Sales Effort Management Under All-or-Nothing Constraint

Longyuan Du,[a] Ming Hu,[b] Jiahua Wu[c]

[a] School of Management, University of San Francisco, San Francisco, California 94117; [b] Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada; [c] Imperial College Business School, Imperial College London, London SW7 2AZ, United Kingdom
Contact: ldu5@usfca.edu, https://orcid.org/0000-0003-1628-7539 (LD); ming.hu@rotman.utoronto.ca, https://orcid.org/0000-0003-0900-7631 (MH); j.wu@imperial.ac.uk, https://orcid.org/0000-0002-2254-1858 (JW)

Abstract. We consider a sales effort management problem under an all-or-nothing constraint. The seller will receive no bonus/revenue if the sales volume fails to reach a predetermined target at the end of the sales horizon. Throughout the sales horizon, the sales process can be moderated by the seller through costly effort. We show that the optimal sales rate is nonmonotonic with respect to the remaining time or the outstanding sales volume required to reach the target. Generally, it has a *watershed* structure, such that for any needed sales volume, there exists a cutoff point on the remaining time above which the optimal sales rate decreases in the remaining time and below which it increases in the remaining time. We then study easy-to-compute heuristics that can be implemented efficiently. We start with a static heuristic derived from the deterministic analog of the stochastic problem. With an all-or-nothing constraint, we show that the performance of the static heuristic hinges on how the profit-maximizing rate fares against the target rate, which is defined as the sales target divided by the length of the sales horizon. When the profit-maximizing rate is higher than the target rate, the static heuristic adopting the optimal deterministic rate is asymptotically optimal with negligible loss. On the other hand, when the profit-maximizing rate is lower than the target rate, the performance loss of any asymptotically optimal static heuristic is of an order greater than the square root of the scale parameter. To address the poor performance of the static heuristic in the latter case, we propose a modified resolving heuristic and show that it is asymptotically optimal and achieves a logarithmic performance loss.

Keywords: revenue management • dynamic programming-optimal control: applications • marketing: salesforce • all-or-nothing constraint

1. Introduction

Freddie Hoyt is the general manager at the Town & Country Chrysler dealership in Levittown, New York. On one typical working day, he was sitting in his office, frowning at the whiteboard hanging on the wall. The whiteboard recorded each person's sales for the month to date. It was the second-to-last day of the month, and they were 16 cars behind the target of 129 cars assigned by Chrysler for the month. Chrysler would pay them from $65,000 to $85,000 if the target was achieved by the end of the month. This commission was so crucial for the dealership that it would essentially determine whether they could make a profit that month (This American Life 2013).

Quota-based compensation contracts, where a seller will receive a lump-sum bonus if and only if her number of sales exceeds a given quota, are prevalent in salesforce management. It is reported that 89% of firms use some variation of a "quota-like" all-or-

nothing mechanism (Oyer 2000). With the all-or-nothing constraint, one crucial question for a decision maker is how to maximize the chance of reaching the target, and, therefore, the payout, by the deadline. In the case of Town & Country, when Freddie saw that the sales were moving slowly, his solution was to push salespeople at the dealership to work harder: "I've got to be at that number … So put your nose right to the ground and come out shooting today, everybody!" (This American Life 2013). Instead of waiting for customers coming to the store, salespeople were asked to call up any lead they might have to try to close deals over the phone. Most worked long hours on the last days of the month.

In general, with a quota-based compensation plan, a seller needs to carefully determine the amount of costly effort she would like to exert over the evaluation period. A delicate balance needs to be struck between effort and the chance of reaching the quota.

A higher level of effort generally leads to a higher sales number, which increases the likelihood of reaching the quota. However, because of the stochastic nature of demand, the costly effort can become futile if the sales fall short.

In this paper, with the context of sales effort management in mind, we study how a seller can exert efforts to maximize her expected profit under an all-or-nothing constraint. Specifically, we consider a model where demand arrives following a Poisson process and the intensity is moderated by the seller through costly effort. The seller's performance is evaluated at the end of the sales horizon. Her payout includes a lump-sum bonus and an over-target commission per unit if the number of sales is greater than or equal to the predetermined target; otherwise, she receives nothing.[1]

We formulate this problem as a dynamic program and study the structure of the optimal policy. As our problem of sales effort management is equivalent to a dynamic pricing problem in which the seller may control sales intensity by varying prices, we compare our results with those obtained in classical revenue-management (RM) settings (see more discussions in Section 2). We show that, when the lump-sum bonus is greater than the per-unit bonus commission, which arguably always holds in practice, the optimal sales rate is not monotonic with respect to the remaining time or the number of sales required to reach the target. The nonmonotonicity in the optimal sales rate differs fundamentally from results in the existing RM literature, where the optimal sales rate is typically monotonic with respect to both time-to-go and remaining inventory (see, e.g., Gallego and van Ryzin 1994). With the all-or-nothing constraint, two opposing forces drive the optimal sales intensity. On the one hand, the all-or-nothing constraint incentivizes the seller to increase the sales, with the hope of reaching the target by the end of the sales horizon. However, the cost will be sunk and the seller will incur a loss if the target is not reached in the end. These two forces together shape the optimal sales intensity and create a turning point in time: For any needed sales volume, the seller will increase the sales intensity as the deadline gets closer, but only to a point. After the turning point, the optimal rate will decrease as time-to-go decreases. Although the two forces that drive the monotonicity of the optimal solution are intuitive, showing the existence of the turning point is not trivial. To that end, we develop a novel approach that leads to a tight bound on the optimal solution. With the carefully constructed bound, we prove that as time-to-go increases, the optimal sales intensity approaches its limit from above, which guarantees that the turning point must be finite.

Next, we propose various heuristics. To evaluate their performance for problems with the all-or-nothing constraint, we first study the deterministic approximation of the problem and seek to come up with a performance bound for the stochastic problem. This approach is motivated by common observations in the RM literature that the optimal deterministic profit is an upper bound for that of the stochastic problem. Surprisingly, we show, with a counterexample, that this statement does not always hold in our setting. With an all-or-nothing constraint, the possibility of a random surge in demand may actually help the seller by pushing sales beyond the target, especially when the sales target is so high that it is not likely to be reached without random demand shocks. Nonetheless, we prove that when some system primitives are scaled up, the optimal profit of the deterministic problem is, indeed, an upper bound for that of the stochastic problem. Thus, we can use it to gauge the performance of heuristics in an asymptotic regime.

We consider an easily implementable, static heuristic, where the sales intensity remains constant throughout the sales horizon. In traditional RM settings, such as Gallego and van Ryzin (1994), a static heuristic that adopts the optimal solution of the deterministic problem is proven to be asymptotically optimal and has a square-root revenue loss. However, with an all-or-nothing constraint, we show that the performance of the static heuristic hinges on how the profit-maximizing rate fares against the target rate, defined as the sales target divided by the length of the sales horizon. When the profit-maximizing rate is higher than the target rate, the static heuristic of the optimal deterministic rate performs extremely well, as the absolute profit loss converges to zero when the target and sales horizon scale up proportionally.

This is intriguing because in many traditional RM problems, the absolute performance loss of the optimal static heuristic is nonnegligible. On the other hand, when the profit-maximizing rate is lower than the target rate, the performance of the static heuristic is compromised. We show that, for the heuristic to be asymptotically optimal, the seller must induce a constant rate higher than the optimal deterministic rate. This is because, with the all-or-nothing constraint, ensuring that the target will almost certainly be reached at the end of the sales horizon is essential for any heuristic to be asymptotically optimal. This can be achieved for a static heuristic only by a rate higher than the optimal deterministic rate to account for demand uncertainty. To find asymptotically optimal static heuristics, we apply the Central Limit Theorem and the results of Poisson tail bounds to find the minimum increase in the sales intensity that will ensure a high probability of the target being reached. Because of the extra cost associated with this positive sales boost, the performance loss is of an order greater than the square root of the scale parameter.

To address the poor performance of the static heuristic when the profit-maximizing rate is lower than the target rate, we investigate periodic resolving heuristics where the seller updates the sales rate by resolving the deterministic problem periodically, with the updated number of sales required to reach the target and the updated remaining time. A standard resolving heuristic, which updates sales rates with the optimal solution to the updated deterministic problem, is proven to be asymptotically optimal in traditional RM problems and yields a logarithmic performance gap (Jasin 2014). However, we show that the same policy is suboptimal with an all-or-nothing constraint. This, again, is because there is no guarantee that the sales target can be reached with the bare optimal deterministic rate, even though it is being updated periodically over the course of the sales horizon. Instead, we propose a *modified* resolving heuristic, where the heuristic adopts the optimal deterministic rate in the early stage, but switches to a higher "full-speed" rate either when the remaining time is short, but the target has not been reached, or when the number of sales is well below that expected. In this heuristic, the time switch to full-speed mode is a crucial parameter. To ensure a good overall performance of the heuristic, we bound the loss in the probability of reaching the target and the cost of extra effort from the higher sales intensity. This allows us to select a switching time that balances the two terms. This modified resolving heuristic is proven to be asymptotically optimal, with a recovered logarithmic performance loss.

2. Literature Review

Our paper is closely related to dynamic pricing problems in RM. In a classical RM setting, a seller seeks to maximize revenue from selling a fixed amount of inventory over a time horizon by varying prices. In sales effort management, the seller seeks to maximize payout from selling products over a given sales horizon by dynamically adjusting the level of effort. We show that our sales effort management problem is equivalent to an RM problem, because the price control can be viewed as the sales intensity control, as long as there is a one-to-one correspondence between price and sales intensity.

Some financial constraints considered in the RM literature resemble the all-or-nothing constraint in our setting. Yuri et al. (2008) consider a dynamic pricing problem where the seller cares about not only the expected revenue, but also the probability of meeting a revenue target as a chance (soft) constraint. They formulate the problem as a continuous-time optimal control problem and study the structural properties of the optimal solution. In contrast, the all-or-nothing constraint in our paper is, by definition, a *hard* constraint on

each sample path, which results in totally different optimal policies. Besbes and Maglaras (2012) study dynamic pricing policies for problems with a series of financial constraints on revenues and sales along the sales horizon. The seller would be penalized for failing to reach the milestone targets. The authors derive heuristics based on the deterministic analog of the stochastic problem and prove their optimality in an asymptotic regime. In contrast to their paper, where the seller's penalties are continuous with respect to revenues/sales, in our setting, the seller's payout function is *discontinuous* around the sales target under an all-or-nothing constraint. This discontinuity plays a crucial role in shaping the optimal policy, as it, along with the seller's sunk costs over the sales horizon, creates both motivation and deterrent for the seller to exert effort, leading to nontrivial distinctions in the optimal strategy compared with those in classical RM problems. More recently, Besbes et al. (2018) build upon the dynamic pricing framework of Gallego and van Ryzin (1994) and study the effect of a debt on pricing, where the seller only collects the residual revenues after the debt is paid off. In a similar context, Ahn et al. (2021) study the optimal policy for a price-taker (as opposed to a price-setter in Besbes et al. 2018) to sell assets at the market price over time under debt obligations. In their settings, the seller's payoff is the positive part of the revenue minus the debt, a continuous function with a kink at the debt. In addition to the fact that the seller's payout function is discontinuous in our setting, our paper also differs from Besbes et al. (2018) and Ahn et al. (2021) in one important respect: The target in our paper is specified with respect to sales volume, rather than revenue. The reason why this seemingly trivial difference is critical is because, when the target is with respect to sales volume, the only way for the seller to reach the target is to sell more products by lowering prices. However, the same cannot be said when the target is with respect to revenue. In this case, in order to increase the chance of hitting the revenue target at the end, the seller can choose either to sell the products in larger quantities by charging a lower price or to charge a higher price and hope that demand produces a favorable result. Besbes et al. (2018) show that the latter strategy, even though it has greater variability, is preferred in a dynamic pricing problem under debt obligations. The structural property of the optimal policy in our paper differs significantly from that in Besbes et al. (2018). Apart from exploring the structural property of the optimal policy for problems with an all-or-nothing sales volume constraint, we also focus on the construction of easy-to-compute heuristics and study their performances.

Driven by the curse of dimensionality, many researchers focus on the construction of heuristics for

dynamic pricing problems. The approach most closely related to our paper is the pricing policy based on the deterministic counterpart of the stochastic problem. Gallego and van Ryzin (1994) show that the optimal profit achieved in the deterministic problem is an upper bound to that of the stochastic problem. When the problem is scaled up with the time and state proportionally growing, a static heuristic that adopts the optimal solution of the deterministic problem is asymptotically optimal and has a performance gap of the square root of the scale parameter. In contrast, we show in this paper that many of the properties associated with the optimal static heuristic in the classical RM setting are overturned in the presence of an all-or-nothing constraint. First, the optimal deterministic profit is no longer a universal upper bound for that of the stochastic problem. This upper bound only holds if the scale of the problem is sufficiently large. Second, a static heuristic based on the optimal deterministic solution is not asymptotically optimal. For a static heuristic to achieve asymptotic optimality, a non-trivial margin needs to be added to the optimal deterministic solution to ensure that the sales target can almost certainly be reached at the end of the sales horizon. Lastly, the optimal static heuristic with the all-or-nothing constraint has a larger-than-square-root performance loss, a typical magnitude in the traditional RM settings.

One common criticism of a static heuristic is that the price is determined upfront, which means that the price has no chance to respond to demand realizations that unfold over time. Because of this, researchers propose to resolve the deterministic problem repeatedly throughout the sales horizon with up-to-date sales and inventory information. Such "resolving" heuristics have been studied extensively in the RM literature with great success (see, e.g., Cooper 2002, Maglaras and Meissner 2006, Jasin and Kumar 2012, and Jasin 2014). In particular, Jasin (2014) shows that a heuristic based on resolving the deterministic problem periodically in the framework of Gallego and van Ryzin (1994) achieves a logarithmic performance loss, which is significantly lower than that of the optimal static heuristic. However, we demonstrate that, with an all-or-nothing constraint, a standard resolving heuristic is not asymptotically optimal. This, again, is owing to the discontinuity around the target in the seller's payout function. We devise an asymptotically optimal modified resolving heuristic that prioritizes cost minimization during the early stage of the sales horizon. However, the priority is switched to a full-speed rate, maximizing the probability of hitting the target, when either the remaining time is short or the number of sales is much lower than expected.

On the application side, our paper is related to the widely studied salesforce compensation problem (see,

e.g., Coughlan 1993). The theoretical root of this literature can be traced back to the principal-agent framework. Under different assumptions about sales-person utility, distribution of sales outcomes, etc., the optimal compensation plan can be nonlinear in the total sales (Basu et al. 1985), linear in the total sales (Lal and Srinivasan 1993), or have a quota-based bonus structure (Oyer 2000). Despite different views on the optimal salesforce compensation plan, sales quotas are prevalent in the industry (Joseph and Kalwani 1998). Part of the reason is that, even though a sales quota might not be optimal under certain circumstances, its performance is quite close to that of the optimal compensation policy (Raju and Srinivasan 1996). In the operations management literature, the hockey-stick phenomenon, which refers to the surge in sales toward the end of an evaluation period as a direct consequence of a sales quota, receives particular attention. Sales fluctuation has an adverse effect on production and inventory planning in a supply chain. To mitigate the hockey-stick phenomenon, Chen (2000) proposes a moving-time-window evaluation schedule to smooth the demand process. Sohoni et al. (2010) show that the reduction of demand variability, along with better information, helps to dampen the hockey-stick phenomenon. It is worth mentioning that the D-contract studied in Sohoni et al. (2010) is a special case of the all-or-nothing contract studied in this paper. However, the focus of our paper is rather different. We explore how to dynamically adjust a seller's effort to maximize the expected payoff under an exogenously given contract.

The all-or-nothing constraint is also prevalent in online crowdfunding, which has recently attracted tremendous interest from researchers. A typical crowdfunding project will follow an all-or-nothing scheme, where the project creator will receive the fund if and only if the goal is reached before the end of the crowdfunding campaign. The papers more related to our work are those that focus on the dynamics of crowdfunding pledging processes. Alaei et al. (2021) and Zhang et al. (2021) study crowdfunding dynamics to find the optimal upfront design of projects, including goals, duration of crowdfunding campaigns, etc. In contrast, we explore, with the terminology of crowdfunding, how a campaign creator can dynamically adjust backers' arrival rate with her costly effort to maximize the cumulative pledged amount at the end of a crowdfunding campaign. Du et al. (2017) study stimulus policies in crowdfunding with network externalities, where project creators can contingently switch or stop offering an incentive scheme for the rest of the campaign to induce customers' pledging, as an optimal stopping-time problem. Stimulus costs of crowdfunding projects, such as promised updates, are incurred only if the

project reaches the goal. In contrast, our paper requires the seller to make costly efforts immediately to increase sales intensity and allows her to increase sales intensity in continuous time. This creates a dilemma for the seller because the spending is wasted if the target cannot be reached by deadline. Closest to ours in terms of high-level research questions, Burtch et al. (2021) study analytically and empirically if and when a creator should send out referral links, and how many, to drive traffic to the campaign throughout the pledging process. The authors focus on the effect of positive network externalities of existing pledgers on future pledging behavior, without considering an all-or-nothing constraint, which is consistent with their empirical context of Indiegogo, which uses the keep-it-all payment scheme rather than the all-or-nothing payment scheme for creators. As mentioned, the presence of an all-or-nothing constraint fundamentally changes the structure of the problem, which is our main focus.

Lastly, the all-or-nothing mechanism is also commonly used in innovation contests, where firms outsource innovation projects to the crowd. Researchers show that the "winner-take-all" scheme—that is, the prize is only awarded to the best performer—is optimal for a wide range of occasions (see, e.g., Kalra and Shi 2001 and Moldovanu and Sela 2001). In this stream, the works closest to ours are those papers that focus on dynamic games and study how contestants dynamically adjust their effort-provision levels over time (see, e.g., Choi 1991, Malueg and Tsutsui 1997, Halac et al. 2017, and Bimpikis et al. 2019). The common recurring theme in this literature is that contestants' effort-provision level gradually decreases over time, with contestants eventually dropping out of the contest as it becomes less likely that they will win the prize. This differs significantly from the structure of the optimal effort level in our paper, which generally first increases and then decreases; this is referred to as the "watershed" structure in our analysis. Furthermore, unlike this stream of literature, which mainly focuses on the design of an optimal policy from the perspective of contest designers—such as the reward structure and information-disclosure policy—we take the contract as exogenously given and study how the seller can maximize the profit by dynamically adjusting her effort levels over the sales horizon.

3. Model

We introduce the model. For concreteness, we discuss the setup in the context of salesforce management: All-or-nothing contracts are prevalent in such a context, as illustrated in This American Life (2013). Consider a seller (say, a salesperson or a car dealer) who

sells a single product over a finite time. We denote T as the length of the sales horizon. Before the start of the sales horizon, the seller enters into an all-or-nothing contract with an upstream supplier, as detailed below.

Definition 1 (All-or-Nothing Contract). Let x be the total sales during the entire sales horizon. The payout to the seller $R(x)$ is given by

$$R(x) = \begin{cases} b + p(x - N) & x \geq N, \\ 0 & x < N, \end{cases} \quad \text{where} \quad b, p \geq 0. \tag{1}$$

In this contract, the sales target N is the goal set for the seller over the sales horizon. If the total sales volume is no less than the predetermined sales target N, the seller will receive a lump-sum payment b and a commission p for each unit sold beyond the target. However, if the total sales volume falls short of the target N, the seller will receive nothing. This is why we refer to it as an "all-or-nothing" contract. Our results still hold for an alternative payment scheme where, even if the seller does not exert any effort, there is a base demand rate and the seller earns a fixed commission for every unit of sales, regardless of whether the target is met, in addition to the all-or-nothing lump-sum bonus and the over-target commission per unit, which can both be zero.

We make some remarks on the all-or-nothing contract. First, many commonly used sales contracts are special cases of the contract given by Definition 1. For example, when $n = 0$ and $b = 0$, the contract is reduced to a regular commission plan, where the seller is paid a constant commission for every unit sold. Sales quotas are reportedly used by 89% of firms, making them the most consistent feature of sales compensation (Oyer 2000). The theoretical rationale for the prevalence of all-or-nothing contracts is that this type of contract is found to be optimal or a piecewise-linear approximation of the optimal contract under a wide range of settings (see, e.g., Basu et al. 1985 and Raju and Srinivasan 1996). Second, the optimal contract design is beyond the scope of our paper. In our discussion, we assume that the contract format and specification, in terms of b, p, and N, are exogenously given. There is a long history of literature on the design of salesforce compensation contracts, mainly based on the principal-agent model. We refer interested readers to Basu et al. (1985), Lal and Srinivasan (1993), and Oyer (2000) for details. Putting the discussion of optimal contracts aside, all-or-nothing contracts are prevalent in practice because of other practicalities. For instance, the bonuses offered by manufacturers are often nonmonetary, such as enhanced tech support, invitation to exclusive events, and access to training programs, which cannot be

discretized. Thus, a linear contract can be difficult to operationalize, should it be theoretically optimal.

Customers arrive and make their purchases according to a Poisson process. We assume that the seller precommits to a fixed price for the product at the beginning of the sales horizon. This price could be the manufacturer's suggested retail price. The intensity of the sales process can be moderated by the seller with costly effort. For instance, it could reflect the amount of effort that the seller exerts to promote the product or the amount of money that the seller spends on advertising.

We denote $c(\lambda) \geq 0$ as the cost rate for the seller if she wants to induce a sales intensity λ. With a cost $c(\lambda_t)$, the sales still have uncertainty, and we denote the moderated sales process as D_t, $\forall\, t \in [0, T]$. The sales process D_t is a nonhomogeneous Poisson process with mean λ_t at time t. Denote by $[\underline{\lambda}, \bar{\lambda}]$ the set of feasible sales rates. We assume the following for the cost function.

Assumption 1 (Cost Function).

  i. $\underline{\lambda} > 0$ *is the cost-free sales rate—that is,* $c(\underline{\lambda}) = 0$;
  ii. $c(\lambda)$ *is invertible, and convex in* λ, $\forall\, \lambda \in [\underline{\lambda}, \bar{\lambda}]$;
  iii. $\lambda^* \in [\underline{\lambda}, \bar{\lambda}]$, *where* $\lambda^* = \arg\max\{\lambda \mid \lambda p - c(\lambda)\}$.

Assumption 1(i) says that there exists a natural cost-free sales rate. Without loss of generality, we assume this rate to be the lower bound of the set of feasible sales rates because there is no incentive for the seller to go for a lower rate. We assume in Assumption 1(ii) the convexity of $c(\cdot)$, which implies that cost expenditure has a diminishing return on sales. This assumption is consistent with common assumptions in the RM and salesforce compensation literature. For instance, Lal and Staelin (1986), Oyer (2000), and Sohoni et al. (2010) assume that the sales response function, which is the inverse of the cost-rate function, is concave. There is also a one-to-one correspondence between cost and sales rate, which allows us to use the sales intensity as our decision variable in the analysis. This follows the convention in the RM literature (see, e.g., Gallego and van Ryzin 1994). Lastly, to avoid the cumbersome discussion of corner solutions, we assume in Assumption 1(iii) that the profit-maximizing sales rate is well defined and within the feasible set. This is without loss of generality and can be achieved by properly defining the feasible set. Intuitively, the optimal sales intensity will neither go below the cost-free rate $\underline{\lambda}$ nor become too large, owing to the convexity of $c(\lambda)$.

The seller is assumed to be risk-neutral. Her objective is to maximize the expected profit over the sales horizon by choosing a nonanticipating policy $u = \{\lambda_t : 0 \leq t \leq T\}$. The value of λ_t at each time t (time-to-go; see more below) is only allowed to depend on the past sales observations. We can formulate the seller's profit-maximization problem as follows: selecting a policy u to maximize

$$\Pi_{opt} := \max_u\ \pi_u = \mathbb{E}_u\left[R\left(\int_0^T \mathrm{d}D_s\right) - \int_0^T c(\lambda_s)\,\mathrm{d}s \right]. \quad \text{(S)}$$

We focus on the model where the seller exerts costly effort to moderate the sales process; however, we emphasize that our model is also applicable when price is the controller. The reason is as follows. In the context of pricing, we can write the inverse demand function as $\mathscr{P}(\lambda) = p_0 - \delta(\lambda)$, where p_0 is the seller's break-even price (e.g., the invoice price from the manufacturer for a dealership), and $\delta(\lambda)$ represents a markup (i.e., $\delta(\lambda)$ is negative) or a markdown (i.e., $\delta(\lambda)$ is positive). Thus, the seller's profit function can be formulated as

$$\Pi_{opt} := \max_u \pi_u = \mathbb{E}_u[R(\int_0^T \mathrm{d}D_s) -$$
$$\int_0^T \delta(\lambda_s)\,\mathrm{d}D_s].$$

For Markov policies, we have $\mathbb{E}_u[\int_0^T \delta(\lambda_s)\,\mathrm{d}D_s] = \mathbb{E}_u[\int_0^T \delta(\lambda_s)\lambda_s\,\mathrm{d}s]$, and, thus, the pricing problem is equivalent to the sales effort problem with "cost" rate $c(\lambda) = \delta(\lambda)\lambda$.[2] Our assumption about the convexity of the discount rate $c(\lambda) = \lambda\delta(\lambda)$ is equivalent to the convention in the RM literature that assumes the revenue rate $\lambda\mathscr{P}(\lambda)$ to be concave.

We also want to highlight here that our objective function is discontinuous around the sales target N because of the all-or-nothing contract. This is fundamentally different from many papers in the RM literature (e.g., Gallego and van Ryzin 1994, Maglaras and Meissner 2006, and Besbes and Maglaras 2012). The discontinuity of the objective function makes our analysis more difficult. It plays a crucial role in shaping the optimal heuristic in this setting, leading to nontrivial distinctions between the optimal strategy and those in classical RM problems.

3.1. Structural Properties of the Optimal Policy

Next, we study how the optimal sales intensity and expected profit evolve over time. We refer to t as the time remaining on the sales horizon (i.e., time-to-go) and $n \geq 0$ as the number of sales required to reach the target. By the principle of optimality, the Hamilton–Jacobi–Bellman (HJB) conditions for the optimal profit can be summarized as follows.

Proposition 1. *Denote by $J_t^*(n)$ the optimal expected profit at state (t, n). It is the solution of*

$$\frac{\partial J_t^*(n)}{\partial t} = \max_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} \left\{ \lambda[J_t^*(n-1) - J_t^*(n)] - c(\lambda) \right\}, \quad (2)$$
$$\forall\ n \geq 1,\ \forall\ t,$$

with boundary conditions $J_0^*(n) = 0$, $\forall n \geq 1$ *and* $J_t^*(0) = b + [\lambda^* p - c(\lambda^*)]t$, $\forall\ t$, *where λ^* is defined in Assumption 1.*

There exists a unique solution to Equation (2), which we denote by $\lambda^*(t, n)$; that is, the profit-maximizing sales rate at the state (t, n), provided that the cost function $c(\lambda)$ satisfies the regularity conditions specified in Assumption 1. We define $\Delta_t(n) := J_t^*(n-1) - J_t^*(n)$. The boundary condition $J_t^*(0)$ is no longer equal to zero, which differs from the norm in the RM literature. This is because of our slightly different definition of n. In our model, n denotes the number of sales required to reach the target, rather than the remaining stock level, as in a typical RM model. Therefore, the marginal revenue for each sale becomes a constant of p once n reaches zero, and the optimal strategy for the seller is to induce the profit-maximizing rate λ^*, which no longer depends on n.

Theorem 1 (Watershed Structure of Optimal Policy). *We have the properties of the optimal policy:*

i. *$J_t^*(n)$ increases in t, but decreases in n;*

ii. *When $b > p$, there exists $0 \leq \tau(n) < \infty$, $\forall n \geq 1$ such that*
   a. *When $t > \tau(n)$, $\Delta_t(n)$ and $\lambda^*(t, n)$ decrease in t; $\Delta_t(m)$ and $\lambda^*(t, m)$ increase in m, for any $m \leq n$.*
   b. *When $t \leq \tau(n)$, $\Delta_t(n)$ and $\lambda^*(t, n)$ increase in t; $\Delta_t(m)$ and $\lambda^*(t, m)$ decrease in m, for any $m \geq n$.*
   Moreover, $\tau(n)$ strictly increases in n, that is, $0 = \tau(1) < \tau(2) < \cdots < \tau(n-1) < \tau(n) < \cdots$;

iii. *When $b \leq p$, $\Delta_t(n)$ and $\lambda^*(t, n)$ increase in t and decrease in n;*

iv. *$\lim_{t \to \infty} \lambda^*(t, n) = \lambda^*$, $\forall n \geq 0$ and $\lim_{t \to 0} \lambda^*(t, n) = \underline{\lambda}$, $\forall n > 0$.*

Theorem 1 summarizes structural properties of $J_t^*(n)$ and $\lambda^*(t, n)$. To give an intuitive understanding of the proposition, we illustrate them with a couple of numerical experiments, as displayed in Figures 1 and 2. The parametric specification of the numerical experiments is given as follows: The initial sales target $N = 20$ and marginal revenue per sale beyond the target $p = 2$. The cost function is assumed to be quadratic and to follow the form $c(\lambda) = (\lambda - 2)^2$. Figure 1 and Figure 2 show the results of numerical experiments where the lump-sum payout $b = 40 > p$ and $b = 1 < p$, respectively.

Theorem 1(i) is intuitive. Everything else being equal, with longer time-to-go or fewer sales required to reach the target, both the expected sales and the probability of reaching the target are higher, leading to a higher optimal expected profit level. The monotonicity of the optimal expected profit $J_t^*(n)$ is also reflected in Figures 1(a) and 2(a). $J_t^*(n)$ is close to zero when t is sufficiently small because there is little chance for the seller to hit the target. As t increases, $J_t^*(n)$ also increases. However, when $b > p$, which arguably always holds in practice, $J_t^*(n)$ is neither convex nor concave in t in general, in contrast to the concavity commonly observed in the classical RM problems. The shape of curves indicates that $J_t^*(n)$ is initially convex in t, but becomes concave after the turning point $\tau(n)$.

Theorem 1 parts (ii)–(iv) characterize the structure of $\tau(n)$ and $\lambda^*(t, n)$. First, Theorem 1(ii) shows that when the lump-sum payment b is greater than the per-unit bonus commission p, the optimal sales rate is nonmonotonic in either the remaining time or the extra number of sales required to reach the target. There exists a cutoff point $\tau(n)$, $\forall n \geq 1$, such that the optimal sales rate $\lambda^*(t, n)$ decreases in t when $t > \tau(n)$; otherwise, $\lambda^*(t, n)$ increases in t. We call this the *watershed* structure. Figure 1 illustrates this watershed structure. When it is very close to the end of the horizon—that is, $t \to 0$—the probability of reaching the target is slim, and, thus, the seller's optimal strategy is to avoid incurring any extra cost by adopting the cost-free rate $\underline{\lambda} = 2$. As the remaining time t increases, the chance of hitting the target improves, and, thus, the seller would respond by exerting a greater amount of effort to induce higher

Figure 1. (Color online) Numerical Results of the Optimal Expected Profit and Optimal Sales Rate when $b > p$

Optimal Expected Profit $J_t^*(n)$ Optimal Sales Rate $\lambda^*(t, n)$
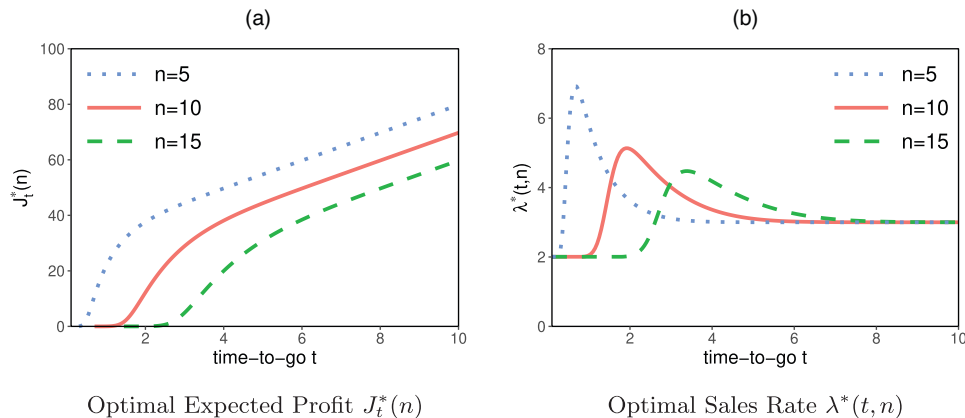
Note. $N = 20$, $b = 40$, $p = 2$, and $c(\lambda) = (\lambda - 2)^2$.

Figure 2. (Color online) Numerical Results of the Optimal Expected Profit and Optimal Sales Rate when $b \le p$

Optimal Expected Profit $J_t^*(n)$ Optimal Sales Rate $\lambda^*(t, n)$
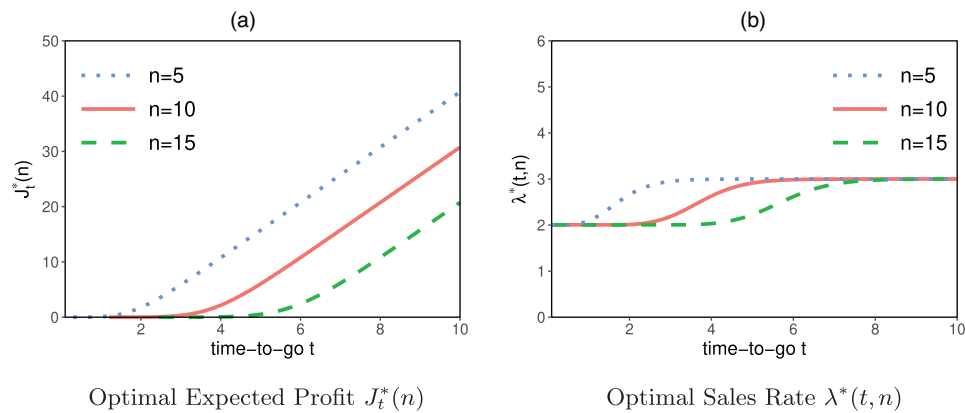
Note. $N = 20$, $b = 1$, $p = 2$, and $c(\lambda) = (\lambda - 2)^2$.

sales. Because of the relatively large lump-sum payout b, it becomes crucial for the seller to hit the sales target, boosting the optimal sales rate higher than the profit-maximizing sales rate λ^*. However, when the remaining time t is sufficiently long to be greater than the turning point $\tau(n)$, the seller can comfortably hit the target without stretching herself, and, thus, she is better off exerting less effort. Hence, the optimal sales intensity decreases in t and converges to λ^* when t is large enough, as shown in Theorem 1(iv).

On the other hand, when the lump-sum payment b is less than or equal to the per-unit bonus commission p, the optimal sales rate monotonically increases in the remaining time and decreases in the extra number of sales required to reach the target, as shown in Theorem 1(iii). Figure 2(b) illustrates the structure of the optimal sales intensity in this case. As in the previous case, it is optimal for the seller to exert little effort by inducing the cost-free rate $\underline{\lambda} = 2$ when the end of the sales horizon is close . As t increases, the optimal sales rate also increases. However, because of the relatively small lump-sum payout b, any extra cost over the profit-maximizing cost rate $c(\lambda^*)$ always outweighs the potential benefit from a greater chance of reaching the target, leading to the convergence of the optimal sales intensity to λ^* when t is sufficiently large.

The watershed structure of the optimal sales rate function differs fundamentally from that of most conventional RM problems. For instance, in the single-product setting, Gallego and van Ryzin (1994) show that the optimal sales rate always decreases in the time-to-go and increases in the remaining inventory. The nonmonotonicity of optimal sales rates under the all-or-nothing constraint is driven by the following two forces, pushing in opposite directions. On the one hand, the all-or-nothing constraint incentivizes the seller to induce a higher sales intensity, with the hope of reaching the target by the end of the sales horizon. On the other hand, the cost is noncontingent,

which means that the seller will incur a loss if the target is not reached in the end. As a result, the seller will refrain from exerting costly effort if the chance of reaching the target is low. These two forces together shape the optimal sales intensity over the course of the sales horizon. For a given required number of sales n, when the remaining time is relatively long—that is, $t > \tau(n)$—the probability of reaching the target at the end of the sales horizon is relatively high. In this case, the benefit of increasing the probability of reaching the target by inducing a higher sales rate, if the remaining time is shorter, outweighs the extra cost, and, thus, the force of reaching the target dominates. But for a given required number of sales n, when $t \le \tau(n)$, the probability of reaching the target is low, and, thus, the cost concern dominates. So, the seller is better off lowering the sales intensity if the remaining time is shorter. Theorem 1 also shows that the cutoff time $\tau(n)$ increases in n—that is, the larger the gap from the target, the earlier the cost concern will become the dominant force.

As time goes by, the random sales process under the optimal policy is driven by the forces mentioned above. As the end of the horizon comes closer (t becomes smaller), the required number of sales to reach the target also becomes smaller (n becomes smaller). Along a sample path, it is likely that the seller ends up not putting in much effort toward the end of the sales horizon if the hope of reaching the target is remote or the seller ends up putting in a lot of effort toward the end, as in the story of This American Life (2013) and as observed in the "hockey-stick" phenomenon (see, e.g., Chen 2000). These possible realizations of a sample path do not contradict the watershed structure, because this structure is a static property for a given required number of sales. Indeed, the nonmonotonicity of the optimal sales rate of $\lambda^*(t, n)$ is consistent with many empirical findings. For instance, Steenburgh (2008) finds that, under a lump-

sum bonus contract, salespeople, in general, respond rationally and work harder, with the hope of reaching the quota. But at the same time, they may also give up if they feel that the quota is simply unreachable.

4. Upper Bound

In general, it is hard to compute the optimal policy for the problem for the following reasons. Theoretically, there is no general closed-form solution for $\lambda^*(t,n)$. Thus, finding the optimal sales rate $\lambda^*(t,n)$ can be computationally intensive, especially for problems on a large scale. Moreover, from the practical side, even if we know the optimal sales rate $\lambda^*(t,n)$, it will be difficult for the seller to constantly adjust the intensity throughout the sales horizon, as $\lambda^*(t,n)$ depends on both the remaining time and the gap between sales and the target. Our solution approach is as follows. First, we come up with a performance bound for Problem (S). Then, we propose various easy-to-compute and implementable heuristics and show their (sub)optimality by comparing the corresponding profits with the performance bound.

Motivated by traditional RM problems, where the optimal profit from the deterministic problem is proven to be an upper bound for that of the stochastic problem, we start with an analysis of the deterministic version of Problem (S). However, as we will show below, perhaps surprisingly, the optimal deterministic profit is no longer a universal upper bound for Problem (S).

4.1. Deterministic Problem

We can formulate the deterministic version of Problem (S) as follows. With the absence of demand uncertainty, there is a one-to-one correspondence between the sales and cost. That is, the sales rate is a function of the cost at time t, which is deterministic. Similar to Problem (S), the decision variable in the deterministic problem is still the sales rate, which, with a bit of abuse of notation, is still denoted by λ_t. Thus, the problem can be formulated as

$$\Pi_D := \max_{\lambda=\{\lambda_t : 0 \le t \le T\}} \pi_D(\lambda) = b + p\left(\int_0^T \lambda_t \, dt - N\right) - \int_0^T c(\lambda_t) \, dt$$

$$\text{s.t.} \int_0^T \lambda_t \, dt \ge N. \qquad (D)$$

With the formulation of Problem (D), we implicitly assume that there exist nontrivial sales rates λ_t, such that the target can be reached at the end of the sales horizon with $\Pi_D \ge 0$. This can be the case when the lump-sum bonus b is sufficiently large or the cost $c(\lambda)$ is not prohibitively high. The problem becomes trivial when this assumption is violated, as the seller would

simply adopt the cost-free sales rate $\underline{\lambda}$ throughout the entire sales horizon and earn zero profit as a result.

Next, we solve for the optimal solution to Problem (D).

Proposition 2 (Optimal Deterministic Rate). *The optimal solution to the deterministic problem is* $\lambda_t = \lambda_D := \max\{N/T, \lambda^*\}, \forall t \in [0,T]$, *where* λ^* *is defined in Assumption 1.*

We make some observations. First, recall that λ^* is the profit-maximizing rate once the target is reached. When $\lambda^* \ge N/T$, inducing a rate of λ^* would allow the seller to meet the target at the end of the sales horizon and, at the same time, maximize profit once the target is reached. Thus, it is optimal for the seller to induce a rate of λ^* throughout the sales horizon. When $\lambda^* < N/T$, the seller needs to induce a rate of at least N/T to reach the target. However, on the other hand, any sales beyond the quota would incur a loss for the seller because the marginal commission p is not high enough to cover the marginal cost (i.e., $c'(N/T) > p$). Hence, the optimal sales rate in this case would be N/T. Note that, under the special case when $p = 0$, Proposition 2 implies that $\lambda_D = N/T$, because $\lambda^* = \underline{\lambda}$. That is, when there is no over-quota commission, it is optimal for the seller to induce sales intensity such that the total sales are equal to the sales target at the end of the sales horizon, which is viable for the deterministic problem.

Second, the result that the optimal deterministic sales rate is static may not be as intuitive as it appears. Recall that we do not require that $b = pN$, so the marginal profit of each sale may differ before and after the target is reached, and, thus, the seller can potentially opt for different sales rates depending on whether the target is reached. However, Proposition 2 shows that it is optimal to choose one single rate throughout the season, owing to the convexity of the cost function.

4.2. A Counterexample

Next, we show with an example that the optimal deterministic revenue no longer serves as a universal upper bound for the expected profit of the stochastic problem when there is an all-or-nothing constraint. That is, it is possible $\Pi_{opt} > \Pi_D$. This is in stark contrast to traditional RM problems (see, e.g., Gallego and van Ryzin 1994, theorem 2) and most operations management problems in which uncertainty cuts into profitability (e.g., the newsvendor problem).

Example 1. Consider an example with the following parameters: $T = 1$, $N = 1$, $p = 5$, and $b = 5$. The cost function is assumed to be $c(\lambda) = (3\lambda - 1)^2$. First, consider the deterministic problem. It is easy to verify that $\lambda^* = 11/18$ and $N/T = 1$. Thus, we have

$\lambda_D = N/T = 1$, and the optimal profit for the deterministic problem is $\Pi_D = 1$.

Next, consider the stochastic problem. Suppose the seller chooses the cost-free rate $\lambda_t = \underline{\lambda} = 1/3$ for the entire sales horizon as a heuristic. Then, the number of sales during the sales horizon follows a Poisson distribution with the rate of 1/3, and the seller's expected profit is given by

$$\mathbb{E}_u\left[R\left(\int_0^T dD_s\right)\right] = \mathbb{E}_u\left[5 + 5\left(\int_0^T dD_s - 1\right)\right]$$

$$\int_0^T dD_s \geq 1\right]\mathbb{P}_u\left[\int_0^T dD_s \geq 1\right]$$

$$= 5\mathbb{E}_u\left[\int_0^T dD_s \middle| \int_0^T dD_s \geq 1\right]\mathbb{P}_u\left[\int_0^T dD_s \geq 1\right]$$

$$= 5\mathbb{E}_u\left[\int_0^T dD_s\right] = \frac{5}{3}.$$

As the above expected profit is achieved by a heuristic policy, we can thus conclude that $\Pi_{opt} \geq 5/3 > \Pi_D = 1$.

The reason why the optimal deterministic profit is no longer guaranteed to be an upper bound for the profit of the stochastic problem is as follows. In traditional RM problems, demand uncertainty always works against the seller: When the seller sets the price for the current sale, she faces future demand uncertainty; if the future demand realization is relatively high (or low), the seller could have set a high (or low) price for the current sale. However, with an all-or-nothing constraint, demand uncertainty has a two-sided impact on the seller's profit and may actually help her. In particular, when she does not expect to reach the sales target in the absence of random demand shocks without exerting any cost—that is, $\mathbb{E}_u[\int_0^T dD_s(\underline{\lambda})] < N$—the seller could refrain from exerting costly effort and rely on pure luck.

A direct implication of Example 1 is that the seller may "game" the system under the all-or-nothing constraint. The purpose of imposing a sales target of N is to motivate the seller to increase sales by exerting greater effort. However, because of the randomness in demand, the seller can be opportunistic by exerting less effort and hoping that the randomness works in her favor to push the sales above the target. As a result, it is crucial to design the contract in a way that appropriately aligns the incentives of multiple parties.

4.3. Asymptotic Analysis
The fact that the optimal profit from the stochastic problem Π_{opt} may be higher than the deterministic profit Π_D makes our analysis more difficult. Because it is generally hard to derive Π_{opt}, we would need a benchmark to gauge the performance of various policies. Fortunately, we are able to re-establish the

inequality $\Pi_{opt} \leq \Pi_D$ in the asymptotic regime when N, T, and b are large enough.

Consider a series of problems, where, for each instance, we scale up the time horizon, sales target, and lump-sum bonus by a factor of θ. That is, for the θ-th problem, we have $T^{(\theta)} = \theta T$, $N^{(\theta)} = \theta N$, and $b^{(\theta)} = \theta b$, and we evaluate the performance of various policies as $\theta \to \infty$. We shall note that similar approaches have been applied in many other applications in RM (see, e.g., Besbes and Maglaras 2012 and Jasin 2014). However, in the context of sales effort management, this construction has a natural interpretation. The original problem can be viewed as evaluating the sales number and rewarding the seller every T time periods, whereas in the θ-th scaled problem, the sales number is evaluated only once at the end of the θT-th time period. So, if we divide the expected profit for the θ-th problem by θ and compare it with that of the original problem, it will shed light on the impact of the length of evaluation windows on the policy's performance.

For the θ-th problem, we denote by $\Pi_{opt}^{(\theta)}$ and $\Pi_D^{(\theta)}$ the optimal profit of the stochastic problem and the deterministic problem, respectively. It is easy to see that the optimal solution of a deterministic problem for any θ remains the same, and $\Pi_D^{(\theta)} = \theta\Pi_D = \theta[b + p(\lambda_D T - N) - c(\lambda_D)T]$. The optimal expected profit of the stochastic problem $\Pi_{opt}^{(\theta)}$ must be computed based on Problem (S), which is a difficult task in general. Fortunately, we show in the following proposition that $\Pi_D^{(\theta)}$ still serves as an upper bound of $\Pi_{opt}^{(\theta)}$ when θ is sufficiently large.

Proposition 3 (Asymptotic Upper Bound). *For any policy* $u = \{\lambda_t : 0 \leq t \leq T\}$, *we have* $\lim_{\theta\to\infty}\Pi_u^{(\theta)}/\Pi_D^{(\theta)} \leq 1$, *where the inequality is strict if* $\lim_{\theta\to\infty}\mathbb{E}_u[\int_0^{\theta T}\lambda_s\,ds]/\theta < N$.

Proposition 3 has two key takeaways. First, when θ is sufficiently large, the optimal deterministic profit always serves as an upper bound. This result allows us to gauge the asymptotic performance of a policy by comparing its expected profit against the optimal deterministic profit. Second, when θ is sufficiently large, any policy with expected sales lower than the target never achieves the upper bound. This is because, if $\lim_{\theta\to\infty}\mathbb{E}_u[\int_0^{\theta T}\lambda_s\,ds]/\theta < N$, the probability that the target will be reached is strictly less than one.

To further explore the implications of Proposition 3, we conduct a series of simulations by scaling up the problem in Example 1 by a factor of θ. In the θ-th scaled problem, the parameters are given by $T = \theta$, $N = \theta$, $p = 5$, and $b = 5\theta$. The cost function is assumed to be $c(\lambda) = (3\lambda - 1)^2$. As a benchmark, the optimal sales intensity in the deterministic problem is $\lambda_D^{(\theta)} = N/T = 1$, with a corresponding profit $\Pi_D^{(\theta)} = \theta$ in the θ-th problem.

Figure 3(a) shows the expected optimal profit of the stochastic problem $\Pi_{opt}^{(\theta)}$ relative to the deterministic profit $\Pi_D^{(\theta)}$ as θ varies. It corroborates our findings in Example 1 and Proposition 3 that the optimal deterministic profit serves as an upper bound for the expected optimal profit of the stochastic problem only when θ is sufficiently large. The curve in Figure 3(b) displays the average optimal sales intensity in the θ-th stochastic problem. The horizontal dashed line in the figure indicates the optimal sales intensity in the deterministic problem, which is always equal to one. The optimal sales intensity in a stochastic problem is calculated by using Equation (2) and is sample-path dependent. To compare it with λ_D, we repeat the simulation 1,000 times and record the average sales intensity during the sales horizon for each θ. Figure 3(b) shows that the average optimal sales intensity in the stochastic problem is lower than λ_D when θ is small. However, with a sufficiently large θ, the average optimal sales intensity eventually converges to λ_D from *above*, which makes sense because the seller wants to ensure the target is met under demand uncertainty.

As mentioned, in the context of sales effort management, θ can be interpreted as the length of an evaluation time window. In this context, Proposition 3 and Figure 3(b) illustrate one potential upside of having a long evaluation time window. That is, with a sufficiently long evaluation window, the rational seller must induce a sufficiently high *average* sales rate to reach the target, following a modified resolving heuristic or, even better, the optimal policy, which alleviates the moral hazard issue in salesforce management. This result contrasts and complements some empirical and analytical studies that show that a sufficiently short evaluation window may be preferred from other perspectives such as psychological factors (Chung et al. 2013) or order smoothing and coordination within a supply chain (Chen 2000).

5. Heuristics

Next, we propose various easy-to-implement heuristics and explore their performance by comparing their profit with the benchmark based on the deterministic problem in the asymptotic regime.
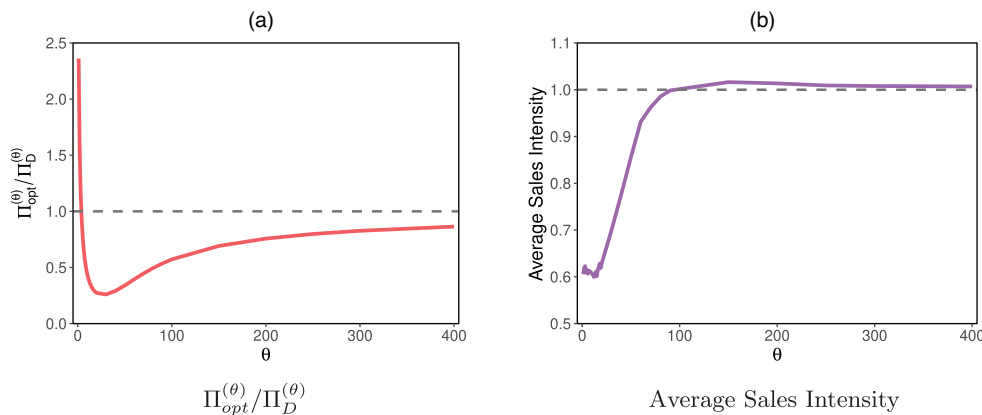
5.1. Static Heuristics

First, we consider a heuristic where the seller induces a static sales rate, denoted as $\lambda_{SH}^{(\theta)}$ for the θ-th problem, throughout the sales horizon. We adopt the solution of the deterministic problem λ_D, which is derived in Proposition 2, as the rate for the static heuristic. This heuristic is easy to implement, as it requires no monitoring of the sales process by the seller during the sales horizon. Moreover, it has a provable performance guarantee in traditional RM problems. Gallego and van Ryzin (1994) show that the static heuristic with a rate of λ_D is asymptotically optimal, with a performance gap smaller than $O(\sqrt{\theta})$. Jasin (2014) shows that this bound is tight.

Next, we investigate the performance of this static heuristic in the stochastic setting when there exists an all-or-nothing constraint. We denote the expected profit of the static heuristic in the θ-th problem by $\Pi_{SH}^{(\theta)}$. We evaluate its performance by comparing $\Pi_{SH}^{(\theta)}$ with $\Pi_D^{(\theta)}$, which serves as an upper bound of the optimal expected profit when θ is sufficiently large, as shown in Proposition 3.

We consider two scenarios, depending on how λ^* fares against N/T. As we show below, the performance of the static heuristic is fundamentally different in each of the two. First, consider the case when $\lambda^* > N/T$. In this case, we have $\lambda_{SH}^{(\theta)} = \lambda^*$ for the static heuristic, as the profit-maximizing sales intensity λ^* can indeed reach the sales target N in the deterministic problem. As we show in the following proposition, this static heuristic is asymptotically optimal and is able to asymptotically eliminate any performance gap in the stochastic setting.

Figure 3. (Color online) Expected Optimal Profit and Average Sales Intensity with Different Scales

Note. $p = 5, b = 5\theta, N = \theta, T = \theta,$ and $c(\lambda) = (3\lambda - 1)^2$.

Proposition 4. *Suppose $\lambda^* > N/T$. We have $\lim_{\theta \to \infty} \left(\Pi_D^{(\theta)} - \Pi_{SH}^{(\theta)} \right) = 0$.*

Proposition 4 delivers the somewhat intriguing message that the static heuristic asymptotically yields no performance loss at all, compared with the optimal policy. That is, a simple static rate policy reduce the performance gap to zero, even though the randomness in demand increases in an absolute term as θ increases. This is in contrast to the performance of the optimal static heuristic in many traditional RM problems, where its performance loss is given by $O(\sqrt{\theta})$ (see, e.g., Gallego and van Ryzin 1994 and Besbes and Maglaras 2012).

The underlying rationale for this seemingly counterintuitive result is as follows. In many traditional RM problems, revenue loss is driven mainly by the mismatch between supply and demand. In particular, when the realized demand is greater than the inventory level, the seller leaves money on the table and could have charged a higher (static) price. On the other hand, when the realized demand is lower than the supply, the seller could have made a greater profit by charging a lower (static) price. The magnitude of the performance loss from adopting a static pricing policy scales at a rate of $O(\sqrt{\theta})$ (see, e.g., Besbes and Maglaras 2012 and Jasin 2014). For a service system, Kim and Randhawa (2017) show that the magnitude of the performance loss of a contingent pricing policy can be reduced to a rate of $O(\theta^{1/3})$, but there will still be a positive gap.

On the contrary, as the price is fixed and the seller controls sales intensity, the main source of performance loss in our model is the potential dire event of the seller gaining nothing if she fails to reach the sales target at the end of the sales horizon. When $\lambda^* > N/T$, this risk asymptotically disappears because the target will almost certainly be reached with a static heuristic of rate λ^* when the scale of the problem is sufficiently large. Along the way, the seller also induces the profit-maximizing sales rate, which eliminates any potential performance loss from excessive effort/cost incurred to reach the target. The preceding two factors guarantee little loss in performance by inducing a rate of λ^* throughout the sales horizon, when the profit-maximizing rate λ^* is higher than N/T.

Next, we analyze the case when $\lambda^* \leq N/T$. In this case, the solution to the deterministic problem is $\lambda_D = N/T$. That is, when there is no uncertainty, the best strategy for the seller is to induce a sales rate such that the target will just be reached at the end of the sales horizon. Given the positive result of Proposition 4 on the static heuristic derived from the deterministic problem, it may be tempting to also use λ_D in the stochastic problem as the rate for the static

heuristic. However, we show in the following proposition that the static heuristic with a rate of λ_D is suboptimal.

Proposition 5. *Suppose $\lambda^* \leq N/T$. Consider any static heuristic with a rate of $\lambda_{SH}^{(\theta)} = \lambda_D + f(\theta)$, which satisfies $\lim_{\theta \to \infty} f(\theta) = 0$.*

i. *The heuristic is asymptotically optimal—that is, $\lim_{\theta \to \infty} \Pi_{SH}^{(\theta)} / \Pi_D^{(\theta)} = 1$—if and only if $\lim_{\theta \to \infty} \sqrt{\theta} f(\theta) = \infty$;*

ii. *When $\lambda^* = N/T$, there exists a static heuristic whose performance loss is bounded by $\Pi_D^{(\theta)} - \Pi_{SH}^{(\theta)} = O(\theta^\epsilon)$, $\forall \epsilon > 0$;*

iii. *When $\lambda^* < N/T$,*

   a. *For any asymptotically optimal static heuristic, we have $\lim_{\theta \to \infty} \frac{1}{\sqrt{\theta}} \left(\Pi_D^{(\theta)} - \Pi_{SH}^{(\theta)} \right) = \infty$;*

   b. *There exists a static heuristic whose performance loss is bounded by $\Pi_D^{(\theta)} - \Pi_{SH}^{(\theta)} = O(\theta^{0.5+\epsilon})$, $\forall \epsilon > 0$.*

Unlike the case when $\lambda^* > N/T$, Proposition 5(i) shows that a static heuristic with a rate of λ_D is no longer asymptotically optimal when $\lambda^* \leq N/T$. As a matter of fact, we show a much stronger result: that a static heuristic can be asymptotically optimal if and only if a positive margin $f(\theta)$ is added to the optimal deterministic rate λ_D. This is because, when there exists an all-or-nothing constraint, the seller receives the payout only when the number of sales reaches the target. Thus, almost certainly reaching the target (i.e., with probability one) is essential for any heuristic to be asymptotically optimal. The optimal deterministic rate of λ_D does not account for demand uncertainty, and, therefore, the seller can only guarantee the target being reached by inducing a rate higher than λ_D. The margin $f(\theta)$ will shrink to zero as θ increases; otherwise, there is a persistent boost beyond the minimum, leading to the heuristic being suboptimal. Having said that, the margin $f(\theta)$ cannot converge to zero too quickly either—that is, with a rate slower than $1/\sqrt{\theta}$—as otherwise, reaching the target cannot be guaranteed.

Under the special case when $\lambda^* = N/T$, the performance loss of a static heuristic can be as low as $O(\theta^\epsilon)$, $\forall \epsilon > 0$, as shown in Proposition 5(ii). More interestingly, Proposition 5(iii) indicates that the performance loss of any asymptotically optimal static heuristic is greater than $\Theta(\sqrt{\theta})$ when $\lambda^* < N/T$.[3] Compared with many RM problems (see, e.g., Gallego and van Ryzin 1994 and Jasin 2014), performance loss of the optimal static heuristic is greater when there is an all-or-nothing constraint. This is because, in order to guarantee the target almost certainly being reached, the seller needs to induce a boost $f(\theta)$ on top of the optimal deterministic sales rate λ_D. Because the boost cannot shrink too fast—that is, $\lim_{\theta \to \infty} \sqrt{\theta} f(\theta) = \infty$—the performance loss, driven by this higher-than-optimal sales intensity, is greater than $\Theta(\sqrt{\theta})$. Having said that, the performance loss of a static heuristic can

be made arbitrarily close to $\Theta(\sqrt{\theta})$, as shown in Proposition 5(iii).

5.2. Dynamic Heuristics
Propositions 4 and 5 show that the performance of the optimal static heuristic is mixed. When $\lambda^* > N/T$, the optimal static heuristic is asymptotically optimal, with little loss in profit. The flip side of the story is not as encouraging. When $\lambda^* < N/T$, some safety quantity on top of the optimal deterministic sales rate is required to ensure that the target will almost certainly be reached at the end of the sales horizon. This boost in the sales rate results in a performance loss greater than $\Theta(\sqrt{\theta})$. To address the inferior performance of static heuristics when $\lambda^* < N/T$, we expand our search space by examining dynamic heuristics in this subsection.

5.2.1. Periodic Resolving Heuristic. The performance loss of static heuristics, when $\lambda^* < N/T$, is mainly driven by the fact that the sales intensity is set at the beginning of the horizon, and, therefore, does not respond to random demand over the course of the sales horizon. Thus, the seller must compensate by inducing a higher-than-optimal sales rate to make sure the total sales will exceed the target. A natural solution to this problem is to resolve the problem periodically during the sales horizon. At various times, we resolve the deterministic problem with the updated number of sales required to reach the target and updated remaining time until the end of the sales horizon. This process gives the seller chances to respond to past demand and to adjust the sales intensity accordingly. We refer to this heuristic as the *periodic resolving heuristic*.

Suppose the heuristic reoptimizes at distinct time points $\mathcal{T} = \{1, 2, \dots, T\}$. We denote $\hat{\lambda}_t, t \in \mathcal{T}$, as the updated sales intensity at time point t. Then, the heuristic can be formally stated below in Algorithm 1.

Algorithm 1 (Periodic Resolving Heuristic)
1. At time-to-go T, set $\hat{\lambda}_T = \lambda_D$.
2. At time-to-go $t > 1$, compute the updated distance to the target, $\hat{n}_t$.
3. **if** $\hat{n}_t \leq 0$ **then**
4.   Set $\hat{\lambda}_t = \lambda^*$.
5. **else**
6.   Compute $\hat{\lambda}_t$ as the solution to Problem (D) with $N = \hat{n}_t$ and $T = t$.
7. **end if**

With the periodic resolving heuristic, the seller will switch to the profit-maximizing sales rate of λ^* once the sales target is reached. Otherwise, at each time point t, the seller would resolve the deterministic problem, as specified in Problem (D), by taking into account the past demand realizations and the time

remaining until the end of the sales horizon. In theory, this periodic resolving heuristic will strike a delicate balance between complexity and performance. Compared with the static heuristic, the resolving heuristic is responsive to past demand realizations; yet, unlike the optimal solution, which requires HJB equations to be solved through backward induction, it only solves a simple deterministic problem for a finite number of times. We show in the following proposition that the periodic resolving heuristic, as outlined in Algorithm 1, is asymptotically suboptimal when there exists an all-or-nothing constraint, though it has been shown to be asymptotically optimal in many RM problems in the absence of such a constraint.

Proposition 6 (Suboptimality of Periodic Resolving Heuristic). *Denote by $\Pi_{RH}^{(\theta)}$ the expected profit of the periodic resolving heuristic. We have $\lim_{\theta \to \infty} \Pi_{RH}^{(\theta)} / \Pi_D^{(\theta)} < 1$.*

Similar resolving heuristics have been studied in many traditional RM problems, with great success. In particular, Jasin (2014) shows that a resolving heuristic is asymptotically optimal and has a logarithmic revenue loss, which significantly reduces the square-root revenue loss from the optimal static heuristic. However, the superior performance of periodic resolving heuristics no longer holds in the presence of an all-or-nothing constraint. The underlying rationale is as follows. Recall that when $\lambda^* < N/T$, the static heuristic with rate λ_D is not optimal because the corresponding probability that the total sales volume will exceed the target at the end of the sales horizon does not converge to one as $\theta \to \infty$. The same situation persists, even if we resolve deterministic problems periodically. The periodic resolving heuristic allows the seller to adjust sales intensities in response to past demand realizations, which potentially reduces excess costs, especially when the number of sales is higher than expected. However, the resulting sales intensity from an updated deterministic problem still fails to account for demand uncertainty in the future. Therefore, the sales target will almost certainly not be reached with optimal deterministic rates only, even though they are being updated periodically based on past demand realizations. This result underscores the distinctiveness of the all-or-nothing constraint as a hard constraint in a stochastic problem.

5.2.2. Modified Resolving Heuristic. Given an all-or-nothing constraint, a superior dynamic policy should not only be responsive to past demand realizations, but also guard against future demand uncertainties. With this observation, we propose a modified resolving heuristic (MRH), which is summarized in Algorithm 2.

Algorithm 2 (Modified Resolving Heuristic)

1. At time-to-go T, set $\lambda_T^{MRH} = \lambda_D$.
2. At time-to-go $t > 1$, find the updated threshold $\hat{n}_t$.
3. **if** $\hat{n}_t \leq 0$ **then**
4.   Set $\lambda_t^{MRH} = \lambda^*$.
5. **else**
6.   Compute $\hat{\lambda}_t$ as the solution to Problem (D) with $N = \hat{n}_t$ and $T = t$.
7.   **if** $t < M\log\theta$ or $|\hat{\lambda}_t - \lambda_D| > \min\left\{\frac{1}{2}(\bar{\lambda} - \lambda_D),\ \lambda_D - \lambda^*\right\}$ **then**
8.   Set $\lambda_t^{MRH} = \bar{\lambda}$.
9.   **else**
10.   Set $\lambda_t^{MRH} = \hat{\lambda}_t$.
11. **end if**
12. **end if**

In Algorithm 2, parameters M and $\bar{\lambda}$ are some chosen constants, independent of θ; see the next section for an exploration of choosing these parameters. The modified resolving heuristic works as follows. Initially, the heuristic is the same as the standard periodic resolving heuristic described in Algorithm 1, where Problem (D) is updated and resolved periodically, and its solution is deployed as the sales rate. This allows the heuristic to respond to past demand realizations and avoid unnecessary excess costs, while still keeping the sales on track to reach the target. However, the heuristic will switch to a sufficiently high sales rate, referred to as the full-speed rate, when either of the following conditions is satisfied: (i) The remaining time is limited (i.e., $t < M\log\theta$), but the sales target has not yet been reached; or (ii) the solution to Problem (D) with parameters updated at time t deviates significantly from the ex ante optimal deterministic rate (i.e., $\lambda_D = N/T$), which can happen only when the sales are substantially lower than those of the deterministic case at time t. When either of these two conditions is satisfied, ensuring the sales target being met becomes the highest priority, and, thus, the seller is better off switching to the full-speed mode. The performance of the modified resolving heuristic is summarized in the following theorem.

Theorem 2 (Logarithm Gap of Modified Resolving Heuristic). *The modified resolving heuristic is asymptotically optimal, and the performance loss is bounded by* $\Pi_D^{(\theta)} - \Pi_{MRH}^{(\theta)} = O(\log\theta)$.

Theorem 2 shows that the performance loss from the modified resolving heuristic is improved to $O(\log\theta)$, which is much lower than that of the optimal static heuristic, an order of greater than $\Theta(\sqrt{\theta})$, as shown in Proposition 5. The superior performance of the modified resolving heuristic is driven by its reduced sales boost, which is still enough to guarantee that the target will almost certainly be reached at the

end of the sales horizon. Instead of adding a constant margin to the optimal deterministic sales rate at the beginning of the sales horizon, as in the optimal static heuristic, the seller only needs to proactively induce a higher-than-optimal rate either when the remaining time is relatively short or when the number of sales is much lower than expected. By doing this, extra costs can be avoided when the number of sales turns out to be higher than expected. In situations when higher sales intensity is indeed required, a costly effort is only incurred for a limited period of time, leading to a much smaller expected performance loss.

The modified resolving heuristic under the all-or-nothing constraint is able to achieve the same performance gap as optimal resolving heuristics in a traditional dynamic pricing setting (Jasin 2014). However, the underlying rationale is different. In the classical RM problem considered by Jasin (2014), resolving the deterministic problem periodically allows the seller to adjust prices in response to *past* demand realization, which reduces the potential mismatch between supply and demand, leading to better performance than that of the optimal static heuristic. However, as we show in Proposition 6, simply adjusting sales rates in response to past sales would be suboptimal when there exists an all-or-nothing constraint. A superior heuristic will also account for *future* demand uncertainties by ensuring that the sales target will almost certainly be reached. With Algorithm 2, we typically only need to adjust the sales rate when the remaining time is less than $O(\log\theta)$ if at all necessary (we show in the proof that the expected time such that $|\hat{\lambda}_t - \lambda_D| > \min\left\{\frac{1}{2}(\bar{\lambda} - \lambda_D), \lambda_D - \lambda^*\right\}$ is less than $O(\log\theta)$, leading to an improved performance gap of $O(\log\theta)$.

Theorem 2 also provides justification and performance assurance for the hockey-stick phenomenon in salesforce management (see, e.g., Chen 2000 and Sohoni et al. 2010). This phenomenon refers to the sales spike near the end of a sales horizon that is commonly observed in a wide range of industries. One possible rationale for this phenomenon is that salespeople refrain from exerting a significant amount of effort in the early stages and let demand unfold naturally. A considerable amount of effort is justified only when the deadline comes closer without the sales target having been reached. This type of mentality closely resembles the modified resolving heuristic in spirit. So, even though the hockey-stick phenomenon may cause difficulties for third parties, such as upstream partners in a supply chain, by generating uneven orders over time, Theorem 2 shows that it may be a strategy with verifiably good performance for the seller herself.

6. Numerical Examples: Performance of Heuristics

In this section, we conduct comprehensive numerical analysis to supplement our theoretical results in Section 5. We evaluate the performance of heuristics by considering a series of problems with different scales. Unless otherwise stated, the parameters are as follows: $N^{(\theta)} = 20\theta$, $T^{(\theta)} = 5\theta$, and $b^{(\theta)} = 40\theta$, in the θ-th problem. The cost function is assumed to be $c(\lambda) = (\lambda - 2)^2$. The same set of θ values are used in the numerical examples. The particular values of θ adopted can be found in the first column of Table 1. The only exception is Figure 3, where, because we need to compute the optimal policy, we only conduct the experiments for $\theta \le 400$ because of computational complexity.

First, we consider static heuristics. As shown in Propositions 4 and 5, the performance of static heuristics depends crucially on how λ^* fares against $N^{(\theta)}/T^{(\theta)}$. Thus, we separate the two cases when $\lambda^* >$

Table 1. Expected Profit of Static Heuristics and Dynamic Heuristics

	Static Heuristics				Dynamic Heuristics			
	SH[a]		MSH[b]		RH[c]		MRH[d]	
θ	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2	6	45	11	34	4	52	7	35
4	10	87	36	57	11	95	30	48
6	12	129	67	75	12	137	57	58
8	15	171	90	105	20	181	88	67
10	15	212	119	126	17	221	113	88
12	7	252	161	128	36	261	152	93
14	14	294	178	171	18	304	184	107
16	15	335	219	177	15	344	217	122
18	43	373	255	184	27	386	259	108
20	39	416	286	206	28	430	291	123
30	−7	620	436	314	52	623	472	177
40	35	823	644	325	151	818	654	228
50	51	1,025	788	464	96	1,026	857	227
60	38	1,230	946	569	231	1,215	1,056	259
70	−11	1,433	1,129	632	123	1,437	1,246	282
80	−37	1,632	1,335	636	121	1636	1,435	354
90	−103	1,829	1,455	830	162	1833	1,604	440
100	58	2,039	1,699	749	322	2,015	1,816	404
200	144	4,048	3,543	1,319	527	4,025	3,793	626
300	0	6,063	5,237	2,299	770	6,004	5,780	725
400	70	8,075	7,363	2,062	793	8,003	7,749	1,094
500	−181	10,083	9,324	2,232	1,759	9,869	9,782	746
600	−709	12,069	11,151	2,983	902	12,013	11,728	1,159
700	63	14,099	13,068	3,358	998	14,044	13,709	1,639
800	−181	16,114	14,975	3,835	2,577	15,846	15,712	1,537
900	−329	18,110	17,184	2,841	2,629	17,887	17,725	1,704
1,000	70	20,120	18,697	5,108	3,105	19,845	19,675	2,270

Note. SD, standard deviation.
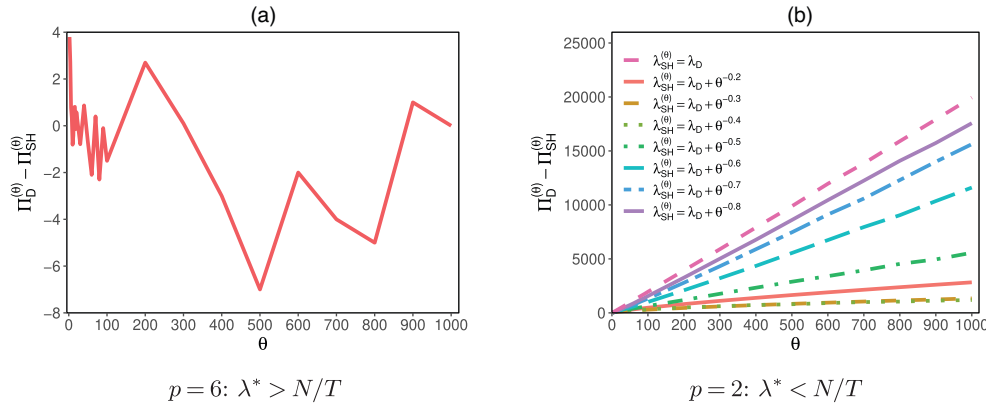[a]Static heuristic with a sales intensity $\lambda_{SH}^{(\theta)} = \lambda_D$.
[b]Modified static heuristic with a sales intensity $\lambda_{MSH}^{(\theta)} = \lambda_D + \theta^{-0.4}$.
[c]Periodic resolving heuristic, as outlined in Algorithm 1.
[d]Modified resolving heuristic, as outlined in Algorithm 2.

$N^{(\theta)}/T^{(\theta)}$ and when $\lambda^* < N^{(\theta)}/T^{(\theta)}$. We compute the expected profit from static heuristics as follows. For the θ-th problem, we repeat simulation experiments for 50,000 times and record the average of the seller's profits as an approximation for $\Pi_{SH}^{(\theta)}$. We then compare it with the optimal deterministic profit $\Pi_D^{(\theta)}$, which serves as an upper bound for the optimal profit of the stochastic problem when θ is large enough, as we show in Proposition 3.

Consider first when the per-unit bonus commission p is six. In this case, we have $\lambda^* = \arg\max\{\lambda: \lambda p - c(\lambda)\} = \arg\max\{\lambda: 6\lambda - (\lambda - 2)^2\} = 5 > N^{(\theta)}/T^{(\theta)} = 4$. Thus, the sales intensity of the static heuristic is set to be $\lambda_{SH}^{(\theta)} = \lambda^* = 5$. We compute the optimal deterministic profit and the expected profit from the static heuristic under various θ and display the differences between the two in Figure 4(a). Some comments are in order. First, we observe that, as θ increases, the profit loss from the static heuristic quickly enters the neighborhood of zero. This is consistent with our statement in Proposition 4 that there is hardly any performance loss for the optimal static heuristic when $\lambda^* > N/T$. Second, the expected profit of the static heuristic is higher than that of the optimal deterministic profit for some values of θ. This corroborates our theoretical result that the optimal deterministic profit is not a universal upper bound for the expected profit from the stochastic problem. The upper-bound statement holds only when θ is sufficiently large.

Next, we consider the case when $p = 2$. In this case, we have $\lambda^* = \arg\max\{\lambda: \lambda p - c(\lambda)\} = \arg\max\{\lambda: 2\lambda - (\lambda - 2)^2\} = 3 < N^{(\theta)}/T^{(\theta)} = 4$. Proposition 5 indicates that static heuristics would perform badly in this case, which is supported by the numerical results, as displayed in Figure 4(b). In particular, we consider eight different sales intensities $\lambda_{SH}^{(\theta)} = \lambda_D + f(\theta)$, where $f(\theta) = \theta^{-\beta}$, and $\beta \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, +\infty\}$, for the static heuristic. We have a couple of observations from Figure 4(b). First, performance losses of the heuristics with rates $\beta \ge 0.5$ are roughly linear in θ and are worse than those of the other three heuristics when θ is large. This is consistent with Proposition 5, in which we show that when the boost in the sales intensity is relatively small (i.e., $\beta \ge 0.5$ in this case), the static heuristic is not asymptotically optimal, and the profit loss would be $\Theta(\theta)$. On the other hand, as long as $\lim_{\theta \to \infty} \sqrt{\theta} f(\theta) = \infty$, which is the case when $\beta < 0.5$, the profit loss would be less than $\Theta(\theta)$. This is reflected by the concave curvature of the corresponding lines in the figure. Second, in our experiments, the case with $\beta = 0.4$ achieves the best asymptotic performance. In general, a good candidate for β for the static heuristic will be just slightly below 0.5. This is because that the Poisson tail bound is subexponential (see Lemma A2 in the online appendix), and, thus, the probability of not hitting the target diminishes more quickly as β increases.

Figure 4. (Color online) Performance of Static Heuristics

Notes. (a) $p = 6$: $\lambda^* > N/T$. (b) $p = 2$: $\lambda^* < N/T$. $b = 40\theta$, $N = 20\theta$, $T = 5\theta$, and $c(\lambda) = (\lambda - 2)^2$.

Next, we evaluate the performance of dynamic heuristics, as discussed in Section 5.2. The setup of the numerical experiments is the same as before. Two extra parameters in the numerical experiments are the constant M and $\bar{\lambda}$ in the modified resolving heuristic. We present the set of results with $M = 10$ and $\bar{\lambda} = 5$ and discuss their impact on the performance of modified resolving heuristics at the end of this section.

The results are summarized in Table 1 and Figure 5. Some comments are in order. First, we observe from Table 1 and Figure 5(d) that both the static heuristic (SH) adopting the optimal deterministic intensity and the standard periodic resolving heuristic (RH) perform poorly, which is consistent with Propositions 5 and 6. Neither of the two heuristics introduces any boost in the sales intensity on top of the optimal deterministic rate (see Figure 5(a)), and, thus, there is a good chance that the target will not be reached at the end of the sales horizon. In fact, in our numerical experiments, the failure rate of either heuristic is consistently higher than 40%, as shown in Figure 5(c). Failing to reach the target is really costly for the seller because the cost incurred over the course of the sales horizon can never be recouped.

Second, Figure 5(d) shows that for both the static heuristic with a rate of $\lambda_D + \theta^{-0.4}$ (MSH) and the modified resolving heuristic (MRH), the profit loss as a percentage of the optimal deterministic profit converges to zero promptly as the scale factor θ increases. This, again, confirms our theoretical results that both a certain modified static heuristic and the modified resolving heuristic are asymptotically optimal. Having said that, we observe that the modified resolving heuristic consistently outperforms the modified static heuristic with a higher average profit level (see Table 1) and smaller profit loss as a percentage of the optimal deterministic profit (see Figure 5(d)). With the modified resolving heuristic, the seller has the chance to respond to past demand realizations and only needs to induce a higher-than-optimal rate when either the remaining time is
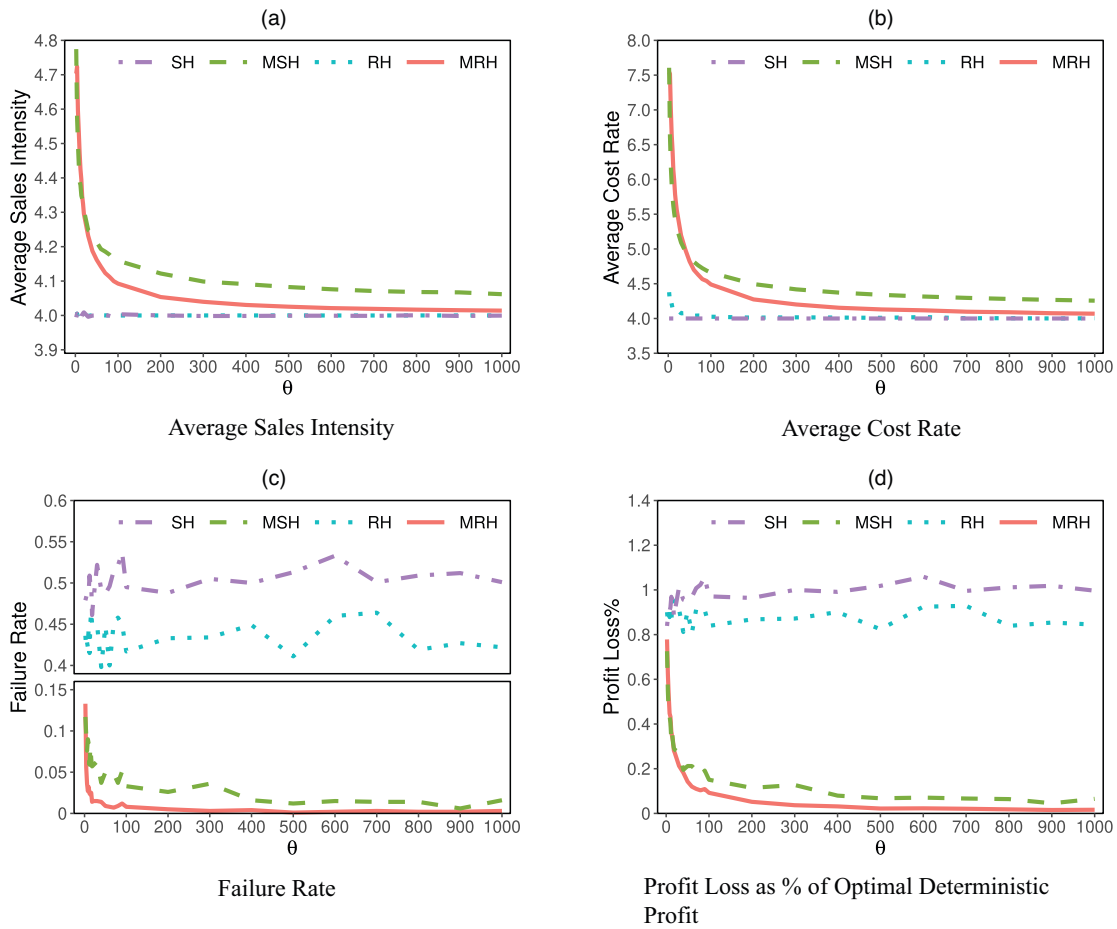
relatively short or the number of sales is well below that expected. Compared with the modified static heuristic, the modified resolving heuristic can produce a greater chance of reaching the target (see Figure 5(c)) with lower sales intensities (see Figure 5(a)).

Finally, we explore the choice of parameters M and $\bar{\lambda}$ for the modified resolving heuristic. Figure 6 displays the profit loss of the heuristic under different parameters with $\theta = 500$. We observe that, for a given $\bar{\lambda}$, the performance loss of the heuristic is nonmonotonic in M in general. Profit loss as a percentage of the optimal deterministic profit initially decreases in M when M is small, but then increases in M as M gets larger. The rationale for this nonmonotonicity is as follows. Recall that Algorithm 2 requires that the sales intensity be switched to the full-speed rate $\bar{\lambda}$ when the remaining time t is less than certain thresholds, such as $M \log \theta$, and the target has not yet been reached. With a small M, the heuristic tends to switch to $\bar{\lambda}$ relatively late in the selling season, if it indeed requires such a switch. Thus, the expected average sales rate and the average cost rate are lower. The flip side, however, is that there is a greater chance of failing to reach the target. On the other hand, the heuristic with a larger M causes the full-speed contingency plan to be implemented earlier, resulting in a higher average sales rate that deviates further from the optimal deterministic rate. However, such a heuristic produces a higher probability that the total sales will reach the target by the deadline. The optimal parameters for the modified resolving heuristic balance these two opposite forces by selecting an expected sales intensity that is just "high enough."

Figure 6 sheds light on the optimal selection of parameters. We denote by $M^*(\bar{\lambda})$ the optimal M such that the profit loss of the heuristic is minimized for any given $\bar{\lambda}$. Figure 6 suggests that $M^*(\bar{\lambda})$ decreases in $\bar{\lambda}$ in general. The practical implication of this observation is that, if the seller has limited maneuvers to

**Figure 5.** (Color online) Performance of Heuristics

Average Sales Intensity

Average Cost Rate

Failure Rate

Profit Loss as % of Optimal Deterministic Profit

Note. $b = 40\theta$, $N = 20\theta$, $T = 5\theta$, and $c(\lambda) = (\lambda - 2)^2$.

increase the sales rate—that is, $\bar{\lambda}$ is relatively small—she is better off implementing a full-speed contingency plan relatively early.

7. Conclusion

We study a sales effort management problem under an all-or-nothing constraint. This constraint plays a

Figure 6. (Color online) MRH: Profit Loss as Percentage of Optimal Deterministic Profit
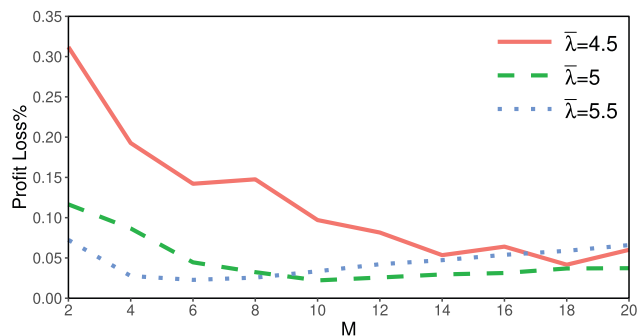
Note. $\theta = 500$, $b = 40\theta$, $N = 20\theta$, $T = 5\theta$, and $c(\lambda) = (\lambda - 2)^2$.

central role in shaping the optimal policy, as it yields two forces pushing in opposite directions. On the one hand, the all-or-nothing constraint incentivizes the seller to increase sales intensity, with the hope of reaching the target by the end of the sales horizon. However, if the target is not reached in the end, the cost will be sunk, and the seller will incur a loss. The optimal sales intensity is thereby nonmonotonic for the remaining time and distance to the target. We characterize the optimal policy for this problem, which generally demonstrates a watershed structure.

We then propose various easily computable and implementable heuristics and study their performance under the asymptotic regime when the target and the sales horizon are scaled up. One key takeaway from our analysis is that, because the seller's revenue is contingent on achieving the sales target, ensuring that the target will almost certainly be reached at the end of the sales horizon is essential for any heuristic to be asymptotically optimal. The performance loss associated with a heuristic hinges on the extra effort required to ensure the target being reached. In particular, we propose an asymptotically optimal modified resolving heuristic with a

logarithmic performance loss, which diminishes the adverse effect of the all-or-nothing constraint. This result also alleviates the moral hazard in the salesforce contract design. It suggests that by having a sufficiently long evaluation window and a proportionally high target, a rational seller can be induced to exert sufficient effort to reach the target, following the modified resolving heuristic or, even better, the optimal policy.

Last, we would like to clarify that our results, as well as insights generated from the analysis, may not be generalizable to the case where price is the decision variable and the target is specified with respect to revenue, rather than sales volume. This is because, as mentioned earlier, when the target is a specific sales volume, the only way for the seller to reach the target is to sell more products by lowering the price. However, the same cannot be said when the target is a certain amount of revenue. In that case, in order to increase the chance of hitting the target at the end of the sales horizon, the seller can choose either to sell the products in larger quantities by charging a lower price or to charge a higher price and hope that demand produces a favorable result. We would expect the latter strategy to be preferable in certain situations when the seller dynamically adjusts prices to hit a revenue target.

Acknowledgments

Endnotes

[1] Our results still hold for a payment scheme where, if the seller exerts no effort, there is a base demand rate, and the seller earns a fixed commission for every unit of sales, regardless of whether the target is met, in addition to the all-or-nothing bonus and the over-target commission per unit, which can be zero.

[2] Our analysis does not require the nonnegativity of $c(\lambda)$.

[3] $f(\theta) = \Theta(g(\theta))$ if there exist $k_1, k_2, \theta_0 > 0$ such that for any $\theta > \theta_0$, we have $k_1 \cdot g(\theta) \le f(\theta) \le k_2 \cdot g(\theta)$.

References

Ahn H-S, Wang D, Wu O (2021) Asset selling under debt obligations. *Oper. Res.* 69(4):1305–1323.

Alaei S, Malekian A, Mostagir M (2021) A dynamic model of crowdfunding. *Oper. Res.* Forthcoming.

Basu AK, Lal R, Srinivasan V, Staelin R (1985) Salesforce compensation plans: An agency theoretic perspective. *Marketing Sci.* 4(4):267–291.

Besbes O, Maglaras C (2012) Dynamic pricing with financial milestones: Feedback-form policies. *Management Sci.* 58(9):1715–1731.

Besbes O, Iancu DA, Trichakis N (2018) Dynamic pricing under debt: Spiraling distortions and efficiency losses. *Management Sci.* 64(10):4572–4589.

Bimpikis K, Ehsani S, Mostagir M (2019) Designing dynamic contests. *Oper. Res.* 67(2):339–356.

Burtch G, Gupta D, Martin P (2021) Referral timing and fundraising success in crowdfunding. *Manufacturing Service Oper. Management* 23(3):676–694.

Chen F (2000) Sales-force incentives and inventory management. *Manufacturing Service Oper. Management* 2(2):186–202.

Choi JP (1991) Dynamic R&D competition under "hazard rate" uncertainty. *RAND J. Econom.* 22(4):596–610.

Chung DJ, Steenburgh T, Sudhir K (2013) Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans. *Marketing Sci.* 33(2):165–187.

Cooper WL (2002) Asymptotic behavior of an allocation policy for revenue management. *Oper. Res.* 50(4):720–727.

Coughlan AT (1993) Salesforce compensation: A review of MS/OR advances. Eliashberg J, Lilien GL, eds. Marketing, Handbooks in Operations Research and Management Science, vol. 5 (Elsevier, Amsterdam), 611–651.

Du L, Hu M, Wu J (2017) Contingent stimulus in crowdfunding. Preprint, submitted August 1, https://dx.doi.org/10.2139/ssrn.2925962.

Gallego G, van Ryzin G (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* 40(8):999–1020.

Halac M, Kartik N, Liu Q (2017) Contests for experimentation. *J. Polit. Econom.* 125(5):1523–1569.

Jasin S (2014) Reoptimization and self-adjusting price control for network revenue management. *Oper. Res.* 62(5):1168–1178.

Jasin S, Kumar S (2012) A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Math. Oper. Res.* 37(2):313–345.

Joseph K, Kalwani MU (1998) The role of bonus pay in salesforce compensation plans. *Indust. Marketing Management* 27(2): 147–159.

Kalra A, Shi M (2001) Designing optimal sales contests: A theoretical perspective. *Marketing Sci.* 20(2):170–193.

Kim J, Randhawa RS (2017) The value of dynamic pricing in large queueing systems. *Oper. Res.* 66(2):409–425.

Lal R, Srinivasan V (1993) Compensation plans for single-and multi-product salesforces: An application of the Holmstrom-Milgrom model. *Management Sci.* 39(7):777–793.

Lal R, Staelin R (1986) Salesforce compensation plans in environments with asymmetric information. *Marketing Sci.* 5(3): 179–198.

Maglaras C, Meissner J (2006) Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing Service Oper. Management* 8(2):136–148.

Malueg DA, Tsutsui SO (1997) Dynamic R&D competition with learning. *RAND J. Econom.* 28(4):751–772.

Moldovanu B, Sela A (2001) The optimal allocation of prizes in contests. *Amer. Econom. Rev.* 91(3):542–558.

Oyer P (2000) A theory of sales quotas with limited liability and rent sharing. *J. Labor Econom.* 18(3):405–426.

Raju JS, Srinivasan V (1996) Quota-based compensation plans for multiterritory heterogeneous salesforces. *Management Sci.* 42(10):1454–1462.

Sohoni MG, Bassamboo A, Chopra S, Mohan U, Sendil N (2010) Threshold incentives over multiple periods and the sales hockey stick phenomenon. *Naval Res. Logist.* 57(6):503–518.

Steenburgh TJ (2008) Effort or timing: The effect of lump-sum bonuses. *Quant. Marketing Econom.* 6(3):235.

This American Life (2013) 129 cars. Accessed August 29, 2021, https://www.thisamericanlife.org/513/transcript.

Yuri L, McGill J, Nediak M (2008) Risk in revenue management and dynamic pricing. *Oper. Res.* 56(2):326–343.

Zhang J, Savin S, Veeraraghavan S (2021) Revenue management in crowdfunding. *Manufacturing Service Oper. Management.* Forthcoming.