DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

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

(Article begins on next page)

HARVARD | BUSINESS | SCHOOL

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

Doug J. Chung
Thomas Steenburgh
K. Sudhir

Working Paper

13-066

January 25, 2013

Do Bonuses Enhance Sales Productivity?

A Dynamic Structural Analysis of Bonus-Based Compensation Plans

Doug J. Chung, Thomas Steenburgh and K. Sudhir[*]

November 2012

Do Bonuses Enhance Sales Productivity?

A Dynamic Structural Analysis of Bonus-Based Compensation Plans

Abstract

We estimate a dynamic structural model of sales force response to a bonus based compensation plan. Substantively, the paper sheds insights on how different elements of the compensation plan enhance productivity. We find evidence that: (1) bonuses enhance productivity across all segments; (2) overachievement commissions help sustain the high productivity of the best performers even after attaining quotas; and (3) quarterly bonuses help improve performance of the weak performers by serving as pacers to keep the sales force on track to achieve their annual sales quotas. The paper also introduces two main methodological innovations to the marketing literature: First, we implement empirically the method proposed by Arcidiacono and Miller (2011) to accommodate unobserved latent class heterogeneity using a computationally light two-step estimator. Second, we illustrate how discount factors can be *estimated* in a dynamic structural model using field data through a combination of (1) an exclusions restriction separating current and future payoff and (2) a finite horizon model in which there is no forward looking behavior in the last period.

Key Words: Sales force compensation, bonuses, quotas, dynamic structural models, two step estimation, discount factors, hyperbolic discounting.

1 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 salespeople in the United States (Zoltners et al. 2008). Sales force costs average about 10% of sales revenues and as much as 40% of sales revenues for certain industries (Heide 1999). In the aggregate, U.S. firms spent over $800 billion on sales forces in 2006, a sum three times larger than advertising spending (Zoltners et al. 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 0.34 (Albers et al., 2010), 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 the 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 1 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.

Thus, this paper has two substantive goals: First, to gain insight on how a firm should design its compensation plan. Should a firm offer quotas and bonuses in addition to commissions? Second, what should be the frequency of bonuses? Should one use a monthly, quarterly or annual bonus? Should one use a quarterly bonus in addition to an annual bonus? 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? As in the education literature, where frequent exams help students to be prepared for the comprehensive final exam; frequent quota-bonus plans may serve as a mechanism to keep the sales force motivated to perform well in the short-run so as to be within striking distance of the overall annual performance quota.

Quotas and bonuses are widely used by firms. According to Joseph and Kalwani (1998), only about 24% of firms use a pure commission-based plan; the rest used some form of quotas. As per the Incentive Practices Research Study (2008) by ZS Associates, 73%, 85% and 89% in the pharma/biotech, medical devices and high tech industries respectively uses quota based compensation. 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.

We begin with a discussion of the relevant theoretical literature. Using the principal agent framework of Holmstrom (1979), Basu et al. (1985) and Rao (1990) find that a combination of salary plus commission (usually nonlinear with respect to sales) is optimal. However, under the assumption of linear exponential utility and normal errors (LEN), Holmstrom and Milgrom (1987) and Lal and Srinivasan (1993) show that a linear commission scheme can achieve the first best outcome. Yet, why do we see quota-bonus plans? 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.

In terms of empirical work, Ferrall and Shearer (1999) and Paarsch and Shearer (2000) estimate static structural models of worker behavior given linear contracts. Oyer (1998) was the first to empirically investigate quota-based plans. Using aggregate sales across different industries in different quarters, he concludes that quota based plans encourage sales people to maneuver the timing of orders and this negative effect overwhelms the positive benefit of quotas. Using reduced form analysis of individual level data, Steenburgh (2008) finds that the net improvement in revenues from effort dominates the inefficiencies induced by inter-temporal dynamic considerations, and shows that an aggregate analysis might have led to the opposite conclusion similar to that of Oyer.

Copeland and Monnet (2008) estimate the first dynamic structural model of worker productivity in a check-sorting environment with nonlinear incentives. Unlike sales force productivity, where we only observe aggregate sales, the outcomes associated with every processed check is observed in their environment; hence they have a qualitatively different model. Our paper shares many similarities with a recent paper by Misra and Nair (2011), who also estimate a dynamic structural model of sales force compensation, although using a very different quota-compensation scheme. In contrast to our focus on quotas with bonuses (plan F in Figure 1), Misra and Nair analyze quotas with floors and ceilings on commissions (plan D in Figure 1). They conclude that quotas reduce performance. This is because of two characteristics of their quotas: First, the quota ceiling (beyond which sales people receive zero additional compensation) limits the effort of the most productive sales people who would normally have exceeded the ceiling. Second, the company followed an explicit policy of ratcheting quotas based on past productivity. This reduces incentives of sales people to work hard in any given period, because hard work is penalized through higher future quotas. In contrast, we find that quotas coupled with bonuses enhance performance. In the plan we consider, the company offers extra overachievement commissions for exceeding quotas and use a group quota updating procedure that minimizes ratcheting

effects. Thus, the two papers offer complementary perspectives that enhance our understanding about how quotas impact performance.

Methodologically, our paper introduces two key ideas to the marketing literature. First, we accommodate latent class heterogeneity within the two-step conditional choice probability (CCP) estimation framework—an issue that has been an econometric challenge for the literature for close to two decades. Misra and Nair sidestep the unobserved heterogeneity issue by estimating each sales person's utility function separately.[1] Although the use of two step estimation approaches have recently gained popularity (Hotz and Miller 1993; Bajari et al. 2007), due to ease of computation relative to traditional nested fixed point estimation approaches (e.g., Rust 1987), their use in empirical applications has been limited by their inability to accommodate unobserved heterogeneity. Arcidiacono and Miller (2011) proposed an algorithm that accommodates latent class heterogeneity within the two-step framework. Our paper introduces this idea to the marketing literature and illustrates its implementation with an empirical application that accommodates unobserved heterogeneity using two-step dynamic structural estimation.[2]

Second, unlike Misra and Nair (2011), and of broader importance to the dynamic structural modeling literature, we *estimate* rather than *assume* discount factors. It is well known in the literature on dynamic structural models that discount factors cannot be identified in standard applications because there are no instruments that provide exclusion restrictions across current and future period payoffs (Rust 1994). Hence the standard approach is to assume discount factors. We illustrate how the bonus setting provides us exclusion restrictions allowing us to estimate discount factors—thus demonstrating how field data can indeed be used to estimate discount factors.[3] Since bonus payoffs occur only at the end of each quarter or year, in the non-bonus periods the probability of achieving quota and receiving bonus will not affect current payoff, but only future payoffs. Only a forward looking person (i.e., one with a non-zero discount factor) would respond to proximity to quota in non-bonus periods. We demonstrate through reduced form evidence that such behavior exists in the data, and then exploit this exclusion restriction to identify the discount factor. In addition, the finite horizon nature of the sales force problem implies that in the last period, the model reduces to a static model, for which it is well-known that utility can be identified. This further facilitates identification of the discount factor. Yao et al. (2012) apply our idea in a cellphone usage context to identify discount factors. Similar to our setting, distance to cellphone plan quota serves as an exclusion restriction in their setting. Similar to the last period of our data where the model reduces to a static model, they have data for a period in which pricing is linear and therefore reduces to a static model.

---

[1] This is similar to the estimating individual level utility functions in conjoint analysis or scanner panel data, when there are a large number of observations per individual. More importantly, the approach requires that sales people will exert effort equally across all customers—an assumption they show is valid in their data, but unlikely to hold in general. Our latent class approach works in the more common situation where there are limited observations per individual.

[2] Finger (2008) and Beauchamp (2010) are two concurrent working papers implementing the Arcidiacono and Miller approach in economics.

[3] Use of exclusion restrictions to empirically infer discount factors is important because prior to this paper, the conventional wisdom was that field data cannot be used for discount factor identification, and one needs to use surveys (e.g., conjoint data as in Dube et al. (2011)).

There are three specific modeling and estimation challenges in the structural estimation of response to compensation plans, especially those with quotas and bonuses. First, we do not observe the effort of the sales force; only the outcome of the agent's effort, i.e., sales, which is correlated with effort. This requires a modeling assumption on the link between sales and effort.[4] Second, compensation plans do not change over time. Here we draw on an empirical insight from Steenburgh (2008) for identification. 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. Thus changes in sales in response to changes in agent's state (distance to quota) facilitate identification. Finally, quotas and bonuses induce dynamic forward looking behavior; an agent choosing effort has to be concerned not just with the current payoff, but the effect of that effort on the likelihood of obtaining a bonus in the future. This requires a dynamic structural model.

We estimate a dynamic structural model of sales force response to various features of the compensation plan using sales force output and compensation data from a Fortune 500 firm that sells office durable goods. This firm used Plan F in Figure 1. In addition, bonuses are provided at two different frequencies: quarterly and annual. As the compensation structure of the focal firm features almost all dimensions in typically used compensation plans, we observe how the sales force responds to these different dimensions of the plan. This rich plan provides us two key benefits: First, the presence of bonuses helps us to identify and estimate discount factors. Second, even though theoretically one can perform counterfactuals of any type of compensation plan if we can estimate structural parameters (other than discount factors) for a sales person with a less rich compensation plan, an analyst or manager should have greater faith in the counterfactuals, based on parameters that were estimated from observed responses to different elements of the compensation plan.

The rest of the paper is organized as follows. Section 2 discusses the institutional details of the compensation plan at the firm, provides a numerical example to give intuition about how bonuses induce effort, and provides some model free evidence that facilitates model building. 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 Descriptive Analysis

We first describe the details of the bonus based compensation plan, followed by a numerical example to clarify how bonuses can help motivate the sales person, and induce inter-temporal effort that assist us with identification. We then provide model-free evidence of forward looking behavior, seasonality, and the absence of sales substitution across quarters.

---

[4] 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.

2.1 The Compensation Plan

The focal firm under study is a highly regarded multinational Fortune 500 company that sells durable office products primarily using its own direct sales force. Each sales agent is given an "exclusive" territory; the firm traditionally does not encourage group work or team cooperation among the sales force. The firm also has an indirect sales force through "rep" firms who do not compete with the direct sales force. They are paid purely on commission, unlike the regular sales force.[5]

Our analysis focuses on sales performance data from 348 sales people from the regular sales force during the three year period 1999-2001. The firm's compensation structure follows the pattern in Plan F of Figure 1 and the details of the compensation schedule for the period of analysis are described in Table 1. We provide descriptive statistics of the data in Table 2.

Every month, sales people receive a fixed monthly salary (average $3,585) and a commission of 1.5% of revenues generated in that month. In the first three quarters, a quarterly lump-sum bonus of $1,500 is paid if the quarterly quotas are met. At the end of the year (i.e., end of the fourth quarter) an annual lump-sum bonus of $4,000 is paid if the annual quota is met. Further, an overachievement commission of 3% is paid for any excess revenues beyond the annual quota. There are no caps on revenues for which an agent could obtain commissions or overachievement commissions. Overall, for a salesperson that meets all quotas, the salary component will be roughly 50% of total compensation.

In building annual and quarterly quotas, the company uses internal metrics called "monthly allocated quotas" to its sales force (based on expected monthly revenues, given seasonality and territorial characteristics), though these are not used for compensation. We do not use these quotas for our modeling and estimation, but use them to benchmark performance in the reduced form analysis.

The most important element in performance evaluation within the firm is the annual quota; i.e., the firm views a salesperson as having a successful year if the annual quota is met. From Table 2, we see that sales people meet their annual quota roughly 50% of the time. All quotas, including the quarterly quotas, are updated annually and reset in January of each year. However, managers at the firm informed us that they were sensitive to the fact that ratcheting quotas based on individual performance could lead to purposefully reduced performance by the sales force. To avoid such adverse impact on sales, the quota adjustment year over year was done only based on group performance, where each individual's current performance would have minimal direct impact on their future quota. Nevertheless, we test for statistical evidence of ratcheting in the data. Table 3 reports the results of the tests. In model 1, we regress the percentage increase in annual quota on whether the salesperson met her quota in the previous year. We find no significant effect of meeting quota in the previous year on current-year quota. In model 2 we regress the logarithm of annual quota on whether a salesperson met her quota in the previous year, this time controlling for salesperson fixed effects. Again, we find no significant effect. Finally, in model 3, we regress the logarithm of quota on the logarithm of previous year's performance relative to the annual quota. Again, we find no significant effects. Thus

---

[5] Such rep firms are the focus of Jiang and Palmatier (2010).

consistent with information from the firm's managers, we find no statistical evidence of ratcheting. We therefore do not model quota ratcheting and treat each salesperson's quarterly and annual quotas as exogenous.

2.2 Numerical Example

Before building an empirical model of sales person's response to a bonus based compensation scheme, we provide a numerical example that illustrates how a quota-bonus scheme can be more effective than pure commissions in generating more effort. The example will also show that a person's distance to quota can induce variations in optimal effort (and generated sales)—providing intuition for our identification strategy.

Let the utility function of the salesperson that trades off effort (e) and income from sales (s), who has sold S units at the beginning of the new period be:

$$U(s,e,S) = -de^2 + rs + BI_{\{s+S \geq Q\}}$$

where $-d$ is the disutility parameter and r is the commission rate $(d > 0, r > 0)$ and B is the bonus for reaching quota (Q). For illustration, let us assume a direct match between sales and effort, i.e., $s=e$.

First, we illustrate the effectiveness of bonuses using a static model. For simplicity assume $S=0$. In the pure commission case with no bonus where $d=1$, $r=10$ and $B=0$, the optimal effort is $e^*=5$. In the bonus case, with $Q=10$, and $B=30$, the optimal effort is higher at $e^*=10$ and the compensation cost to the firm is \$130. To achieve the same level of sales and effort ($e^*=10$) from a pure commission plan, the commission rate r has to increase to 20 and costs more for the firm at \$200. Figure 2a illustrates these results graphically. Thus, the quota-bonus plan in this context induces more effort and sales for the same level of compensation.[6]

Second, we show how distance to quota induces variation in effort. Let $d=2$, $r=10$, $Q=10$, and $B=30$. Consider three scenarios of distance to quotas: $S=0$ (far away from quota), $S=5$ (moderately close to quota) and $S=7$ (close to quota). Figure 2b shows that the optimal effort levels are $e^*=2.5$, 5 and 3, respectively, for the three cases and the salesperson exerts maximum effort at $S=5$, when moderately far away from quota, all else being equal. We use this variation of optimal levels of effort (thus, sales) within agent across time to identify agents' preferences in our main model.

2.3 Model Free Analysis

We consider three features of the data that informs model development. First, we provide evidence of forward looking behavior induced by bonuses and hence the need to develop a dynamic model. Second, we show the presence of seasonality in the data, which therefore requires us to

---

[6] Since this analysis is in partial equilibrium framework, we are not claiming that a bonus based plan is the optimal compensation plan. We simply illustrate that a non-linear plan can be more efficient than a linear plan, suggesting that the optimal compensation plan is possibly non-linear and the bonus plan can be a way to approximate the optimal non-linear incentive plan.

accommodate this in the model. Third, we show that there is little evidence of sales substitution across quarters by sales agents—allowing us to abstract away from modeling potential timing games played by sales agents.

Forward Looking Behavior

As discussed, bonuses provide an exclusion restriction in that it does not impact current payoffs, but only future payoffs. To the extent that a sales agent's sales performance is affected by variables relating to proximity to bonuses, this is evidence of forward looking behavior.[7] But proximity to bonus quota will impact performance only if agents have a reasonable chance of making quota. Figure 3 shows the graph of the probability of reaching the annual quota, conditional on the cumulative fraction of annual quota ($\%AQ$) achieved till November. It is clear that there is very little chance of achieving quota if $\%AQ < 0.5$. We therefore divide agents by their state $\%AQ < 0.5$ and $\%AQ > 0.5$ to test if the state affects sales and estimate regressions on sales performance in November as a function of their states. Table 4 reports the results of the regressions. Consistent with forward looking behavior, the state $\%AQ$ is significant only for agents with $\%AQ > 0.5$.[8]

For additional evidence of forward looking behavior, we show scatter plots and the best fitting nonparametric smoothed polynomial (and its 95% confidence interval) of sales revenues normalized by monthly allocated quotas in the quarterly bonus months (March, June, September, December) against percentage of quota attained by the previous month in Figure 4a. For March, June and September, the x axis is the percentage of quarterly quota completed ($\%QQ$), while for December, the x axis is the percentage of annual quota completed ($\%AQ$). The vertical dotted line shows the $\%QQ$ and $\%AQ$ at which the salespeople, on average, achieve their monthly allocated quotas.

Two key elements stand out from Figure 4a. First, across the board there is little reduction in effort when salespeople are close to achieving quota, due to the overachievement commission rate. Second, there is a steady increase over time in the $\%QQ$ and $\%AQ$ threshold beyond which sales people reach their monthly targets. The threshold is about 25% in March, 35% in June, 45% in September and close to 70% in December. Early in the year, even if below targets, salespeople still have hopes of receiving the large annual bonus by working hard. As one gets closer to year-end, chances of reaching quota becomes less likely, and sales people respond by reducing effort even at higher levels of $\%AQ$ and $\%QQ$. As annual bonuses should have no impact on current payoffs in March, June or September, but only on future payoffs, this is suggestive of forward looking behavior.

The next set of graphs presented in Figure 4b, shows the same relationship in the pre-bonus months (February, May, August and November) and provides additional evidence for forward looking behavior. In the early months, February, May and even August, at all levels of $\%QQ$, the salesperson on average sells above the monthly allocated quota. This is because hard work (and some good luck in

---

[7] A similar argument is made in providing evidence of forward looking behavior by students in the textbook market by Chevalier and Goolsbee (2009).

[8] The qualitative conclusions are robust in the range of thresholds of $\%AQ$ from 0.4 to 0.6. Note that if there were no overachievement commission, agents very close to quota may reduce their effort.

the form of positive sales shocks) may give a reasonable chance of attaining the smaller quarterly targets. However, in November, only at a very high level of *%AQ*, does the salesperson sell above the monthly allocated quota, because one has a very limited chance of making up the large gap in just two months. In the pre-quarterly bonus months, the immediate future quarterly bonus impacts behavior, even though it has no impact on current payoff; indicating more conclusive forward looking behavior.

This evidence leads to a natural question. Should the large annual bonus be split into a quarterly bonus (as in other months) and an annual bonus? The quarterly bonus can prevent sales people from giving up in November, even if they do not have a chance of reaching the annual quota. But with such a quarterly quota, early in the year, agents may have limited incentive to work hard after reaching quarterly quotas. How these two issues tradeoff is an empirical question, which we subsequently address in the counterfactual analysis.

Seasonality

Figures 5a graphs the average revenues over the months for the regular sales force. There are clear peaks at the end of each quarter. These peaks could be either due to seasonality or bonuses at the end of each quarter. Figure 5b shows the index of indirect sales revenue (ISR) for the pure commission indirect sales force.[9] As the ISR index is not contaminated by bonuses, we use it to control for seasonality and isolate the effect of bonuses on sales person effort and revenues. To build intuition for how the ISR index can help control for seasonality and isolate effort, see Figure 5c which graphs the average revenues of the direct sales force and multiples of the ISR index. At a multiple of around 50, the ISR virtually mimics the average revenues, making the revenues from the commissioned and bonus sales force close to identical. This suggests that bonuses are not effective in inducing additional effort. When the ISR index has a multiple of 30 or 40, even after the overall seasonality is accounted for, there is gap in revenues that we interpret as induced by effort. It is interesting that these gaps are larger at the end of the quarter, suggesting the value of bonuses in inducing effort. We empirically estimate the multiple for the ISR index in order to control for seasonality.[10]

Sales substitution across quarters

One possibility is that salespeople giving up at the end of the quarter may be doing so to increase the odds of hitting quotas in subsequent quarters by simply not booking the sales in the current quarter. If this were true, then one should see a positive relation between the first month of the quarter and the previous quarter's percentage cumulative performance to quota (*%QQ* or *%AQ*). Figure 6 shows an illustrative example. Let us assume that a salesman's second-quarter quota is 300 units. And suppose she had sold 120 units till the end of May (*%QQ* = 40%). Then she would have

---

[9] For the indirect sales force, the firm only provided us with an index of revenues. The indirect sales force revenue (ISR) index is set to a base of 1 for January 1999.

[10] From Table 6a, we know the ISR multiple for our model is about 30, given the average of lagged annual quota is 1639. While we acknowledge that accounts handled by the direct and indirect sales force are likely systematically different, our maintained assumption is that the seasonality multipliers on sales are identical across the two types of accounts.

an incentive to pull-in sales from the next quarter – first case in Figure 6. Now suppose she has sold 240 units by the end of May ($\%QQ = 80\%$). In this case, she would have incentive to push-out sales to the next quarter – second case in Figure 6. If this were the case, the quota-bonus incentive scheme would be encouraging salespeople to play timing-games, where they merely shift sales across quarters, rather than producing more sales through greater effort. We test for statistical evidence of such time-shifting if any. Table 5 shows the regression results of sales against the previous months' distance-to-quota. If timing games were played by the sales force, the percentage cumulative performance to quota in the previous month and the sales of the first month of the quarter would be positively correlated. We find this relation to be insignificant and very small in magnitude, suggesting little substitution across quarters. Thus, given the lack of statistical evidence of time-shifting, we do not model sales substitution across quarters.[11]

3   Model

Based on the model-free evidence, we build a dynamic model of sales force response to the quota-based compensation scheme. The timing of the model 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 describe the model in five parts: (i) the compensation plan (ii) the sales agent's utility function (iii) the state transitions (iv) effort as a function of state variables and (v) the optimal effort choice by the sales agent.

3.1   Sales Response Model

We model the sales revenue function (S_{it}) for salesperson i at time t in two parts: (1) a base level of sales independent of effort, parameterized by demand shifters (z_{it}^D) and (2) sales induced due to effort (e_{it}) parameterized by effort shifters that include territory and salesperson characteristics (z_{it}^E).

$$S_{it} = f(z_{it}^D) + e_{it}(z_{it}^E) + \varepsilon_{it} \tag{1}$$

where ε_{it} is an additive sales revenue shock, not anticipated by the salesperson when choosing effort.

As discussed, the market potential varies across territories and across time. To account for the cross-sectional variation in market potential, we use annual quota from the previous year ($AQ_{i,y-1}$).

---

[11] Managers of the focal firm stated that they monitor sales agents extensively to prevent such time shifting of sales.

To account for seasonality of demand across months, we use the ISR index, (ISR_t). We also include an interaction between the two variables to allow for seasonality to have a larger impact on larger territories.

For effort shifters in e_{it}, we use the following variables: Given that effort is a function of demand shifters, we include both $AQ_{i,y-1}$ and ISR_t in z_{it}^E. As discussed in the motivation, the salesperson's state with respect to achieving quota will have an impact on the effort they expend. We therefore use the cumulative percentage of quarterly and annual quota completed till time t ($\%QQ_{it}$, $\%AQ_{it}$) as variables that affect effort. In addition, we allow a time-invariant salesperson specific variable, tenure with the firm (τ) to moderate the level of effort. We allow for interactions between ISR_t and $\%QQ_{it}$, $\%AQ_{it}$ in the effort function, thus, effort policy function is different for different months. While it would be ideal to estimate separate effort policy functions for each month, there are not typically enough degrees of freedom in the data to do this for each month. We balance the flexibility/degrees of freedom tradeoff with our approach. Although this approach seems somewhat restrictive, we show via simulations in the Appendix that it does a good job in adequately recovering the true parameters.[12]

Note that unlike the demand shifter function *f*, which is common across all salespeople, the effort function will vary across salespeople. Specifically, we allow for salespeople to belong to one of multiple discrete segments, hence these effort functions will be estimated at the segment level. We estimate the effort function non-parametrically, by using Chebyshev polynomials of the variables described above.

3.2   Compensation Plan

The compensation plan has three components. They are: (1) the monthly salary w_{it}, (ii) end-of quarter bonus, B_{iqt} for achieving the corresponding quarterly quota Q_{iqt}, and end of year bonus B_{iyt} for achieving the corresponding annual quota Q_{iyt} (3) commission rate r_{it} per dollar worth of sales and an overachievement commission rate, r'_{it} given 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 $\psi_{it} = \{w_{it}, Q_{iqt}, Q_{iyt}, B_{iqt}, B_{iyt}, r_{it}, r'_{it}\}$.

3.3   Sales person's per-period utility

In each period t, sales person i receives positive utility of wealth W_{it} earned based on realized sales and a disutility $C(e_{it}; \theta_i)$ from exerting effort e_{it}. Thus the utility function is defined as:

$$U\left(e_{it}, S_{it}; \psi_i, \theta_i, \gamma_i\right) = E\left[W\left(S_{it}; \ \psi_i\right)\right] - \gamma_i \operatorname{var}\left[W\left(S_{it}; \ \psi_i\right)\right] - C\left(e_{it}; \theta_i\right)$$

where γ_i and θ_i are the risk aversion and disutility parameters respectively for salesperson i.[13]

---

[12] We thank an anonymous reviewer for suggesting this idea.

[13] In the case of the CARA utility function (exponential utility function) with normal errors and a linear compensation plan, this functional form represents the certainty equivalent utility of the agent. Here we consider the utility function to be a second order approximation to a general concave utility function with constant level of risk-aversion.

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,

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

$$B_{it} = I_{qt}I\left(z_{i1t} + \frac{s_{it}(e_{it}(z_{it}^E), z_{it}^D; \alpha_i) + \varepsilon_{it}}{Q_{iqt}} > 1\right)B_{qt} + I_{yt}I\left(z_{i2t} + \frac{s_{it}(e_{it}(z_{it}^E), z_{it}^D; \alpha_i) + \varepsilon_{it}}{Q_{iyt}} > 1\right)B_{yt}$$

$$C_{it} = (s_{it}(e_{it}(z_{it}^E), z_{it}^D; \alpha_i) + \varepsilon_{it})r_{it}$$

$$OC_{it} = I_{yt}I\left(z_{i2t} + \frac{s_{it}(e_{it}(z_{it}^E), z_{it}^D; \alpha_i) + \varepsilon_{it}}{Q_{iyt}} > 1\right)\left(z_{i2t}Q_{iyt} + s_{it}(e_{it}(z_{it}^E), z_{it}^D; \alpha_i) + \varepsilon_{it} - Q_{iyt}\right)r_{it}'$$

where z_{i1t} and z_{i2t} are the percentage of quarterly and annual quotas completed respectively by salesperson i until time t. I_{qt} and I_{yt} are indicators for whether time t is a quarterly or annual bonus period.

In our empirical analysis, we use a quadratic functional form for the disutility function; specifically, $C\left(e;\theta_i\right) = \theta_i e^2$. Thus the set of structural parameters of the salesperson's utility function that needs to be estimated are $\mu_i = \left(\theta_i, \gamma_i\right)$.

3.4   State Variables

As discussed, 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 commissions tomorrow. To incorporate the dynamics of the model we consider the following stochastic state variables, the percentage of annual quota completed, the percentage of quarterly quota completed. These state variables evolve as follows:

1. Percentage of quarterly quota completed ($\%QQ$)

$$z_{i1t} = \begin{cases} 0, & \text{if } t \text{ is start of quarterly quota period} \\ z_{i1(t-1)} + \dfrac{S_{i(t-1)}}{Q_{iqt}}, & \text{otherwise} \end{cases}$$

2. Percentage of annual quota completed ($\%AQ$)

$$z_{i2t} = \begin{cases} 0, & \text{if } t \text{ is start of annual quota period} \\ z_{i2(t-1)} + \dfrac{S_{i(t-1)}}{Q_{iyt}}, & \text{otherwise} \end{cases}$$

Other state variables would include time varying demand shifters, ISR index and territory characteristics, for which we use previous year's annual quota. Naturally, the time varying indirect sales is a one-to-one mapping to period type and hence includes information about different periods.

11

We use tenure with the focal firm (τ) as an individual state variable that impacts effort. These state variables are collected in a state vector $z_{it}^E = \left\{ z_{i1t}, z_{i2t}, IS_t, AQ_{i(y-1)}, \tau_i \right\}$.

3.5 Optimal Choice of Effort

Given the parameters of the compensation scheme ψ, and the state variables and their transitions, each sales agent would choose an effort level conditional on her states to maximize the discounted stream of expected future utility flow. Alternatively, if this value function is below the reservation wage, the salesperson may choose to leave the firm.

The stream of utility flow, under the optimal effort policy function, conditional on staying at the firm, can be represented by a value function,

$$V\left(z;\psi,\Omega\right) = \max_e U\left(e,z;\psi,\mu\right) + \delta E\left[\max_{e'} U\left(e',z';\psi,\mu\right) + \delta E\left[\max_{e''} U\left(e'',z'';\psi,\mu\right) + ...\right]\right]$$

where $\Omega = \left(\mu,\delta\right)$ is the set of primitives or structural parameters of the underlying utility function μ, and the discount factor δ. The expectation of the value function is taken with respect to both the present and future sales shocks.

4 Estimation

Traditionally, the nested fixed-point algorithm (NFXP) developed by Rust (1987) is used to estimate dynamic models. However, NFXP estimators are computationally burdensome as one has to solve the dynamic program numerically over each guess of the parameter space for every iteration. The two-step estimation first introduced by Hotz and Miller (1993) and extended by Bajari et al. (2007) can serve to reduce the computational burden. In this approach, the model estimation proceeds in two steps. In the first step, we estimate the conditional choice probabilities of choosing a certain action as a flexible nonparametric function of state variables. Then, in the second step, these conditional choice probabilities are used to estimate the structural parameters of the sales agent's utility function.

Until recently, it was believed that the accurate estimation of conditional choice probabilities for an agent is impractical when there is unobserved heterogeneity. Arcidiacono and Miller (2011) propose an EM–Algorithm based approach to accommodate unobserved heterogeneity in the first step of the two step estimation procedure. We provide one of the first applications of this approach – illustrating the empirical validity of the approach in practical applications. We now discuss the details of the two step estimation procedure.

4.1 Step 1

In this step, we need to estimate a flexible non-parametric mapping between observable states and actions of the sales person; this requires a non-parametric model of the monthly effort function $e_i(z_{it}^E)$, that links effort and states in equation (1). We model the effort function non-parametrically as a combination of basis functions of the state variables. Thus the non-parametric effort function is:

$$e_{it} = \sum_{\ell=1}^{L} \rho_\ell(z_{it}^E)\lambda_{i\ell} \tag{2}$$

where the ℓth basis function is $\rho_\ell(z_{it}^E)$. In this application, the ℓth basis function is the ℓth order Chebyshev polynomial.[14]

From equations (1) and (2) we have the following sales response function to estimate.

$$S_{it} = f(z_{it}^D) + \sum_{\ell=1}^{L} \rho_\ell(z_{it}^E)\lambda_{i\ell} + \varepsilon_{it}$$

For z_{it}^D, which is a subset of z_{it}^E, (from now on referred to as z_{it}), we use two variables: (1) lagged annual quota for salesperson i, (2) the indirect sales force revenue (ISR) index. We use the direct linear effect of these variables to control for cross sectional variations of territory characteristics and temporal variations in monthly seasonality. The interaction effects of these variables with the other state variables go into the polynomial function in (2).[15]

The lagged annual quota takes into account general territory characteristics that are likely to be generated with limited effort, i.e., market size. The revenues from the indirect sales force capture market seasonality, independent of the nonlinear nature of the compensation plan. We assume that the revenue shocks (ε_{it}), come from an i.i.d. normal distribution.[16]

If one could estimate the sales response and effort response function at the level of each individual, we can simply obtain the individual level parameters of the effort and sales policy function by maximizing the log likelihood of the sample such as

$$\hat{\Theta}_i = \arg\max_{\Theta_i} \sum_{t=1}^{T} \log\left\{ L_i\left(S_{it} - f(z_{it}^D; \alpha_i) - \sum_{\ell=1}^{L} \rho_\ell\left(z_{it}^E\right)\lambda_{i\ell} \right) \right\} \tag{3}$$

---

[14] Chebyshev polynomials are a sequence of orthogonal polynomials, defined by the recurrence relation: $T_0(x) = 1$, $T_1(x) = x$ and $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$. By using the orthogonal Chebyshev polynomials, to approximate the policy function, rather than approximate the policy function with the standard 1, x, x^2 polynomials, we avoid multicollinearity issues. In our approximation of the effort policy function (reported in Table 6b), we use up to third order Chebyshev polynomials; i.e., $T_0(x) = 1$, $T_1(x) = x$, $T_2(x) = 2x^2 - 1$ and $T_3(x) = 4x^3 - 3x$. For more details on Chebyshev polynomials, we refer the reader to "Numerical Methods in Economics," Kenneth L. Judd, MIT Press, 1998.

[15] We use estimations based on simulated data to check that separating z_{it}^D from the effort policy function indeed recovers the true parameter estimates, relative to a model where the main effects of z_{it}^D are directly included into the effort policy function. We thank an anonymous reviewer who encouraged us to assess the robustness of this assumption.

[16] Interestingly, Mirrlees (1975) shows that the first-best can be approximated, even if not attained with normal errors in the sales response function in a one shot model setting if unbounded punishments are feasible—unlikely in general and particularly in a sales force context.

where the vector $\Theta_i = \{\alpha_i, \lambda_i, \sigma_i\}$ contains the set of parameters of the sales response and effort policy functions and the distribution of sales shocks, where

$$L_i(\varepsilon) = \frac{1}{\sigma_i (2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma_i}\right)^2} \tag{4}$$

We accommodate unobserved heterogeneity by allowing for discrete segments. Assume that sales person i belongs to one of K segments, $k \in \{1,...,K\}$ with segment probabilities $q_i = \{q_{i1},..., q_{iK}\}$. Let the population probability of being in segment k be π_k. Let $\mathcal{L}(S_{it} \mid z_{it}, k; \Theta_k)$ be the likelihood of individual i's sales being S_{it} at time t, conditional on the observables z_{it}, and the unobservable segment k, given segment parameters Θ_k. Then the likelihood of observing sales history S_i over the time period $t = 1...T$, given the observable history z_i, and the unobservable segment k is given by:

$$L_k(S_i \mid z_i; \Theta_k, \pi_k) = \pi_k \left(\prod_{t=1}^{T} \mathcal{L}_{ikt}\right) \tag{5}$$

where $\mathcal{L}_{ikt} = \mathcal{L}(S_{it} \mid z_{it}, k; \Theta_k)$. As noted earlier we assume the distribution of the revenue shocks to be normally distributed and hence use the normal likelihood for equation (5) as in equation (4). The parameter $\Theta_k = \{\alpha_k, \lambda_k, \sigma_k\}$ is the vector of segment level parameters of the sales response and effort policy function where each λ_k is the parameters that index the effort policy for segment k and σ_k is parameter for the distribution of the revenue shocks for segment k.

By summing over all of the unobserved states $k \in \{1,...,K\}$, we obtain the overall likelihood of individual i:

$$L(S_i \mid z_i; \Theta, \pi) = \sum_{k=1}^{K} L_k(S_i \mid z_i; \Theta_k, \pi_k)$$

and hence the log-likelihood over the N sample of individuals becomes

$$\sum_{i=1}^{N} \log\left(L(S_i \mid z_i; \Theta, \pi)\right) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k \prod_{t=1}^{T} \mathcal{L}_{ikt}\right) \tag{6}$$

Directly maximizing the log-likelihood in (6) is computationally infeasible because the function is not additively separable so we use the approach of Arcidiacono and Jones (2003) and Arcidiacono and Miller (2011) to iteratively maximize the expected log-likelihood in equation (7)

14

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{t=1}^{T}q_{ik}\log\mathcal{L}\left(S_{it}\mid z_{it}\,k;\Theta_k\right) \qquad (7)$$

where q_{ik} is formally defined below as the probability that individual i is of segment type k given parameter values $\Theta=\{\Theta_1,...,\Theta_K\}$, where $\Theta_k=\{\alpha_k,\lambda_k,\sigma_k\}$ and segment probabilities $\pi=\{\pi_1,...,\pi_k\}$, conditional on all of the observed data of individual i.

$$\Pr\left(k\mid S_i,z_i;\Theta,\pi\right)=q_{ik}\left(S_i,z_i;\Theta,\pi\right)=\frac{L_k\left(S_i\mid z_i;\Theta_k,\pi_k\right)}{L\left(S_i\mid z_i;\Theta,\pi\right)} \qquad (8)$$

The iterative process is as follows: We start with an initial guess of the parameters Θ^0 and π^0. A natural candidate for such starting values would be to obtain the parameters from a model without unobserved heterogeneity and slightly perturbing those values.[17] Given the parameters $\{\Theta^m,\pi^m\}$ from the mth iteration, the update of the $(m+1)^{\text{th}}$ iteration is as follows

    a) Compute $q_{ik}^{(m+1)}$ using equation (8) with Θ^m and π^m

    b) Obtain $\Theta^{(m+1)}$ by maximizing (7) evaluated at $q_{ik}^{(m+1)}$

    c) Update $\pi^{(m+1)}$ by taking the average over the sample such that

$$\pi_k^{(m+1)}=\frac{1}{N}\sum_{i=1}^{N}q_{ik}^{(m+1)}$$

We iterate (a) through (c) till convergence.

For the basis functions in the effort policy, we use Chebyshev polynomials of state variables to approximate effort as described in Footnote 14. From the estimation, we obtain the vector of parameters for the basis functions (λ), the vector of parameters for the sales policy (α), and the parameters of the revenue shocks (σ) for each segment k. Also we obtain the population segment probabilities (π) for each segment. Thus, this procedure gives us the sales revenue function $\hat{S}(.)$ and effort policy function $\hat{e}(.)$ for each segment. We use these segment level policies to obtain the structural parameters of each segment.

4.2 Step 2

The key idea of the two-step estimation is that in the first stage we observe the agent's optimal actions. Using these observed optimal actions we are able to construct estimates of the value function, which enables us to estimate the primitives of the model that rationalize these optimal actions.

---

[17] We started the initial values from one tenth of the standard error from the parameter values obtain from a single segment model. The initial values of the segment probabilities were set equally across segments.

Let the value function of a representative agent at state z that follows an action profile e, conditional on the compensation plan ψ, the sales profile S and the primitives of the utility function and discount parameters $\Omega = (\mu, \delta)$ be represented as

$$V\left(z;e;\psi,S,\Omega\right) = E\left\{\sum_{t=0}^{T} D(t)U(e(z_t), z_t, \varepsilon_t; \mu) \mid z_0 = z; \psi, S, \Omega\right\} \tag{9}$$

where $D(t) = \begin{cases} 1, & \text{if } t = 0 \\ \delta^t, & \text{otherwise} \end{cases}$ is the discount factor, and the expectation operator would be over the present and future sales shock ε_t.

Using the estimated sales and effort policy function and the distribution of the sales shocks in the first stage, we are able to forward simulate the actions of sales agents to obtain the estimate of the value function. The detailed simulation procedure is as follows.

    a) From initial state of z_t calculate the optimal actions as $e(z_t)$

    b) Draw sales shock ε_t from $f(\varepsilon)$

    c) Update state z_{t+1} using the realized sales $s(e(z_t)) + \varepsilon_t$

    d) Repeat (a)-(c) until $t = T$

By averaging the sum of the discounted stream of utility flow over multiple simulated paths we can get the estimate of the value function $\tilde{V}\left(z;e(z);\psi,S,\Omega\right)$.[18]

Let $e^s(z)$ be any deviation policy from a set of feasible policies that is not identical to the optimal policy and, by using the same simulation method proposed above, let the corresponding estimate of the value function be called the sub-optimal value function $\tilde{V}\left(z;e^s\left(z\right);\psi,S,\Omega\right)$. Since $e(z)$ by definition is the effort policy and thus at an optimum, then any deviations from this policy rule would generate value functions of less or equal value to that of the optimal level.

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

$$Q(v;\psi,S,\Omega) = V(z;e(z);\psi,S,\Omega) - V(z;e^s(z);\psi,S,\Omega)$$

where $v \in \mathcal{V}$ denotes a particular $\{z, \ e^s(z)\}$ combination.[19] Then if $e(z)$ is the optimal policy, the function $Q(v;\psi,S,\Omega)$ would always have value of greater or equal to zero. Thus our estimate of the underlying structural parameters Ω would satisfy,

---

[18] For each segment, we drew four hundred simulation draws over each period and computed the value functions.
[19] 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

$$\hat{\Omega} = \arg\min \int (\min\{Q(v; \psi, S, \Omega), 0\})^2 dH(v)$$

where $H(v)$ is the distribution over the set $\mathcal{V}$ of inequalities. Our empirical counterpart to $Q(v; \psi, S, \Omega)$ would be $\tilde{Q}(v; \psi, \hat{S}, \Omega) = \tilde{V}(z; \hat{e}(z); \psi, \hat{S}, \Omega) - \tilde{V}(z; \hat{e}^s(z); \psi, \hat{S}, \Omega)$ and as a result our estimates of the structural parameters are obtained from minimizing the objective function in equation (10).[20]

$$\frac{1}{N_I} \sum_{j=1}^{N_I} \left(\min\{\tilde{Q}(v_j; \psi, \hat{S}, \Omega), 0\} \right)^2 \tag{10}$$

The above procedure is performed for each segment with the segment specific effort policies obtained in Step 1. This allows us to estimate the structural parameters for each segment.[21] In practice, since the objective function in equation (10) is relatively flat with respect to changes in the discount factor, it is difficult to pin down the discount parameter with traditional gradient based optimization. We therefore perform a grid search to estimate the discount factor. Specifically, first we estimate the structural parameters by minimizing the objective function in equation (10) over a grid of discount factors. Second, we simulate data with the estimated structural parameters associated with each discount factor to compute the mean absolute percentage error (MAPE) per period with respect to the observed data. We choose the final estimates as the discount factor and the associated structural parameters with the lowest MAPE.[22]

4.3 Identification

There are a couple of major identification challenges. First, we do not observe effort. Hence the link between effort and sales cannot be identified non-parametrically. Second, in dynamic structural models, it is typically impossible to identify discount factors separately from the utility function. Below, we discuss how we address these issues.

Realized sales are a function of demand shifters, effort and additive sales shocks. Conditional on observed demand shifters and 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 assume a deterministic (but flexible) relationship between effort and observable states (*%QQ* and *%AQ* and demand shifters) for each segment. Finally, as we do not observe effort, we need a strictly

---

chose to draw a deviation policy from a normal distribution with mean zero and quarter of the variance from the revenue shock distribution, i.e. $e^s(z) = e(z) + \eta$.

[20] We drew two hundred deviation strategies to construct the objective function and hence $N_I = 200$.

[21] In addition, we used a second set of moment inequalities to reflect the participation constraint that employees continued to work at a firm because they at least obtained a reservation value (normalized to zero); i.e., $\min\{V(z; e(z); \psi, S, \Omega), 0\}$. It turns out these inequalities are non-binding and do not impact our estimates.

[22] The standard errors were computed using subsample bootstrap methods where we subsampled the data and computed our estimates multiple times to construct a bootstrap distribution (Efron, 1979).

monotonic parametric relationship between sales and effort. As we estimate a flexible relationship between observable states and effort, we model the relationship between sales and effort to be linear.

The discount factor is not identified separately from the utility function in standard dynamic structural models because typically there are no variables that do not affect contemporaneous utility, but only future utility. (Rust 1994; Magnac and Thesmar 2002). In the absence of such an exclusion restriction, this implies that if an agent exerts low effort in a period, it is not possible to distinguish if this is due to high disutility for effort, or because they discount future utilities very heavily.[23]

Two aspects of our setting allow us to identify utility functions separately from the discount factor. First, we have a finite horizon setting, where at the end of the year, the quotas are reset and all agents start with a fresh quota for the following year. This means that every December, the agent faces a static optimization problem, conditional on the sales agent's state (*%AQ*). Utility parameters are well identified for a static model, and hence the agent's choice in the last period should allow us to non-parametrically identify the agent's utility function. Given this, variation in sales (that is monotonically linked to effort) in the last period and variations in wealth should help identify the effort disutility and risk aversion coefficient within the utility function.

Second, the bonus setting generates exclusion restrictions between current and future utility; i.e., we have instruments in non-bonus periods that do not affect current utility, but only future utility. As we demonstrated with reduced form evidence earlier, the fact that an agent's performance in November is related to his proximity to the annual bonus that will only be given in December indicates forward looking behavior. We also demonstrated other evidence of how agents respond to quarterly or annual quotas even though they do not affect current payoffs. This allows us to estimate the discount factor.

Beyond these conceptual arguments, we designed a simulation study to demonstrate that the structural parameters of the utility function and discount factors can be identified using our empirical strategy. The simulation is described in the appendix.

5   Results

We first report the first stage estimates of the demand shifters and effort policy function for the sales response model; then we report estimates of structural parameters of sales agents' utility functions from the second stage estimation. We then perform several counterfactual simulations to address the substantive questions we seek to answer.

5.1   First Stage Estimates

The parameter estimates for the demand shifters in the sales response function is reported in Table 6a. We find that only the interaction term between lagged annual quota and indirect sales

---

[23] Another reason why an agent might exert low effort is that they may have wrong expectations about the transition density of future states, i.e., they may be very pessimistic about future good states. Like other dynamic structural modeling papers, we assume rational expectations for the transition densities of states. In this case, this translates into a rational expectations assumption on sales shocks.

revenue are statistically significant. Thus larger markets tend to have a bigger sales multiplier independent of effort in high demand periods.

We estimate segment level effort policy functions by estimating the non-parametric relationship between sales and state variables through Chebyshev polynomials of the state variables. We estimated up to fourth order Chebyshev polynomials with alternative number of segments and choose the best fitting model based on the Bayesian Information Criterion (BIC). The best fit was for the model with up to third order Chebyshev polynomials, allowing for three segments. The estimates of the best fitting polynomial function and the standard deviations of the revenue shocks for each segment are reported in Tables 6b and 6c. As the coefficients associated with the Chebyshev polynomials have no intuitive meaning, for intuition, we show graphs of the effort policy function for the three segments as a function of percentage annual quota (*%AQ*) for select months in Figure 7a. *%AQ* is normalized across sales agents, such that 1 implies at quota and 0.9 indicates 10% below quota and 1.1 indicates 10% above quota.[24]

Table 7 shows the share of the three segments and their descriptive characteristics. Segment 2 is the largest with a share of 47%; Segments 1 and 3 have shares of 32% and 21% respectively. The average tenure with the firm is not very different across segments at approximately 12 years. Segment 3 has the highest annual quotas, followed by Segment 2 and Segment 1. Interestingly, Segments 2 and 3 with larger quotas achieve their quota targets more often than Segment 1 which has trouble meeting quota.

Figure 7a shows the Segment 3 exerts the most effort and is the most productive segment, and Segment 1 exerts the least effort and is the least productive segment. This is consistent with the allocated quotas and percentage of time quotas are achieved in Table 7. We also see a positive relationship between exerted effort and *%AQ* for all months shown. As for *%QQ*, we see an increasing but concave relationship in March implying that once a sales person is way above the quarterly quota she starts to gradually slow down. Given that the average states in March for each segment were 0.55, 0.58, 0.62, respectively, not a lot of sales people are in the position to slow down. Effort in December does not fall off even if the sales person has already reached or exceed quota *(%AQ>1)*, likely due to the overachievement commissions in preventing sales people from lowering effort after achieving quota. Our results are consistent with Steenburgh (2008), who finds that sales people "give up" when far away from achieving quota, such as for all segments in our case, but do not slow down much once quota is reached.

---

[24] The average coefficient of variation in the annual quotas *across* agents (i.e., $CV = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_i^{AQ}}{\mu_i^{AQ}}$ where μ_i^{AQ} and σ_i^{AQ}

are the mean and the variance of the annual quotas over three years for agent i) is 0.21. In contrast, the average coefficient of variation of monthly states (distance to annual quota) within agents over the three years is substantially larger at 3.12. Thus, our primary source of identification comes due to changes in sales in response to changes in states *within* agents.

Figure 7b shows the effect of tenure on effort for all segments. Sales people in segment 2 and 3 initially increase effort with experience, but this tapers off with time. This is probably due to the fact that in the early years of their careers, they want to work hard not only for monetary payments from increased wages but also other intangible incentives such as promotions or transfers to better job titles. However, after a certain amount of years, these intangibles don't matter as much and the effort levels tend to taper off. Interestingly, Segment 1, the lowest productivity segment, does not gain in productivity from experience.

The two-step estimation procedure relies on correct estimation of optimal actions in the first stage. We have shown that the estimated optimal effort policy makes intuitive sense in the illustration above. However, one limitation of this approach is that the unobserved error components in the first stage estimation might be serially correlated across time, which would lead to biased estimates of the structural parameters in the second stage (Bajari et al. 2007). We mitigate this problem by accounting for permanent heterogeneity via methods proposed by Arcidiacono and Miller (2011) and by using a very flexible set of polynomials of state variables to capture most of the persistence in sales.

5.2 Second-Stage Structural Parameter Estimates

The first column of Table 8 shows the estimates of the dynamic model with exponential discounting. We find the long-term discount factor δ to be 0.91 and highly significant. Frederick, Loewenstein and O'Donoghue (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. For purely monetary values, the estimated discount factor seems rather low. But as Frederick, Loewenstein and O'Donoghue (2002) point out, for behavioral aspects such as pain and thus in our case effort, the discount factors tend to be low and, hence, our estimate appears to be reasonable.

The parameters for disutility of effort are negative and significant for all three segments. Their relative magnitudes are consistent with the effort policy functions estimated in the first stage. Segment 3, which produces the greatest sales on average, has the lowest disutility for effort. Segment 1, which has the lowest sales, has the greatest disutility. The risk aversion coefficients for all segments are insignificant showing no direct evidence of risk aversion by the sales agents. This may be because in the range of incomes earned by the sales force, risk aversion is not a serious concern. The estimated model fits the observed sales revenue data reasonably well with a mean absolute percentage error (MAPE) of 11%.

Hyperbolic discounting

The constant exponential discounting (Samuelson 1937) model is the normative standard for intertemporal behavior. Researchers in marketing and economics, therefore, routinely assume such behavior among agents in estimating dynamic structural models. Yet, the psychology and behavioral economics literature has shown strong evidence that hyperbolic discounting (Thaler 1981, Ainslie 1992, Laibson 1997) explains agent behavior better in many settings. Under hyperbolic discounting, agents discount

the immediate future from the present more heavily than the same time interval starting at a future date. It is typically represented by the quasi-hyperbolic discount function $D(t) = \beta \delta^t$ (Phelps and Pollak (1968), Elster (1979), Laibson (1997, 1998) where $\beta < 1$ is the short-run *present bias factor* and δ is the *long-term discount factor*. Hyperbolic discounting can lead to time inconsistent preferences and preference reversals. For example, a hyperbolic discounter can prefer \$100 today (t=0) to \$120 in a year (t=1), yet prefer \$120 in two years (t=2) to \$100 in one year (t=1).

For a given level of disutility for effort, the exponential and hyperbolic discount models differ in their predictions for inter-temporal effort under the current bonus compensation scheme. Under exponential discounting, the agent's effort would be smoother over bonus and non-bonus periods, while under hyperbolic discounting, the agent would concentrate effort in the bonus period. In our finite horizon setting, the utility function is uniquely identified off the last period, where there is no forward looking behavior and the problem reduces to a static model. Given knowledge of the utility parameters, we are able to exploit the differences in predictions between hyperbolic and exponential discounting to estimate a present bias factor in addition to the long-term discount factor where the discount function in equation (9) would be given as

$$D(t) = \begin{cases} 1, & \text{if } t = 0 \\ \beta \delta^t, & \text{otherwise} \end{cases}.$$

We report the results of the model with hyperbolic discounting in the second column of Table 8. The parameters of the utility function are virtually identical with the exponential discounting model. However, the hyperbolic discounting model has a better overall fit with a MAPE of 10.7%.

The end of the quarter or year effect where a salesperson's effort is driven by certain deadlines can be confounded with the effect of hyperbolic discounting. Yet, we partially account for these effects by controlling for seasonality through the performance of the commission only sales force who are also likely to have the deadline effect, if present.[25] Hence, we use the estimates from the hyperbolic discounting model to compute the following counterfactual analysis.

5.3   Assessing the value of a dynamic structural model

How important is it to model the dynamics of salesperson behavior? In a static model, any effort would be attributed to current payoff, not accounting for the large future bonuses. This will downward bias the salesperson's disutility parameters and overstate the effects of compensation on productivity. The third column of Table 8 reports the estimates of the myopic model – discount factor set to zero. As expected, the disutility parameters are smaller in magnitude relative to the forward looking model for all segments. For Segment 3 the downward bias is as much as 22%. The myopic model also has a

---

[25] Another confounding effect would be the multidimensional aspects of effort where closing of effort may not be too costly and is done at the end of the quarter or year. However, given the product characteristics of the focal firm it is difficult to assume that closing of such deals systematically occur at the end of the quarter or year.

poorer fit: a MAPE of 18.8% relative to the MAPE of 11.0% and 10.7% for the exponential and hyperbolic dynamic model, respectively.

We next compare the revenue and effort predictions between the dynamic and myopic models. To isolate the effects of forward looking behavior, we simulate based on the disutility estimates from the dynamic model, but set the discount parameters to zero for the myopic model. Figure 8 compares the predicted revenues and effort of the myopic and dynamic models. The myopic agent has systematically lower revenues because she does not take into account the effect of future bonuses and overachievement commission in current effort. In contrast, the forward looking agent anticipates that in an uncertain environment, there is a chance of bad shocks later, which may prevent getting to the quota, so they prepare for such a rainy day by working harder early on so that they are within striking target of quota even if a bad sales shock occurred.

The effort graph in Figure 8 enables us to isolate out the sales revenue cyclicality and focus on the differences in effort across dynamic and myopic agents. The myopic salesperson concentrates much more effort in the bonus period, but the forward looking sales person smoothes effort over time, given the uncertainty in future demand shocks. The effort peaks in the bonus periods are not as pronounced for the dynamic agent. The observed effort smoothing is similar to consumption smoothing by forward looking consumers facing uncertain incomes in the development economics literature.

5.4   Counterfactual 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. The overall change in revenues under the alternative conditions is reported in Table 9 and the effect by segment in Table 10. Second, we compare the role of bonus frequency— how quarterly and annual bonuses affect performance. In running the counterfactual simulation studies, given a regime change, we randomly chose 25 representative sales agents from each segment (total of 75 salespeople) and simulated 1,000 simulation paths for each of them to obtain the productivity of each segment. Then, we weighted each segment based on the estimated segment sizes reported in Table 7.

Value of Nonlinear Incentive Components

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 20.8% greater with the current compensation plan compared to a pure commission plan. Not surprisingly, given the lower incentives in the absence of quotas, bonuses and overachievement commissions, Table 10 shows that all segments suffer (about equally) from substantially poorer performance. Even after adjusting commission rates to be higher to make total compensation identical to the current plan, revenues are 4% higher with the current compensation plan, suggesting that the non-linear plan shifts salespeople

22

into high power areas of incentives. To understand the role of different components, we next investigate counterfactuals based on individual components—one at a time.

Value of Overachievement Compensation

When overachievement commissions are eliminated, not surprisingly, overall revenues drop by 13.3%. Even accounting for the additional commission costs, profits are lower by about 2% (assuming gross margin of 33%).

To gain insight into how overachievement commissions impact sales, Figure 9a compares the effort level of sales agents who eventually meet/do not meet the annual quota. For those who met the annual quota, the effort level does not decline even when close to the quota because of the overachievement commission. In contrast, those who did not meet quota decrease effort towards the end of the year as they are unlikely to meet quota and therefore overachievement commission has no impact on their earnings. Thus overachievement commission provides the incentives for the most productive sales people even if they have already met quota (or likely to meet quota). Not surprisingly, Table 10 indicates that overachievement commissions have the greatest impact on Segments 3, the most productive segment. Revenues drop by about 17% for this segment, while the effect on the least productive segment is substantially smaller at 7%. Overall, we conclude that over-achievement commission rates motivate high-performing sales people to continue exerting high levels of effort when they are close to and exceed quota, but have limited impact on the low performing sales-people.

Value of Cumulative Annual Quota

Next we investigate the effect of cumulative annual quota by replacing the cumulative quota with just a fourth quarter quota. To study this, we remove the overachievement commission (which is based on reaching the annual quota) and split the total annual bonus payments across all four quarters. Overall, revenues drop by 15.8%. This decrease is greater than the 13.3%, when we just dropped the overachievement commissions. Thus the cumulative annual quota induces sales agents to exert greater effort and raise revenues by 2.5%.

How does the cumulative annual quota work? On the negative side, the cumulative annual quota reduces the incentive of salespeople who have not had success in the first three quarters to continue to put in effort in the fourth quarter. But on the positive side, the cumulative annual quota causes salespeople who reach their quarterly quota in earlier quarters to continue to extend themselves in order to get them closer to the annual target—this additional effort early on, insure themselves against negative shocks in future quarters and allows them to be within striking distance of the annual quota. On balance, the positive effect of the latter outweighs the negative effect of the former.

We also consider the case where we split the annual quota into a quarterly bonus and an annual bonus so that people do not "give up" in the last quarter when they are far away from quota. While this did increase the effort in the last quarter, it reduced revenues overall because sales people did not put in as much effort earlier in the year to be within striking distance of annual quota, because it is not as large. Total revenues drop by 1%.

23

Thus these results show that cumulative annual quota has a significant impact on salespeople by inducing the agents to work harder even when they reach their quarterly targets; again motivating the high performing salespeople to remain productive and bring additional sales.

Quota-Bonus Frequency

We next investigate the value of quarterly bonuses relative to annual bonuses. Figure 9b shows the comparison of effort between the current plan and when quarterly bonuses are eliminate and only the annual bonus is left. Effort drops consistently across the year when there are no quarterly quotas. Overall revenues fall by 5%. Even in December, when there is the annual bonus on the table, revenue falls by 2% and effort falls by 4%. Thus annual bonus and over-achievement commissions have less of an impact on year-end performance without quarterly bonuses. Why?

The quarterly bonus induces sales agents to work harder in a given quarter. But it also helps them achieve the annual quota by helping them stay on track of their annual goal. Without quarterly bonus, sales agents do not have much incentive to work hard early on. This lack of incentive leads them to be farther away from the annual quota by December. Annual bonuses and over-achievement commission have little impact on effort as sales agents are more likely to give up meeting annual quota.

The impact of quarterly bonuses also differs across the three segments of consumers. Table 10 indicates that quarterly bonuses have relatively minimal impact on Segment 3, the most productive segment, but very high impact on Segment 1. In effect, quarterly bonuses are needed as pacers to the less productive sales people than for the most productive sales people.

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 behavioral psychology (Heath et al., 1999). The basic idea is that achieving short-term goals make achieving long-term goals more feasible. Our analysis show that the short term goals are more valuable to the least productive segment; i.e., in education terms it may imply that weaker students gain more by periodic testing, relative to stronger students who would study independent of exams.

Discussion

Overall, our results provide empirically grounded substantive insight on the role of various elements of nonlinear compensation to improve productivity. The numerical example in Section 2.2 shows that quotas and bonuses serve as important goals for the average performers, inducing them to increase their effort; our empirical analysis shows support for the role of the quota and bonus as a motivational goal and stretch incentives respectively. However, it is important to recognize that other elements of the compensation plan are critical in improving the performance of both high performers and laggards. First in this setting, managers recognize that quota updating based on the salesperson's own performance in the previous year can induce ratcheting effects that reduce productivity among the highest performing salespeople who do not book sales much higher than their quota in order to avoid

higher quotas next year. By updating quotas, based not on an individual's performance, but only those of a larger group of salespeople, managers minimized ratcheting effects. In fact, empirically, we found little statistical evidence of effort shading due to ratcheting effects. Second, high performers, who are most likely to reach quota will reduce effort upon reaching quota if there are no additional incentives. The firm used overachievement commissions and cumulative annual quotas to induce the high performing sales agents to continue to put in effort and excel in their sales performance. In combination with avoiding ratcheting effects, the firm was thus able to keep their best sales agents to continue high levels of performance. Finally, quarterly bonuses serve as pacers to help the least productive segment motivated to help them within striking distance of their long-term goals. By shedding empirically grounded insight on the differential motivational roles of different incentive components of a nonlinear compensation plan for star, average and laggard sales agents, this paper contributes substantively to literature on role of nonlinearities in quota bonus plans and explains its widespread popularity in practice (Joseph and Kalwani 1998).

6   Conclusion

Personal selling is a primary marketing mix tool for most B2B firms to generate sales, yet there is little research on how the compensation plan motivates the sale-force and affects performance. This paper develops and estimates a dynamic structural model of sales force response to a compensation plan with various components: salary, commissions, lump-sum bonus for achieving quotas, and different commission rates beyond achieving quotas. Our analysis helps us assess the impact of (1) different components of compensation and (2) the differential importance of periodic bonuses on performance on different segments of sales people.

We find that the quota-bonus scheme used by this firm increases performance of the sales force by serving as intermediary goals and pushing employees to meet targets. Features such as overachievement compensation reduce 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 distance of their annual quotas. In the absence of quarterly bonuses, failure in the early periods to accomplish targets caused agents to fall behind more often than in the presence of quarterly bonuses. Thus, quarterly bonus serves as a valuable sub-goal which helps the sales force stay on track in achieving their overall goal; they are especially valuable to low performers. In contrast, overachievement commissions increase performance among the highest performers.

In this empirical application, we introduce two important methodological innovations to the marketing literature. First, we are the first empirical implementation of the Arcidiacono and Miller (2011) approach to accommodate unobserved heterogeneity within a two-step estimation framework. Second, we demonstrate that discount factors can be estimated in naturally occurring field data using appropriate exclusion restrictions—and overturn the conventional wisdom that estimation of discount factors requires augmentation with survey data.

We now discuss limitations of the paper, which provide promising avenues for future research. First, effort tends to be multi-dimensional and one possibility is that quotas and bonuses force people to focus on the effort that lead to final sales in bonus periods, while agents may focus on earlier stages of the selling process in non-bonus periods. Such a multidimensional effort cannot be identified merely from sales data. We hope data from CRM databases which track customer stages through the selling process can help shed insight on this issue. We believe this is an exciting area for future research.

Second, compensation contracts can serve to select the right type of sales people. We do not address selection issues. One possibility is to use a longer panel of sales people's performance that includes attrition information. If there were variation in contracts that affected employee retention, that could also help address this problem. One needs more work on scenarios with richer contracts. For example, one could study peer effects on sales performance and selection effects when firms shift from individual to team based compensation (Chan et al., 2009).

In summary, this paper provides a rigorous framework to empirically understand 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. Our analysis helps obtain a number of useful substantive insights. Nevertheless, the issues raised above provide an interesting agenda for future work.

References

Albers, Sonke, Murali K. Mantrala, and Shrihari Sridhar (2010), "Personal Selling Elasticities: A Meta-Analysis," *Journal of Marketing Research*, 47-5, pp. 840-853.

Ainslie, G. (1992), *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*, Cambridge University Press.

Arcidiacono, Peter and John B. Jones (2003), "Finite Mixture Distributions, Sequential Likelihood, and the EM Algorithm," *Econometrica,* 71-3, pp. 933-946.

Arcidiacono, Peter and Robert A. Miller (2011), "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, forthcoming.

Assmus, Gert, John U. Farley and Donald R. Lehmann (1984), "How Advertising Affects Sales: Meta-Analysis of Econometric Results," *Journal of Marketing Research*, 21-1, pp. 65-74.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007), "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75-5, pp. 1331-1370.

Bangert-Drowns, Robert L., James A. Kulik and Chen-Lin C. Kulik (1991), "Effects of frequent classroom testing," *Journal of Educational Research*, 85-2, pp. 89-99.

Basu, Amyia K., Rajiv Lal, V. Srinivasan and Richard Staelin (1985), "Sales-force Compensation Plans: An Agency Theoretic Perspective," *Marketing Science*, 8-3, pp. 324-342.

Beauchamp, Andrew (2010), "Abortion Supplier Dynamics," Working Paper, Duke University.

Bijmolt, Tammo H.A., Harald J. Van Heerde and Rik G.M. Pieters (2005), "New Empirical Generalizations on the Determinants of Price Elasticity," *Journal of Marketing Research* 42-2, pp. 141–56.

Chan, Tat Y., Jia Li, and Lamar Pierce (2009), "Competition and Peer Effects in Competing Sales Teams," Working Paper, Washington University, St. Louis.

Chevalier, Judith and Austan Goolsbee (2009), "Are Durable Goods Consumers Forward-Looking? Evidence from College Textbooks," *The Quarterly Journal of Economics*, 124-4, pp. 1853-1884,

Copeland, Adam and Cyril Monnet (2009), "The Welfare Effects of Incentive Schemes," *Review of Economic Studies*, 76-1, pp. 96-113.

Darmon, Rene (1997), "Selecting Appropriate Sales Quota Plan Structures and Quota-setting Procedures," *Journal of Personal Selling & Sales Management*, 17-1, pp. 1-16.

Efron, Bradley (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, pp. 1-26.

Elster, Jon (1979), *Ulysses and Sirens: Studies in Rationality and Irrationality*, Cambridge University Press.

Ferrall, Christopher, and Bruce Shearer (1999), "Incentives and Transactions Costs within the Firm: Estimating an Agency Model Using Payroll Records," *Review of Economic Studies*, 66-2, pp. 309-338.

Finger, Stephen R. (2008), "Research and Development Competition in the Chemicals Industry," Working Paper, Duke University.

Frederick, Shane, George Loewenstein and Ted O'Donoghue (2002), "Time Discounting and Time Preference: A Critical Review", *Journal of Economic Literature*, 40-2, pp. 351-401.

Heath, Chip, Richard P. Larrick and George Wu (1999), "Goals as Reference Points," *Cognitive Psychology*, 38, pp. 79-109.

Heide, Christen P. (1999), *Dartnell's 30th Sales Force Compensation Survey*, Dartnell Corporation; 30th edition.

Holmstrom, Bengt (1979), "Moral Hazard and Observability," *Bell Journal of Economics*, 10-1, pp. 74-91.

Holmstrom, Bengt and Paul Milgrom (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55-2, pp. 303-328.

Hotz, V. Joseph. and Robert A. Miller (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60-3, pp. 497-529.

Jiang, Renna and Robert W. Palmatier (2009) "Structural Estimation of a Moral Hazard Model: An Application to Business Selling," Working Paper, UC Davis.

Joseph, Kissan and Manohar Kalwani (1998), "The Role of Bonus Pay in Sales force Compensation Plans," *Industrial Marketing Management*, 27-2, pp. 147-159.

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

Kim, Son Ku (1997), "Limited Liability and Bonus Contracts," *Journal of Economics and Management Strategy*, 6-4, pp. 899-913.

Laibson, David (1997), "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112-2, pp. 443-477.

Laibson, David (1998), "Life-Cycle Consumption and Hyperbolic Discount Functions," *European Economic Review*, 42, pp. 861-871.

Lal, Rajiv and V. Srinivasan (1993), "Compensation Plans for Single-and Multi-product Sales forces: An Application of the Holmstrom – Milgrom Model," *Management Science*, 39-7, pp. 777-793.

Magnac, Thierry and David Thesmar (2002), "Identifying dynamic discrete choice models," *Econometrica*, 70-2, pp. 801-816.

Mirrlees, James A. (1975), "The Theory of Moral Hazard and Unobservable Behaviour: Part I," Mimeo, Oxford. (Published in 1999 in the *Review of Economic Studies*, 66, pp. 3-21.)

Misra, Sanjog and Harikesh Nair (2011), "A Structural Model of Sales-force Compensation Dynamics: Estimation and Field Implementation," *Quantitative Marketing and Economics*, 9-3, pp. 211-25.

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

Oyer, Paul (2000), "A Theory of Sales Quotas with Limited Liability and Rent Sharing," *Journal of Labor Economics*, 18, pp. 405-426.

Paarsch, Harry, J. and Bruce Shearer (2000), "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records," *International Economic Review*, 40, pp. 59-92.

Park, Eun-Soo (1995), "Incentive Contracting under Limited Liability," *Journal of Economics and Management Strategy*, 4, pp. 477-490.

Phelps, Edmund. S. and Robert Pollak (1968), "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, 35, pp. 185-199.

Raju, Jagmohan S., and V. Srinivasan (1996), "Quota-Based Compensation Plans for Multi-Territory Heterogeneous Sales-Forces," *Management Science*, 42, pp. 1454-1462.

Rao, Ram C. (1990), "Compensating Heterogeneous Sales-forces: Some Explicit Solutions", *Marketing Science*, 9, pp. 319-342.

Rust, John (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, pp. 999-1033.

Rust, John (1994), "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, Volume 4, ed. by R.E. Engle and D. McFadden, Amsterdam: Elsevier-North Holland, Chapter 14, pp. 3082-3139.

Samuelson, Paul (1937), "A Note on Measurement of Utility," *Review of Economic Studies*, 4, pp. 155-161.

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

Sudhir, K. (2001), "Structural Analysis of Manufacturer Pricing in the Presence of Strategic Retailer," *Marketing Science*, 20-3, pp. 244-264.

Thaler, Richard H. (1982), "Some Empirical Evidence on Dynamic Inconsistency," *Economic Letters*, 8, pp. 201-206.

Yao, Song, Carl F. Mela, Jeongwen Chiang, and Yuxin Chen (2012) "Determining Consumers' Discount Rates With Field Studies," *Journal of Marketing Research*, forthcoming

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

Table 1: Firm's Compensation Plan

Type	Description	Payment period
Quarterly Bonus	$1500 Awarded if quarterly revenue exceeds quarterly quota	Mar, Jun, Sep
Annual Bonus	$4000 Awarded if annual revenue exceeds annual quota	Dec
Base Commission	About 1.5%[*] paid in proportion to the revenue generated each month	Every month
Overachievement Commission	About 3%[*] paid in proportion to the total cumulative revenue surpassing the annual quota	Dec

[*]*These numbers are approximate for confidentiality reasons.*

Table 2: Descriptive statistics

No. salespeople	348		
Average salary (USD)	$3,585		
Average tenure (years)	11.8		
	Average Quota '000 USD	% Achieving Quota	Average Sales '000 USD
Quarter 1	232.4	51.1	273.0
Quarter 2	374.2	49.8	445.0
Quarter 3	397.1	42.8	407.0
Quarter 4	-	-	565.6
Annual	1,639.3	49.9	1,690.6

Table 3: Testing for Ratcheting Effects[*]

Dependent Variable	Model 1 $\dfrac{AQ_y - AQ_{y-1}}{AQ_{y-1}}$	Model 2 $\log(\boldsymbol{AQ_y})$	Model 3 $\log(\boldsymbol{AQ_y})$
$I(\%AQ_{y-1}^{Dec})$	0.0364	0.0033	
	(0.0229)	(0.0299)	
$\log(\%AQ_{y-1}^{Dec})$			0.0347
			(0.0196)
Year fixed effects	Yes	Yes	Yes
Sales person fixed effects	NA	Yes	Yes

[*]*y represents each year in this analysis.*

Table 4: Sales Performance in November

	$\%AQ{<}0.5$	$\%AQ{>}0.5$
	Estimate	Estimate
Intercept	0.05^{***}	0.06^{***}
	(0.018)	(0.0123)
$\%AQ$	0.06	0.04^{***}
	(0.049)	(-.0126)

[***]: $p<0.01$

31

Table 5: Testing for Sales Substitution across Quarters

	Model 1	Model 2
Other months of quarter	168.87***	101.76***
	(34.57)	(34.35)
First month of quarter	147.79***	101.86***
	(34.87)	(34.52)
(Other months of quarter)*(Previous month % distance-to-quota)	91.09***	56.58***
	(5.62)	(5.90)
(First month of quarter)*(Previous month % distance-to-quota)	2.59	0.15
	(7.02)	(6.93)
ISR (indirect revenue) index		38.72***
		(2.26)
Salesperson fixed effects	Yes	Yes

*** : $p<0.01$

Table 6a: Parameter Estimates – Sales Response

Lagged annual quota	0.002
	(0.005)
Indirect sales	-6.735
	(5.554)
Indirect sales*Lagged annual quota	0.022***
	(0.003)

*** : $p<0.01$

Table 6b: Parameter Estimates – Effort Policy Function

	Segment 1	Segment 2	Segment 3
ρ_0	-18.63	**-74.23****	**-263.00*****
	(26.09)	(32.74)	(89.28)
$\rho_1(z_1)$	**141.10*****	**146.42*****	**312.36****
	(44.32)	(55.41)	(137.85)
$\rho_2(z_1)$	-23.54	-21.67	**-148.39*****
	(17.09)	(22.37)	(63.23)
$\rho_3(z_1)$	6.41	6.98	15.87

	(3.94)	(5.14)	(13.09)
$\rho_1(z_2)$	**102.26*****	**241.73*****	**458.59*****
	(38.00)	(44.93)	(117.20)
$\rho_2(z_2)$	-22.13	**-62.91*****	-87.28
	(18.67)	(21.89)	(55.50)
$\rho_3(z_3)$	**8.09***	**11.34*****	15.90
	(4.39)	(4.54)	(10.79)
$\rho_1(z_1)\rho_1(z_2)$	-114.52	-31.98	110.00
	(75.28)	(82.95)	(181.61)
$\rho_1(z_1)\rho_2(z_2)$	-16.60	0.27	**-113.83*****
	(19.77)	(18.40)	(40.33)
$\rho_2(z_1)\rho_1(z_2)$	9.07	-10.14	**86.81*****
	(14.63)	(17.08)	(36.85)
$\rho_1(z_1)\rho_1(ISR)$	**-20.00****	8.12	42.42
	(8.95)	(11.31)	(29.62)
$\rho_1(z_2)\rho_1(ISR)$	-1.11	**16.76*****	**20.10*****
	(2.68)	(3.18)	(8.53)
$\rho_1(z_1)\rho_1(z_2)\rho_1(ISR)$	9.52	**-43.91*****	**-89.11****
	(17.62)	(18.88)	(44.87)
$\rho_1(z_1)\rho_1(AQ_{y-1})$	**-0.04****	**-0.03****	**-0.07*****
	(0.02)	(0.02)	(0.03)
$\rho_1(z_2)\rho_1(AQ_{y-1})$	-0.01	**-0.06*****	-0.02
	(0.01)	(0.01)	(0.01)
$\rho_1(z_1)\rho_1(z_2)\rho_1(AQ_{y-1})$	0.04	**0.05*****	0.02
	(0.04)	(0.03)	(0.04)
$\rho_1(\tau)$	-0.53	**9.00*****	**18.53*****
	(1.38)	(2.14)	(6.90)
$\rho_2(\tau)$	0.05	**-0.27*****	**-0.52****
	(0.05)	(0.07)	(0.26)
$\rho_3(\tau)$	0.001	**0.002*****	0.004
	(0.000)	(0.001)	(0.003)

*****: p<0.01, ****: p<0.05, *: p<0.1

Table 6c: Revenue Shock Distribution – Standard Deviation

	Segment 1	Segment 2	Segment 3
Sigma	81.61	143.72	279.35

Table 7: Descriptive Characteristics of Segments

	Segment1	Segment2	Segment3
Share	0.32	0.47	0.21
Tenure[*]	11.5	12.2	11.5
Achieve quarterly quota - Q1	0.46	0.54	0.58
Achieve quarterly quota - Q2	0.38	0.55	0.62
Achieve quarterly quota - Q3	0.31	0.49	0.53
Achieve annual quota	0.30	0.57	0.64
Average annual quota[**]	1,201.4	1,615.7	2,363.7
Average December revenue[**]	130.2	273.0	559.1

[*]Tenure is measured in years [**]Average quotas and revenues are indicated in USD(K)

Table 8: Structural Parameters

	Exponential	Hyperbolic	Myopic
Present bias		0.8***	
factor		(0.03)	
Long-term	0.91***	0.92***	
discount factor	(0.02)	(0.02)	
Segment 1			
Disutility	-0.241***	-0.240***	-0.199***
	(0.005)	(0.005)	(0.006)
Risk	-0.0001	-0.0001	-0.0001
Aversion	(0.0015)	(0.0020)	(0.0003)
Segment 2			
Disutility	-0.118***	-0.119***	-0.092***
	(0.002)	(0.038)	(0.005)
Risk	-0.0001	-0.0001	-0.0001
Aversion	(0.0004)	(0.0005)	(0.0005)
Segment 3			
Disutility	-0.061***	-0.061***	-0.048***
	(0.002)	(0.002)	(0.002)
Risk	0.0000	0.0000	0.0000
Aversion	(0.0003)	(0.0002)	(0.0004)
MAPE	0.110	0.107	0.188

***: $p<0.01$

Table 9: Impact of Alternative Bonus Plans on Sales Revenues

Counterfactual	Change in Revenues
1a. Only Pure Commissions	-20.8%
1b. Only Pure Commissions (adjusted to equal payout with bonus)	-3.8%
2a. No Bonus (Only Commissions + Overachievement Commission)	-9.3%
2b. No Bonus (Commissions adjusted to equal payout with bonus)	-1.5%
3. No overachievement commissions	-13.3%
4a. Cumulative Annual Quota replaced with quarterly quota	-4.2%
4b. Annual Bonus split into Quarterly and Annual Bonus	-1.0%
5a. Remove quarterly bonus	-4.6%

Table 10: Impact of Alternative Bonus Plans on Sales Revenues by Segment

% decrease from different components	Seg1	Seg2	Seg3
Pure commission	17.9%	21.0%	21.4%
Without overachievement	7.0%	12.6%	17.1%
Without quarterly bonus	10.0%	4.5%	2.0%

Figure 1: Types of Incentive Compensation Schemes
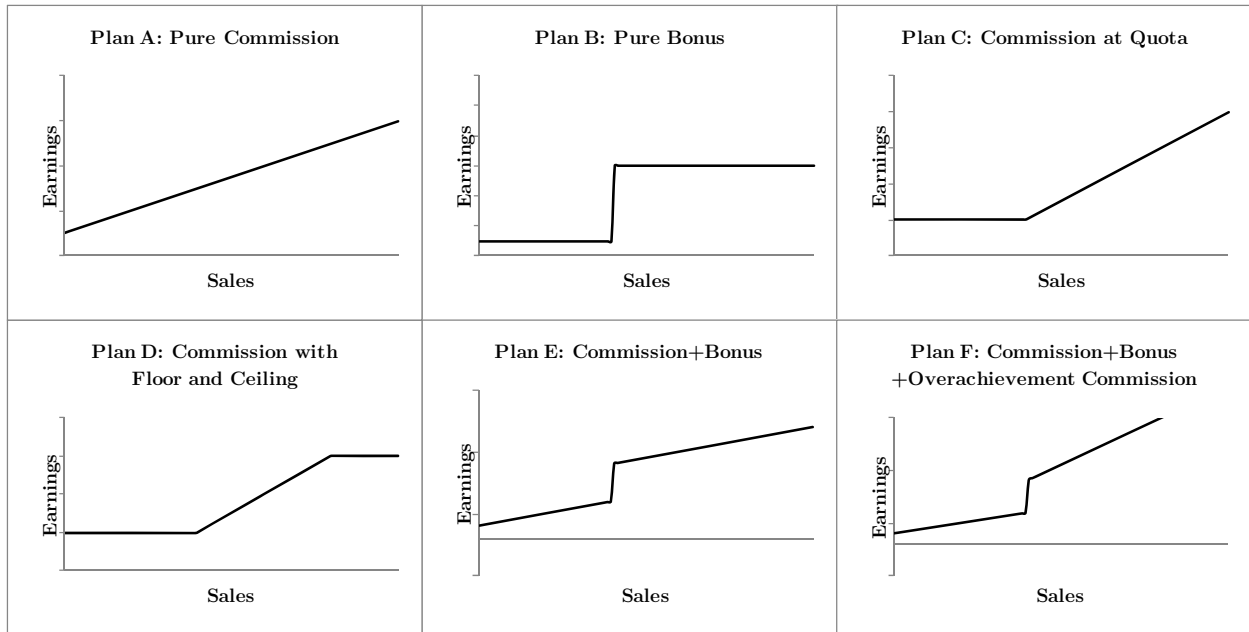
Figure 2a: How Quotas and Bonus Serve as Stretch Goals

$d=1, r=10, B=0,$
$e^*=5$

$d=1, r=10, B=30$
$e^*=10$

d=1, r=20, B=0
$e^*=10$
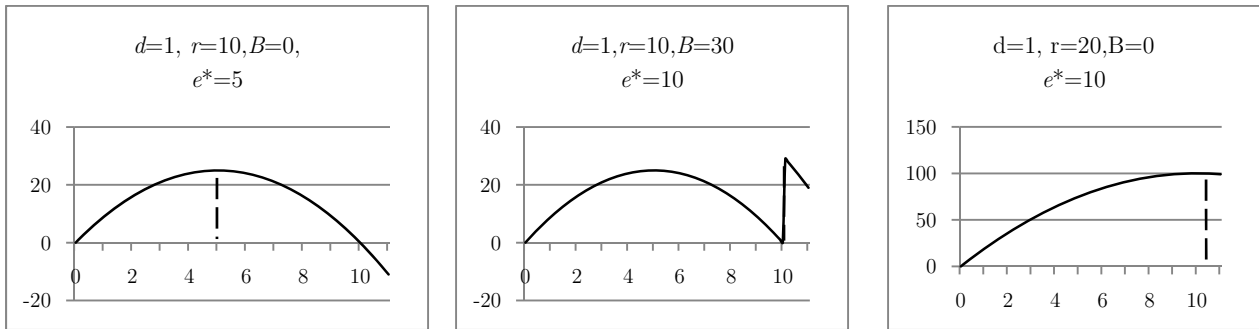
Figure 2b: Effort as a Function of Distance to Quotas

$d=2, r=10, B=30$
$S=0, e^*=2.5$

$d=2, r=10, B=30$
$S=5, e^*=5$

$d=2, r=10, B=30$
$S=7, e^*=3$
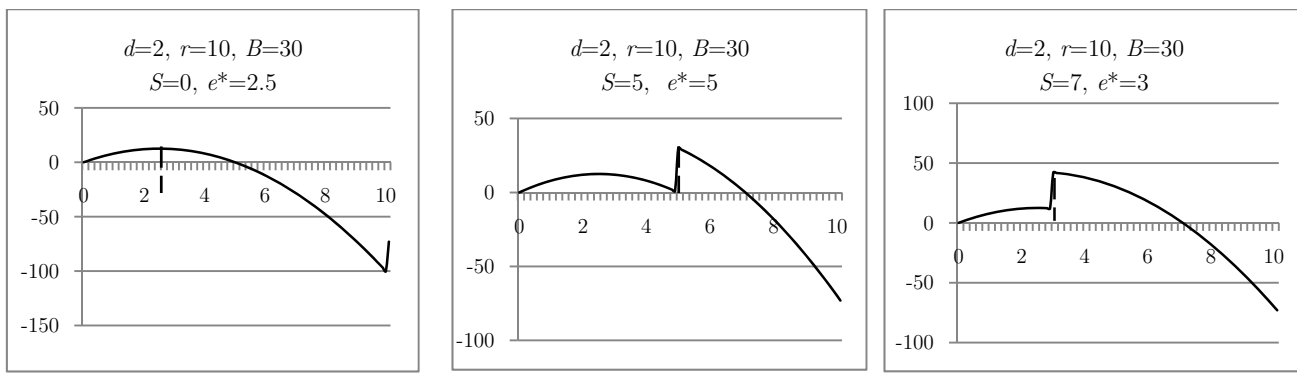
Figure 3: Fraction of People Achieving Quota in December

$\%AQ$ at end of October

36
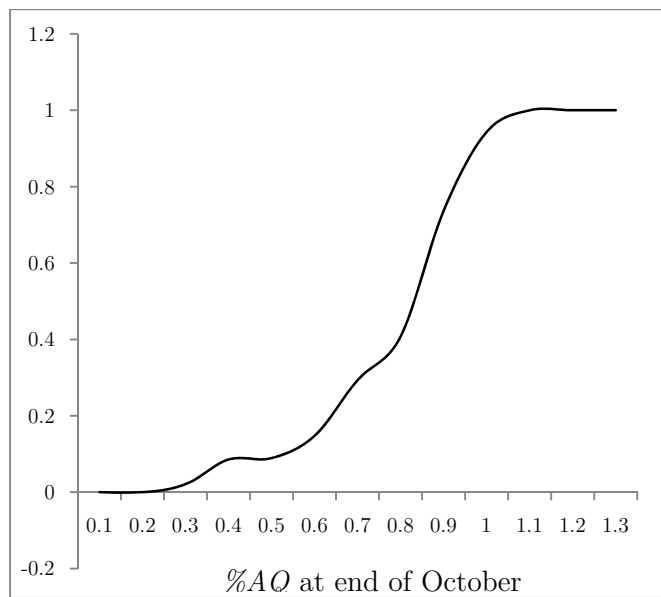
Figure 4a: Sales and Percentage Quota Achieved – Bonus Months
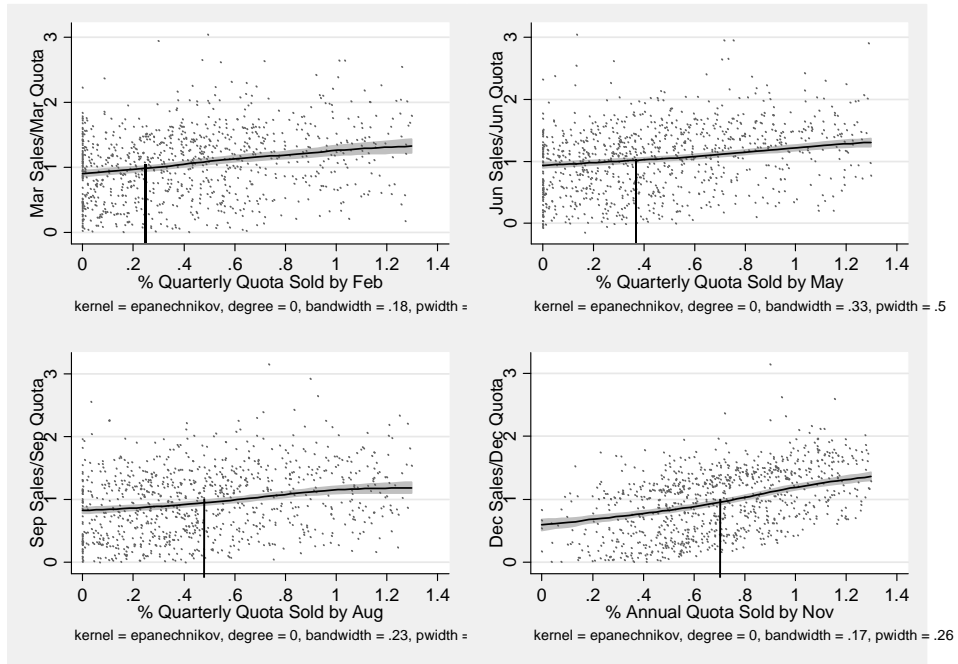
Figure 4b: Sales and Percentage Quota Achieved – Pre-Bonus Months
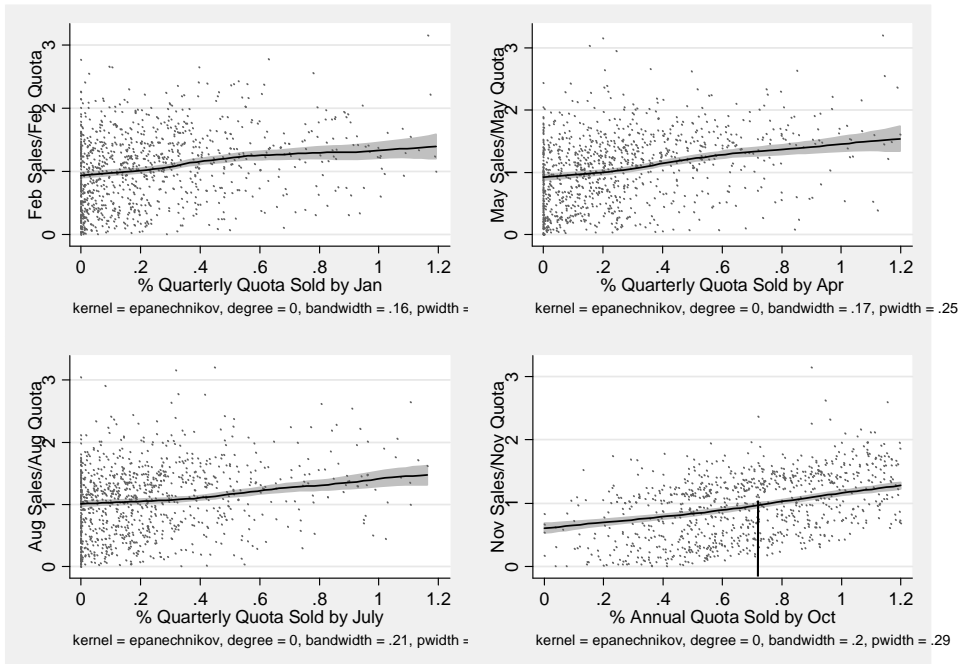
Fig 5a: Revenues from Regular Sales Force
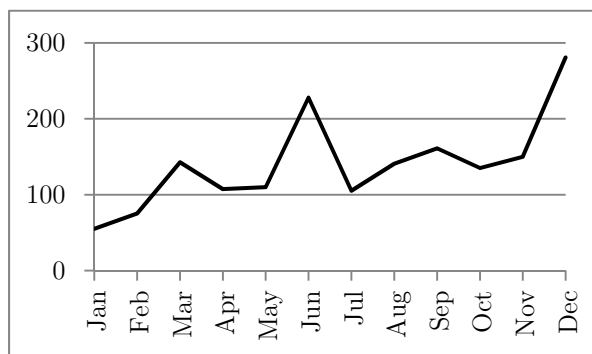
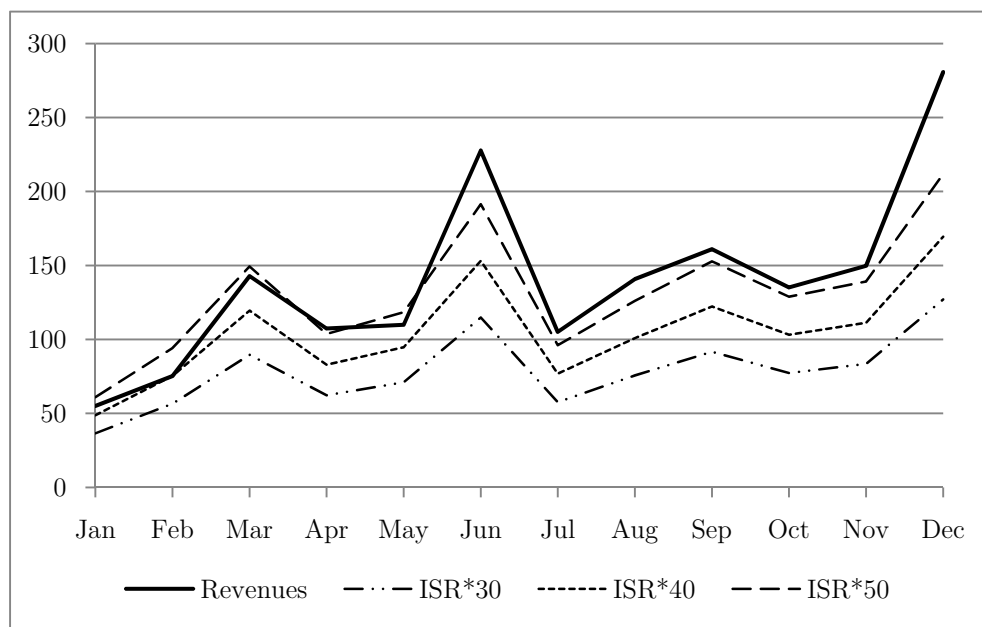**Fig 5b: Indirect Sales Revenue (ISR) Index[26]**
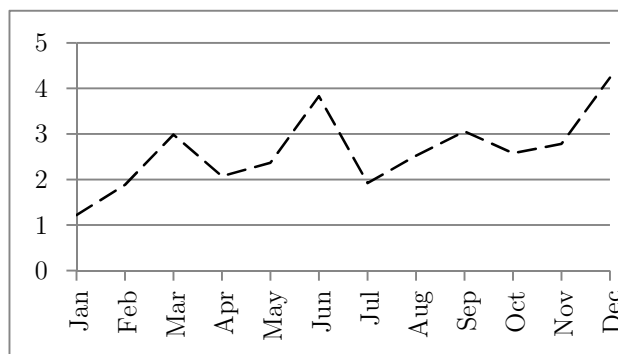
Fig 5c: Revenues from Regular Sales Force and ISR Index Multiples

---

[26] The indirect sales force revenue (ISR) index is on a base of 1 reflecting the indirect sales in January 1999. Here we have averaged the index across the three years from 1999-2001.
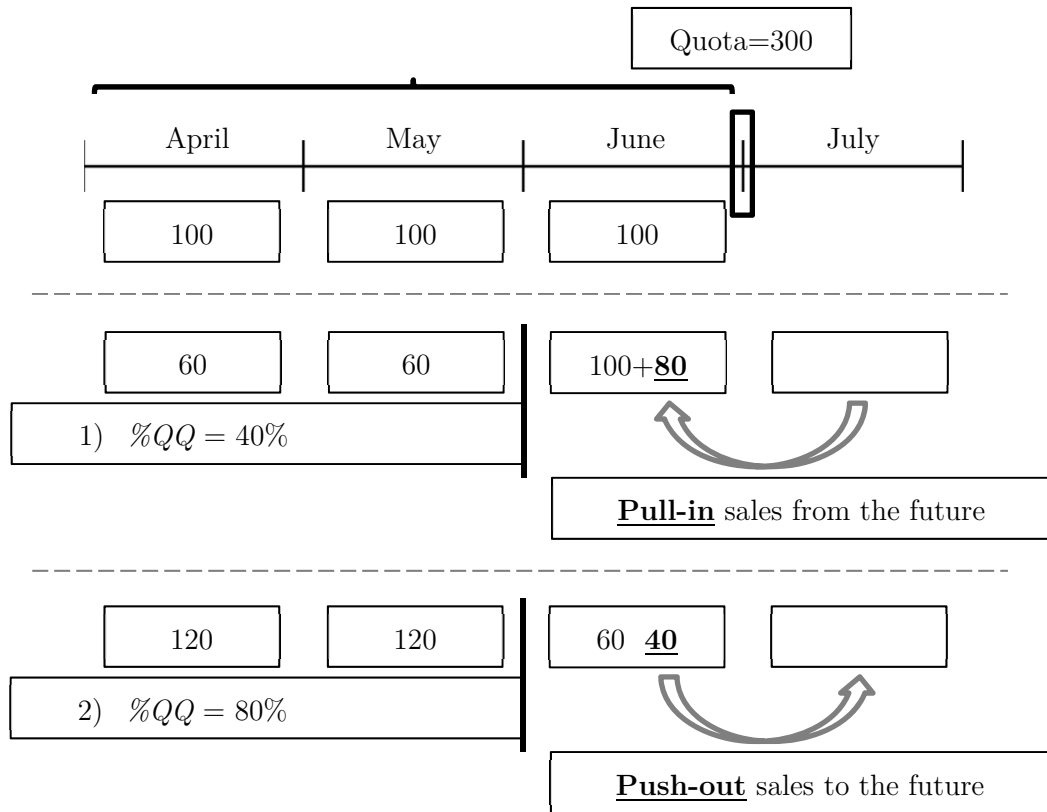
Fig 6: Sales Substitution across Quarters

Quota=300

April	May	June	July

| 100 | 100 | 100 |

| 60 | 60 | 100+**80** | |

1) *%QQ = 40%*

Pull-in sales from the future

| 120 | 120 | 60 **40** | |

2) *%QQ = 80%*

Push-out sales to the future

Figure 7a: Effort Policy by Segment as a Function of % Quota

July

December

March
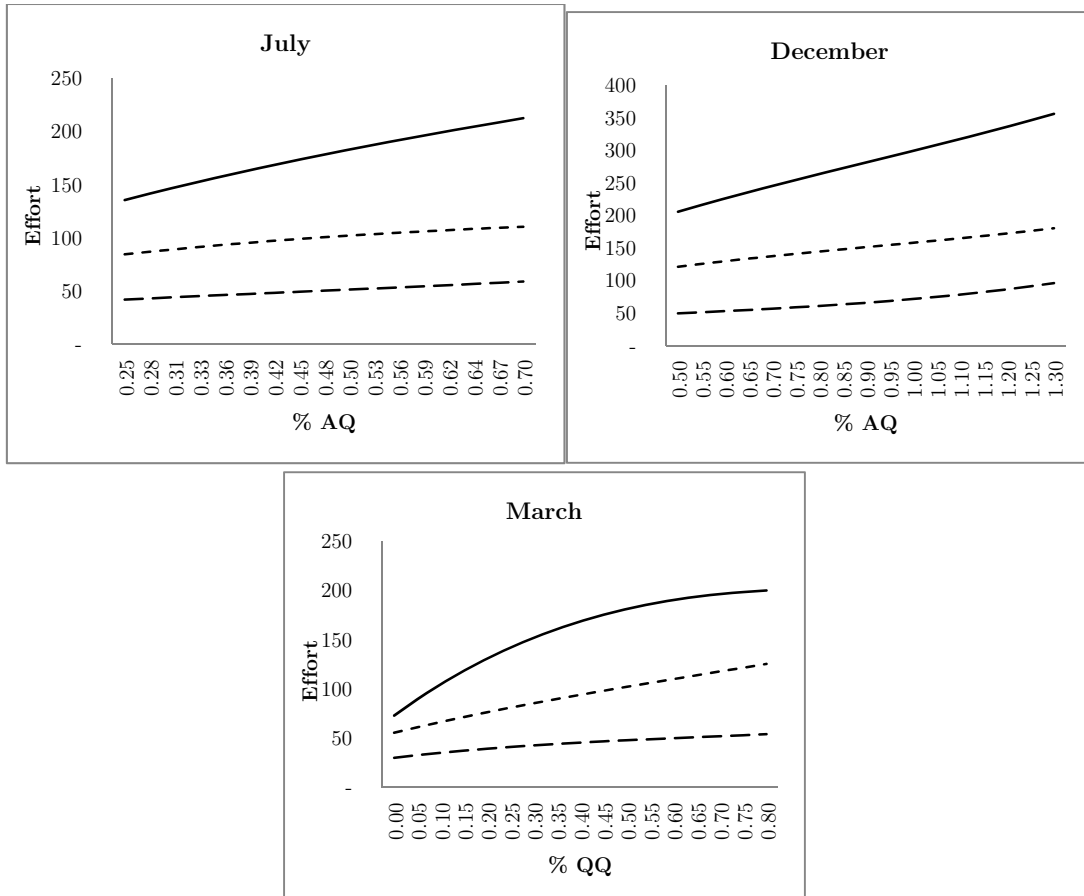
Figure 7b: The Effect of Tenure on Effort

The Effect of Tenure
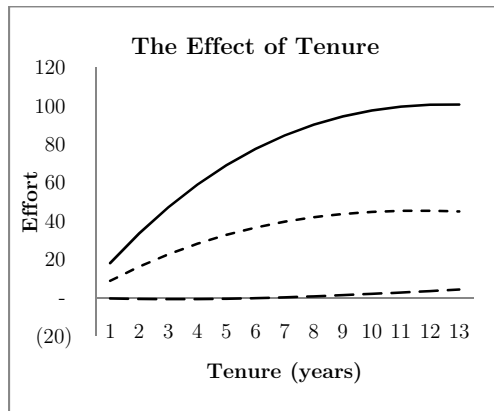
Figure 8: Simulated Revenue & Effort– Static vs. Dynamic
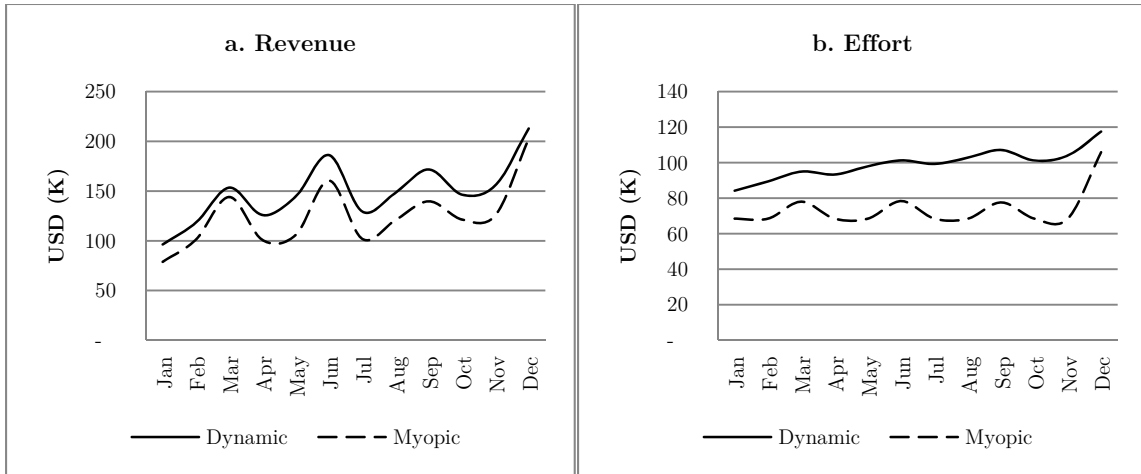

a. Revenue

b. Effort
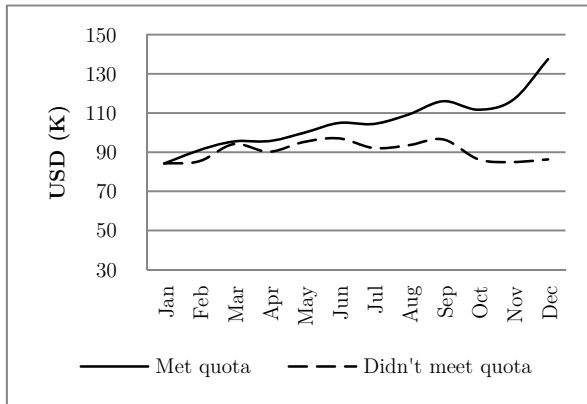
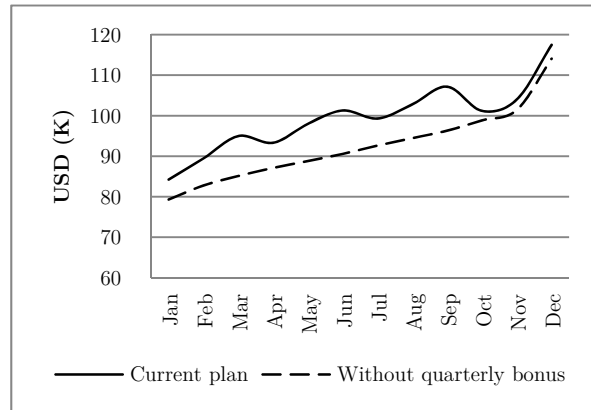**Figure 9a: Overachievement Commission**

Figure 9b: Quarterly Quotas

Appendix

A.1. Simulation Study to identify structural parameters controlling for seasonality

Let realized sales be a function of effort, function of demand shifter z_t and a normal random sales shock ε_{it} such that,

$$y_{it} = h(z_t) + e_{it} + \varepsilon_{it}, \ \varepsilon_{it} \sim N(0, \sigma^2).$$

To illustrate identification, we consider a simulation scenario with seasonality and quarterly quotas. The demand shifter z_t represents seasonality index comparable to the ISR index in our empirical setting. For the purpose of the simulation, we assume the function $h(.)$ is linear in the seasonality index i.e., $h(z_t) = \alpha z_t$. Let salesperson's utility be given as $u_{it} = -de_{it}^2 + W_{it}$, where $W_{it} = BI_{\{t=3,6,9,12\}}I_{\{s_{it}+y_{it}>Q\}}$ is the wealth earned from bonus payment B for achieving quarterly quota Q. The cumulative sales state within quarter evolves as follows,

$$s_{it} = \begin{cases} 0, \ \text{if} \ t \ \text{is start of quota period} \\ s_{i(t-1)} + y_{it}, \quad\quad\quad \text{otherwise} \end{cases}.$$

We report the results of the simulation with seasonality and quarterly quotas using three years of data. Let the vector of seasonality index for each month throughout the year be given by (2, 3, 5, 1, 2, 3, 3, 4, 7, 2, 3, 5). We set bonus at $B=60$ and quarterly quota at $Q=30$. Consistent with the model in the paper, we estimate the first stage estimates using our specification in equation (1) where we structure the per-period policy function to be affected through the effects of z (ISR index) and its interaction with s (distance to quarterly quota).

We varied the simulated number of individuals from 50 to 1000. The true values and the estimates and standard errors for each simulation are reported in Table A-1. Even with 200 individuals (relative to our sample size of 350), we are able to recover the disutility parameter and the discount factor with reasonable precision. We therefore conclude that our approach to estimating the policy function using interactions between z and s can indeed recover the structural parameter estimates.

A.2. Simulation Study to identify hyperbolic discount factors

We consider a similar model to that presented above, but with the addition of a present bias factor consistent with the hyperbolic discounting model. Again, we varied the simulated number of individuals from 50 to 1000. The true values and the estimates and standard errors for each simulation are reported in Table A-2. Again, the disutility parameters, discount factors and the standard deviation of the sales shocks are all estimated precisely even with 200 individuals, lending confidence to the identification arguments in Section 4.3.

Table A-1: Simulation Results – Seasonality

No. individuals	True Values		
	d=-0.1	*δ=0.9*	*σ=10*
50	-0.132	0.892	10.282
	(0.052)	(0.019)	(0.320)
200	-0.109	0.898	10.043
	(0.019)	(0.005)	(0.150)
400	-0.103	0.899	9.899
	(0.011)	(0.002)	(0.104)
1000	-0.109	0.898	9.958
	(0.007)	(0.002)	(0.065)

Table A-2: Simulation Results – Discount Factors

No. individuals	True Values			
	d=-0.1	*β=0.6*	*δ=0.9*	*σ=10*
50	-0.159	0.590	0.887	10.282
	(0.055)	(0.241)	(0.327)	(0.320)
200	-0.109	0.599	0.898	10.043
	(0.043)	(0.009)	(0.010)	(0.150)
400	-0.097	0.600	0.901	9.899
	(0.018)	(0.003)	(0.004)	(0.104)
1000	-0.107	0.599	0.899	9.958
	(0.012)	(0.002)	(0.002)	(0.065)