informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 67, No. 4, April 2021, pp. 2211–2230
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Incentive Design for Operations-Marketing Multitasking

Tinglong Dai,[a] Rongzhu Ke,[b,*] Christopher Thomas Ryan[c]

[a] Carey Business School, Johns Hopkins University, Baltimore, Maryland 21202; [b] School of Economics, Zhejiang University, Hangzhou, Zhejiang 310058, China; [c] Sauder School of Business, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada

*Corresponding author

Contact: dai@jhu.edu, https://orcid.org/0000-0001-9248-5153 (TD); rzke455@zju.edu.cn, https://orcid.org/0000-0001-6378-3223 (RK); chris.ryan@sauder.ubc.ca, https://orcid.org/0000-0002-1957-2303 (CTR)

Received: March 11, 2019
Revised: January 22, 2020
Accepted: March 18, 2020
Published Online in Articles in Advance: August 4, 2020

https://doi.org/10.1287/mnsc.2020.3651

Abstract. A firm hires an agent (e.g., a store manager) to undertake both operational and marketing tasks. Marketing tasks boost demand, but for demand to translate into sales, operational effort is required to maintain adequate inventory. The firm designs a compensation plan to induce the agent to put effort into both marketing and operations while facing *demand censoring* (i.e., demand in excess of available inventory is unobservable). We formulate this incentive-design problem in a principal-agent framework with a multitasking agent subject to a censored signal. We develop a bang-bang optimal control approach, with a general optimality structure applicable to a broad class of incentive-design problems. Using this approach, we characterize the optimal compensation plan, with a bonus region resembling a "mast" and "sail" such that a bonus is paid when either all inventory above a threshold is sold or the sales quantity meets an inventory-dependent target. The optimal mast-and-sail compensation plan implies nonmonotonicity, where the agent can be less likely to receive a bonus for achieving a better outcome. This gives rise to an ex post moral hazard issue where the agent may "hide" inventory to earn a bonus. We show that this ex post moral hazard issue is a result of demand censoring. If available information includes a waiting list (or other noisy signals) to gauge unsatisfied demand, no ex post moral hazard issues remain.

History: Accepted by Vishal Gaur, operations management.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2020.3651.

Keywords: marketing-operations interface • multitasking • moral hazard • retail operations • optimal control

1. Introduction

The impetus for studying the interface of operations and marketing is the contention that each function cannot be managed without careful consideration of the other (Shapiro 1977, Ho and Tang 2004). This reality is acutely apparent in retail settings, where a store manager oversees both operational and marketing tasks. A search on the online employment platform Monster.com returns over 8,500 retail store manager job listings with "multitasking" as a core skill. Echoing this requirement, DeHoratius and Raman (2007, p. 523) contend that the store manager is a "*multitasking agent* who allocates effort to different activities based on the rewards that accrue from, and the cost of pursuing, each of these activities" (emphasis added). Stated differently, the multitasking store manager must allocate effort across both functions, and the effectiveness of this "balance of effort" is critical to the success of the store.

We focus on two activities of a store manager: (1) marketing (i.e., bolstering customer demand) and (2) operations (i.e., ensuring that inventory can be put in the hands of customers, where it belongs, instead of being misplaced, damaged, spoiled, or stolen through mismanagement).[1] In light of these two competing areas of focus, how does one design compensation plans to get the most out of one's store managers? Such a question is a critical concern for those running decentralized retail chains. The challenge of compensation design for retail store managers is the subject of business school case studies (Krishnan and Fisher 2005) and empirical research (DeHoratius and Raman 2007).

The vast majority of compensation models considers single-tasking agents, most prominently in the salesforce compensation literature (see Section 2 for more details). When it comes to multitasking agents, the research typically restricts attention to *linear contracts* in settings where the outcomes of each task are *perfectly observable*. The former belies an interest in nonoptimal contracts (the optimality of linear contracts is only established in very restrictive settings), whereas the latter is unsuitable for our setting. To translate demand generated through marketing effort into sales requires sufficient inventory, an outcome of operational effort. When inventory is insufficient, unmet demand is lost and unobservable, a phenomenon known as *demand censoring*. Accordingly, the outcomes of the associated tasks in our setting lack observability.

Demand censoring is widely seen in practice and is well studied in economics (e.g., Conlon and Mortimer 2013), marketing (e.g., Anupindi et al. 1998), and operations management (e.g., Besbes and Muharremoglu 2013). Its negative implications for sales performance are well known, largely because censoring complicates the forecasting of demand and the planning of inventory. However, the effect of demand censoring on contract design has not been studied in the multitasking setting.

A major takeaway of this paper is that demand censoring—a defining feature of the interplay between operations and marketing—has inherent and perplexing implications for compensation design. We come to this conclusion as follows. Practically implementable compensation plans typically have simple structures. Two prime examples are quota-bonus contracts and linear commission contracts in salesforce compensation. A pivotal property of these contracts, in addition to being easily understood by salespeople, is that they are *monotone*, meaning that an increase in sales weakly increases compensation. It would strike a salesperson as strange if an additional sale *reduced* his or her compensation. However, establishing the monotonicity of *optimal* contracts proves difficult.

In single-tasking salesforce compensation, researchers have examined the optimality of quota-bonus and linear commission contracts. Rogerson (1985), for example, shows monotonicity using the so-called first-order approach. This approach—a standard procedure used in deriving the optimal compensation plan in moral hazard problems—is not without controversy. Laffont and Martimort (2009, p. 200) state that "the first-order approach has been one of the most debated issues in contract theory" and that "when the first-order approach is not valid, using it can be very misleading." In particular, the convex distribution function condition (CDFC), often assumed in the moral hazard principal-agent literature to support the first-order approach, is satisfied by essentially *no* familiar distributions. (Recent work by Ke and Ryan (2018a, b) attempts to establish the monotonicity of the optimal contract without using the first-order approach.) The validity of the first-order approach is particularly troubling in a multitasking setting, with a multidimensional effort and a multidimensional output signal.[2]

To overcome these technical challenges, we develop a "bang-bang" optimal control approach that applies to a broad class of incentive-design problems that significantly relaxes conditions needed to establish optimality. This approach allows for most of the commonly used families of distributions on both the operational and marketing sides. Using this approach, we characterize the optimal compensation plan for a multitasking agent subject to a censored signal. The optimal compensation plan we derive is analogous to the quota-bonus contracts of the salesforce literature, except now a *bonus region* for sales and inventory realizations exists: if sales and inventory realize in this region, a bonus is granted; otherwise, the store manager gets only his or her salary. Concretely, we find an optimal compensation plan for the multitasking store manager under the monotone likelihood-ratio property (MLRP), which is commonly assumed in the principal-agent literature (see Laffont and Martimort 2009, pp. 164–165), that consists of a base salary and a bonus paid to the store manager when either (1) inventory does not clear and the sales quantity exceeds an inventory-dependent threshold or (2) inventory clears and the realized inventory level exceeds a threshold.

Intriguingly, the structure of such a bonus region, which resembles a "mast" and "sail" (see Figure 2(a)), gives rise to inherent nonmonotonicity of the optimal compensation plan—given the same sales outcome, scenarios exist in which the store manager receives the bonus at some inventory level but no longer so at a higher inventory level. In other words, ceteris paribus, the store manager seems to be penalized for better inventory performance. This nonintuitive result can be understood as follows. When inventory is cleared, the realized demand is unobservable and capped by the inventory level. The firm's observed sales quantity is a *lower bound* on realized demand. Given the same sales quantity, as inventory increases, the sales manager no longer clears the inventory. The observed sales quantity is *equal to* (as opposed to a lower bound of) realized demand. Increased inventory is informative of the store manager *not* exerting high marketing effort. This informational reasoning justifies a loss of the bonus.

Our derivation of the optimal compensation plan restricts attention to ex ante moral hazard (i.e., the agent's effort after entering into the compensation plan is not observable). This focus is standard in the literature—the vast majority of the moral hazard literature ignores any ex post moral hazard (i.e., after exerting effort, the agent does not manipulate the realized outcome). A nonmonotone optimal compensation plan evokes the speculation that, in certain cases, a store manager may "hide" inventory to represent a stockout to the firm, thereby hiding a potential deficiency in marketing effort. In other words, in the presence of the additional consideration of ex post moral hazard, demand censoring further confounds operations-marketing multitasking.

Of course, the incentive to hide inventory could be monitored by the company. However, the need for such careful monitoring runs against the principle of effective incentive design: if incentives are appropriately designed, employees have the "right" incentives to manage their own behavior. If monitoring can capture the overstating of inventory losses (i.e., ex post moral hazard), can we not also monitor operational and marketing effort (i.e., ex ante moral hazard)? This reveals an *agency conundrum*: because of the firm's inability to monitor customer intentions (i.e., not observing all of demand because of inventory shortfalls), it is unable to design intuitive compensation schemes that preclude the need for the monitoring of employee intentions, either their conscientiousness in sales and operational activities or their honesty in representing the level of inventory in the store. This conundrum has important implications for incentive design in the retail setting.

We believe that this agency conundrum (the result of demand censoring and nonmonotone contracts) is at the core of multitasking with censored signals. Further analysis and numerical investigations show that several intuitive monotone compensation plans fail to be optimal. A natural first idea, given the two output signals (demand and inventory), is to give the store manager a bonus if each signal meets some minimum threshold. We call such compensation plans *corner* compensation plans because the two thresholds form a corner in the outcome space. The logic of corner compensation plans finds its trace in practice (e.g., Krishnan and Fisher 2005, DeHoratius and Raman 2007) and is in line with known results in the single-tasking contract theory literature (e.g., Oyer 2000). Nonetheless, we show that such plans cannot be optimal and furthermore exhibit natural cases where they perform arbitrarily poorly. Other simple (and monotone) compensation plans, such as linear compensation, do not fare any better in our numerical experiments.

Our resolution of the agency conundrum is also telling. The trap of both ex ante and ex post moral hazard is not overcome by further monitoring of employees but instead by improved monitoring of customer intentions, even to a modest degree. If the firm can noisily gauge unsatisfied demand, for example, through a waiting list where an unknown but nonzero proportion of unsatisfied demand is recorded, an optimal compensation plan can be constructed to handle both ex ante and ex post moral hazard issues. Remarkably, going from complete demand censoring to partial demand censoring greatly alleviates the challenge of managing inventory, here indirectly through incentive design.

Taken together, our results allude to a novel connection between customer intention and employee effort. The visibility of customer behavior (i.e., customer demand) and the visibility of employees' behavior (i.e., their effort) are linked through employee compensation. Monitoring employee behavior in order to improve employee effort is unnecessary; improved monitoring of customers can suffice. This interplay between customer behavior and operational planning goes to the very heart of what makes the operations-marketing interface compelling to study.

2. Related Literature

The retail operations literature has empirically documented the importance of incentive design for store managers. DeHoratius and Raman (2007) empirically study store managers as multitasking agents who function as both an inventory-shrinkage controller and a salesperson. DeHoratius and Raman (2007) substantiate the view that store managers make their effort decisions across both job functions in response to incentives. Krishnan and Fisher (2005) provide a process view of the range of a retail manager's responsibilities and detail the impact of incentive design on operational and marketing efforts, counting spoilage and shrinkage control as crucial areas of managerial control. To the best of our knowledge, our paper is the first analytical treatment of optimal incentive design for a multitasking store manager. Accordingly, we are the first to provide an optimal benchmark to assess losses due to demand censoring and multitasking. Our findings shed light on the nature of the relationship between marketing and operations, an issue that has inspired a voluminous literature (e.g., Shapiro 1977; Ho and Tang 2004; Jerath et al. 2007, 2017).

Salesforce compensation has been studied in the economics, marketing, and operations management literature (see, e.g., Lal and Srinivasan 1993, Raju and Srinivasan 1996, Oyer 2000, Misra et al. 2004, Herweg et al. 2010, Jain 2012, Jain et al. 2019, Chen et al. 2020, Long and Nasiry 2020). Much of this literature focuses on two types of contracts, linear commission and quota bonus (i.e., the salesperson receives a bonus for meeting a sales quota). The optimality of linear commission contracts has various caveats—its primary justification assumes a normally distributed outcome and a constant absolute risk-aversion (CARA) agent utility. By contrast, the optimality of quota-bonus contracts has followed from less restrictive conditions, namely risk neutrality, limited liability, and a general outcome distribution. (Limited liability captures an agent's aversion to downside risk and can be viewed as a type of risk aversion.) We follow the latter tradition and derive optimal contracts in a spirit similar to quota-bonus contracts, although with important differences.

The salesforce compensation literature had, until recently, assumed that unlimited inventory meets the demand generated by the salesperson. A recent stream of literature (Chu and Lai 2013; Dai and Jerath 2013, 2016, 2019) incorporates demand censoring because of limited inventory into the single-tasking model.

We study the compensation of a store manager who undertakes operational effort to increase the realized inventory level, in addition to marketing effort to influence demand. As a result, our optimal compensation plan exhibits a structure that does not immediately generalize the well-studied quota-bonus contract from the single-tasking setting. Indeed, we show several "intuitive" generalizations, including that corner compensation plans are not optimal and can perform poorly relative to the optimal compensation plan.

Our paper also relates to the accounting and economics literature (e.g., Holmstrom and Milgrom 1991, Feltham and Xie 1994, Dewatripont et al. 1999) on multitasking. The seminal work here is Holmstrom and Milgrom's (1991) model of a multitasking agent whose job consists of multiple concurrent activities that jointly produce a multidimensional output. They focus on a linear compensation scheme and show that varying the weights of the compensation plan elicits changes in the agent's effort allocation. Our work departs from this setting in several ways. First, our multidimensional output affects the principal's utility in a *nonlinear* fashion. Second, our paper derives the *optimal* compensation plan, whereas the literature (following Holmstrom and Milgrom 1991) mostly assumes a linear compensation scheme without careful justification of optimality. Third, the observability of our multidimensional output signal is *imperfect*, microfounded through demand censoring. By contrast, the literature typically assumes perfect observability on all dimensions of the output signal. As we reveal, unobservability provides rich managerial implications.

Thematically, demand censoring plays a significant role in our analysis.[3] By providing a novel connection between understanding customer demand and designing compensation plans, our paper enriches a stream of literature on demand censoring (e.g., Huh et al. 2011, Besbes and Muharremoglu 2013, Feiler et al. 2013, Rudi and Drake 2014, Jain et al. 2015). In particular, Besbes and Muharremoglu (2013) show an exploration–exploitation trade-off in a multiperiod inventory control problem without moral hazard. They show in the case of a discrete demand distribution that the lost-sales indicator voids the need for active exploration. Jain et al. (2015) study another multiperiod inventory control problem (also without moral hazard) and numerically show that the timing information of stockout can help recover much of the inefficiency from demand censoring. Related to this literature, we show that a noisy signal of the lost demand can resolve ex post moral hazard issues.

Finally, we advance the methodology of principal-agent theory by using a bang-bang approach to solve risk-neutral, limited-liability moral hazard problems with finitely many actions. Although optimal control is a classical tool in economics, marketing, and operations (see, e.g., Sethi and Thompson 2000; Crama et al. 2008), to our knowledge, the application of this type of logic in the moral hazard literature is limited. We model the risk-neutral setting, which makes the problem linear, so optimality is based on *extremal* solutions with a bang-bang structure. We explore this approach generally to provide a methodological understanding of the approach that we believe may be of separate interest for contract theory researchers. Our work applies results from a particularly cogent presentation of optimization in L_∞ spaces in Barvinok (2002, sections III.5 and IV.12). This general setup treats linear optimal control as a special case.

3. Model

Consider a multitasking store manager (the agent) hired by a firm (the principal) to make operational effort e_o and marketing effort e_m. We assume that e_o and e_m take on at most finitely many values. Operational effort concerns increasing available inventory, and marketing effort concerns increasing demand. The principal cannot directly observe the effort choices of the store manager; they can only be indirectly inferred by observing inventory and demand realizations.

Let us be more precise about the mechanics of operational effort and realized inventory. The firm supplies the store manager with an initial inventory level $\bar{I}$. The realized inventory $I \le \bar{I}$ is all that is available to meet demand. The difference $\bar{I} - I$ is unavailable to meet demand because of a variety of factors, including theft, damage, spoilage, and misshelving. Operational effort stochastically affects these factors to improve realized inventory. Until Section 9.2, we focus on the underlying incentive issues for effectively handling a given stock of inventory (i.e., $\bar{I}$ is exogenous).

The cumulative distribution function of realized available inventory I is $F(i|e_o)$ with probability density function $f(i|e_o)$, where $i \in [0, \bar{I}]$. We denote demand by Q, its cumulative distribution function by $G(q|e_m)$, and its probability density function by $g(q|e_m)$ for $q \in [0, \bar{Q}]$, where $\bar{Q}$ is an upper bound on demand.[4] These assumptions imply that operational effort does not affect demand and that marketing effort does not affect the realization of available inventory. Accordingly, for every effort level, the random variables I and Q are independent. Both density functions f and g are continuous functions of their first argument.

We make a standard assumption (see, e.g., Grossman and Hart 1983, Rogerson 1985) that the output distributions $f(I|e_o)$ and $g(Q|e_m)$ satisfy the monotone likelihood-ratio property (MLRP); that is,

$$\frac{f(i|e_o)}{f(i|\hat{e}_o)} \text{ nonincreasing in } i \text{ for } e_o < \hat{e}_o \text{ and}$$
$$\frac{g(s|e_m)}{g(s|\hat{e}_m)} \text{ nonincreasing in } s \text{ for } e_m < \hat{e}_m. \tag{1}$$

The MLRP implies that a better inventory (demand) outcome is more informative of the fact that the store manager has exerted operational (marketing) effort. The MLRP is satisfied by most of the commonly used families of distributions.

The (random) sales outcome is denoted $S \triangleq \min\{I, Q\}$. To reflect the phenomenon of demand censoring, we assume that both the firm and the store manager observe the realized inventory level and sales outcome, but neither can observe the realized demand in excess of the realized inventory level. We assume that $\bar{Q} \geq \bar{I}$ to allow for the possibility that demand is censored at its highest level.

The store manager is effort averse. His or her disutility from exerting efforts (e_o, e_m) is given by $c(e_o, e_m)$. We assume that $c(e_o, e_m)$ is increasing in both dimensions of effort.

The firm designs a compensation plan $w(I, S)$ to maximize its total expected revenue less the total expected compensation to the store manager. We assume that both the firm and the store manager are risk neutral but with limited liability, bounding w below by $\underline{w}$ and above by $\bar{w}$. The lower bound on compensation ($\underline{w}$) is normalized to zero without loss. The latter ($\bar{w}$) implies that the firm is budget constrained and cannot compensate beyond $\bar{w}$. This budget $\bar{w}$ is known to both the firm and the store manager. Assuming an upper bound for the compensation level is fairly common in the contract theory literature (e.g., Holmstrom 1979, Innes 1990, Arya et al. 2007, Jewitt 2008, Bond and Gomes 2009). In particular, Bond and Gomes (2009, p. 177) provide a variety of motivations for it, such as "a desire to limit the pay of an employee to less than his/her supervisor." We take $\bar{w}$ as given until Section 9.1, in which we generalize the upper bound on $w(i, s)$ to be a more general resource constraint that is an integrable function of i and s.

The sequence of events is as follows. First, the firm offers a compensation plan $w(i, s)$ to the store manager, who either takes it or leaves it. Second, if the compensation plan is accepted, the store manager chooses an operational effort e_o and a marketing effort e_m. Both efforts are exerted simultaneously. Third, inventory I and demand Q outcomes are realized simultaneously, and inventory and sales $S = \min\{Q, I\}$ are observed. Each unit of met demand yields the principal a margin of r, unmet demand is lost and unobserved, and unused inventory is salvaged at a return normalized to zero. Fourth, the firm compensates the store manager according to $w(\cdot, \cdot)$. Because initial inventory $\bar{I}$ is given, the cost of procuring inventory is sunk. Accordingly, we may formulate the firm's problem as

$$\max_{w, e_o^*, e_m^*} \quad r\mathbb{E}[S|e_o^*, e_m^*] - \mathbb{E}[w(I, S)|e_o^*, e_m^*] \tag{2a}$$
$$\text{s.t.} \quad S = \min\{Q, I\}, \tag{2b}$$
$$\mathbb{E}[w(I, S)|e_o^*, e_m^*] - c(e_o^*, e_m^*) \geq \underline{U}, \tag{2c}$$
$$\mathbb{E}[w(I, S)|e_o^*, e_m^*] - \mathbb{E}[w(I, S)|e_o, e_m]$$
$$\geq c(e_o^*, e_m^*) - c(e_o, e_m) \text{ for all } (e_o, e_m), \tag{2d}$$
$$0 \leq w(i, s) \leq \bar{w} \text{ for all } (i, s), \tag{2e}$$

where the expectation $\mathbb{E}[\cdot|e_o, e_m]$ is taken over the joint distribution of I and S at effort levels e_o and e_m. The participation constraint (2c) ensures that the store manager's expected net payoff is no lower than a reservation utility $\underline{U}$, and the incentive compatibility (IC) constraint (2d) ensures that choosing (e_o^*, e_m^*) over all other effort levels is optimal for the store manager.

We refer to problem (2) as a *multitasking store manager* problem. This problem is conceptually challenging. Indeed, it is a bilevel optimization problem with an *infinite*-dimensional decision variable w. Deriving the form of an optimal compensation plan $w(\cdot, \cdot)$ requires a methodical exploration of optimality conditions in this setting.

4. A Bang-Bang Optimal Control Approach

In this section, we study a general class of risk-neutral moral hazard problems with finite agent action sets. Our approach applies more broadly than the multitasking store manager setting, so we describe it in a general notation not overly specific to its use in this paper.

Consider a moral hazard problem between one principal and one agent. The agent has a finite set of actions $\mathcal{A} = \{\vec{a}^{\,1}, \vec{a}^{\,2}, \ldots, \vec{a}^{\,m}\}$; we use the arrow notation $\vec{a}$ to denote a vector. In the multitasking setting, this assumption implies a finite number of operational effort levels e_o, a finite number of marketing effort levels e_m, and that each action $\vec{a} \in \mathcal{A}$ is a *pair* of efforts $\vec{a} = (e_o, e_m)$.

The agent incurs a cost $c(\vec{a})$ for taking action $\vec{a} \in A$, where we assume that $c(\vec{a})$ is increasing in $\vec{a}$. The output is a vector $\vec{x} \in \mathcal{X}$, where $\mathcal{X}$ is a compact subset of $\mathbb{R}^n$, for some integer n.[5] The random output X has a density function $f(\vec{x}\,|\vec{a})$, where $f(\cdot|\vec{a})$ is in $L^1(\mathcal{X})$ for all $\vec{a} \in \mathcal{A}$ and $f(\vec{x}\,|\vec{a}) > 0$ for all $\vec{x} \in \mathcal{X}$ and $\vec{a} \in \mathcal{A}$; the notation $L^1(\mathcal{X})$ denotes the space of all absolutely integrable functions on $\mathcal{X}$ with respect to the Lebesgue measure on $\mathbb{R}^n$. This general formulation allows the signals to be correlated and depend on combinations of efforts.

The principal offers the agent the wage contract $w: \mathscr{X} \to \mathbb{R}$ that pays out according to the realized outcome. The principal values outcome $\vec{x} \in \mathscr{X}$ according to the valuation function $\pi: \mathscr{X} \to \mathbb{R}$. The agent has limited liability and must receive a minimum wage of $\underline{w}$ almost surely. We normalize $\underline{w}$ to zero. Moreover, the principal has a constraint that tops compensation out at $\bar{w}$; that is, $w(x) \le \bar{w}$ for almost all $x \in \mathscr{X}$. Finally, the agent has a reservation utility $\underline{U}$ for his or her next-best alternative.

Both the principal and agent are risk neutral. The expected utility of the principal is denoted $V(w, \vec{a}) \triangleq \int_{x \in \mathscr{X}} (\pi(\vec{x}) - w(\vec{x})) f(\vec{x}|\vec{a}) dx$, and the expected utility of the agent is $U(w, \vec{a}) \triangleq \int_{\vec{x} \in \mathscr{X}} w(\vec{x}) f(\vec{x}|\vec{a}) d\vec{x} - c(\vec{a})$. We formulate the moral hazard problem as

$$\max_{w, \vec{a}} \quad V(w, \vec{a}) \tag{3a}$$

$$\text{s.t.} \quad U(w, \vec{a}) \ge \bar{U}, \tag{3b}$$

$$U(w, \vec{a}) - U(w, \vec{a}^i) \ge 0 \text{ for } i = 1, 2, \ldots, m, \tag{3c}$$

$$0 \le w \le \overline{w}. \tag{3d}$$

Following the two-step solution approach developed by Grossman and Hart (1983), we suppose that an implementable target action $\vec{a}^*$ has been identified. This approach reduces the problem to

$$\min_{w} \quad \int_{\vec{x} \in \mathscr{X}} w(\vec{x}) f(\vec{x}|\vec{a}^*) d\vec{x} \tag{4a}$$

$$\text{s.t.} \quad \int_{\vec{x} \in \mathscr{X}} w(\vec{x}) f(\vec{x}|\vec{a}^*) d\vec{x} \ge \underline{U}, \tag{4b}$$

$$\int_{\vec{x} \in \mathscr{X}} R_i(\vec{x}) w(\vec{x}) f(\vec{x}|\vec{a}^*) d\vec{x} \ge c(\vec{a}^*) - c(\vec{a}^i) \text{ for } i \in \{1, 2, \ldots, m\} \text{ such that } \vec{a}^i \ne \vec{a}^*, \tag{4c}$$

$$0 \le w \le \overline{w}, \tag{4d}$$

where we use the fact that $V(w, \vec{a}) = \mathbb{E}[\pi(\vec{x})|\vec{a}^*] - \int_{\vec{x} \in \mathscr{X}} w(\vec{x}) f(\vec{x}|\vec{a}^*) d\vec{x}$, drop the constant $\mathbb{E}[\pi(\vec{x})|\vec{a}^*]$ from the objective, convert to a minimization problem, and simplify the constraint (3c) by defining

$$R_i(\vec{x}) \triangleq 1 - \frac{f(\vec{x}|\vec{a}^{\,i})}{f(\vec{x}|\vec{a}^*)} \tag{5}$$

for $i = 1, 2, \ldots, m$. Finally, we drop the IC constraint for $\vec{a}^i = \vec{a}^*$ because this constraint is always satisfied with equality.

A *bang-bang contract* is a feasible solution to (4), where $w(\vec{x}) \in \{0, \overline{w}\}$ for almost all $\vec{x} \in \mathscr{X}$.

Theorem 1. *An optimal bang-bang contract for* (4) *exists.*

Next, we characterize when an optimal bang-bang contract takes the value of zero and when it takes the value $\bar{w}$. This characterization is associated with a *trigger* value of a weighted sum of appropriately defined covariances of the contract with the likelihoods of outcomes under different actions. Our analysis uses tools found in Barvinok (2002, section IV.12).

Theorem 2. *There exist nonnegative multipliers* ω_i *and a target* t *such that an optimal solution to* (4) *of the following form exists:*[6]

$$w^*(\vec{x}) = \begin{cases} \overline{w} & \text{if } \sum_{i=1}^{m} \omega_i R_i(\vec{x}) \ge t, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where $\sum_{i=1}^m \omega_i = 1$ *holds.*

Let $B \triangleq \{\vec{x} \in \mathscr{X} : \sum_{i=1}^m \omega_i R_i(\vec{x}) \ge t\}$ denote the *bonus region* of the compensation plan w^*. In other words, $w^*(\vec{x})$ evaluates to $\bar{w}$ inside B and zero outside B.

The contract in (6) has a compelling economic interpretation. Consider the condition

$$\sum_{i=1}^{m} \omega_i R_i(\vec{x}) \ge t \tag{7}$$

that defines the bonus region B. Because the ω_i values are nonnegative and sum to one, the left-hand side is a weighted sum of likelihood ratios that can be viewed as a measure of the information value (or informativeness) of outcome $\vec{x}$ for determining if the agent took the target action $\vec{a}^*$. For the given outcome $\vec{x}$, larger values of $R_i(\vec{x})$ are associated with actions $\vec{a}^i$, where the outcome $\vec{x}$ is less likely under action $\vec{a}^i$ than action $\vec{a}^*$. Thus, the larger $\sum_{i=1}^m \omega_i R_i(\vec{x})$ is, the less likely the agent is to have deviated from $\vec{a}^*$. The trigger condition (7) rewards outcomes whose informativeness exceeds the given threshold t. The weights ω_i fine-tune how we measure this informativeness and are determined by solving a dual problem that "prices" the significance of deviations to different actions.

In light of this logic, we refer to contracts of the form (6) as *information-trigger* contracts (or simply *trigger* contracts). If the information value (as measured by $\sum_{i=1}^m \omega_i R_i(\vec{x})$) exceeds some trigger value, the agent is rewarded for that outcome.

The proof of Theorem 2 derives ω_i and t from solving a dual optimization problem. However, another approach is to solve a restricted class of the primal moral hazard problem (4) where contracts are information-trigger contracts of the form (6). If w is an information-trigger contract,

$$V(w, \vec{a}^*) = \bar{w} \int_{\substack{\vec{x} \in \mathscr{X} \text{ subject to} \\ \sum_{i=1}^m \omega_i R_i(\vec{x}) \ge t}} f(\vec{x}|\vec{a}^*) d\vec{x} = \bar{w} \mathbb{P}\left[\sum_{i=1}^{m} \omega_i R_i(\vec{X}) \ge t\right]$$

and

$$\int_{\substack{\vec{x}\in\mathcal{X}\text{ subject to}\\ \sum_{i=1}^{m}\omega_i R_i(\vec{x})\geq t}} R_i(\vec{x})w(\vec{x})f(\vec{x}|\vec{a})d\vec{x} = \bar{w}\mathbb{E}\left[R_i\left(\vec{X}\right)\middle|\sum_{i=1}^{m}\omega_i R_i\left(\vec{X}\right)\geq t\right],$$

where $\mathbb{P}[\cdot]$ is the probability measure, and $\mathbb{E}[\cdot]$ is the expectation operator associated with $f(\cdot|\vec{a}^*)$. By using this notation, the restriction of (4) over trigger contracts of the form (6) is

$$\min_{\omega,t} \quad \bar{w}\mathbb{P}\left[\sum_{i=1}^{m}\omega_i R_i\left(\vec{X}\right)\geq t\right] \tag{8a}$$

$$\text{s.t.} \quad \bar{w}\mathbb{P}\left[\sum_{i=1}^{m}\omega_i R_i\left(\vec{X}\right)\geq t\right] - c(\vec{a}^*)\geq \overline{U}, \tag{8b}$$

$$\bar{w}\mathbb{E}\left[R_i\left(\vec{X}\right)\middle|\sum_{i=1}^{m}\omega_i R_i\left(\vec{X}\right)\geq t\right]\geq c\left(\vec{a}^*\right)-c\left(\vec{a}^i\right) \quad \text{for } i\in\{1,2,\ldots,m\}, \tag{8c}$$

$$\sum_{i=1}^{m}\omega_i = 1, \tag{8d}$$

$$\omega_i \geq 0 \text{ for all } i\in\{1,2,\ldots,m\}. \tag{8e}$$

The next result relates optimality in this problem to the original problem (4).

Theorem 3. *Problem* (8) *has the same optimal value as* (4). *Moreover, an optimal solution to* (8) *corresponds to an optimal solution to* (4).

Theorem 3 says that it suffices to solve the finite-dimensional problem (8) to solve the original moral hazard problem.

5. Analyzing the Multitasking Store Manager Problem

We now study the store manager problem introduced in Section 3 using the bang-bang approach of Section 4.

5.1. General Optimality Structure

Theorem 2 applies directly to the store manager problem (2). A critical object needed to define information-trigger compensation plans of the form (6) is the joint distribution of S and I. Demand Q and inventory I are assumed to be independent, and hence, deriving their joint distribution is straightforward. Deriving the joint distribution of the sales and inventory is more difficult because of demand censoring. The following lemma provides the joint cumulative distribution function $\Pr(I\leq i, S\leq s|e_o,e_m)$ of I and S.

Lemma 1. *The joint cumulative distribution function*

$$\Pr(I\leq i, S\leq s|e_o,e_m) = \begin{cases} F(s|e_o)+G(s|e_m)\,[F(i|e_o)-F(s|e_o)] & \text{if } s<i,\\ F(i|e_o) & \text{if } s=i.\end{cases}$$

Before deriving the joint probability density function, we briefly discuss the domain of compensation plans. Note that $D\triangleq\{(i,s):0\leq s\leq i \text{ and } 0\leq i\leq\bar{I}\}$ is the domain of any feasible compensation plan because of demand censoring. We also denote by $D^{\mathrm{NSO}}\triangleq\{(i,s)\in D: s<i\}$ and $D^{\mathrm{SO}}\triangleq\{(i,s)\in D: s=i\}$ the regions of the domain where no stockout occurs and where stockout occurs, respectively. For simplicity, we denote by $w(i)$ the compensation level when $s=i$; that is, we shorten $w(i,i)$ to $w(i)$.

The underlying measure of tuples (i,s) is absolutely continuous when $s<i$, whereas along the 45° line for each i, a point mass of weight $1-G(i|e_m)$ at (i,i) is present. The joint probability density function of S and I is thus $h(i,s|e_o,e_m)=f(i|e_o)g(s|e_m)$ for $s<i$ and $h(i,i|e_o,e_m)=f(i|e_o)(1-G(i|e_m))$ when $s=i$.

Given this density function, using (5), we represent the ratio function $R_{e_o,e_m}(i,s)$ as

$$R_{e_o,e_m}(i,s) = 1-\frac{\mathbb{I}[i>s]f(i|e_o)g(s|e_m)+\delta(i=s)f(i|e_o)(1-G(i|e_m)}{\mathbb{I}[i>s]f\left(i|e_o^*\right)g\left(s|e_m^*\right)+\delta(i=s)f\left(i|e_o^*\right)\left(1-G\left(i|e_m^*\right)\right)},$$

where $\mathbb{I}[\cdot]$ is the indicator function, and $\delta(i=s)$ is a Dirac function at i. We describe an optimal information-trigger compensation plan in two different scenarios: (1) where $i>s$ (no stockout) and (2) where $i=s$ (stockout), by defining appropriate ratio functions. In the nonstockout (NSO) case,

$$R^{\mathrm{NSO}}_{e_o,e_m}(i,s) = 1-\frac{f(i|e_o)g(s|e_m)}{f\left(i|e_o^*\right)g\left(s|e_m^*\right)}, \tag{9}$$

and in the stockout (SO) case,

$$R^{\mathrm{SO}}_{e_o,e_m}(i) = 1-\frac{f(i|e_o)(1-G(i|e_m)}{f\left(i|e_o^*\right)\left(1-G\left(i|e_m^*\right)\right)}. \tag{10}$$

Theorem 2 implies that an optimal compensation plan takes the following form:

$$w^*(i,s) = \begin{cases} w^{\mathrm{NSO}}(i,s) & \text{if } (i,s)\in D^{\mathrm{NSO}},\\ w^{\mathrm{SO}}(i) & \text{if } (i,s)\in D^{\mathrm{SO}},\end{cases} \tag{11}$$

where

$$w^{\mathrm{NSO}}(i,s)=\begin{cases}\overline{w} & \text{if } \sum_{e_o,e_m}\omega_{e_o,e_m}R^{\mathrm{NSO}}_{e_o,e_m}(i,s)\geq t,\\ 0 & \text{otherwise,}\end{cases}$$

$$w^{\mathrm{SO}}(i)=\begin{cases}\overline{w} & \text{if } \sum_{e_o,e_m}\omega_{e_o,e_m}R^{\mathrm{SO}}_{e_o,e_m}(i)\geq t,\\ 0 & \text{otherwise,}\end{cases}$$

for some choice of t and nonnegative ω_{e_o,e_m} satisfying $\sum_{e_o,e_m}\omega_{e_o,e_m}=1$.

Recall that B denotes the bonus region of the information-trigger compensation plan w^* defined in (6). We adopt that notation here and refine it further by setting

$$B^{\mathrm{NSO}}\triangleq\left\{(i,s)\in D^{\mathrm{NSO}}:\sum_{e_o,e_m}\omega_{e_o,e_m}R^{\mathrm{NSO}}_{e_o,e_m}(i,s)\geq t\right\},\quad(12)$$

$$B^{\mathrm{SO}}\triangleq\left\{(i,s)\in D^{\mathrm{SO}}:\sum_{e_o,e_m}\omega_{e_o,e_m}R^{\mathrm{SO}}_{e_o,e_m}(i,s)\geq t\right\}.\quad(13)$$

A key observation here is that the bonus region B^{NSO} is possibly a full-dimensional subset of the NSO region of the domain D^{SO}, whereas the bonus region B^{SO} is a one-dimensional set along the 45° line D^{SO}.

5.2. Mast-and-Sail Compensation Plans

Throughout the rest of this paper, we make more concrete the structure of the optimal compensation plan (11) in a special multitasking setting with two levels—*high* (H) and *low* (L)—for each of the operational and marketing efforts. In the notation of Section 4, $\mathcal{A}=\{(e_o^H,e_m^H),(e_o^H,e_m^L),(e_o^L,e_m^H),(e_o^L,e_m^L)\}$. We also assume the target action is (e_o^H,e_m^H), that is, for the store manager to make his or her best effort in both operations and marketing. For a discussion of scenarios where other effort levels may be targeted, see Section OA.6 of the online appendix.

We first look into the structure of B^{NSO} under these assumptions.

Proposition 1. *A nonincreasing and continuous function s^* and $i_s\in(0,\bar{I}]$ exist such that $B^{NSO}=\{(i,s): i\geq i_s$ and $s^*(i)\leq s<i\}$.*[7]

Because $s^*(i)$ is a nonincreasing and continuous function of i on its domain, the bonus region resembles the one shown in Figure 1(a). We call the shape of this region a *sail*.

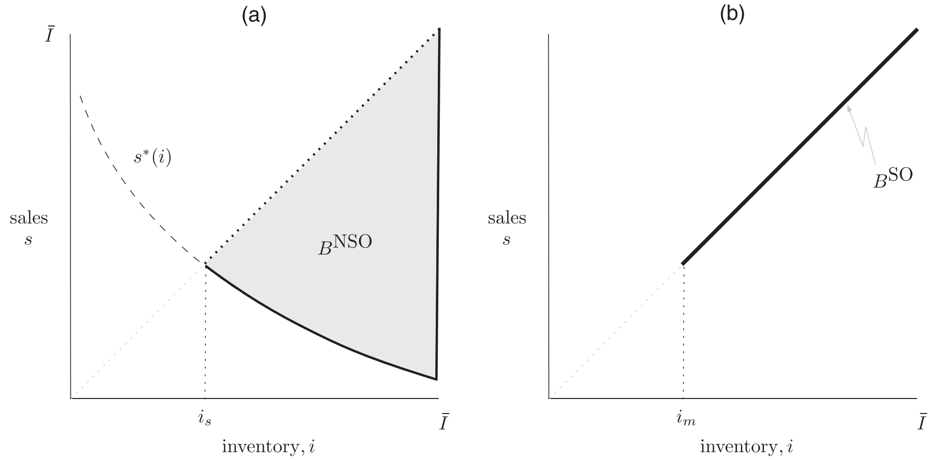
Proposition 2. *An inventory value $i_m\in(0,\bar{I}]$ exists such that $B^{SO}=\{(i,s): s=i\geq i_m\}$.*

Figure 1(b) gives a visualization of the bonus region B^{SO}. We call this region a *mast*.

Taken together, the bonus region of the optimal compensation plan w^* defined in (11) is the union of the regions in Propositions 1 and 2. Figure 2, (a) and (b), illustrates two of the possible structures of this union that result from the regions failing to "overlap" perfectly. When $i_s>i_m$, the bonus region has an inherent nonconvexity at (i_s,i_s), as illustrated in Figure 2(a). The shape in this figure makes clear our usage of the phrase *mast and sail* to describe the bonus region of an optimal compensation plan. When $i_s<i_m$, the mast and sail (with what looks like a mast that is too short for its sail) overlap to form a region that is not closed, as illustrated in Figure 2(b). The next result shows that only the structure seen in Figure 2(a) is possible.

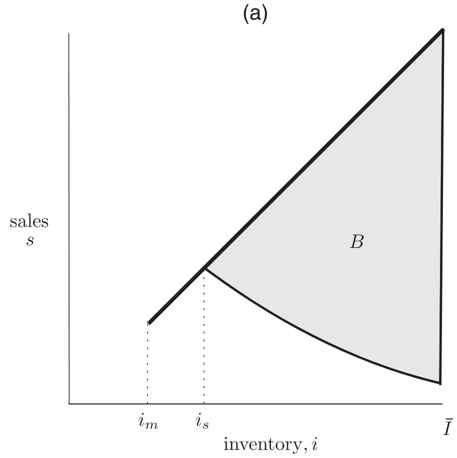
Proposition 3. *In every optimal compensation plan w^* of the form (11), we have $i_s^*\geq i_m^*$.*

Figure 1. Illustrations of the Sail and Mast Regions

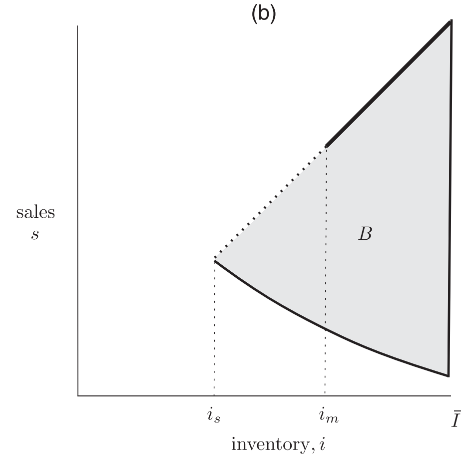


The "sail" bonus region B^{NSO}. The "mast" bonus region B^{SO}.

Figure 2. Two Possible Structures of the Union of the Mast and Sail Bonus Regions



A "mast-and-sail" bonus region.



A non-closed bonus region.

5.3. Are Marketing and Operational Outcomes Complements or Substitutes?

Mast-and-sail compensation plans have several interesting properties. We discuss a key one (nonmonotonicity) in the next section. Here we examine whether operational and marketing outcomes act as complements or substitutes.

Proposition 4. *As i increases,* (a) *if $i_m \le i < i_s$, the minimum sales quantity required for the store manager to qualify for the bonus strictly increases ("moving up the mast"), and* (b) *if $i \ge i_s$, the minimum sales quantity $s^*(i)$ required for the store manager to qualify for the bonus decreases ("slipping down the sail").*

Proposition 4(a) reveals that in the mast part of the bonus region (i.e., the region with $i < i_s$), which corresponds to stockout scenarios, a high realized inventory level has to be accompanied by a high sales outcome. The complementarity in the compensation plan takes such an extreme form that the sales threshold is exactly equal to the inventory outcome. Because the inventory is not sufficiently high, the firm expects the agent to generate a high enough demand to clear all the inventory to demonstrate that the agent has exerted sufficient marketing effort.

In the sail part of the bonus region (i.e., the region with $i \ge i_s$), as inventory increases, the minimum sales quantity to receive the bonus *decreases*. Intuitively, if inventory is high, one might expect the firm to only reward *higher* marketing effort to clear the inventory. Slipping down the sail seems to suggest that lower marketing efforts are also tolerated, precisely when the store has a lot of inventory. Might this realization send the wrong signal to store managers, namely that they can slack off in marketing when they keep a lot of inventory in the store? Can slipping down the sail induce a "slipping" in marketing effort?

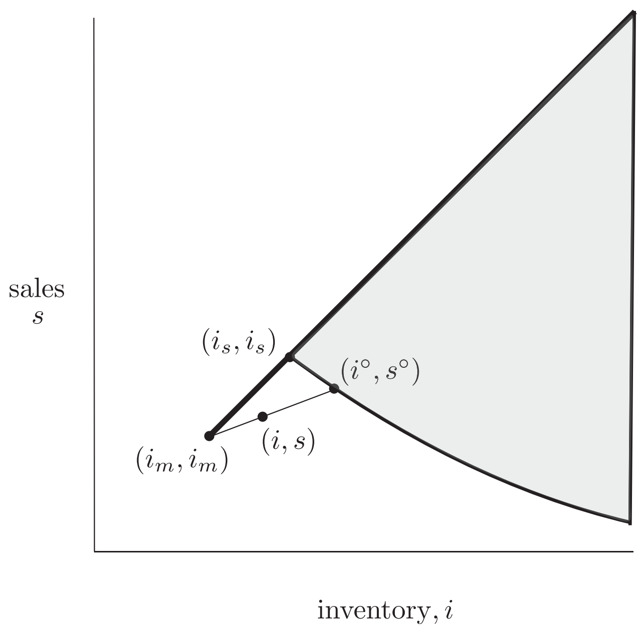
To see that this is not the case, slipping down the sail occurs only when inventory meets the minimum threshold i_s, indicating a sufficiently high likelihood that a significant operational effort has been invested. Slipping down the sail is not meant as an enticement for low marketing effort; rather, it comes as an acknowledgment that high marketing effort may still result in low demand, and because inventory effort is already likely to be high, such unlucky outcomes should not be overly penalized. An upward-sloping sail heightens penalties for unlucky marketing outcomes, which for an agent who has already invested significant operational effort is a deterrent for investing even in the marketing effort needed to clear inventory.

More technically, the optimality of slipping down the sail is connected to the MLRP. The MLRP suggests that the informative value of the observed signal (i, s) increases in both i and s. Note that the sail part corresponds to the scenarios without stockouts, so the true demand is equal to the observed sales quantity. Thus, the firm can infer the same likelihood that the store manager has exerted both operational and marketing efforts based on either (1) a low inventory level and a high demand outcome or (2) a high inventory level and a lower demand outcome. For this reason, operations and marketing act as substitutes in the optimal compensation plan.

6. Ex Post Moral Hazard in Mast-and-Sail Compensation Plans

In this section, we explore the monotonicity of the mast-and-sail structure. To make things concrete, and because a compensation plan has two arguments (i and s), we start with carefully defining monotonicity.

Figure 3. An Illustration of the Lack of Joint Monotonicity of an Optimal Compensation Plan



We say that $w(i,s)$ is *monotone in i* if $w(i',s) \le w(i'',s)$ for every $(i',s),(i'',s) \in D$ and $i' \le i''$. Similarly, we say that $w(i,s)$ is *monotone in s* if $w(i,s') \le w(i,s'')$ for every $(i,s'),(i,s'') \in D$ and $s' \le s''$. Finally, we say that $w(i,s)$ is (strictly) *jointly monotone* if $w(i',s') \le w(i'',s'')$ for all $(i',s'),(i'',s'') \in D$ with $i' < i''$ and $s' < s''$.[8]

Proposition 5. *For an optimal compensation plan w^* of the form* (11), (a) *w^* is monotone in s*, (b) *if $i_s = i_m$, w^* is also monotone in i and jointly monotone, and* (c) *if $i_s > i_m$, w^* is neither monotone in i nor jointly monotone.*

Figure 3 provides the intuition for this result. Proposition 5(a) concerns monotonicity in the vertical direction, which clearly holds in the figure because we never move beyond the 45° line in the vertical direction. Proposition 5(b) concerns monotonicity in the horizontal direction. Moving a short distance horizontally from the bottom corner (i_m, i_m) of the mast drops the store manager's compensation from having the bonus to losing the bonus. Lastly, Proposition 5(c) concerns moving northeast in the graph. As shown in Figure 3, a move from the corner (i_m, i_m) of the mast to the point (i°, s°), where $i^\circ = i_s + \epsilon$ for some positive ϵ and $s^\circ = s^*(i^\circ)$, again drops the bonus for the store manager.

As discussed at length in the Introduction, to say an optimal compensation plan is not monotone in every sense is somewhat nonintuitive. Indeed, as seen in Figure 3, the store manager could be worse off for achieving strictly better inventory and sales outcomes. When stockout occurs along the mast part of the bonus region, the realized demand is censored by the inventory level. The firm's observed sales quantity is only a *lower bound* of the realized demand. The store manager might have made significant marketing effort that realized in a high demand level, which (possibly unluckily) available inventory was not able to meet. Given the same sales quantity, as inventory increases, the firm no longer experiences stockouts. The observed sales quantity is *equal to* (as opposed to a lower bound of) the realized demand. Thus, an increased realized inventory level may be informative of the fact that the store manager *has not* exerted high marketing effort. In other words, to encourage greater marketing effort, the firm is rewarding the possibility of a high demand realization when inventory stocks out. When the uncertainty surrounding realized demand (as opposed to sales) is removed, better performance is required to warrant the bonus.

Interesting as the preceding nonmonotonicity property is, it may raise implementability concerns. If inventory can be hidden from the principal ex post, which does not seem entirely inconceivable, the compensation plan becomes faulty. To see this possibility concretely, suppose that a store manager realizes inventory and sales (i', s') with $i' > i_m$ but does not receive a bonus. This scenario occurs, for instance, when $i' \in (i_m, i_s)$ and $i' < s' < s^*(i')$. If the store manager could hide some of the realized inventory (or claim that it was "shrunk") to reveal an output of (i', i'), he or she would receive a bonus. In other words, the store manager is effectively rewarded for disposing of inventory, meaning that the ex post moral hazard issue is inherent with contracts that are nonmonotone in inventory.

In practice, inventory can neither be freely hidden nor perfectly observed. The ex post moral hazard problem of manipulating inventory is thus similar to *costly state falsification* problems studied in the accounting and economics literature (see, e.g., Lacker and Weinberg 1989, Beyer et al. 2014). Monitoring ex post manipulation by the firm is itself limited, and penalties are difficult to enforce. Indeed, the setup of our problem supposes that the marketing and operational efforts of the store manager are not observable to the firm. This setup suggests that monitoring of inventory is limited for the same reasons. To the extent that a careful accounting of realized inventory and assessment of store manager effort are confounded, the nonmonotonicity of the mast-and-sail compensation plan is an endemic issue. If, under some method, inventory realizations can be observed and operational effort remains hidden, the nonmonotonicity of the mast-and-sail compensation plan is less of a concern. However, we are unaware of such tools for use in practice.

In the rest of this paper, we search for alternate compensation plans that resolve the ex post moral hazard issue of inventory manipulation.

7. Monotone (But Not Optimal) Compensation Plans

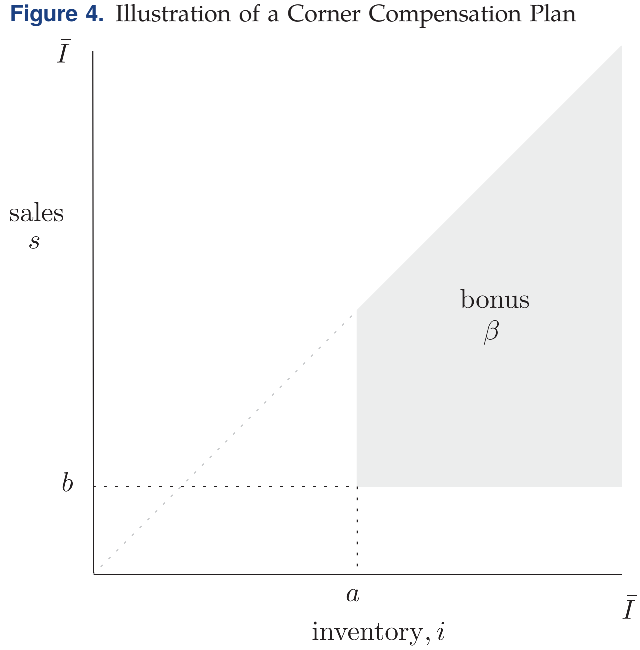
In the preceding two sections, we have characterized an optimal mast-and-sail compensation plan, but these compensation plans suffer from nonmonotonicity, limiting their practicality in the presence of ex post moral hazard over the hiding of inventory.[9] This issue motivates interest in exploring the performance of classes of implementable compensation plans that are monotone.

There are two natural candidates for implementable compensation plans. The first is a bonus compensation plan where a bonus is given if both a sales and inventory quota are met, termed a *corner* compensation plan in the Introduction. The other candidate is a modification of the mast-and-sail plan by snipping the mast to remove the nonconvexity of the bonus region (and thus ensuring monotonicity) and linearizing the downward-sloping s^* function defining the sail in Proposition 1. Our goal in studying these two candidate solutions is to assess the extent of optimality loss associated with monotonicity using the optimal mast-and-sail compensation plan as the benchmark.

7.1. Corner Compensation Plans

Corner compensation plans build on the logic of the quota-bonus compensation plans that are optimal in the risk-neutral setting in the salesforce compensation literature (Oyer 2000; Dai and Jerath 2013, 2016). A corner compensation plan (a, b) is one where any outcomes (i, s) with $i \geq a$ *and* $s \geq b$ earn a bonus $0 \leq \beta \leq \bar{w}$. See Figure 4 for an illustration.

Figure 4. Illustration of a Corner Compensation Plan



The next result shows that the mast-and-sail compensation plans often perform *strictly* better than corner compensation plans. We say that f and g satisfy the *strict* MLRP; that is, (1) holds with weak inequalities replaced by strict inequalities. Many commonly studied families of distributions (e.g., binomial, exponential, lognormal, normal, and Poisson) satisfy the strict MLRP.

Proposition 6. *Given a multitasking store manager problem described in Section 5.2, with the further restriction that f and g satisfy the strict MLRP and the agent earns positive rents, an optimal compensation plan cannot be a corner compensation plan.*

Assuming positive rents for the agent is common in the literature (e.g., Oyer 2000, Dai and Jerath 2013). The situation in which the agent earns no rents yields a first-best contract whereby the incentive issue does not have any "bite" and is thus less interesting as an incentive problem. Using similar reasoning as the proof of this proposition, one can show that the best corner compensation plan with bonus $\bar{w}$ outperforms every other corner compensation plan. Accordingly, we focus on corner compensation plans with bonus $\bar{w}$. Moreover, observe that compensation plans rewarding sales *only* are achieved by setting $a = b$, and those rewarding inventory *only* are achieved by setting $b = 0$. Single-tasking compensation plans are special cases of corner compensation plans and so are (weakly) dominated by the optimal corner compensation plan.

Additional analytical performance bounds are hard to come by, in no small part because of the challenging nature of computing the parameters of the optimal compensation plan. The difficulty is that the weights ω_i and t in (12) and (13) must be computed to get a sense of the shape of the mast and sail. Problem (8) and Theorem 3 provide our best hope for computing ω_i and t in general. However, (8) is a challenging optimization problem and, to our knowledge, does not readily admit analytical characterizations that can be used to provide bounds. For this reason, we primarily use numerical calculations to further compare various compensation plans.

To numerically quantify the performance loss of corner compensation plans, we need to describe the structure of an *optimal* corner compensation plan. Luckily, the analysis under a corner compensation problem greatly simplifies, as evidenced by the following simple result.

Proposition 7. *The expected wage payout of the corner compensation plan (a, b) is $\bar{w}(1 - F(a|e_o^*)(1 - G(b|e_m^*)$, where $(1 - F(a|e_o^*)(1 - G(b|e_m^*)$ is the probability of paying out the bonus and where (e_o^*, e_m^*) is the target effort level to be implemented.*

Given this characterization of expected wage payout, problem (2) evaluated at the corner compensation plan (a, b) becomes (after some basic simplifications)

$$\max_{a,b,a\geq b} \quad r\mathbb{E}[S|e_o^H, e_m^H] - \bar{w}\left[1-F\left(a|e_o^H\right)\right]\left[1-G\left(b|e_m^H\right)\right] \tag{14a}$$

$$\text{s.t.} \quad \left[1-F\left(a|e_o^H\right)\right]\left[1-G\left(b|e_m^H\right)\right] - \left[1-F\left(a|e_o^L\right)\right]\left[1-G\left(b|e_m^L\right)\right] \geq \frac{c\left(e_o^H, e_m^H\right) - c\left(e_o^L, e_m^L\right)}{\bar{w}}, \tag{14b}$$

$$\left[F\left(a|e_o^L\right) - F\left(a|e_o^H\right)\right]\left[1-G\left(b|e_m^H\right)\right] \geq \frac{c\left(e_o^H, e_m^H\right) - c\left(e_o^H, e_m^L\right)}{\bar{w}}, \tag{14c}$$

$$\left[1-F\left(a|e_o^H\right)\right]\left[G\left(b|e_m^L\right) - G\left(b|e_m^H\right)\right] \geq \frac{c\left(e_o^H, e_m^H\right) - c\left(e_o^L, e_m^H\right)}{\bar{w}}, \tag{14d}$$

assuming that we look at the setting where the store manager earns positive rents (as discussed after Proposition 6). Optimal solutions to (14) are relatively easy to characterize, depending on which of the constraints are slack or tight. The next result follows from this reasoning.

Proposition 8. *If constraint (14b) is tight at optimality, the optimal corner compensation plan (a, b) has a and b satisfy the equation*

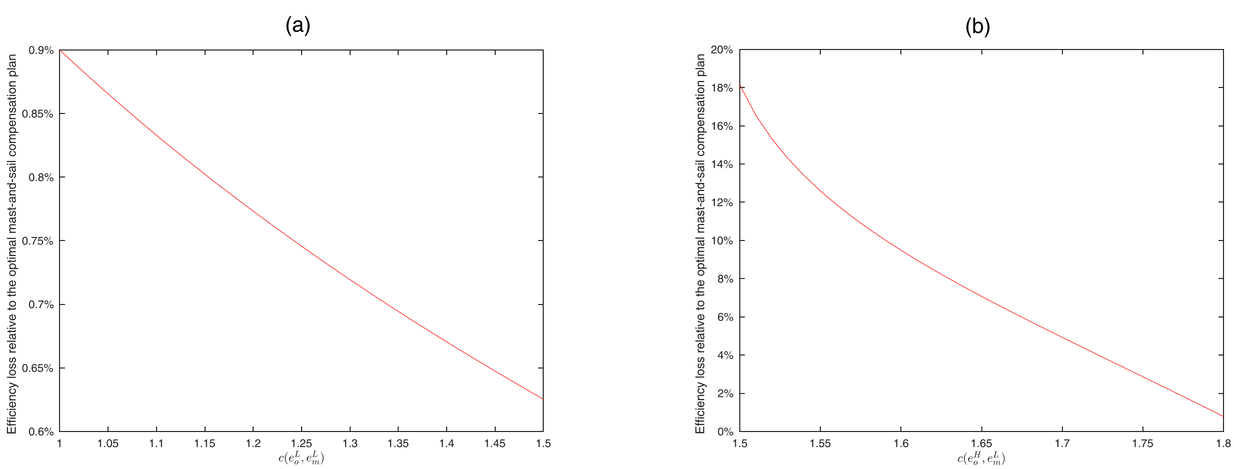
$$\frac{\mathcal{H}^f\left(a|e_o^H\right)}{\mathcal{H}^g\left(b|e_m^H\right)} = \frac{\mathcal{H}^f\left(a|e_o^L\right)}{\mathcal{H}^g\left(b|e_m^L\right)}, \tag{15}$$

where $\mathcal{H}^f(a|e_o^*) = \frac{f(a|e_o^*)}{1-F(a|e_o^*)}$ is the hazard rate for density f, and $\mathcal{H}^g$ is the hazard rate of density g. By contrast, if (14b) does not bind, then $a = b$, where a is characterized by setting either constraint (14c) or constraint (14d) to be tight.

This structure assists us in running numerical experiments to evaluate the performance of optimal corner compensation plans. For an illustration of how to use these results, see a concrete numerical example in Section OA.3 of the online appendix. Here we present two representative and *contrasting* scenarios in Figure 5, (a) and (b). Figure 5(a) shows that the performance of the corner compensation plan is close to optimal (within 1%) when the marketing and operational activities are highly complementary in terms of the agent's cost structure. By contrast, Figure 5(a) shows that when the marketing and operational activities are not sufficiently complementary, the performance of the corner compensation plan is far from optimal, with a gap of up to 18%.

In certain cases, the corner compensation plan fails to induce the target action achievable under the optimal compensation plan. Example 1 provides one such example. Although the corner compensation plan may lead to a lower expected compensation than the optimal compensation plan, the firm's expected sales quantity is also lower because of the store manager's lower effort than the desired one. Thus, under a sufficiently high unit revenue (so that the target action entails high effort in both operational and marketing activities), the firm's expected profit is higher under the mast-and-sail compensation plan. Indeed, for this type of scenario, we can show that the efficiency loss under the corner compensation plan increases linearly in the unit revenue. In other words,

Figure 5. (Color online) Performance of Optimal Corner Compensation Plan vs. Optimal Mast-and-Sail Compensation Plan



Parameters: $\bar{w} = 10$, $c(e_o^H, e_m^L) = c(e_o^L, e_m^H) = 3.0$, $c(e_o^H, e_m^H) = 3.5$, $c(e_o^L, e_m^L)$ between 1 and 1.5.

Parameters: $\bar{w} = 10$, $c(e_o^L, e_m^L) = 1$, $c(e_o^L, e_m^H) = 1.8$, $c(e_o^H, e_m^H) = 3.5$, $c(e_o^H, e_m^L)$ between 1.5 and 1.8.

Note. We assume that $F(i|e_o) = (H(i))^{e_o}$ and $G(s|e_m) = (L(s))^{e_m}$, where $H(i) = i$ and $L(s) = s$.

the worst-case loss in performance of corner compensation plans is *arbitrarily* large.

Example 1. Consider the following instance in which $e_o \in \{e_o^L, e_o^H\}$ and $e_m \in \{e_m^L, e_m^H\}$, where $e_o^L = e_m^L = 1$ and $e_o^H = e_m^H = 2$. The target action is $(e_o^H, e_m^H) = (2, 2)$. The cost function is $c(e_o^H, e_m^H) = 3.1$, $c(e_o^H, e_m^L) = 1$, $c(e_o^L, e_m^H) = 1.6$, and $c(e_o^L, e_m^L) = 0.1$. The resource constraint for the firm is $\bar{w} = 10$. For this instance, we can show that the firm can use a mast-and-sail compensation plan with $\omega^*_{e_o^L, e_m^L} = 0, \omega^*_{e_o^H, e_m^L} = 0.8602$, $\omega^*_{e_o^L, e_m^H} = 0.1398$, and $t^* = 0.1817$ to induce the target action, under which the store manager's probability of receiving the bonus is 58.70%. However, no corner compensation plan exists that can induce the target action. Indeed, the best that the corner compensation plan can achieve is to induce (e_o^H, e_m^L) with parameters of $a^* = b^* = 0.6186$, under which the store manager's probability of receiving the bonus is 23.55%. We illustrate the firm's expected profits under both types of compensation plans in Figure 6 as a function of the per-unit revenue rate r.

7.2. Modifying Mast-and-Sail Compensation Plans for Implementability

In Section 7.1, we used a *single-tasking logic* to construct and evaluate corner compensation plans, with the sales-quota-bonus compensation plan being the simplest case, and found that its performance depends on the store manager's cost structure and can be far from optimal. We now switch gears to using our mast-and-sail compensation plan as inspiration for designing ex post implementable compensation plans. We do so in two directions: (1) removing the mast (i.e., setting $i_m = i_s$) and (2) linearizing the downward-sloping function s^* in Proposition 1. Removing the mast ensures monotonicity of the compensation plan, and linearizing s^* makes communicating the compensation plan to sales managers easier in practice.[10]

Figure 6. (Color online) Expected Profits Under Optimal the Mast-and-Sail and Corner Compensation Plans vs. Revenue Rate r

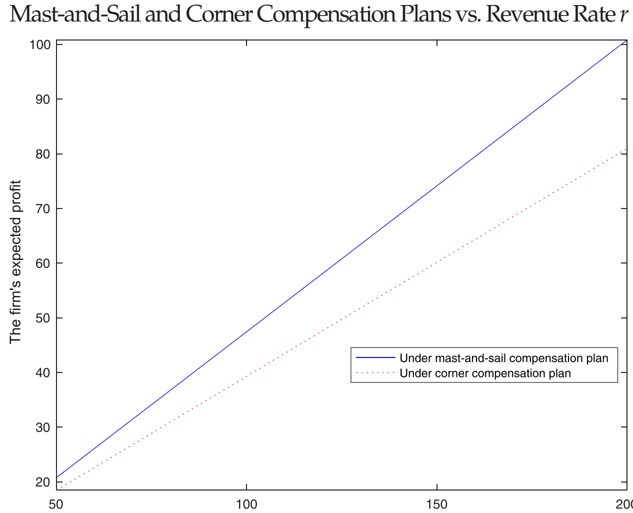
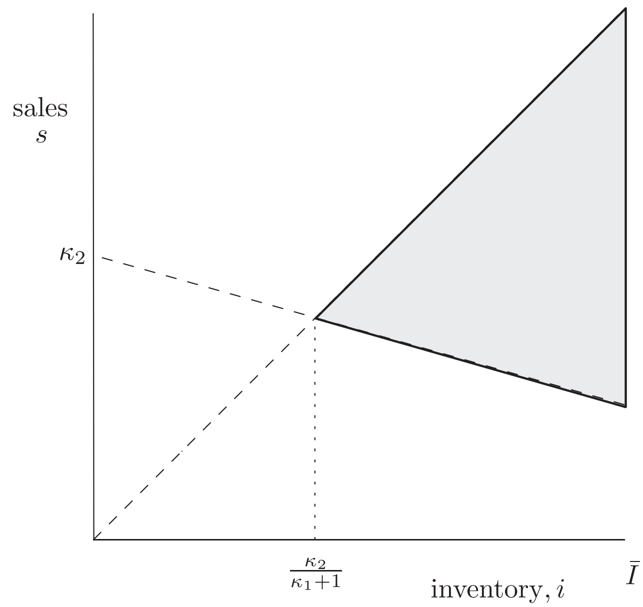


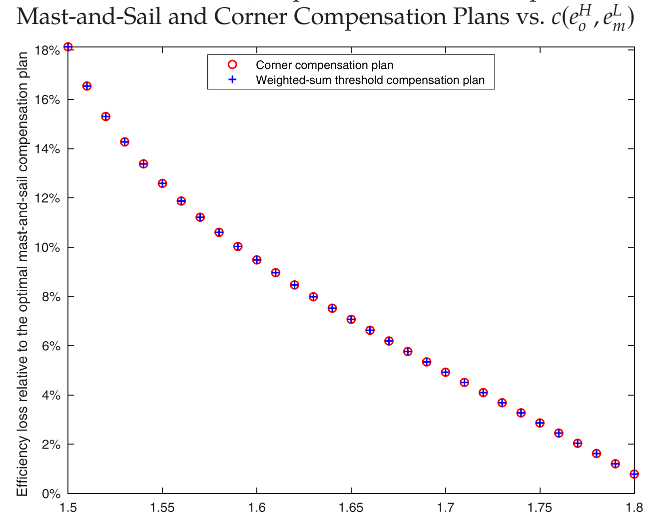
Figure 7. An Illustration of a Weighted-Sum-Threshold Compensation Plan



Together this effort amounts to finding the best compensation plan where the bonus region is characterized by a downward-sloping line (creating a "triangular sail" like that in Figure 7). We call such compensation plans *weighted-sum threshold compensation plans* because the payout of the bonus is determined by the weighted sum of the sales quantity and inventory level. Specifically, the agent receives a bonus if the realized sales quantity s and inventory level i satisfy $s + \kappa_1 \cdot i \geq \kappa_2$, for some $\kappa_1, \kappa_2 \geq 0$.[11] To find the optimal weighted-sum threshold compensation plan, one searches over the values of κ_1 and κ_2 that give the best payoff to the firm.

Numerical results (see Figure 8) show that the optimal weighted-sum threshold compensation plans

Figure 8. (Color online) Expected Profits Under Optimal the Mast-and-Sail and Corner Compensation Plans vs. $c(e_o^H, e_m^L)$



also perform poorly (indeed, as poorly as corner compensation plans) in bad cases (losses of up to 18% in this example). We conclude that the *loss due to monotonicity* captured by the ex post moral hazard issue that afflicts mast-and-sail contracts has no easy fix. The next section shows, however, with some additional information, that this issue can be resolved.

8. Resolving Ex Post Moral Hazard Through Gauging Unsatisfied Demand[12]

In previous sections, we assumed that any demand in excess of inventory could not be observed. We now consider a more general setting where partial information is revealed when demand exceeds sales. In particular, we assume that some random (and unknown) fraction of customers who do not receive the product express interest via a waiting list (or some other method of capturing unsatisfied demand). We introduce a new random variable[13]

$$\Theta := \Lambda(Q - I) + I, \tag{16}$$

where Λ is a continuous random variable distributed on $(0, 1)$. When Q, I, and Λ realize outcome (q, i, λ), where $q > i$, λ can be interpreted as the fraction of customers who sign the waiting list when facing stockout. We call Λ the *random fraction of captured demand* (or simply the *fraction*).

We define a new random variable

$$Z := Z(I, Q, \Theta) = \begin{cases} Q & \text{if } Q \le I \\ \Theta & \text{if } Q > I \end{cases} \tag{17}$$

that captures what demand information can be observed. We cannot observe Q when $Q > I$, and we cannot observe Θ when $Q \le I$. We assume that the conditional density function $\gamma(\theta|q, i)$ for all q, i such that $q > i$ is known to both the firm and the store manager. This assumption amounts to knowing the probability density function φ of the fraction Λ, because in this case $\gamma(\theta|q, i) = \frac{1}{q-i}\varphi(\frac{\theta-i}{q-i})$.

The firm and store manager observe I and Z. When $Z = \Theta$, the product $\Lambda(Q - I)$ can be observed (because I is also observable), but knowledge of this product does not reveal Λ or Q directly. That is, the proportion of unsatisfied customers who sign up for the waiting list is not observable. The derived signal Z captures intermediate degrees of demand censoring. In the classical censoring case, Θ is precisely I with Λ a constant at zero. Accordingly, Z becomes the random variable S studied in earlier sections. Similarly, the situation where demand is observed sets $\Theta = Q$ with Λ a constant at one; this full-information case is explored in Section OA.4 of the online appendix.

Because the firm and the store manager observe I and Z, the compensation plan w is a function of I and Z. The same bang-bang methodology applies to this new setting because the underlying problem remains linear in w. The optimal compensation plan is therefore defined by characterizing its bonus region where the store manager receives $\bar{w}$ using the methodology of Section 4.

To get a sense of the bonus region, we need to understand the domain of w. According to (17), two regions of the domain need to be considered: (1) the *no lost sales* (NLS) region (where $Z \le I$) and (2) the *lost-sales* (LS) region (where $Z > I$). As before, we may construct the bonus region in the two "chunks" of the underlying domain, the NLS region and the LS region. The joint density function of (I, Z) can be expressed over these two regions as follows:

$$h(i, z|e_o, e_m) := \begin{cases} f(i|e_o)g(z|e_m) & \text{if } z \le i, \\ \displaystyle\int_{q=i}^{\bar{Q}} \gamma(z|q, i)g(q|e_m)\, dq f(i|e_o) & \text{if } z > i. \end{cases} \tag{18}$$

Using the joint density function in (18), we define the bonus regions in the NLS and LS regions in terms of likelihood-ratio functions as follows:

$$R^{\text{NLS}}_{e_o, e_m}(i, z) = 1 - \frac{g(z|e_m)}{g(z|e_m^*)} \cdot \frac{f(i|e_o)}{f(i|e_o^*)}, \tag{19}$$

$$R^{\text{LS}}_{e_o, e_m}(i, z) = 1 - \underbrace{\frac{\int_{q=i}^{\bar{Q}} \gamma(z|q, i)g(q|e_m)dq}{\int_{q=i}^{\bar{Q}} \gamma(z|q, i)g(q|e_m^*)dq}}_{(*)} \cdot \frac{f(i|e_o)}{f(i|e_o^*)}, \tag{20}$$

where e_o^* and e_m^* are the target effort levels.

The NLS bonus region has a structure analogous to the NSO bonus region described in Proposition 1. When $z \le i$, we have $z = q = s$, and the bonus region is precisely $B^{\text{NLS}} = \{(i, s) : i \ge i_s \text{ and } s^*(i) \le s \le i\}$. The LS region is more complex because of the dependence of Θ on both I and Q. Further assumptions are required to derive an interpretable structure.

Assumption 1. *The random variable Θ defined in (16) with conditional density function $\gamma(\theta|q, i)$ is such that $\frac{\partial \log \gamma(\theta|q,i)}{\partial q}$ is nondecreasing in θ for every i.*

This assumption is the MLRP of Θ with respect to changes in q, given every inventory realization i. As we have said before, this assumption is common in the contract theory literature and is satisfied when, for example, Λ is uniformly distributed (among other distributions).

Proposition 9. *Under Assumption 1, we have the following cases:*

a. *If* $\int_i^{\bar{Q}} \gamma(z|q,i)g(q|e_m)dq$ *is unbounded for all i, then*

$$B^{LS} = \{(i,z) : i \geq i_s \text{ and } z > i\}, \text{ and} \tag{21}$$

b. *There exists a continuous function* ℓ^* *defined on* $[0, i_s)$ *such that*[14]

$$B^{LS} = \{(i,z) : i < i_s \text{ and } \ell^*(i) \leq z \leq \bar{Q}\} \cup \{(i,z) : i \geq i_s \text{ and } i < z \leq \bar{Q}\}. \tag{22}$$

An important fact used in this proof is that $\Lambda \in (0,1)$. We can infer that $Q = I$ (i.e., stockout occurs) when $Z = I$ because this implies that $\Lambda(Q - I) = 0$ and Λ never realizes to zero. In other words, when the signal Θ is equal to I, no demand is lost. This ensures that the LS and NLS bonus regions meet at the same point on the 45° line (i_s).

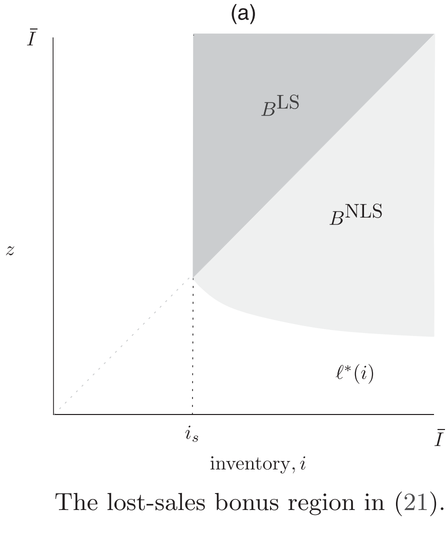
Figure 9 gives a visualization of the bonus regions B^{LS} and B^{NLS} in cases (a) and (b) of Proposition 9. Figure 9(a) was generated assuming that Λ is a power-law distribution with cumulative distribution function λ^α, where $\alpha \leq 1$. It is straightforward that this class of distributions (which includes the uniform distribution) satisfies the conditions of Proposition 9(a). We generated Figure 9(b) by considering the case in which Λ was distributed so that $-\log \Lambda$ is a Gamma(α, β) distribution. This ensures that $\Lambda \in (0,1)$. It is straightforward to check that the conditions of Proposition 9(b) are satisfied for this setting. To generate the figure, we took $\alpha = 2$ and $\beta = 1/2$.

The two different cases for the bonus region have interesting implications for the question of monotonicity and ex post moral hazard. Observe that in Figure 9(a), the bonus region is jointly monotone in z and i. This case is not trivial because it includes the uniform distribution and general power-law distributions for the fraction Λ of captured demand. The intuition here is that by including waiting list information, the old mast region disappears. The waiting list reveals sufficient information about marketing effort so that it is no longer necessary to reward low sales outcomes. The essential message here is that the bonus region is now monotone; therefore, the ex post moral hazard issue of hiding inventory no longer exists. Indeed, the feasible region B^{LS} is monotone in z, so even if we allow *downward* manipulation of the waiting list, there is no incentive to do so.

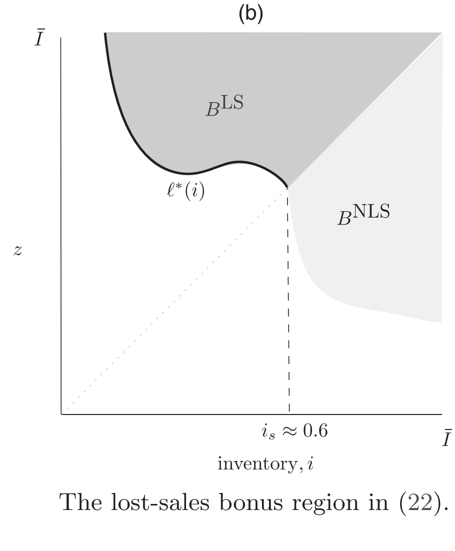
By contrast, in the bonus region in Figure 9(b), the waiting-list signal is sufficiently correlated with marketing effort to offer bonuses if the waiting list is sufficiently large, even when the sales target i_s is not met. This scenario may lead to bonus regions that are *not* jointly monotone in z and i, a result coming from the fact that Assumption 1 does not require monotonicity properties of Θ with respect to changes in i. It is important to point out that it does, nonetheless, preclude any ex post moral hazard issue of hiding inventory. In the NSO region of the domain ($s < i$), the store manager has no incentive to hide inventory because of monotonicity in i: the bonus region is monotone below the 45° line. The area above the 45° line captures scenarios with lost sales, so no inventory is left to hide. Assuming also that the signal Θ cannot be manipulated by the store manager (e.g., the waiting list captures that the unique identity of customers can be directly verified by the firm), no ex post moral hazard arises.

Of course, one may still argue that compensation plans with bonus regions such as in Figure 9(b) are not completely intuitive. In practice, a store manager might wonder why a lower level of sales requires a *smaller* waiting-list signal to get a bonus, whereas a higher sales level requires a *larger* bonus. The fact that

Figure 9. Illustrations of Lost-Sales Bonus Region B^{LS}



The lost-sales bonus region in (21).



The lost-sales bonus region in (22).

no scope exists for ex post manipulation does not change the fact that it could be hard to explain such compensation plans to store managers. This lack of joint monotonicity above the 45° line can be removed under the following assumption.

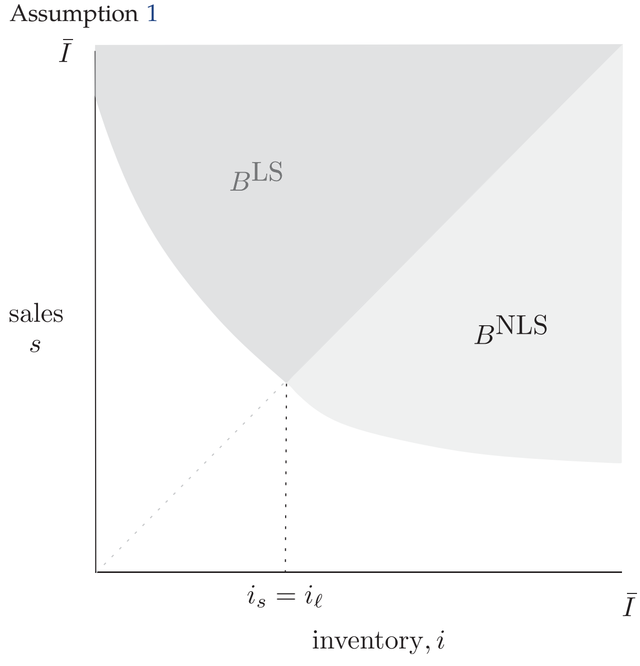
Assumption 2. *The random variable Θ defined in (16) with conditional density function $\gamma(\theta|q,i)$ is such that $\frac{\partial \log \gamma(\theta|q,i)}{\partial q}$ is nondecreasing in i for every θ.*

Proposition 10. *Under Assumptions 1 and 2, a continuous and nonincreasing function $\ell^*(i)$ exists such that $B^{LS} = \{(i,z): i \le i_s \text{ and } \ell^*(i) \le z \le \bar{Q}\} \cup \{(i,z): i \ge i_s \text{ and } i \le z \le \bar{Q}\}$, implying that $B^{NLS} \cup B^{LS}$ has the double-sail structure depicted in Figure 10.*

It is straightforward to see that the resulting optimal contract $w^*(i,z)$, where $w^*(i,z) = \bar{w}$ when $(i,z) \in B^{\text{NLS}} \cup B^{\text{NLS}}$ and zero otherwise, is jointly monotone. In other words, under the waiting-list approach for gauging unsatisfied demand (and given the preceding technical conditions), an optimal double-sail compensation plan exists that is monotone. This avoids the ex post moral hazard hiding of inventory that afflicted the mast-and-sail compensation plan and has a more intuitive structure than what we see in Figure 9(b).

One may ask how restrictive Assumptions 1 and 2 are on the distribution of the signal Θ. Note that distributions that satisfy the condition of Proposition 9(a) fail this condition but nonetheless give rise to jointly monotone bonus regions. We saw in Figure 9 that gamma distributions can give rise to scenarios with nonmonotone lost-sales bonus regions. However, one can also check that nonmonotonicity does not hold for all parameter values. Indeed, an algebra exercise can verify that when Λ is such that $-\log \Lambda$ is a Gamma(α,β) distribution, where $\alpha - 1 \ge \frac{4}{e^2+1}(\frac{1}{\beta}+1)$, Assumptions 1 and 2 hold, and the resulting contract is monotone (by Proposition 10).

Figure 10. The Double-Sail Bonus Region $B^{\text{NLS}} \cup B^{\text{LS}}$ Under Assumption 1



9. Further Discussions

In this section, we include some additional discussion on the flexibility of our analytical framework. In particular, we are able to relax some of the assumptions of the base model that were included for ease of discussion and presentation. Although not central to our managerial takeaways regarding the connection between demand censoring, nonmonotonicity, and ex post moral hazard, we nonetheless consider these extensions worthy of further discussion.

9.1. The Role of $\bar{w}$ and More General Resource Constraints

The role of the upper bound $\bar{w}$ on compensation is a delicate one. As mentioned in Section 3, the assumption is not uncommon in the literature and has been justified elsewhere. However, because $\bar{w}$ is exogenous to the model, a question remains as to how to interpret it. Can the firm set $\bar{w}$? If so, how high or low should it be set? How does $\bar{w}$ change the optimal compensation plan?

Changing $\bar{w}$ does not change the optimality of the mast-and-sail compensation plan but may change the relative length of the mast and shape of the sail and the probability of the agent receiving the bonus under the target actions. If we view $\bar{w}$ purely as a choice of the firm and consider its optimization over the choice of $\bar{w}$, larger choices of $\bar{w}$ are obviously better. Indeed, $\bar{w}$ only enters in the constraint $w(i,s) \le \bar{w}$, so increasing $\bar{w}$ can only improve the objective value of the firm. This slope is a slippery one. If the choice of $\bar{w}$ is unconstrained, it will be sent to infinity. When $\bar{w} = +\infty$, an optimal compensation plan need not exist in general. This issue is discussed at length in the economics literature (see, e.g., Chu and Sappington 2009). We find it natural that, in practice, a natural upper bound for $\bar{w}$ would exist that avoids this theoretical issue. One possible justification is provided in Section OA.8 of the online appendix.

We consider here a more general upper bound than $\bar{w}$. Let $m(i,s)$ be the available resources for compensation by the firm when outcome (i,s) prevails. That is, constraint $w(i,s) \le \bar{w}$ is replaced by constraint $w(i,s) \le m(i,s)$ for almost all (i,s). As an example of $m(i,s)$, consider the following description from DeHoratius and Raman (2007, p. 521): "BMS [the company they study] store managers were offered a bonus for generating sales that ranged from 0.2% to 5% of the sales dollars above store-specific targets." In this case, $m(i,s)$ is a fixed proportion of the store revenue less a store-specific target. In other words, $m(i,s)=\alpha\cdot r\cdot s-C$, where r is the per-unit revenue, and C denotes the store-specific target. The range of α in this case is from 0.2% to 5%.

Our model can be adjusted to the setting with resource constraint $w(i,s)\le m(i,s)$, assuming that $m(i,s)$ is an L^1 function. Define a new variable $\beta(i,s)$, where $w(i,s)=\beta(i,s)m(i,s)$ and $\beta(i,s)\in[0,1]$ for almost all (i,s). The new function β can be interpreted as the percentage of the resource given to the store manager as a bonus. The problem becomes

$$\max_{\beta} \quad r\int_i\int_s sf(i|e_o^*)g(s|e_m^*)dids - \int_i\int_s \beta(i,s)\,m(i,s)f(i|e_o^*)g(s|e_m^*)dids \tag{23a}$$

$$\text{s.t.} \quad \int_i\int_s \beta(i,s)m(i,s)f(i|e_o^*)g(s|e_m^*)dids - c(e_o^*,e_m^*)\ge \underline{U}, \tag{23b}$$

$$\int_i\int_s \beta(i,s)m(i,s)f(i|e_o^*)g(s|e_m^*)dids - \int_i\int_s \beta(i,s)m(i,s)f(i|e_o)g(s|e_m)dids \ge c(e_o^*,e_m^*)-c(e_o,e_m) \text{ for all } (e_o,e_m) \tag{23c}$$

$$0\le\beta(i,s)\le 1 \text{ for all } (i,s). \tag{23d}$$

This problem is of the form (4), interpreting $f(\bar{x}|\bar{a}^*)$ in that formulation as $m(i,s)f(i|e_o)g(s|e_m)$. Thus, an optimal bang-bang contract exists for (23) with a similarly nice structure.

Proposition 11. *If m is an L^1 function, nonnegative multipliers ω_i and a target t exist such that an optimal solution to (23) of the following form exists:*[15]

$$w^*(i,s)=\begin{cases} m(i,s) & \text{if } \sum_{e_o,e_m}\omega_{e_o,e_m}R_{e_o,e_m}(i,s)\ge t,\\ 0 & \text{otherwise,}\end{cases}$$

where

$$R_{e_o,e_m}(i,s) = 1-\frac{\mathbb{I}[i>s]f(i|e_o)g(s|e_m)+\delta(i=s)f(i|e_o)(1-G(i|e_m))}{\mathbb{I}[i>s]f(i|e_o^*)g(s|e_m^*)+\delta(i=s)f(i|e_o^*)(1-G(i|e_m^*))}.$$

Under the compensation plan specified by Proposition 11, the store manager's compensation is monotone nondecreasing in s as long as $m(i,s)$ does not decrease in s.

9.2. Endogenizing Initial Inventory

In this section, we endogenize the choice of initial inventory $\bar{I}$. Whether it is more natural for $\bar{I}$ to be under the control of the firm or the store manager is a matter of debate. In this paper, we analyze the former. This perspective particularly applies in settings where the firm oversees a large chain of stores where ordering is done centrally. Allowing the store manager to decide the initial inventory level gives rise to additional incentive issues that go beyond our scope.

To set the benchmark, suppose that we can ignore the incentive compatibility constraint of the store manager and pay him or her a constant wage to meet the minimum utility $\underline{U}$ to work at effort level (e_o^H,e_m^H). Under this assumption, the firm's problem is $\max_{\bar{I}} r\times \mathbf{E}[S|\bar{I}]-C(\bar{I})$, where $C(\cdot)$ is a convex increasing cost for procuring inventory $\bar{I}$, and $\mathbf{E}[\cdot]$ is the conditional expectation given inventory $\bar{I}$. Thus, the optimal inventory level is the classical newsvendor solution I^{NV} that solves $r\frac{d}{d\bar{I}}\mathbf{E}[S|I^{\mathrm{NV}}]=C'(I^{\mathrm{NV}})$.

As for the second-best compensation plan, where the store manager's incentives must be taken into consideration, the firm's inventory decision becomes

$$\max_{\bar{I}} r\mathbf{E}[S|\bar{I}]-W(\bar{I})-C(\bar{I}), \tag{24}$$

where

$$\begin{aligned} W(\bar{I}) \triangleq \min_{w}\quad & \mathbb{E}\left[w(I,S)|e_o^H,e_m^H\right]\\ \text{s.t.}\quad & \mathbb{E}\left[w(I,S)|e_o^H,e_m^H\right]-c\left(e_o^H,e_m^H\right)\ge\underline{U},\\ & \mathbb{E}\left[w(I,S)|e_o^H,e_m^H\right] - \mathbb{E}\left[w(I,S)|e_o,e_m\right]\ge c\left(e_o^H,e_m^H\right) - c(e_o,e_m) \text{ for all } (e_o,e_m),\\ & 0\le w(i,s)\le\bar{w} \text{ for all } (i,s),\end{aligned}$$

is the optimal value function when (e_o^H,e_m^H) is the target effort level.

Proposition 12.

a. *Under the assumption that (e_o^H,e_m^H) is the target effort level and f and g satisfy the MLRP, we have $\frac{d}{d\bar{I}}W(\bar{I})<0$ for all $\bar{I}$. That is, an increase in $\bar{I}$ leads to a decrease in the expected payout to the store manager.*

b. *The firm's optimal inventory level, by accounting for the multitasking store manager problem, is higher than that in the newsvendor problem, which helps the firm achieve a lower expected payment to the store manager than otherwise.*

The intuition behind Proposition 12(a) is as follows. Because the firm is more likely to pay a bonus if the inventory is cleared (as long as sales are greater than i_m), increasing inventory reduces the chance of inventory clearing and thus the chance of paying out the bonus.

Proposition 12(b) entails optimizing the initial inventory level $\bar{I}$. Let I^* be the optimal inventory choice in (24). The first-order condition of (24) yields the necessary optimality condition for I^*:

$$r\frac{d}{d\bar{I}}\mathbf{E}[S|I^*]-\frac{d}{d\bar{I}}W(I^*)=C'(I^*). \tag{25}$$

In light of (24) and (25) and because $\frac{d}{d\bar{I}}W(I^*)<0$ (by Proposition 12) and C is a convex increasing function, we can conclude that $I^* > I^{NV}$. In other words, the firm overinvests in inventory as compared with the classical newsvendor setting without agency issues. This result echoes the view of Dai and Jerath (2013, 2019) that a higher inventory level mitigates the possibility of demand censoring and hence benefits the firm by reducing the complications in contract design.

10. Conclusion

In this paper, we have examined incentive issues at the intersection of operations and marketing. We show that the censoring of marketing outcomes (i.e., demand censoring) gives rise to a vexing incentive issue of both ex ante and ex post moral hazard. Addressing ex ante moral hazard alone leads to an optimal compensation plan that does not overcome the ex post issue. Only by providing for an additional signal of unsatisfied demand (e.g., via a waiting list) can we construct a compensation plan that both is optimal and resolves the ex ante and ex post moral hazard issues.

Taken together, our research provides a compelling narrative linking customer and employee behavior in the retail setting. Because of its inability to monitor customer intentions (i.e., not observing all of demand), for the firm to design intuitive compensation schemes, it has to monitor employee intentions (i.e., their conscientiousness in sales and operational activities or in accurately representing the level of inventory in the store). In effect, an arm's-length company cannot stop an employee from using its lack of understanding of customer demand to benefit employee compensation. Employees incur a rent from the company's lack of visibility over customers. Only additional information about customer intentions removes this rent-seeking opportunity.

Our novel methodology (i.e., the bang-bang optimal control approach) transcends the limitations of classical solution approaches in contract theory and is applicable to a broad set of incentive design problems. On the application side, we hope that our work will inspire future research into this immensely exciting venue. For example, instead of having a single manager in charge of both operations and marketing, one could imagine two managers, each responsible for one of the tasks. The nature of the relationship between these two managers, and their compensation plans, should also include aspects of customer demand and behavior. Another scenario involves a store manager who has operational responsibilities not only for execution (i.e., maintaining inventory) but also for making operational decisions, such as inventory stocking levels (see, e.g., Sen and Alp 2020). Stockouts—and the associated incentive complications—will be likely even more prevalent in a decentralized supply chain because of double marginalization. An interesting research question would be whether agency issues make optimal order quantities more or less conservative.

Outside of the retail sector, numerous settings involve multitasking between operations and marketing activities. In a global health setting, for example, private agencies are often engaged in delivering and administering vaccines to children residing in some of the hardest-to-reach places in the world. The success of their work depends on not just effective campaigns to raise public awareness of the importance of vaccination (marketing), but also delicately managing a cold chain system essential for the storage and transportation of vaccines from freezer to freezer (operations). Our paper has implications for incentive design problems in those settings.

Acknowledgments

The authors are grateful to the department editor, the associate editor, and three anonymous reviewers for the constructive review process. The authors also appreciate Anil Arya, Maqbool Dada, Nicole DeHoratius, Vishal Gaur, Nagesh Gavirneni, Jonathan Glover, Xiaoyang Long, Suresh Sethi, Christopher S. Tang, Fuqiang Zhang, and Gaoqing Zhang for their valuable suggestions. Seminar participants at Boston College, Cornell University, Duke University, Hong Kong Baptist University, INSEAD, National University of Singapore, Singapore Management University, University of Alberta, University of California Los Angeles, University of Florida, University of Michigan, University of Rochester, University of Texas at Austin, University of Texas at Dallas, University of Virginia, University of Wisconsin–Madison, Wake Forest University, and Washington University in St. Louis provided helpful comments for various versions of this paper. The authors also benefited from interacting with session participants at the 2018 MSOM Conference (Dallas, Texas), the 2019 INFORMS Marketing Science Conference (Rome, Italy), the 2019 POMS International Conference (Brighton, United Kingdom), and the INFORMS 2019 Annual Meeting (Seattle, Washington). The authors contributed equally, with names listed in alphabetical order.

Endnotes

[1] Among the main sources of the discrepancy of recorded inventory and available inventory are shrinkage and misplacement (Ton and Raman 2004, Atalı et al. 2009). Beck and Peacock (2009) estimate that retailers around the globe suffer a $232 billion annual loss from inventory shrinkage.

[2] The multitasking literature (e.g., Holmstrom and Milgrom 1991, Feltham and Xie 1994, Dewatripont et al. 1999) focuses on deriving optimal parameters of linear compensation schemes, without establishing their optimality.

[3] The operations management and accounting literature (e.g., Chao et al. 2009; Baiman et al. 2010; Krishnan and Winter 2012, section 8.1; Nikoofal and Gümüş 2018) has studied similar settings where the outcome of a product is determined by the weakest of its several components, such as demand and inventory in our setting.

[4] Consistent with most of the moral hazard literature, we use a constant support for both demand and inventory outcomes. If either support moves with effort, the well-known *Mirrlees argument* applies:

the firm can detect, with a positive likelihood, that the store manager has deviated from the desired action (Mirrlees 1999).

[5] The compactness condition of $\mathcal{X}$ is not overly restrictive. If the original space of signals is unbounded, for instance, a transformation of the signal could make the signal space compact. For instance, tasking the transformation $\frac{e^x}{1+e^x}$ of the original signal x, in each dimension, can achieve the desired goal.

[6] This assumes that the function $\sum_{i=1}^{m} \omega_i R_i(\vec{x})$ has zero mass at the cutoff t. If positive mass exists at the cutoff, a lottery with payouts on zero and $\bar{w}$ can characterize an optimal contract. We assume zero mass at the cutoff to avoid this additional complication.

[7] Observe that B^{NSO} can be empty under this definition when the function s^* only takes values in $(\bar{I}, \bar{Q}]$, in which case $i_s = \bar{I}$, which is not optimal. For this reason, in figures such as Figure 1(a), we restrict the vertical axis to be between zero and $\bar{I}$, as opposed to zero and $\bar{Q}$.

[8] We say "strictly" here because we require strict improvement in both the inventory and sales outcomes. Note that allowing $i' = i''$ or $s' = s''$ in the definition of joint monotonicity is a case that can be handled by one of the two earlier definitions of monotonicity.

[9] It is worth noting that the nonmonotonicity issue disappears if demand is fully observed. We explore this issue in Section OA.4 of the online appendix, where we also explore the deadweight loss due to demand censoring.

[10] In Section OA.5 of the online appendix, we additionally explore nonmonotone approximations of mast-and-sail compensation plans for the purpose of examining what aspects of the mast-and-sail structure drive optimality. There we show via extensive numerical experiments that there is little optimality loss by replacing s^* by a linear function. By contrast, removing the mast can have a significant impact.

[11] We restrict attention to nonnegative κ_1 and κ_2 to ensure that the triangular sail is described by a downward-sloping line and the 45° line. Recall that the function s^* in mast-and-sail compensation plans was downward sloping. Moreover, an upward-sloping triangular sail itself is nonmonotone, despite its simplicity, and thus is susceptible to the ex post moral hazard of hiding inventory. For these reasons, we restrict attention to nonnegative κ_1 and κ_2.

[12] We thank an anonymous reviewer for suggesting this direction of analysis to tackle ex post moral hazard.

[13] In fact, our analysis allows for Θ to be defined by a general function of Γ, Q, and I that satisfies sufficient monotonicity properties. We study the linear version because of its intuitive nature.

[14] Recall that it is possible in theory for $i_s = \bar{I}$, and hence, the region B^{LS} need not cross the 45° line. In other words, $\ell^*(i) \in (\bar{I}, \bar{Q}]$ for all i. We ignore this degenerate case because here the store manager does not get a bonus even when $\bar{I}$ is sold, which cannot be optimal because it does not implement high efforts.

[15] Again, this assumes that the function $\sum_{i=1}^{m} \omega_i R_i(\vec{x})$ has zero mass at the cutoff t.

References

Anupindi R, Dada M, Gupta S (1998) Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Sci.* 17(4):406–423.

Arya A, Glover J, Mittendorf B (2007) Aligning incentives by capping bonuses. Antle R, Gjesdal F, Liang PJ, eds. *Essays in Accounting Theory in Honour of Joel S. Demski* (Springer, New York), 169–182.

Atalı A, Lee HL, Özer Ö (2009) If the inventory manager knew: Value of visibility and RFID under imperfect inventory information. Working paper, Stanford University, Palo Alto, CA.

Baiman S, Netessine S, Saouma R (2010) Informativeness, incentive compensation, and the choice of inventory buffer. *Accounting Rev.* 85(6):1839–1860.

Barvinok A (2002) *A Course in Convexity* (American Mathematical Society, Providence, RI).

Beck A, Peacock C (2009) *New Loss Prevention: Redefining Shrinkage Management* (Palgrave Macmillan, Basingstoke, UK).

Besbes O, Muharremoglu A (2013) On implications of demand censoring in the newsvendor problem. *Management Sci.* 59(6): 1407–1424.

Beyer A, Guttman I, Marinovic I (2014) Optimal contracts with performance manipulation. *J. Accounting Res.* 52(4):817–847.

Bond P, Gomes A (2009) Multitask principal-agent problems: Optimal contracts, fragility, and effort misallocation. *J. Econom. Theory* 144(1):175–211.

Chao GH, Iravani SMR, Savaskan RC (2009) Quality improvement incentives and product recall cost sharing contracts. *Management Sci.* 55(7):1122–1138.

Chen K-Y, Li S, Rong Y (2020) The behavioral promise and pitfalls in compensating store managers. *Management Sci.* Forthcoming.

Chu LY, Lai G (2013) Salesforce contracting under demand censorship. *Manufacturing Service Oper. Management* 15(2):320–334.

Chu LY, Sappington DEM (2009) Implementing high-powered contracts to motivate intertemporal effort supply. *RAND J. Econom.* 40(2):296–316.

Conlon CT, Mortimer JH (2013) Demand estimation under incomplete product availability. *Amer. Econom. J. Microeconomics* 5(4): 1–30.

Crama P, Reyck BD, Degraeve Z (2008) Milestone payments or royalties? Contract design for R&D licensing. *Oper. Res.* 56(6): 1539–1552.

Dai T, Jerath K (2013) Salesforce compensation with inventory considerations. *Management Sci.* 59(11):2490–2501.

Dai T, Jerath K (2016) Impact of inventory on quota-bonus contracts with rent sharing. *Oper. Res.* 64(1):94–98.

Dai T, Jerath K (2019) Salesforce contracting under uncertain demand and supply: Double moral hazard and optimality of smooth contracts. *Marketing Sci.* 38(5):852–870.

DeHoratius N, Raman A (2007) Store manager incentive design and retail performance: An exploratory investigation. *Manufacturing Service Oper. Management* 9(4):518–534.

Dewatripont M, Jewitt I, Tirole J (1999) The economics of career concerns, Part II: Application to missions and accountability of government agencies. *Rev. Econom. Stud.* 66(1):199–217.

Feiler DC, Tong JD, Larrick R (2013) Biased judgment in censored environments. *Management Sci.* 59(3):573–591.

Feltham GA, Xie J (1994) Performance measure congruity and diversity in multi-task principal/agent relations. *Accounting Rev.* 69(3):429–453.

Grossman SJ, Hart OD (1983) An analysis of the principal-agent problem. *Econometrica* 51(1):7–45.

Herweg F, Muller D, Weinschenk P (2010) Binary payment schemes: Moral hazard and loss aversion. *Amer. Econom. Rev.* 100(5): 2451–2477.

Ho TH, Tang CS (2004) Introduction to the special issue on marketing and operations management interfaces and coordination. *Management Sci.* 50(4):429–430.

Holmstrom B (1979) Moral hazard and observability. *Bell J. Econom.* 10(1):74–91.

Holmstrom B, Milgrom P (1991) Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *J. Law Econom. Organ.* **7**(Sp):24–52.

Huh WT, Levi R, Rusmevichientong P, Orlin JB (2011) Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Oper. Res.* 59(4):929–941.

Innes RD (1990) Limited liability and incentive contracting with ex-ante action choices. *J. Econom. Theory* 52(1):45–67.

Jain S (2012) Self-control and incentives: An analysis of multiperiod quota plans. *Marketing Sci.* 31(5):855–869.

Jain A, Rudi N, Wang T (2015) Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Oper. Res.* 63(1):134–150.

Jain T, Hazra J, Gavirneni S (2019) Ranking-based versus outcome-linked incentive systems: Designing an effective compensation system. Working paper, Cornell University, Ithaca, NY.

Jerath K, Kim S-H, Swinney R (2017) Product quality in a distribution channel with inventory risk. *Marketing Sci.* 36(5): 747–761.

Jerath K, Netessine S, Zhang ZJ (2007) Can we all get along? incentive contracts to bridge the marketing and operations divide. Working paper, Columbia University, New York.

Jewitt I (2008) Moral hazard with bounded payments. *J. Econom. Theory* 143(1):59–82.

Ke R, Ryan CT (2018a) Monotonicity of optimal contracts without the first-order approach. *Oper. Res.* 66(4):1101–1118.

Ke R, Ryan CT (2018b) A general solution method for moral hazard problems. *Theoretical Econom.* 13(3):1425–1481.

Krishnan H, Winter RA (2012) The economic foundations of supply chain contracting. *Foundations Trends Tech. Inform. Oper. Management* 5(3–4):147–309.

Krishnan J, Fisher M (2005) Store Level Execution at Wawa: Case Study. Working paper, University of Pennsylvania, Philadelphia.

Lacker JM, Weinberg JA (1989) Optimal contracts under costly state falsification. *J. Political Econom.* 97(6):1345–1363.

Laffont JJ, Martimort D (2009) *The Theory of Incentives: The Principal-Agent Model* (Princeton University Press, Princeton, NJ).

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.

Long X, Nasiry J (2020) Wage transparency and social comparison in salesforce compensation. *Management Sci.*, ePub ahead of print April 16, https://doi.org/10.1287/mnsc.2019.3461.

Mirrlees JA (1999) The theory of moral hazard and unobservable behaviour: Part I. *Rev. Econom. Stud.* 66(1):3–21.

Misra S, Pinker EJ, Shumsky RA (2004) Salesforce design with experience-based learning. *IIE Trans.* 36(10):941–952.

Nikoofal ME, Gümüş M (2018) Quality at the source or at the end? Managing supplier quality under information asymmetry. *Manufacturing Service Oper. Management* 20(3):498–516.

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.

Rogerson WP (1985) The first-order approach to principal-agent problems. *Econometrica* 53(6):1357–1367.

Rudi N, Drake D (2014) Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. *Management Sci.* 60(5):1334–1345.

Sen A, Alp O (2020) Delegation of stocking decisions under asymmetric demand information. *Manufacturing Service Oper. Management*, ePub ahead of print March 6, https://doi.org/10.1287/msom.2019.0810.

Sethi SP, Thompson GL (2000) *Optimal Control Theory: Applications to Management Science and Economics* (Springer, New York).

Shapiro BP (1977) Can marketing and manufacturing coexist? *Harvard Bus. Rev.* 55(5):104–114.

Ton Z, Raman A (2004) The effect of product variety and inventory levels on misplaced products at retail stores: A longitudinal study. Working paper, Harvard Business School, Boston.