B_{qit} for achieving the corresponding quarterly quota Q_{qit} , and end of year bonus B_{yit} for achieving the corresponding annual quota Q_{yit} (3) commission rate r_{it} per dollar worth of sales and an overachievement commission rate, r_{it}' received at the end of the year for sales over and above the annual quota for each individual i at time t . We represent the compensation plan for a salesperson i by the vector $\psi_i = \{w_i, Q_{qi}, Q_{yi}, B_{qi}, B_{yi}, r_i, r_i'\}$.

3. 2 Sales agent's per-period utility

In each period, sales agent i receives a negative disutility from exerting effort $C(e; \theta_i)$ to obtain sales and a positive utility of wealth W_{it} based on the level of sales. The level of effort that the sales person will exert depends on the agent's state variables s_{it} which we will describe later. Thus the utility function is defined as:

$$U(e, y_{it}; \psi_i, \theta_i) = E[W(y_{it}; \psi_i)] - \gamma \text{var}[W(y_{it}; \psi_i)] - C(e; \theta_i) \quad ^5$$

where γ is the risk aversion coefficient for the agent. In reality, we do not observe effort, but only observe sales y_{it} . Hence we need to make assumptions about how effort and sales are related. We specify sales revenue y_{it} as a parametric *monotonically increasing* function of unobservable effort e_{it} and observable industry demand shifters and individual characteristics z_{it} with an *additive* idiosyncratic revenue shock ε_{it} . This shock is not observed by the sales agent when choosing the level of effort. Thus the sales production function for individual i at time t takes the form,

$$y_{it} = y(e_{it}, z_{it}; \alpha_i) + \varepsilon_{it} \quad (1)$$

where α represents the sensitivity to industry and individual observable characteristics z_{it} .

⁵ In the case of a CARA utility function (exponential utility function), this functional form represents the certainty equivalent utility of the agent.

Given the sales levels, and the compensation plan, the wealth for individual i W_{it} can be computed. W_{it} arises from four components, the per period salary component w_{it} , the lump-sum bonus component B_{it} , the commission component C_{it} , and the overachievement commission component OC_{it} . The detailed expressions of wealth is as follows,

$$W_{it} = w_{it} + B_{it} + C_{it} + OC_{it}$$

$$B_{it} = 1_q \cdot 1_{(s_{1it} + y(e_{it}, z_{it}; \alpha_i) + \varepsilon_{it} > Q_{qit})} \cdot B_{qt} + 1_y \cdot 1_{(s_{2it} + y(e_{it}, z_{it}; \alpha_i) + \varepsilon_{it} > Q_{yit})} \cdot B_{yt}$$

$$C_{it} = r_{it} \cdot y(e_{it}, z_{it}; \alpha_i)$$

$$OC_{it} = 1_y \cdot 1_{(s_{2it} + y(e_{it}, z_{it}; \alpha_i) + \varepsilon_{it} > Q_{yit})} \cdot r'_{it} \cdot (s_{2it} + y(e_{it}, z_{it}; \alpha_i) + \varepsilon_{it} - Q_{yit})$$

where s_{1it} and s_{2it} are the cumulative sales within the quarterly and annual quota periods. And 1_q and 1_y are indicator functions that take the value of 1 if t is at the quota-period-end for the quarterly and annual quotas and zero otherwise.

In our empirical analysis, we use a quadratic functional form for the disutility function; specifically, $C_i = \theta_i e^2$.

3. 2.2 State Transitions

In a linear compensation model with only commission incentives the sales agent faces a decision where there is a tradeoff between the disutility of effort and positive utility of commission from sales in the current period. However, the nonlinearity of the compensation scheme with quotas and bonuses introduces dynamics into the sales agent's behavior because there is an additional tradeoff between the disutility of effort today and a higher probability of lump-sum bonus and overachievement commission tomorrow. To incorporate the dynamics of the model we implement three state variables, the distance to annual quota, the distance to quarterly quota, and period type. While the first two variables evolve stochastically, conditional on the effort levels and revenue shocks of the previous periods, the period type evolves in a purely deterministic manner.

Specifically, the state variables evolve as follows:

1. Cumulative sales after the start of a quota-period – quarterly

$$s_{1it} = \begin{cases} 0, & \text{if } t \text{ is start of quarterly quota period} \\ s_{1i(t-1)} + y_{i(t-1)}, & \text{other wise} \end{cases}$$

2. Cumulative sales after the start of a quota-period – annual

$$s_{2it} = \begin{cases} 0, & \text{if } t \text{ is start of annual quota period} \\ s_{2i(t-1)} + y_{i(t-1)}, & \text{other wise} \end{cases}$$

3. Period type

$$s_{3it} = \begin{cases} 0, & \text{if } t = 1, 13, 25, \dots \\ s_{3i(t-1)} + 1, & \text{other wise} \end{cases}$$

3.3 Optimal Choice of Effort

Given the parameters of the compensation scheme ψ , and the state transitions, each sales agent would choose an effort level conditional on her states to maximize the discounted stream of expected future utility flow. The stream of utility flow, under the optimal effort policy function, and the behavioral notion of quasi-hyperbolic discounting can be represented by a value function,

$$V(s; \psi, \Omega) = \max_e U(e, s; \psi, \Omega) + \beta \cdot E[\max_e U(e, s'; \psi, \Omega) + \delta \cdot E[\max_e U(e, s''; \psi, \Omega) + \dots]]$$

or differently put as,

$$V(s; \psi, \Omega) = \max_e U(e, s; \psi, \Omega) + \beta \cdot E[V^\delta(s'; \psi, \Omega)]$$

where

$$V^\delta(s'; \psi, \Omega) = \max_e U(e, s'; \psi, \Omega) + \delta \cdot E[V^\delta(s''; \psi, \Omega)]$$

where β and δ are the hyperbolic parameters, and the expectation is taken with regards to the present and future sales shocks. The key idea of hyperbolic discounting is that individuals discount the immediate future from the present more than they do for the same time interval starting at a future date. The most frequently given example is the preference reversal shown between two delayed rewards. An individual may prefer 100\$ today to 120\$ in a year but may also prefer 120\$ in two years to 100\$ in a year. We apply the notion of hyperbolic discounting to capture the behavioral aspects regarding the effort choice of sales agents. We

shall see later in the results section that this assumption is quite candid and fits the model very well.

4. Estimation

Traditionally, nested fixed-point algorithm (NFXP) developed by Rust (1987) has been used to estimate such dynamic models. It is well known that NFXP estimators are computationally burdensome as one has to solve the dynamic program numerically over each guess of the parameter space for every iteration.

To reduce the computational burden, we use the recently developed two-step estimation method first introduced by Hotz and Miller (1993) and extended by Bajari, Benkard, and Levin (2007) to estimate the model parameters. In this approach, the model estimation proceeds in two steps. In the first step, the conditional choice probabilities of choosing a certain action as a function of state variables are estimated in a flexible non-parametric manner. Then, in a second step, these conditional choice probabilities are used to estimate the structural parameters of the agent's utility function. For this approach to work, it is critical that the conditional choice probabilities are estimated accurately in the first step. It was generally believed until recently that the accurate estimation of conditional choice probabilities for an agent is impractical when there is unobserved heterogeneity.

Arcidiacono and Miller (2008) recently developed an EM-Algorithm based approach to estimate the first stage conditional choice probability even when there is unobserved heterogeneity. Using this approach, we model unobserved individual level heterogeneity with relatively little additional computational complexity. This approach is in contrast to Misra and Nair (2008), who address the issue of unobserved heterogeneity by estimating the dynamic model at the level of each sales agent. This is feasible because they treat each observation as the sales at each client; and each sales agent has a large number of clients. Our approach however is more general and can be applied to a wider set of scenarios where there are limited observations on any one sales agent. We now discuss the two steps of the estimation procedure.

4.1 Step – One

Instead of estimating the parameters of the effort policy function for each individual, we estimate them via unobserved segment-types and hence obtain the likelihood of each individual belonging to each segment.

Let us assume that an underlying unobservable state $k \in \{1, \dots, K\}$ exists and let $\mathcal{L}(y_{it}|s_{it}, z_{it}, k; \Theta_k, \pi)$ be the likelihood of individual i 's sales being y_{it} at time t , conditional on the observables s_{it}, z_{it} , and the unobservable k , given parameters Θ_k and segment transition probabilities π . If we assume unobserved heterogeneity to be permanent, and hence allow for only initial segment probabilities, then the likelihood of individual i being in state k (and hence be of segment type k ; since unobserved heterogeneity is permanent) can be represented as,

$$L_k(y_i|s_i, z_i; \Theta_k, \pi) = \prod_{t=1}^T \pi_k \mathcal{L}_{ikt}$$

where π_k denotes the probability that an individual belongs to segment-type k . By summing over all of the unobserved states $k \in \{1, \dots, K\}$, we obtain the overall likelihood of individual i :

$$L(y_i|s_i, z_i; \Theta, \pi) = \sum_{k=1}^K L_k(y_i|s_i, z_i; \Theta_k, \pi)$$

Then, by Bayes' rule the probability that individual i is of segment-type k is given as

$$\Pr(k|y_i, s_i, z_i; \Theta, \pi) = q_{ik}(y_i, s_i, z_i; \Theta, \pi) = \frac{L_k(y_i|s_i, z_i; \Theta_k, \pi)}{L(y_i|s_i, z_i; \Theta, \pi)} \quad \dots \quad (1)$$

of which we represent as q_{ik} . Then as shown in Arcidiacono and Jones (2003), the maximum likelihood estimator of the sample must solve,

$$\sum_{i=1}^N \sum_{k=1}^K \Pr(k|y_i, s_i, z_i; \hat{\Theta}, \hat{\pi}) \frac{\partial \ln (L_k(y_i|s_i, z_i; \hat{\Theta}, \hat{\pi}))}{\partial \Theta} = 0$$

so that

$$\hat{\Theta} = \operatorname{argmax} \sum_{i=1}^N \sum_{k=1}^K \sum_{t=1}^T q_{ik} \log \mathcal{L}(y_{it}|s_{it}, z_{it}, k; \Theta, \pi) \quad \dots \quad (2)$$

Directly maximizing equation (2) can be computationally infeasible. As an alternative we perform the following iterative steps.

- (a) Guess initial parameters for Θ^0 and π^0
- (b) Compute q_{ik}^1 using equation (1) with Θ^0 and π^0 from above

- (c) Obtain Θ^1 by maximizing (2) evaluated at q_{ik}^1 and π^0
(d) Update π^1 by taking the average over the sample such that

$$\pi_k = \frac{1}{N} \sum_{i=1}^N q_{ik}$$

- (e) Iterate (a) – (d) till convergence

For the functional form of the effort policy function, we use Chebyshev polynomials of state variables to approximate effort.⁶ As a result, we obtain the estimated effort policy $\hat{e}_k(s)$ the sales policy $\hat{y}_k(e, z)$, and the distribution of the segment specific error term ε_k , $\hat{F}_k(\varepsilon)$.⁷

4.2 Step – Two

With the estimated parameters of the effort policy function and hence the sales policy function, and the distribution of the revenue shocks in the 1st stage, we then forward simulate the actions of sales agents to obtain the estimate of the value function $\tilde{V}(s; \hat{e}(s); \psi, \Omega)$ for state s . If $e(s)$ is the effort policy and thus at an optimum, then any deviations from the policy rule would generate value functions of less or equal value to that of the optimal level. We thus simulate a suboptimal effort policy rule $\hat{e}^s(s)$ and obtain the corresponding estimate of the sub-optimal value function $\tilde{V}(s; \hat{e}^s(s); \psi, \Omega)$.⁸

Let us define the difference in the two value functions as,

$$Q(x; \psi, \Omega) = \tilde{V}(s; \hat{e}(s); \psi, \Omega) - \tilde{V}(s; \hat{e}^s(s, x); \psi, \Omega),$$

where x denotes a suboptimal policy. Then if $e(s)$ is the optimal policy, the function $Q(x; \psi, \Omega)$ would always have value greater or equal to zero. Thus our estimate of the underlying structural parameters Ω would satisfy,

$$\hat{\Omega} = \operatorname{argmin} \int (\min\{Q(x; \psi, \Omega), 0\})^2 dH(x)$$

⁶ For reference, see “Numerical Methods in Economics”, Kenneth L. Judd, MIT Press, 1998. We use a fourth order Chebyshev approximation method.

⁷ We assume a normal distribution for the parametric form of the revenue shock distribution.

⁸ As indicated in Bajari, Benkard, and Levin (2007), there are multiple ways to draw these suboptimal policy rules. Although the method of selecting a particular perturbation will have implications for efficiency the only requirement necessary for consistency is that the distribution of these perturbations has sufficient support to yield identification. We chose to draw a deviation policy from a normal distribution with mean zero and quarter of the variance from the revenue shock distribution at a particular state. i.e. $e^s(s) = e(s) + \varepsilon$

where $H(x)$ is the distribution over the set χ of inequalities. The above estimation is performed for each segment with the segment specific effort policies obtained in Step1. This allows us to estimate the structural parameters for each segment.

4.3. Identification

We provide a brief and informal discussion of identification. Realized sales is a function of effort and additive sales shocks. Given multiple observations of sales at different states, we can separately identify non-parametrically the density of sales shocks and a deterministic function of effort. We make a parametric assumption about a strictly monotonic relationship between sales and effort because it is not possible to identify this relationship non-parametrically.

Further, we assume a deterministic (but highly flexible) relationship between effort and observable states (distance to quarterly and annual quotas and the month) at the level of each segment. Thus variation in sales (which is monotonically linked to effort) as a function of distances to quota over time helps identify the effort disutility parameter. The risk parameter is identified by differences in response to variations in wealth levels over time.

The discount factor is typically not identified in dynamic choice models because instruments affect both the current utility and future utility (Rust 1994; Magnac and Thesmar 2002). In our model, the states serve as identifying instruments. In non-bonus periods, current payoffs are unaffected by the current states, but only affects the future payoff through the value function. Such instruments are not typically available in most applications of discrete choice models, but these help us identify discount factors in our model.

5. Results

We begin with the first stage estimates of the effort policy function; then we report estimates of structural parameters of sales agents' utility functions from the second stage. In doing so, we perform a grid search over the different hyperbolic parameters, deltas and betas to identify the optimal discount factors that fit the data the best. We then perform several counterfactual simulations to address the substantive questions we seek to answer.

5.1 First Stage Estimates

As discussed, we obtain segment level effort policy functions by estimating the non-parametric relationship between sales and state variables through Chebyshev polynomials of

the state variables (distance to annual and quarterly quotas). In addition, we also allow for observed characteristics (e.g., tenure) to affect the effort policy function.

Using the Bayesian Information Criterion (BIC) for model selection, we choose a three-segment model. We do not report the parameter estimates themselves, given that the coefficients associated with the Chebyshev polynomials have no intuitive meaning. However, to get some intuition for the effort policy function, we show graphs of the effort policy function (which maps into equivalent dollar revenues) for the three segments as a function of distance to annual quota in Figure 5. We report the share of the three segments and some descriptive characteristics of these segments in Table 3. Segment 1 is the largest with a share of 56%, with Segments 2 and 3 each having shares of 32% and 12% respectively. The average tenure of the sales force within each segment is not very different. The quarterly and annual quotas for Segment 1 are about 32% larger than those of Segment 2, which in turn is about 13% larger than those of Segment 3. Interestingly, however Segment 1 with large quotas meets the quotas far more often than Segment 2, while Segment 3 hardly ever meets the annual quota (4%). This suggests that quotas reflect some measure of the potential of the market, and the sales force in segment 3 is in charge of low potential markets.

From Figure 5, we see a positive relationship between induced effort and the annual distance-to-quota (DTQ). Segment 1 representing about 56% of the sales force, show an increasing relationship between December sales (effort) at low DTQ, but the rate decreases as the sales person gets close to quota. Interestingly, average December sales does not fall off as even if the person has already reached or exceed quota ($DTQ > 1$). In fact the person continues to produce roughly maximum sales even after being 20% above quota. In fact, peak December sales is at 20% above quota.

This finding is consistent in shape with the inverted-U relationship reported in Steenburgh (2008). However, the steepness of the drop in effort is not as dramatic as that observed in Misra and Nair (2008), perhaps due to the fact that the overachievement bonus mitigates any possible reduction in effort.

The smaller segment 2 however shows a linear relationship between distance to quota and December sales. This group is motivated to achieve as much sales as possible and do not reduce effort even on reaching quota. However, they are less productive on average than the sales force in segment 1, partly because their territories have less potential as suggested by the quota numbers in Table 3.

Segment 3 shows a pattern similar to Segment 1, but is considerably less productive and gives up much before they reach the quota. In fact maximum December sales are for those who have reached 45% of quota. This suggests that bonuses are not a huge factor for this small segment of the sales force, In fact only 4% of the sales force actually achieves the annual quota supporting our conclusion.

5.2 Discount Factor

We performed a grid search over the exponential discount factors. Conditional on each discount factor, we estimate the structural parameters of the model. Then, with the estimated structural parameters, we simulated revenues of sales agents and compare them with the actual revenues. The best fitting discount factor based on the mean absolute percentage errors between actual and predicted revenues was 0.9. Given that this is for a month, this turns out to be an annual discount factor of 0.31, which is considerably lower than a discount rate based on monthly discount factor.

We also performed a grid search over each set of hyperbolic parameters. Table 4 presents the mean absolute percentage errors (MAPE) associated with each set of hyperbolic parameters. We can see that a beta of 0.7 and a delta of 0.95 fits the model the best with regards to MAPE. Thus our estimates show a distinct present bias. Given that these are based on monthly data, the annual discount rate is effectively 0.42. Frederick and Loewenstein (2002) have a comprehensive summary of the estimated discount factors from previous studies. The summary shows that the estimated discount factors vary extensively ranging from as low as a mere 0.02 to no discounting at all with a discount factor of 1. Our estimates of 0.7 and 0.95 seems reasonable given that range. For pure monetary value, this discount factor seems rather low but as Frederick and Loewenstein (2002) have indicated, for behavioral aspects such as pain and thus in our case effort, the discount factors tend to be low and hence our estimates do not appear unreasonable.

5.3 Second-Stage Structural Parameter Estimates

Table 5 reports the structural parameter estimates of the sales agent's utility function. Overall, the disutility parameters for all three segments are negative and highly significant. These estimates are consistent with the effort policy functions estimated in the first stage. Segment 1, which produces the greatest sales on average, has the lowest average disutility parameter. Segment 3, which has the lowest productivity, has the greatest disutility

parameter. The risk aversion coefficients for all segments are insignificant showing no direct evidence of risk aversion by the sales agents. Perhaps this is because in the range of incomes for the sales force, risk aversion is not a serious concern for the sales force .

How well do the estimates of the structural model fit the observed behavior? We compare the actual average revenues with predicted average revenues over an annual basis based on the dynamic structural model. As we can see from Figure 5 the estimated model fits the actual data extremely well with a MAPE of 7.7%.

5. 4 Assessing the value of a dynamic structural model

How important are dynamics in our structural model? How well would the model fit the data without the dynamics? To address this, we compare the revenue and effort predictions from a static versus a dynamic model for a sales agent. For the static model we simply set the two hyperbolic discount factors to zero. Figure 6a shows the comparison of the revenues generated when sales agents are forward looking versus myopic. We find that the revenues are systematically lower for the myopic sales agent. However, we do find that sales increases systematically at the end of every quarter and end of year even when agents are myopic because of bonuses and overachievement commissions.

Does the sales revenue peaks at the end of each quarter arise due to demand cyclicity as opposed to greater effort on the part of the sales agent? To assess this, we graph the effort levels in the static and dynamic setting in Figure 6b. What is most interesting is that for the myopic consumer, the effort levels do not respond as much to bonuses at the end of the quarter as they do for dynamic agents. This suggests that any cyclicity in revenues that we observe in the myopic model of Figure 6a is mostly due to demand cyclicity (as captured by our indirect agent sales levels).

To get a qualitative understanding of why effort is not affected as much by quotas and bonuses at the end of bonus periods for myopic agents verses dynamic agents, we assess the overall utility (Wealth from compensation- Disutility from work) against effort for a representative sales agent belonging to segment 1 as a function of their current state.

For a myopic sales agent, it is easy to see that in non-bonus periods ($t=1, 2, 4, 5, \dots, 11$), since there is no lump-sum bonus and with no future to worry about (due to myopia), the sales agent would simply tradeoff the negative disutility of effort with the positive utility of expected wealth from commissions in a way where the marginal utility from expected wealth

equals the marginal cost of disutility. With a convex disutility function and a linear commission rate the expected total utility would be concave in effort and the optimal effort level would be the same throughout the non-bonus periods, regardless of agent's state.

In bonus periods, the optimal effort would differ by the state of the world that each agent is in. Figure 7a – 7c shows the relationship between expected utility and effort when the state (DTQ) is far from the quota, moderately close from the quota, and very close from the quota, respectively. If the quota is out of reach as in figure 7a, then the sales agent would give up reaching the quota. However, this does not mean that she will exert no effort at all. The agent will behave as if she was in a non-bonus period and exert effort to the degree where the marginal utility from commission offsets the marginal disutility of effort. When the quota is moderately achievable, as in figure 7b, then the agent would exert extra effort compared with non-bonus periods so as to achieve the quota. Finally, if the quota is very close, then the agent would exert effort not only to the level where the quota is met but also to the level where the overachievement commission offsets the disutility of effort, as in figure 7c. The 3 dimensional plot of the optimal effort conditional on the states in the bonus period is presented in Figure 8.

This gives us the intuition for why myopic sales agents exhibit lower effort in bonus and non-bonus periods throughout the year. A forward-looking sales agent would want to be at a better state in future bonus periods and hence exert greater effort even in non-bonus periods. With such extra effort in every period, the agent remains more likely to be at a state close to quota when they reach the end of the year.

5.5 Counter-Factual Simulations

We now perform a series of counterfactual simulations that address the two sets of substantive questions we wish to answer. First, we address the issue of how valuable different components of the compensation plan are. Second, we assess the importance of quota-bonus frequency in a compensation plan.

We now perform a series of counterfactual simulations to take advantage of our structural model to evaluate the impact and importance of the different components of the compensation plan.

Value of Quotas and Bonuses

We compare changes in revenues and profits when the firm moves from the current compensation plan to a pure commission-only plan. We consider two cases: (1) where the commission rate is the same as the current commission rate; and (2) a higher commission rate is such that total compensation is exactly equal to the current compensation. We find that the revenues are about 18.5 % greater with the current compensation plan compared to a pure commission plan; even when adjusting commission rates to be higher to make total compensation identical to current levels, we find that revenues are about 4.6% higher.

Value of Overachievement Compensation

We compare changes in revenues and profits when the firm eliminates the over-achievement commission rate, which motivates sales people who are close to reaching their quota to continue exerting effort. Figure 9a shows comparative revenues with and without overachievement commissions. Overall revenues drop by 6.7% and even accounting for the additional commission costs, profits are lower by 6.1% (assuming gross margin of 33%). As can be seen the revenues begin to systematically drop early on in the year without the over-achievement commission.

To verify the crucial role of the overachievement commissions Figure 9b plots the effort level of sales agents who met and didn't meet the annual quota, respectively. For those of who met the annual quota, we can see that the effort level does not decline even when close to the quota. One would think that when a sales agent is closed to achieving the quota she would likely slow down. But because every sales amount surpassing the quota gets rewarded with a higher commission rate the sales agent keeps on exerting effort. For those who didn't meet quota, they are likely to have incurred consecutive negative sales shocks early on in the year. Since the probability of meeting quota is relatively low they start to decrease effort. For these people the overachievement commission plays no role and hence we can see that they exert minimal level of effort towards the year-end. We can see that the key role of the overachievement commission is to provide incentives for the sales force to keep on exerting effort even if they have already met quota (or highly likely of meeting quota).

Value of Cumulative Annual Quota

Figure 10 represents the average revenue per month when we take away the overachievement commission and the annual quota is replaced with a fourth quarter quota such that the amount

is equal to the annual quota and the 4th quarter quota is set as the difference between the annual quota and the sum of three previous quarterly quotas. As one can see, monthly average revenues systematically drop starting from early on due to not having the cumulative annual quota at the end of the year. Overall, revenues drop by 9.9%. This decrease is greater than the counterfactual in figure 9, the case where we just drop the overachievement commissions, showing that the cumulative nature of annual quotas are induce sales agents to exert greater effort and raise revenues by 2.2%.

Quota-Bonus Frequency

We next investigate the value of quarterly bonuses relative to annual bonuses. We show in figure 11a, how revenues respond when quarterly bonuses are eliminated. We find a consistent reduction in performance over the entire year without quarterly quotas. Overall revenues fall by 13.6%. More interestingly, although we have kept the year-end bonus constant (annual bonus + overachievement commission), year-end revenue falls by 7% and effort level falls by 14%. Thus we see that annual quotas and over-achievement commissions have less of an impact on year-end performance without quarterly bonuses. Why?

Not only does the quarterly quota induce sales agents to work harder in a given quarter but it also helps them achieve the annual quota by giving the incentive to stay on track of their annual goals (see Figure 11b). Although the annual bonus has the largest monetary value among bonuses, it also is temporally the furthest away and hence discounted a lot by sales agents early in the year. The quarterly bonus acts as an incentive early on in the year and helps the agents get to a state where the annual quota is achievable at the end of the year. When the quarterly quota is removed, sales agents no longer have the incentive to work hard early on. And by the time they get to later periods, it is more likely that they are quite a distance away from achieving the annual quota. At this point, annual quotas and over-achievement commissions have little impact on sales force performance because sales agents likely give up meeting the quota. Consistent with this outcome, we find that even when we sum-up all the bonuses (3 quarterly bonuses + annual bonus) and give it to agents at the end of the year on achieving the annual quota we see a 5% reduction in effort and approximately 4% decline in revenues and profit, supporting our claim about the importance of quarterly quotas.

To the best of our knowledge, there has been no analysis to-date on what is the appropriate frequency of quota and bonuses. There has been some descriptive work in the

education literature on how frequent testing affects academic performance (for an extensive survey, see Bangert-Drowns et al. 1991) and some experimental work in the behavioral psychology (Heath, Larrick and Wu, 1999). The basic idea is that achieving short-term goals make achieving long-term goals more feasible. But this does not mean that one may increase the frequency of quota-bonuses indiscriminately. There is a tradeoff in terms of payouts and performance. In the education literature, more frequent testing takes time away from classroom learning, but also motivates students to learn earlier material better and helping them do better on later material. In a similar way being on target with the sub-goal early on makes it more likely that one will be closer to achieving longer-term targets and therefore increases performance.⁹

6 . Conclusion

Even though personal selling is a primary marketing mix tool for most B2B firms to generate sales, there is little research on how the compensation plan used to motivate the sales force affects performance. This paper developed and estimated a dynamic structural model of sales force response to a compensation plan with many components: salary, commissions, lump-sum bonus for achieving quotas, and different commission rates beyond achieving quotas. Our analysis helped us assess the impact of (i) different components of compensation and (2) frequency of quotas and bonuses on performance.

There has been a fair amount of controversy on the value of quotas and bonuses in the literature. Overall, we find that the quota-bonus scheme used by this firm increased performance effectiveness of the sales force. Thus quotas serve as stretch goals overall in order to push employees to accomplish targets. Yet, features such as overachievement compensation reduces the problems associated with sales agents slacking off when they get close to achieving their quota.

Further, quarterly bonuses serve as a continuous evaluation scheme to keep sales agents within striking targets of their annual quotas. In the absence of quarterly bonuses, a failure in the early periods to accomplish targets caused agents to fall behind more often than in the presence of quarterly bonuses. Thus, the quarterly bonus serves as a valuable sub-goal

⁹ Unlike the rest of the analysis where the value of bonuses appear less important due to the low hyperbolic discount factors, quarterly bonuses appear relatively more valuable because of their immediacy, relative to the annual bonus.

which helps the sales force stay on track on their overall goal. A key finding of the paper is that annual bonuses are not as effective for sales forces without quarterly bonuses.

We used recent innovations in the two-step dynamic structural model estimation to accommodate unobserved heterogeneity in sales force response. The approach is flexible, yet computationally feasible with minimal additional burden compared to traditional two-step methods. Features of our data allow us to separate seasonality in sales due to quotas as opposed to underlying consumer demand seasonality. This enables us to get better estimates of the response to compensation, because demand peaks that coincide with quota periods may be wrongly interpreted as a by-product of compensation.

We now discuss limitations of the paper, which provide promising avenues for future research. First, the paper has not included potential problems due to ratcheting. Preliminary analysis suggests ratcheting affects are weak in our data, because the firm does not simply raise quotas in response to above quota performance. The increase if any combines not only one's own individual performance, but also the performance of others. This implies that when one reduces effort on reaching quota for fear of getting higher quotas next year, the first order loss is the loss of commission; the additional sales' impact on next year's quota is muted by other comparable sales agent's performance.

A second issue is that effort tends to be multi-dimensional and one possibility is that quotas and bonuses may force people to focus on the effort that lead to final sales, while in other periods, they may focus on earlier stages of the selling process. Our current model does not allow for such a multi-dimensional selling process because it is not possible to identify such a multidimensional effort merely from the sales data. Nevertheless new data from CRM databases which track customer stages through the selling process may be able to shed light on this issue. We believe this is an exciting area for future research.

A third issue is how compensation contracts serve as a selection mechanism to draw the right type of sales people into the sales force. This paper does not address the selection issues. By looking at a longer panel of sales people's performance, one can use attrition information to be able to shed light on this issue. Further, looking at scenarios where contracts varied over time, one can shed better light on this issue. Papers that have looked at varying contracts over time typically have focused on only contracts with linear commission rates (e.g., Paarsch and Shearer 1998). One needs more work on scenarios with richer contracts.

In summary, this paper has shed important insights on how the sales force responds to a very rich compensation structure involving many components of compensation: salary, commissions, quota and bonuses at quarterly and annual frequencies. How employees respond to such rich compensation structures with bonuses, a reality at many firms, has not been investigated at all in the literature. This paper illustrates a rigorous framework to analyze this problem and obtains useful substantive insights. Nevertheless, the issues raised above provide an interesting agenda for future work.

References

- Albers, Sönke, Murali K. Mantrala and Shrihari Sridhar (2008): A Meta-Analysis of Sales Force Response Elasticities, Report 08-100, Marketing Science Institute.
- Bajari, Patrick and Benkard, C. Lanier, and Levin, Jonathan, “Estimating Dynamic Models of Imperfect Competition”, *Econometrica*, 2007.
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1991) Effects of frequent classroom testing, *Journal of Educational Research*, 85, pp. 89-99.
- Bijmolt, T.H.A., H.J. Van Heerde, R.G.M. Pieters. 2005. New Empirical Generalizations on the Determinants of Price Elasticity. *Journal of Marketing Research* 42(May) 141–56.
- Copeland, A. and Monnett, C. (2008), "The Welfare Effects of Incentive Schemes," *Review of Economic Studies*, forthcoming.
- Darmon, Rene (1997), “Selecting Appropriate Sales Quota Plan Structures and Quota-setting Procedures,” *Journal of Personal Selling and Sales Management*.
- Elster, Jon, “Ulysses and Sirens: Studies in Rationality and Irrationality” *Cambridge University Press*, 1979
- Fehr, Ernst and Götte, Lorenz, “Do Workers Work More if Wages are High? Evidence from a Randomized Field Experiment”, *American Economic Review*, 2007
- Frederick, Shane and George Loewenstein and Ted O’Donoghue, “Time Discounting and Time Preference: A Critical Review”, *Journal of Economic Literature* 40, 2002
- Heath, Chip, Richard P. Larrick and George Wu (1999) "Goals as Reference Points," *Cognitive Psychology* **38**, 79 – 109
- Holmstrom, Bengt and Milgrom, Paul, “Aggregation and Linearity in the Provision of Intertemporal Incentives”, *Econometrica*, March, 1987
- Hotz, V. J. and Miller, R. A., “Conditional Choice Probabilities and the Estimation of Dynamic Models”, *Review of Economic Studies*, 1993

Joseph, Kissan and Manohar Kalwani (1998), "The Role of Bonus Pay in Sales force Compensation Plans", *Industrial Marketing Management*.

Judd, Kenneth L., "Numerical Methods in Economics", *MIT Press*, 1998

Laibson, David, "Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics* 112, 1997

Laibson, David, "Life-Cycle Consumption and Hyperbolic Discount Functions", *European Economic Review* 42, 1998

Lal, Rajiv and V. Srinivasan, "Compensation Plans for Single-and Multi-product Sales forces: An Application of the Holmstrom – Milgrom Model", *Management Science*, 1993

Larkin, I. (2006). "The Cost of High-Powered Incentive Systems: Gaming Behavior in Enterprise Software Sales, *Working Paper*, Harvard Business School.

Magnac, Thierry and D. Thesmar, "Identifying dynamic discrete choice models", *Econometrica*, 2002, 70:801-16

Misra, Sanjog and Harikesh Nair (2008), "Quota Dynamics and the Intertemporal Allocation of Sales – Force Effort", Working Paper, Stanford University.

Oyer, Paul (1998), "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality", *Quarterly Journal of Economics*.

Paarsch, Harry, J. and Bruce Shearer (1999), "The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry," *Journal of Human Resources*, 34, 643-667.

Phelps, E. S. and Robert Pollak, "On Second-Best National Saving and Game-Equilibrium Growth", *Review of Economic Studies* 35, 1968

Rust, John, "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, 1987.

Samuelson, Paul, "A Note on Measurement of Utility", *Review of Economic Studies* 4, 1937

Steenburgh, Thomas, "Effort or Timing: The Effect of Lump-Sum Bonuses", *Quantitative Marketing and Economics* 6, no. 3 (September 2008): 235-256.

Sudhir, K., "Competitive Pricing Behavior in the US Auto Market: A Structural Analysis," *Marketing Science*, Winter, 2001

Thaler, Richard H., "Some Empirical Evidence on Dynamic Inconsistency", *Economic Letters* 8, 1982

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

Table 1: Firm's Compensation Plan

Type	Description
Quarterly Bonus	Awarded if quarterly revenue exceeds quarterly quota – Q1, Q2, Q3
Annual Bonus	Awarded if annual revenue exceeds annual quota – Dec
Base Commission	Paid in proportion to the revenue generated each month – Every month
Overachievement Commission	Paid in proportion to the total cumulative revenue surpassing the annual quota – Dec

Table 2: Descriptive statistics – Sales force under study

	size	Average tenure (months)	Average full- year quota (\$K)	Average quarterly quota (\$K)
Sales force in study	280	134.9	1,467.1	303.1

Table 3: Descriptive Characters via Segments

	segment1	segment2	segment3
Share (%)	56	32	12
Tenure*	133	139	132
Average Quarterly Quota**	415	316	278
Average Annual Quota**	1,658	1,263	1,113
% time quarterly quota is met	49	38	21
% time annual quota is met	56	40	4

* Tenure is measured in months

** Average quotas are indicated in USD(K)

Table 4: Optimal Discount Factor – Model Fit**Mean Absolute Percentage Error by Discount Factors**

		□				
		0.9	0.95	0.97	0.98	0.99
□	0.6	0.080	0.079	0.091	0.096	0.098
	0.7	0.081	0.077	0.091	0.100	0.102
	0.8	0.089	0.083	0.087	0.104	0.104
	0.9	0.090	0.092	0.094	0.106	0.102

Table 5: Parameter Estimates

	Segment 1	Segment 2	Segment 3
<i>risk aversion coefficient</i>	-0.0000 (0.0002)	0.0001 (0.0002)	-0.0003 (0.0002)
<i>disutility</i>	-0.122 (0.002)	-0.180 (0.038)	-0.207 (0.022)

Figure 1a: Distribution of Incentive Compensation Scheme; Joseph & Kalwani (1998)

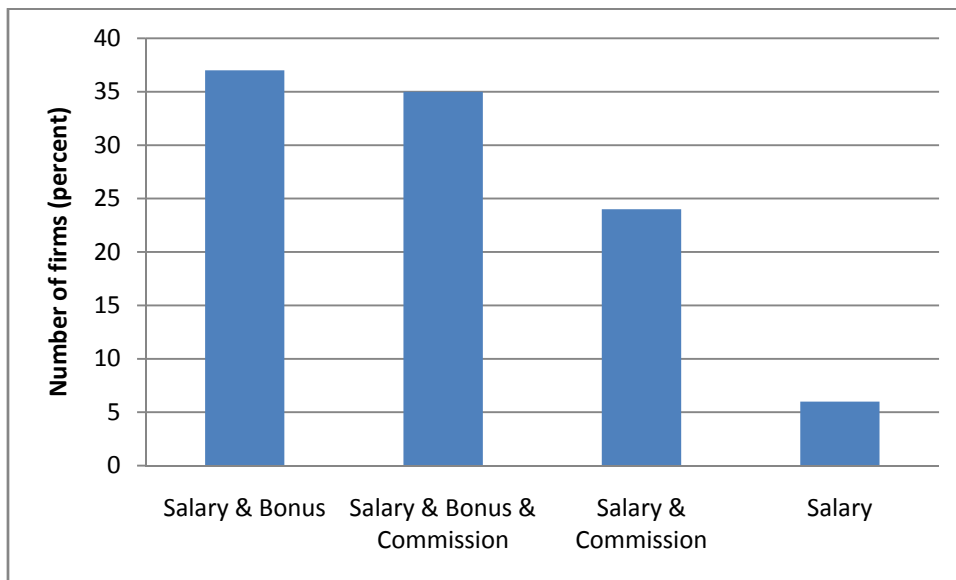


Figure 1b: Plots of Incentive Compensation Schemes

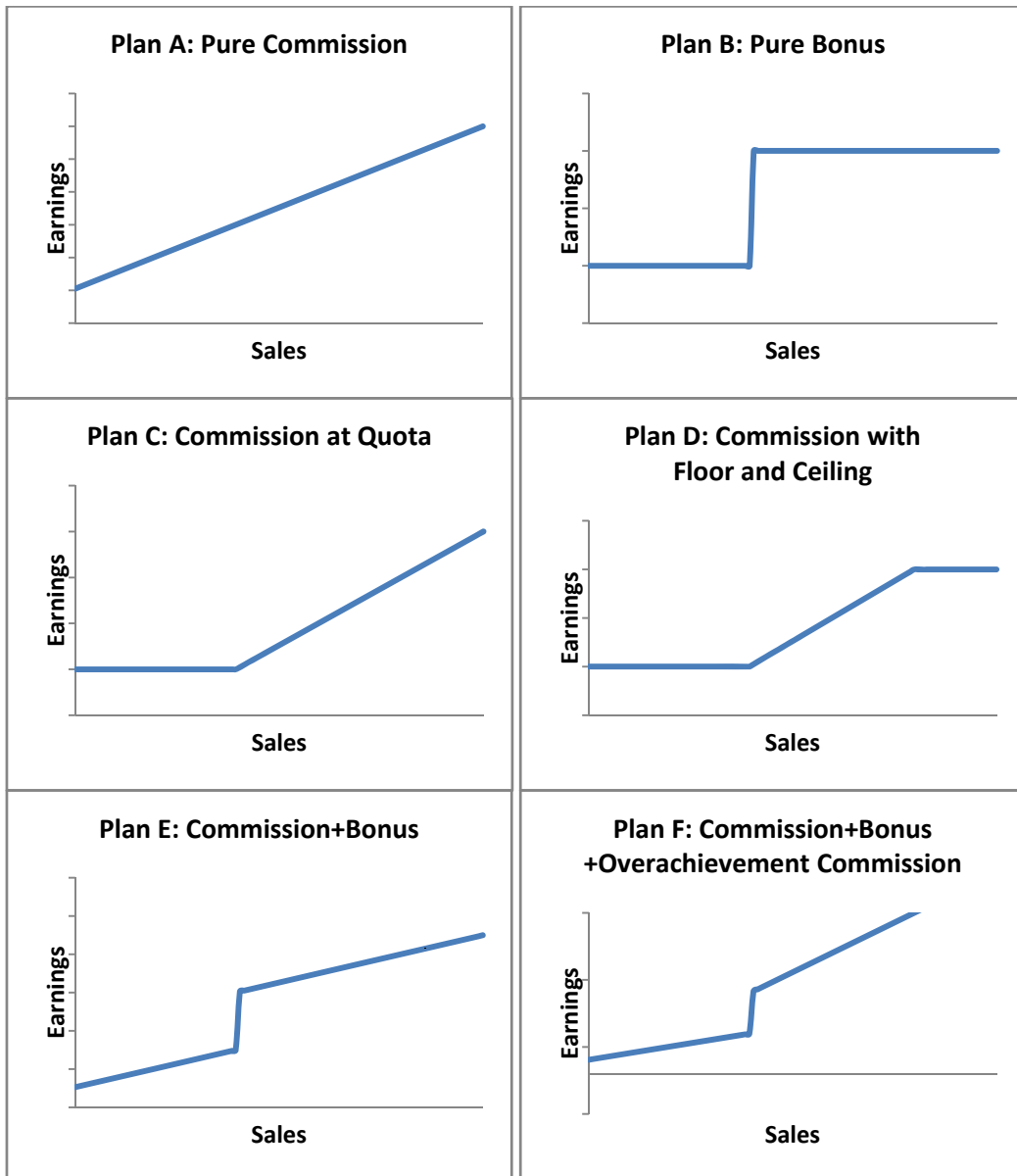


Figure 2a: Monthly Average Sales – Aggregate

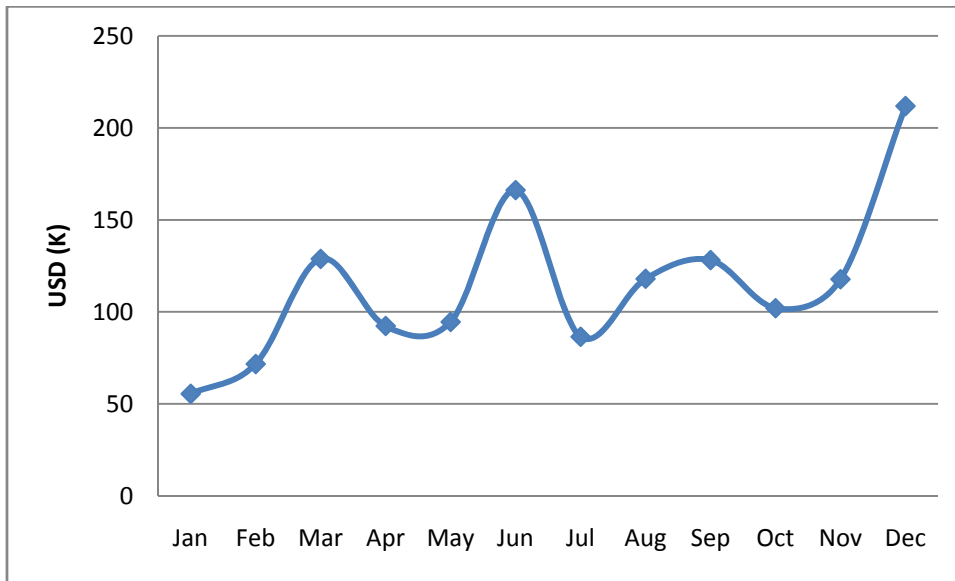


Figure 2b: Average Sales per Month – Indirect Sales Force

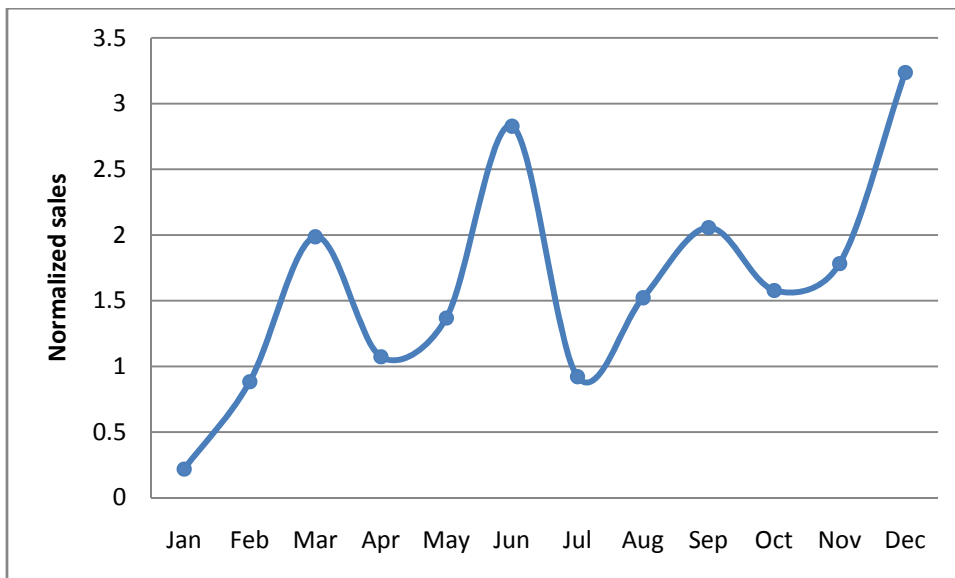


Figure 2c: Percentage of Fiscal Year-End Periods – 2000

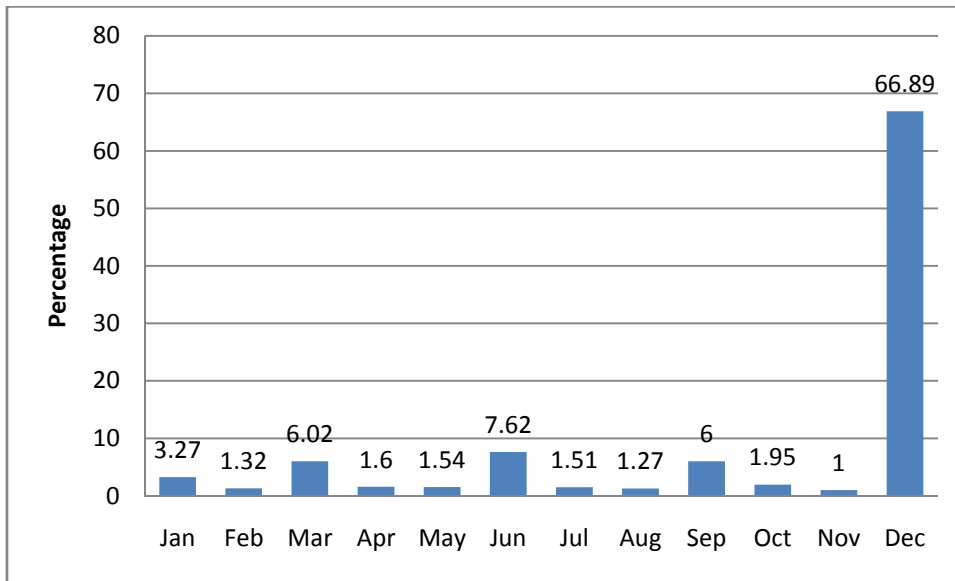
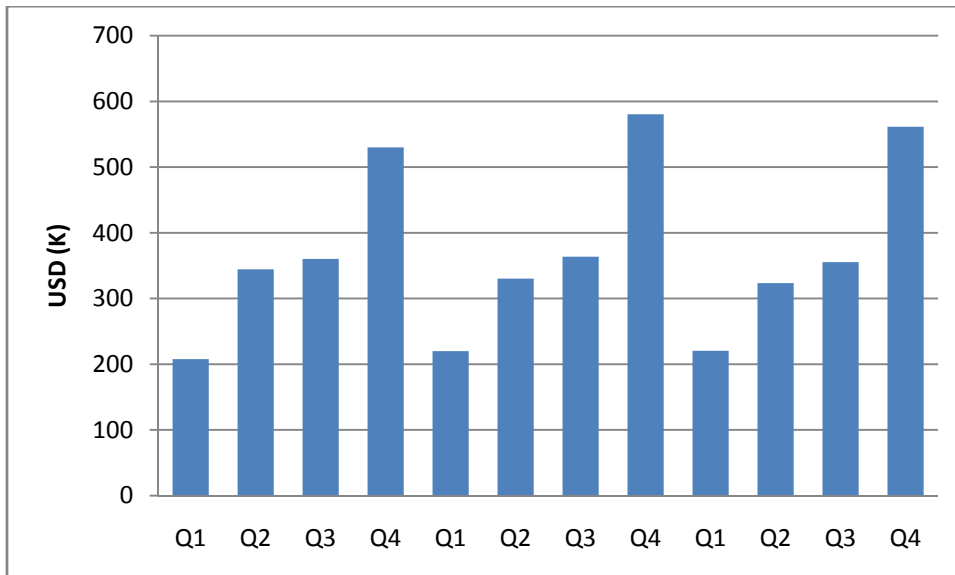


Figure 3: Average Quota per Period¹⁰



¹⁰ Fourth quarter quota is represented as the difference of the annual quota and the sum of past three quarterly quotas.

Figure 4: Quota as a stretch goal

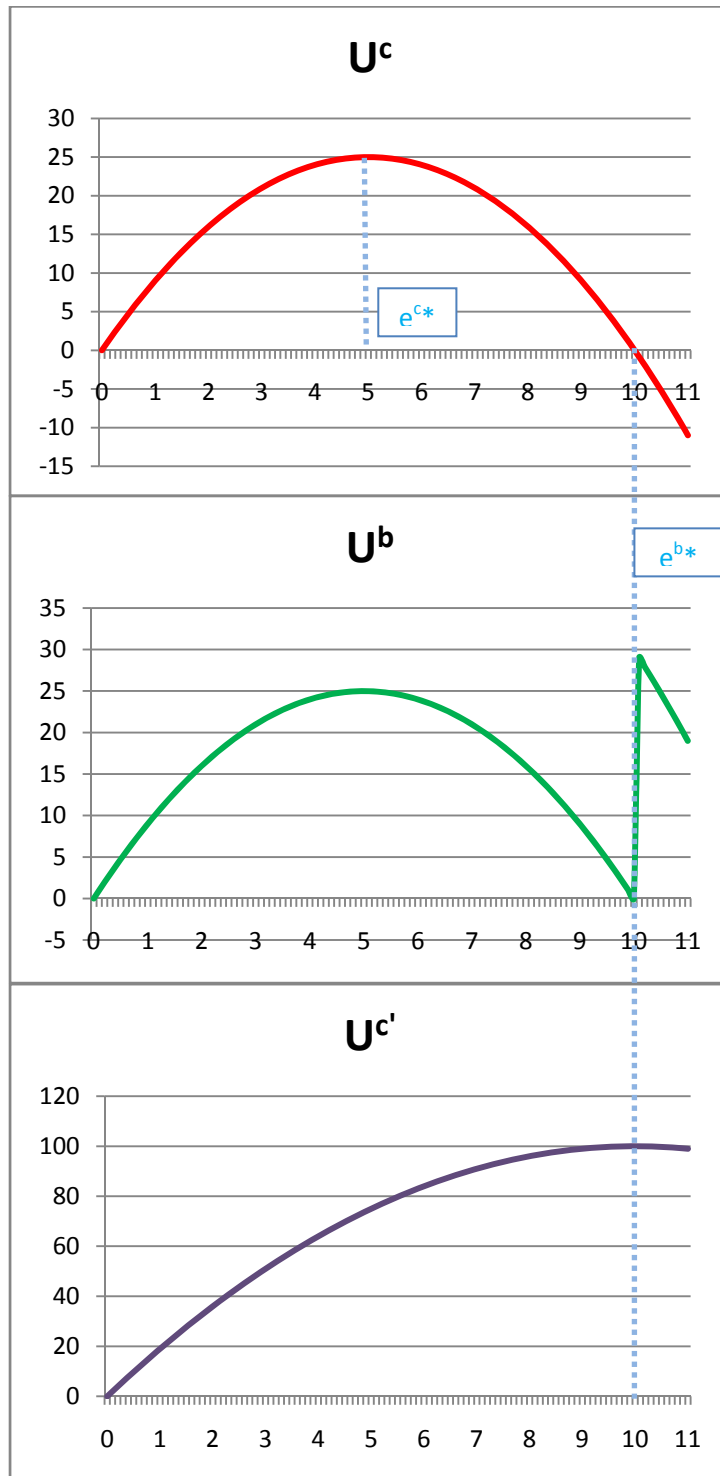


Figure 5: Effort Policy as a function of distance to quota

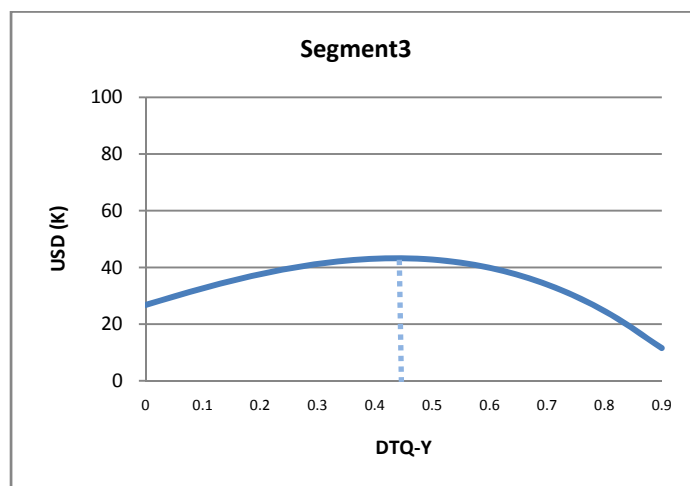
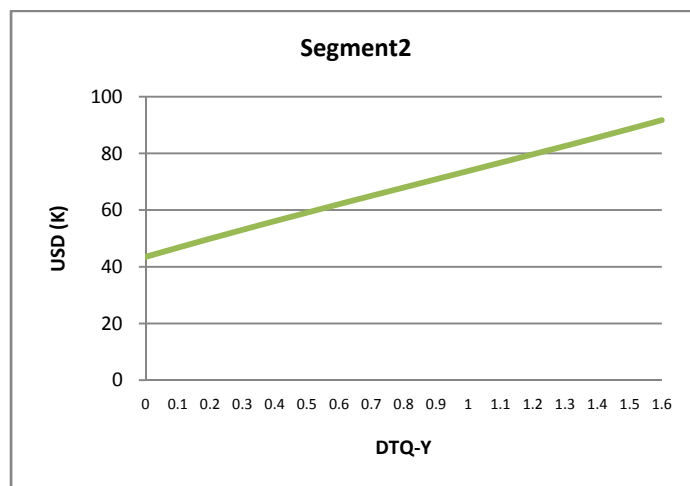
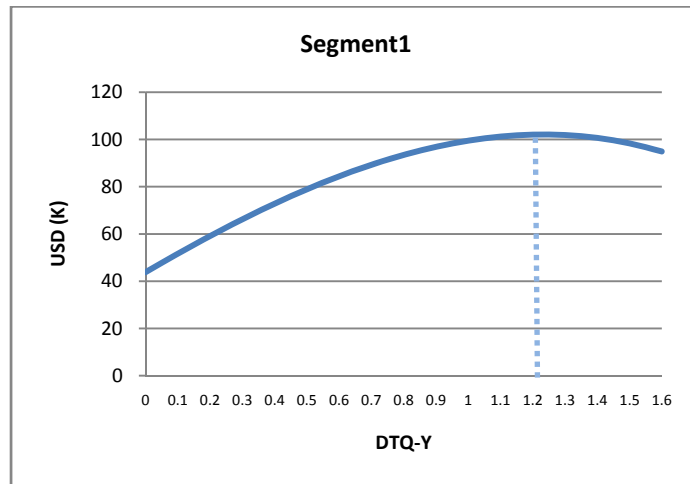


Figure 5: Simulated vs. Actual Revenue

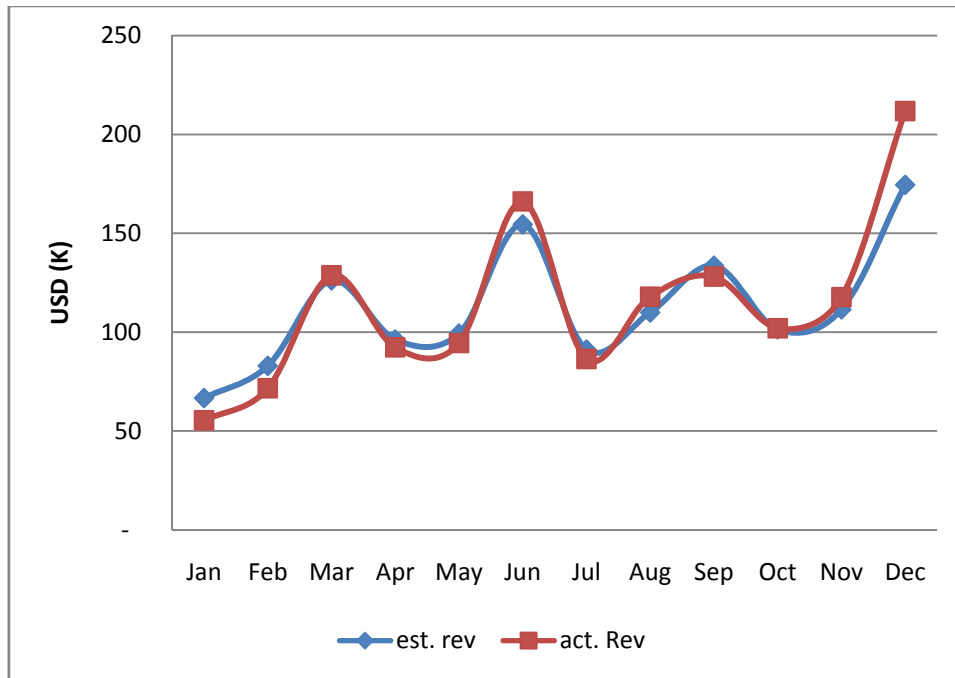


Figure 6: Simulated Revenue & Effort– Static vs. Dynamic

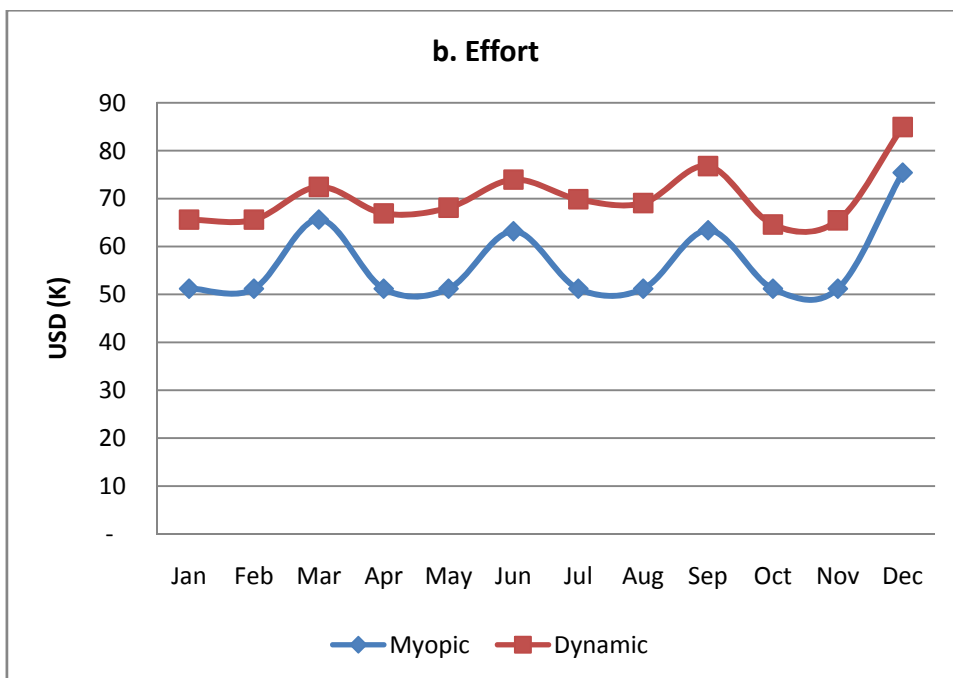
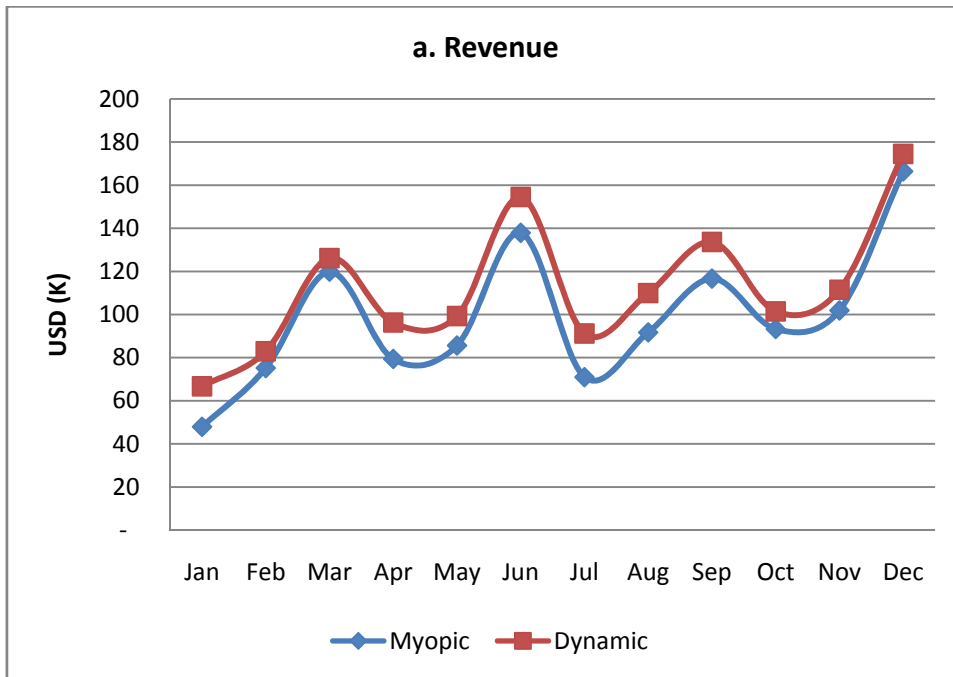


Figure 7: Expected Utility – Bonus periods – Static

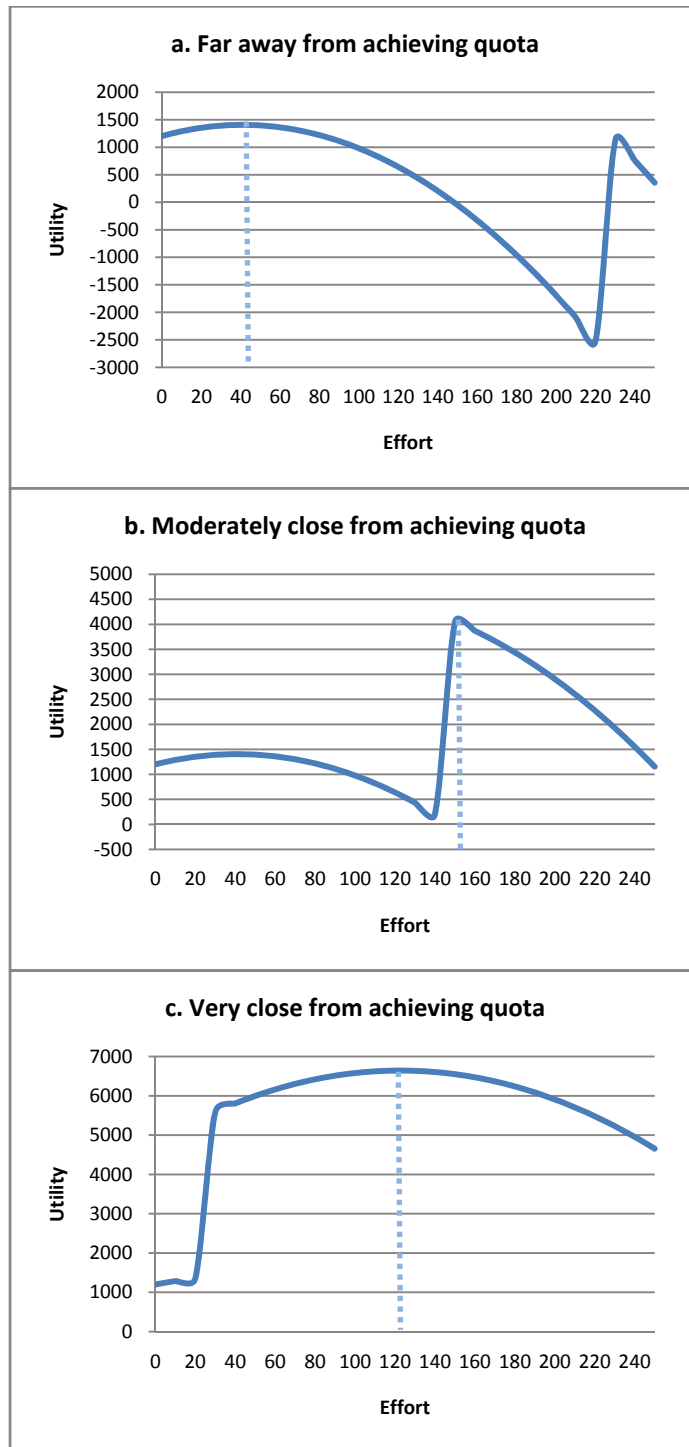


Figure 8: Expected Utility – Bonus periods – Static – 3D

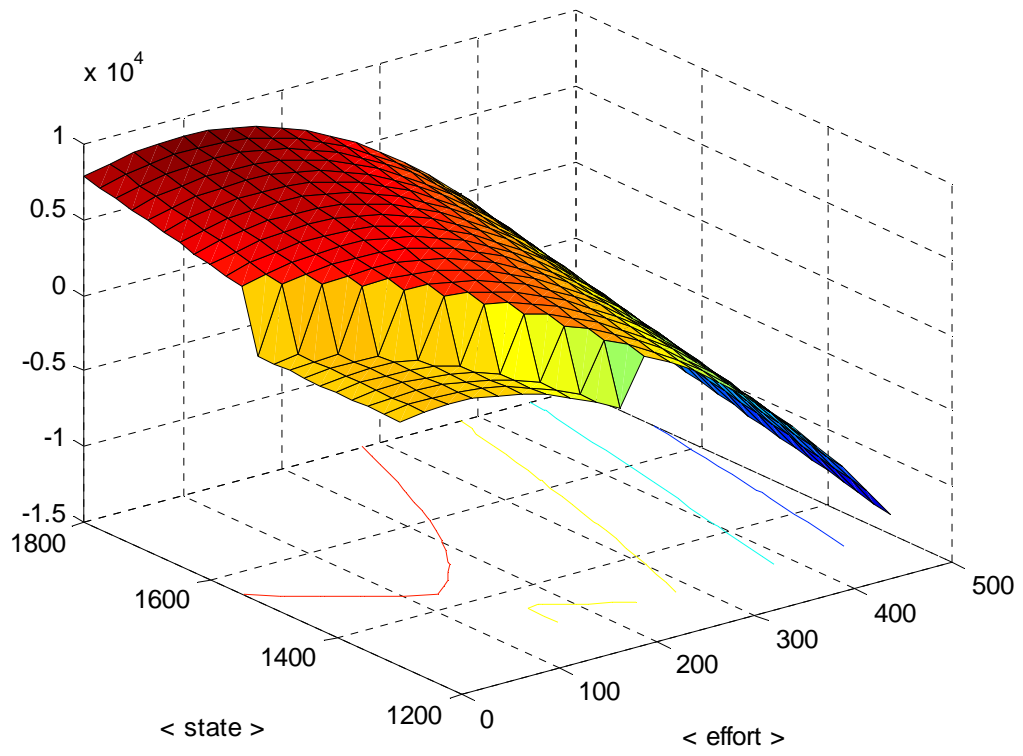
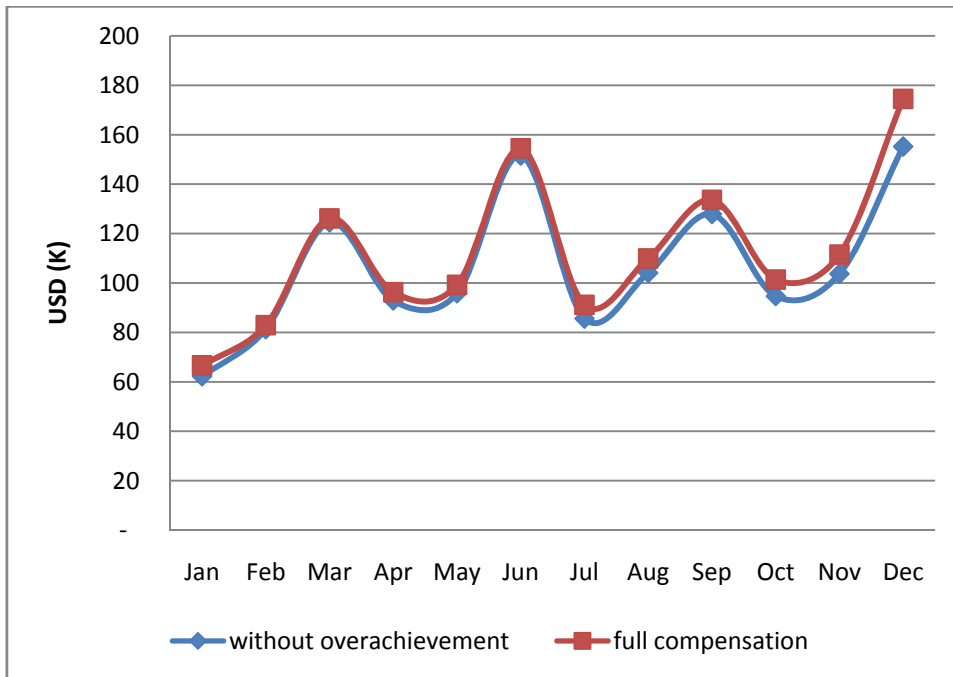


Figure 9a: Monthly Revenues – No overachievement commission



Increase in Annual Revenues with Overachievement Compensation: 6.7%

Figure 9b: Effort level of Agents who met and didn't meet the Annual Quota

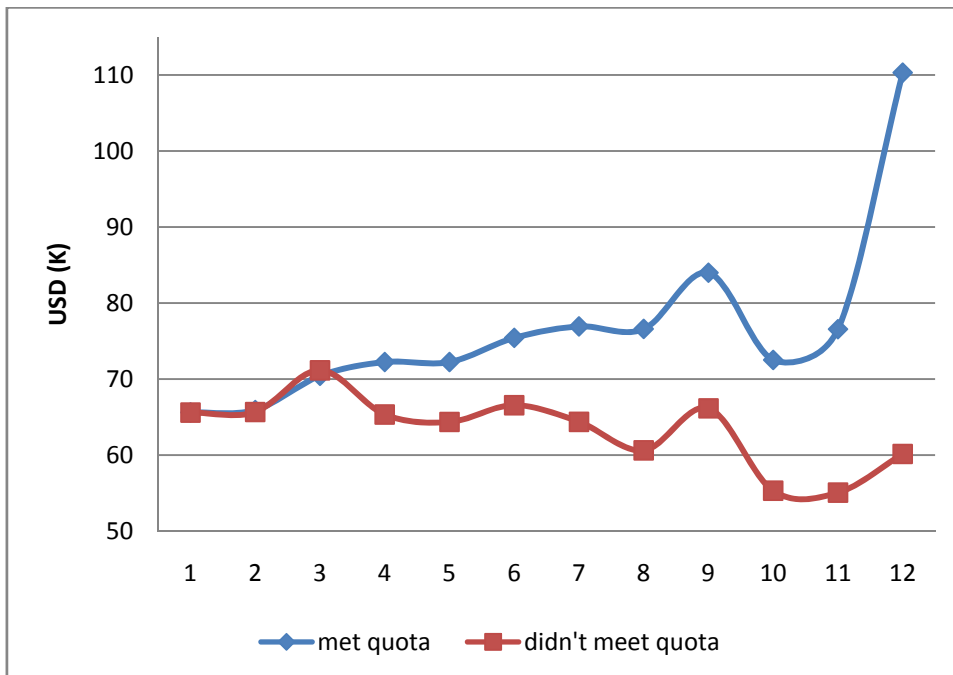
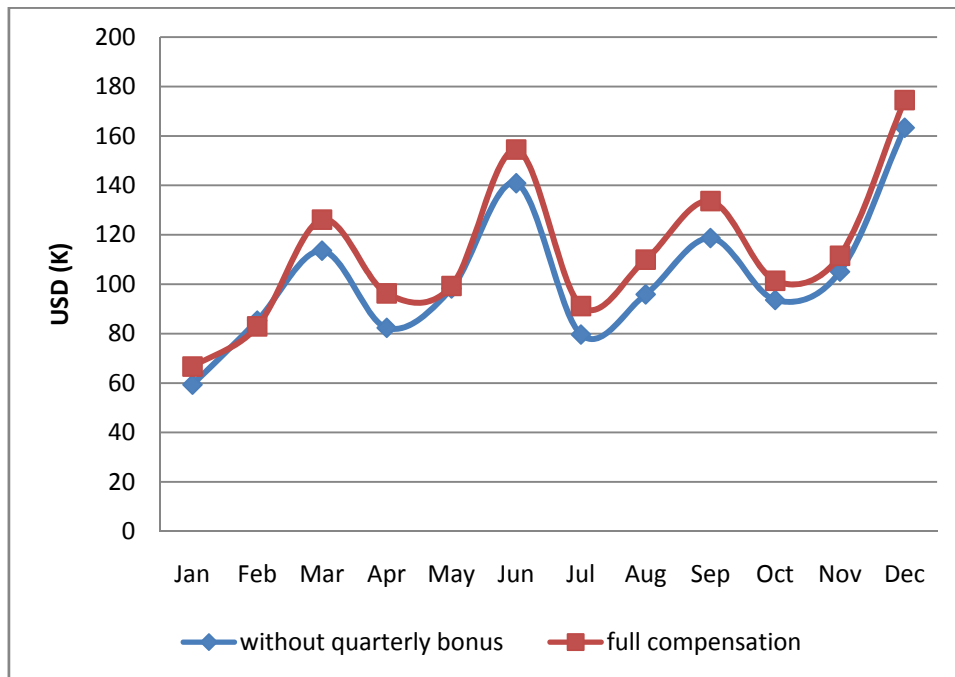


Figure 10: Monthly Revenue

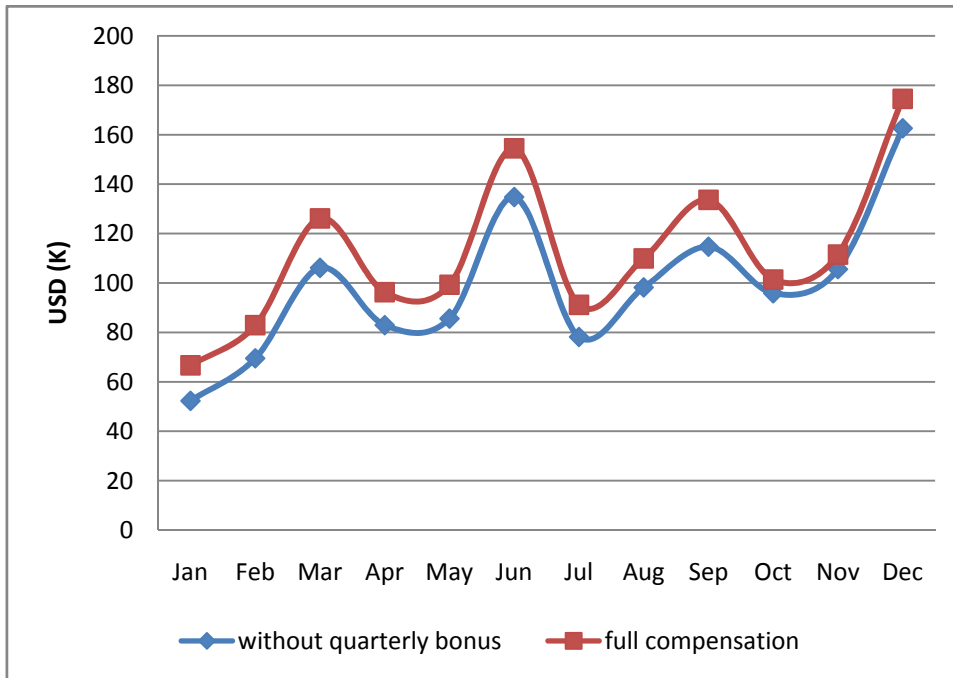
– Change Annual Quota to Quarterly Quota + No overachievement commission



Increase in Annual Revenues with Cumulative Annual Quota + Overachievement Commission: 9.9%

Net Increase in Annual Revenues due to Cumulative Annual Quota alone: 2.2%

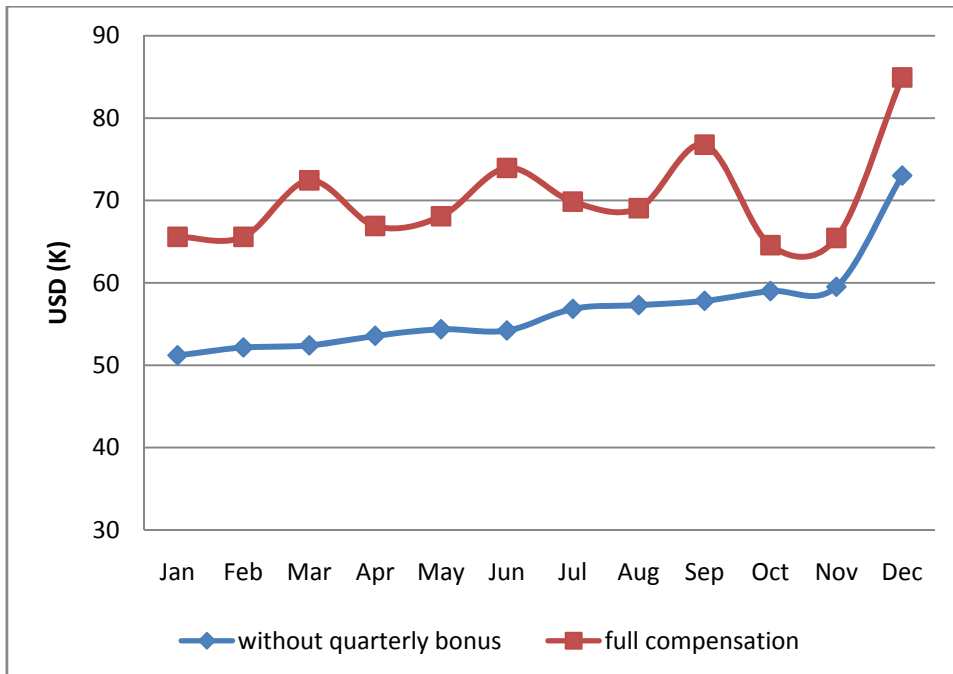
Figure 11a: Monthly Revenue – No Quarterly Bonus



Increase in Annual Revenues due to quarterly bonus 13.6%

Increase in December Revenues due to quarterly bonus (at identical year-end bonus): 7%

Figure 11b: Average Effort – No Quarterly Bonus



Increase in Effort due to quarterly bonus: 14%