

Incentive Design for Operations-Marketing Multitasking

Tinglong Dai* Rongzhu Ke[†] Christopher Thomas Ryan[‡]

*Carey Business School, Johns Hopkins University, Baltimore, Maryland 21202, dai@jhu.edu

[†]Department of Economics, Hong Kong Baptist University, Kowloon, Hong Kong, rongzhuke@hkbu.edu.hk

[‡]Sauder School of Business, University of British Columbia, Vancouver, Canada, chris.ryan@sauder.ubc.ca

Forthcoming in *Management Science*

A firm hires an agent (e.g., 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 non-monotonicity, 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 this *ex post* moral hazard issue is a result of demand censoring. If available information includes a waitlist (or other noisy signals) to gauge unsatisfied demand, no *ex post* moral hazard issues remain.

Key words: 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 single 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 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 efforts 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: (a) marketing (i.e., bolstering customer demand) and (b) operations (i.e., ensuring inventory can be put in the hands of customers, where it belongs,

instead of being misplaced, damaged, spoiled, or stolen through mismanagement).¹ How to design compensation plans to get the most out of their store managers, in light of these two competing areas of focus? 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 consider single-tasking agents, most prominently in the salesforce compensation literature (see Section 2 for more detail). When it comes to multi-tasking 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 non-optimal 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 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 2012). Its negative implications for sales performance is 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* their 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.” The first-order 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

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

issues in contract theory” and “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. The validity of the first-order approach is particularly troubling under 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 her salary. Concretely, we find an optimal compensation plan for the multitasking store manager, under the monotone likelihood ratio property (MLRP) that 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 (i) inventory does not clear and the sales quantity exceeds an inventory-dependent threshold or (ii) inventory clears and the realized inventory level exceeds a threshold.

Intriguingly, the structure of such a bonus region gives rise to inherent non-monotonicity 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 non-monotone optimal compensation plan evokes the speculation that, in certain cases, a store

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.

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 due to 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 in 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 this agency conundrum (the result of demand censoring and non-monotone contracts) is at the core of multitasking with censored signals. Further analysis and numerical investigations show 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 such plans cannot be optimal and furthermore exhibit natural cases where they perform arbitrarily poorly. Other simple (and monotone) compensation plans, like 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 waitlist 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., their 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 decision 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).

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; Chen et al. 2019; Long and Nasiry 2019). 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 distributions. (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 unlimited inventory meets demand generated by the salesperson. A recent stream of literature (Chu and Lai 2013; Dai and Jerath 2013, 2016, 2019) incorporates demand censoring due to 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 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. In particular, Besbes and Muharremoglu (2013) show an exploration–exploitation trade-off in a multi-period 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 multi-period inventory control problem (also without moral hazard) and numerically show the timing information of stock-out can help recover much of the inefficiency from demand censoring. Related to this literature, we show 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 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.

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 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 $I \leq \bar{I}$ is all that is available to meet demand. The difference $\bar{I} - I$ is unavailable to meet demand due to 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 operational effort does not affect demand and 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,

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

The MLRP implies 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 both the firm and 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 $\bar{Q} \geq \bar{I}$ to allow for the possibility that demand is censored at its highest level.

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

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).

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 both the firm and 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 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 simultaneously realized 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

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

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 the store manager’s expected net payoff is no lower than a reservation utility \underline{U} , and the incentive compatibility (IC) constraint (2d) ensures 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 $\mathcal{A} = \{\vec{a}^1, \vec{a}^2, \dots, \vec{a}^m\}$; 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 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 $\vec{a} \in \mathcal{A}$, where we assume $c(\vec{a})$ 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 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 $w : \mathcal{X} \rightarrow \mathbb{R}$ that pays out according to the realized outcome. The principal values outcome $\vec{x} \in \mathcal{X}$ according to the valuation function $\pi : \mathcal{X} \rightarrow \mathbb{R}$. 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, $w(x) \leq \bar{w}$ for almost all $x \in \mathcal{X}$. Finally, the agent has a reservation utility \underline{U} for her next-best alternative.

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

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

$$\text{subject to} \quad U(w, \vec{a}) \geq \underline{U} \quad (3b)$$

$$U(w, \vec{a}) - U(w, \vec{a}^i) \geq 0 \text{ for } i = 1, 2, \dots, m \quad (3c)$$

$$0 \leq w \leq \bar{w}. \quad (3d)$$

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

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

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.

$$\text{subject to } \int_{\vec{x} \in \mathcal{X}} w(\vec{x}) f(\vec{x}|\vec{a}^*) d\vec{x} \geq \underline{U} \quad (4b)$$

$$\int_{\vec{x} \in \mathcal{X}} R_i(\vec{x}) w(\vec{x}) f(\vec{x}|\vec{a}^*) d\vec{x} \geq c(\vec{a}^*) - c(\vec{a}^i) \text{ for } i \in \{1, 2, \dots, m\} \text{ such that } \vec{a}^i \neq \vec{a}^* \quad (4c)$$

$$0 \leq w \leq \bar{w}, \quad (4d)$$

where we use the fact that $V(w, \vec{a}) = \mathbb{E}[\pi(\vec{x})|\vec{a}^*] - \int_{\vec{x} \in \mathcal{X}} w(\vec{x}) f(\vec{x}|\vec{a}^*) d\vec{x}$, 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

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

for $i = 1, 2, \dots, m$. 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 $w(\vec{x}) \in \{0, \bar{w}\}$ for almost all $\vec{x} \in \mathcal{X}$.

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

Next, we characterize when an optimal bang-bang contract takes the value of 0 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*

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

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

Let $B \triangleq \{\vec{x} \in \mathcal{X} : \sum_{i=1}^m \omega_i R_i(\vec{x}) \geq t\}$ 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

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

that defines the bonus region B . Because the ω_i are nonnegative and sum to 1, the left-hand side is a weighted sum of likelihood ratios that can be viewed as a measure of the information value (or

⁶ This assumes 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 0 and \bar{w} can characterize an optimal contract. We assume zero mass at the cutoff to avoid this additional complication.

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 through 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,

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

and

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

where $\mathbb{P}[\cdot]$ is the probability measure and $\mathbb{E}[\cdot]$ is the expectation operator associated with $f(\cdot | \vec{a}^*)$.

Using this notation, the restriction of (4) over trigger contracts of the form (6) is

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

$$\text{subject to} \quad \bar{w} \mathbb{P} \left[\sum_{i=1}^m \omega_i R_i(\vec{X}) \geq t \right] \geq \underline{U} \tag{8b}$$

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

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

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

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 store manager problem introduced in [Section 3](#) using the bang-bang approach of the previous section.

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 $D \triangleq \{(i, s) : 0 \leq s \leq i \text{ and } 0 \leq i \leq \bar{I}\}$ is the domain of any feasible compensation plan because of demand censoring. We also denote by $D^{\text{NSO}} \triangleq \{(i, s) \in D : s < i\}$ and $D^{\text{SO}} \triangleq \{(i, s) \in D : s = i\}$ the regions of the domain where no stockout occurs and 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

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

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: (i) where $i > s$ (no stockout) and (ii) where $i = s$ (stockout), by defining appropriate ratio functions. In the nonstockout (NSO) case,

$$R_{e_o, e_m}^{\text{NSO}}(i, s) = 1 - \frac{f(i|e_o)g(s|e_m)}{f(i|e_o^*)g(s|e_m^*)}, \quad (9)$$

and in the stockout (SO) case,

$$R_{e_o, e_m}^{\text{SO}}(i) = 1 - \frac{f(i|e_o)(1 - G(i|e_m))}{f(i|e_o^*)(1 - G(i|e_m^*))}. \quad (10)$$

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

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

where

$$w^{\text{NSO}}(i, s) = \begin{cases} \bar{w} & \text{if } \sum_{e_o, e_m} \omega_{e_o, e_m} R_{e_o, e_m}^{\text{NSO}}(i, s) \geq t \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad w^{\text{SO}}(i) = \begin{cases} \bar{w} & \text{if } \sum_{e_o, e_m} \omega_{e_o, e_m} R_{e_o, e_m}^{\text{SO}}(i) \geq t \\ 0 & \text{otherwise,} \end{cases}$$

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

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

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

A key observation here is that the bonus region B^{NSO} is possibly a full-dimensional subset of the nonstockout 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 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 her best effort in both operations and marketing. For a discussion of scenarios where other effort levels may be targeted, see Section OA.6.

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 $B^{\text{NSO}} = \{(i, s) : i \geq i_s \text{ and } s^*(i) \leq s < i\}$.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 $B^{\text{SO}} = \{(i, s) : s = i \geq i_m\}$.*

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 like Figure 1(a), we restrict the vertical axis to be between 0 and \bar{I} , as opposed to 0 and \bar{Q} .

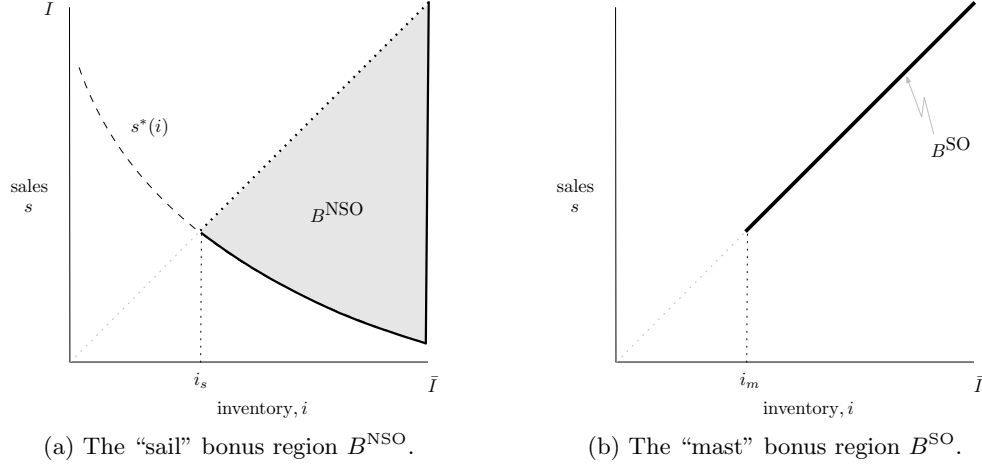


Figure 1 Illustrations of the “sail” and “mast” regions.

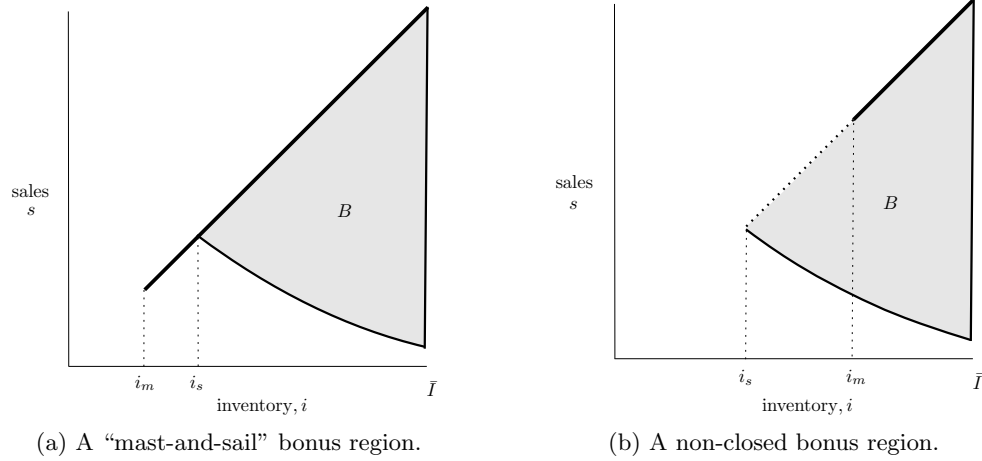


Figure 2 Two possible structures of the union of the “mast” and the “sail” bonus regions.

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. Figures 2(a) and 2(b) illustrate 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 only the structure seen in Figure 2(a) is possible.

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

5.3. Are Marketing and Operational Outcomes Complements or Substitutes?

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

PROPOSITION 4. *As i increases, (i) if $i_m \leq i < i_s$, the minimum sales quantity required for the store manager to qualify for the bonus strictly increases (“moving up the mast”), and (ii) if $i \geq i_s$, the minimum sales quantity $s^*(i)$ required for the store manager to qualify for the bonus decreases (“slipping down the sail”).*

Proposition 4(i) 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 the agent has exerted sufficient marketing effort.

In the “sail” part of the bonus region (i.e., the region with $i \geq i_s$), 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 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 it is not the case, “slipping down the sail” only occurs when inventory meets the minimum threshold i_s , indicating a sufficiently high likelihood of 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 the informative value of the observed signal (i, s) increases in both i and s . Note the “sail” part corresponds to the scenarios without stockouts so that 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 (i) a low inventory level and a high demand outcome or (ii) a high inventory level and a lower demand outcome. For this reason, operations and marketing act as substitutes in the optimal compensation plan.

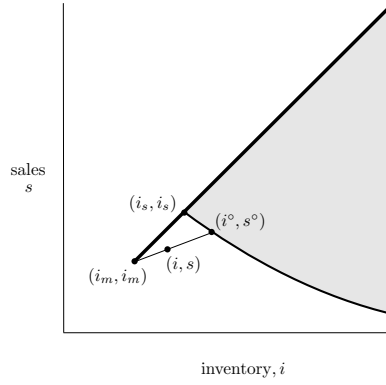


Figure 3 An illustration of the lack of joint monotonicity of an optimal compensation plan.

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. We say $w(i, s)$ is *monotone in i* if $w(i', s) \leq w(i'', s)$ for every $(i', s), (i'', s) \in D$ and $i' \leq i''$. Similarly, we say $w(i, s)$ is *monotone in s* if $w(i, s') \leq w(i, s'')$ for every $(i, s'), (i, s'') \in D$ and $s' \leq s''$. Finally, we say $w(i, s)$ is (strictly) *jointly monotone* if $w(i', s') \leq w(i'', s'')$ 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), (i) w^* is monotone in s , (ii) if $i_s = i_m$, w^* is also monotone in i and jointly monotone, and (iii) if $i_s > i_m$, w^* is neither monotone in i nor jointly monotone.*

Figure 3 provides the intuition for this result. Part (i) concerns monotonicity in the vertical direction, which clearly holds in the figure because we never move beyond the 45° line in the vertical direction. Part (ii) 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, part (iii) 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^o, s^o) , where $i^o = i_s + \epsilon$ for some positive ϵ and $s^o = s^*(i^o)$ 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

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.

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 above non-monotonicity 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 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 it was “shrunk”) to reveal an output of (i', i') , she would receive a bonus. In other words, the store manager is effectively rewarded for disposing of inventory, meaning *ex post* moral hazard issue is inherent with contracts that are non-monotone 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 marketing and operational effort of the store manager are not observable to the firm. This setup suggests 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 non-monotonicity 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 non-monotonicity 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 the 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 previous two sections, we have characterized an optimal mast-and-sail compensation plan, but these compensation plans suffer from non-monotonicity, 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.

⁹ It is worth noting that the non-monotonicity issue disappears if demand is fully observed. We explore this issue in [Section OA.4](#), where we also explore the dead-weight loss due to demand censoring.

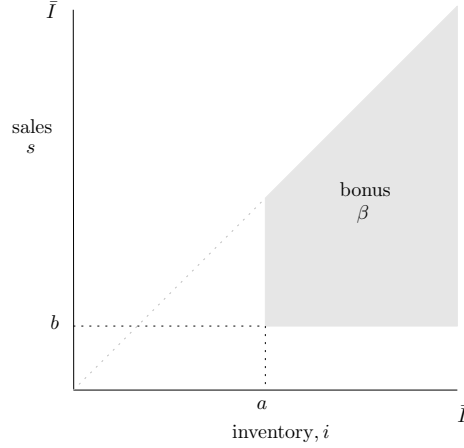


Figure 4 Illustration of a corner compensation plan.

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 plans in the introduction. The other candidate is a modification of the mast-and-sail by snipping “mast” to remove non-convexity of the bonus region (and thus assuring 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.

The next result shows the mast-and-sail compensation plans often perform *strictly* better than corner compensation plans. We say 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, log-normal, 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 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, 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)

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

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

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

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

assuming 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:*

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

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 . On the other hand, if (14b) does not bind, then $a = b$, where a is characterized by setting either constraint (14c) or (14d) to be tight.

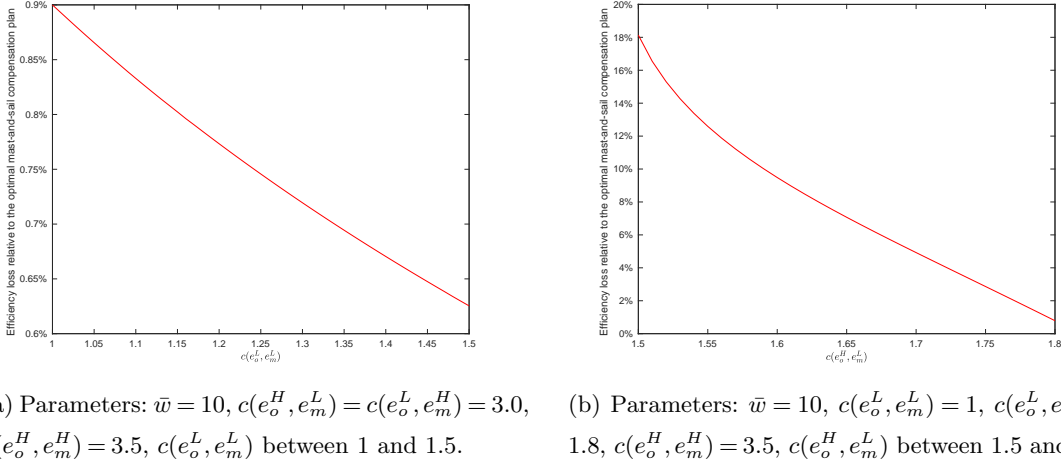


Figure 5 Performance of optimal corner compensation plan versus optimal mast-and-sail compensation plan. We assume $F(i|e_o) = (H(i))^{e_o}$ and $G(s|e_m) = (L(s))^{e_m}$, where $H(i) = i$ and $L(s) = s$.

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](#). Here, we present two representative and *contrasting* scenarios in [Figures 5\(a\)](#) and [5\(b\)](#). [Figure 5\(a\)](#) shows 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\(b\)](#) shows 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 the efficiency loss under the corner compensation plan increases linearly in the unit revenue. In other words, the worst-case loss in performance of corner compensation plans is *arbitrarily* large.

EXAMPLE 1. Consider the following instance where $e_o \in \{e_o^L, e_o^H\}$ 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 has $\bar{w} = 10$. For this instance, we can show 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

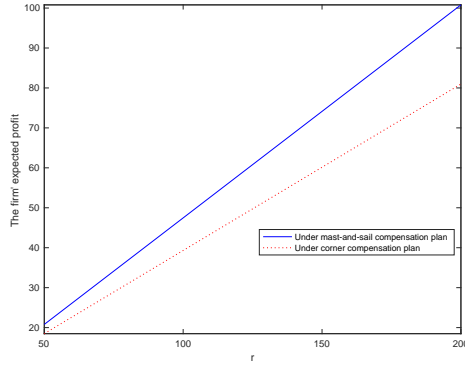


Figure 6 Expected profits under the mast-and-sail and corner compensation plans versus revenue rate r .

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 the previous subsection, 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 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: (i) removing the mast (i.e., setting $i_m = i_s$ and (ii) linearizing the downward-sloping function s^* in [Proposition 1](#). Removing the mast assures monotonicity of the compensation plan and linearizing s^* makes communicating the compensation plan to sales managers easier in practice.¹⁰

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 the optimal weighted-sum threshold compensation plans

¹⁰ In [Section OA.5](#) we additionally explore non-monotone 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 the “triangular sail” is described by a downward-sloping line and the 45° line. Recall the function s^* in mast-and-sail compensation plans was downward sloping. Moreover, an upward-sloping triangular sail itself is non-monotone, despite its simplicity, and thus susceptible to the *ex post* moral hazard of hiding inventory. For these reasons, we restrict attention to nonnegative κ_1 and κ_2 .

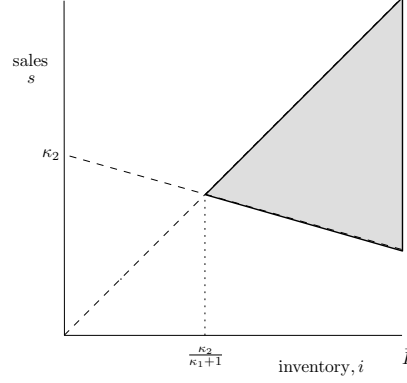


Figure 7 An illustration of a weighted-sum-threshold compensation plan.

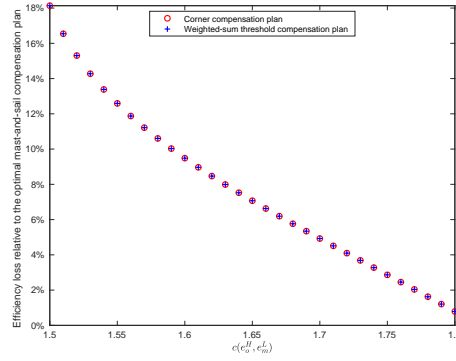


Figure 8 Expected profits under the mast-and-sail and corner compensation plans versus $c(e_o^H, e_m^L)$

also perform poorly (indeed, as poorly as corner compensation plans) in bad cases (losses of up to 18% in this example). We conclude the “loss due to monotonicity” that is 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, this issue can be resolved.

8. Resolving Ex Post Moral Hazard through Gauging Unsatisfied Demand¹²

In previous sections, we assumed any demand in excess of inventory cannot be observed. We now consider a more general setting where partial information is revealed when demand exceeds sales. In particular, we assume some random (and unknown) fraction of customers who do not receive the product express interest via a waitlist (or some other method of capturing unsatisfied demand). We introduce a new random variable¹³

$$\Theta := \Lambda(Q - I) + I, \quad (16)$$

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 due to its intuitive nature.

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

We define a new random variable,

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

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

The firm and store manager observe I and Z . When $Z = \Theta$, the product $\Lambda(Q - I)$ can be observed (since I is also observable) but knowledge of this product does not reveal Λ or Q directly. That is, the proportion of unsatisfied customers that sign up for the waitlist 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 0. 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 1; this full information case is explored in [Section OA.4](#) of the online appendix.

Because the firm and 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, as 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: (i) the “no lost sales” (NLS) region (where $Z \leq I$) and (ii) 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:

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

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:

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

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

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

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

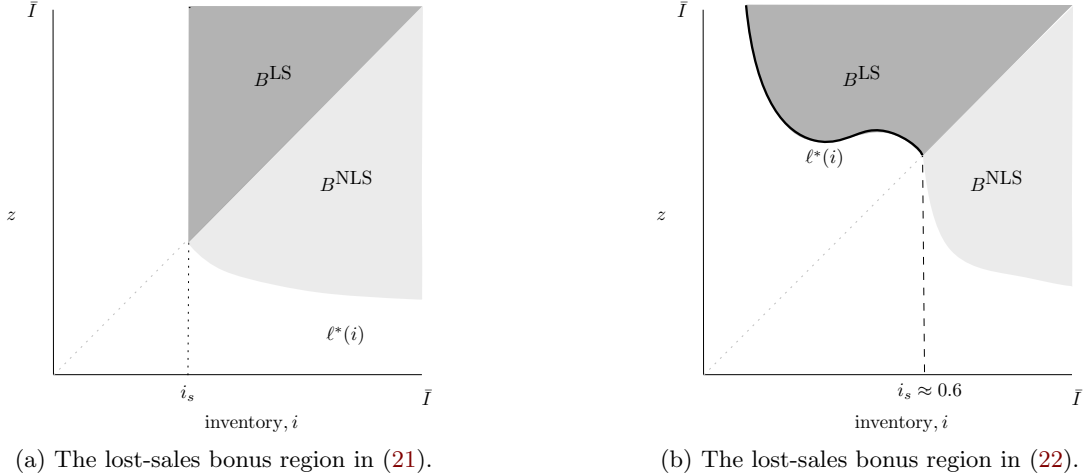


Figure 9 Illustrations of lost sales bonus region B^{LS} .

PROPOSITION 9. *Under [Assumption 1](#), we have the following cases:*

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

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

(b) there exists a continuous function ℓ^* defined on $[0, i_s)$ such that¹⁴

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

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.

An important fact used in this proof is that $\Lambda \in (0, 1)$. When we can infer that $Q = I$ (i.e., stockout occurs) when $Z = I$ because this implies $\Lambda(Q - I) = 0$ and Λ never realizes to 0. In other words, when the signal Θ is equal to I , no demand is lost. This ensures 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 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 $\text{Gamma}(\alpha, \beta)$ distribution. This ensures $\Lambda \in (0, 1)$. 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 waitlist information, the old “mast” region disappears. The waitlist 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, *ex post* moral hazard issue of hiding inventory no longer exists. Indeed, the feasible region B^{LS} is monotone in z and so even if we allow *downward* manipulation of the waitlist, there is no incentive to do so.

By contrast, in the bonus region in Figure 9(b), the waitlist signal is sufficiently correlated with marketing effort to offer bonuses if the waitlist 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 nonstockout region of the domain ($s < i$), the store manager has no incentive to hide inventory, due to monotonicity in i : the bonus region is monotone below the 45° line. The area above the 45° line captures scenarios with lost sales, and so no inventory is left to “hide.” Assuming also that the signal Θ cannot be manipulated by the store manager (e.g., the waitlist 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* waitlist signal to get a bonus, whereas a higher sales level requires a *larger* bonus. The fact that no scope exists for *ex post* manipulation does not change the fact it could be

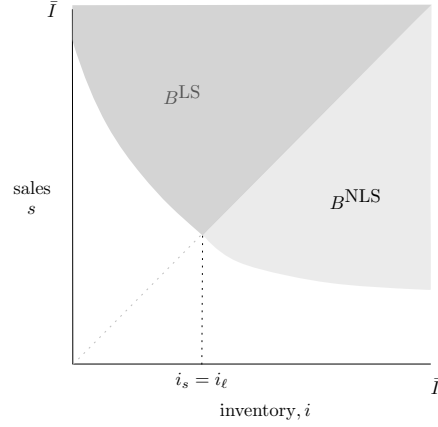


Figure 10 The “double sail” bonus region $B^{\text{NLS}} \cup B^{\text{LS}}$ under **Assumption 1**.

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 $B^{\text{LS}} = \{(i, z) : i \leq i_s \text{ and } \ell^*(i) \leq z \leq \bar{Q}\} \cup \{(i, z) : i \geq i_s \text{ and } i \leq z \leq \bar{Q}\}$, implying that $B^{\text{NLS}} \cup B^{\text{LS}}$ has the “double sail” structure depicted in **Figure 10**.*

It is straightforward to observe that the resulting optimal contract is $w^*(i, z)$, where $w^*(i, z) = \bar{w}$ when $(i, z) \in B^{\text{NLS}} \cup B^{\text{LS}}$ and 0 otherwise, is jointly monotone. In other words, under the waitlist approach for gauging unsatisfied demand (and given the above 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 non-monotone lost-sales bonus regions. However, one can also check that non-monotonicity does not hold for all parameter values. Indeed, an algebra exercise can verify that when Λ is such that $-\log \Lambda$ is a $\text{Gamma}(\alpha, \beta)$, where $\alpha - 1 \geq \frac{4}{e^2 + 1}(\frac{1}{\beta} + 1)$, **Assumptions 1** and **2** hold and the resulting contract is monotone (by **Proposition 10**).

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, non-monotonicity, 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 of 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 $w(i, s) \leq \bar{w}$, and 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](#).

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 $w(i, s) \leq \bar{w}$ is replaced by constraint $w(i, s) \leq m(i, s)$ for almost all (i, s) . As an example of $m(i, s)$, consider the following description from DeHoratius and Raman (2007): “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 as 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 ranges from 0.2% to 5%.

Our model can be adjusted to the setting with resource constraint $w(i, s) \leq m(i, s)$, assuming $m(i, s)$ 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

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

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

$$\int_i \int_s \beta(i, s) m(i, s) f(i|e_o^*) g(s|e_m^*) di ds - \int_i \int_s \beta(i, s) m(i, s) f(i|e_o) g(s|e_m) di ds \quad (23c)$$

$$\begin{aligned} &\geq c(e_o^*, e_m^*) - c(e_o, e_m) \text{ for all } (e_o, e_m) \\ 0 &\leq \beta(i, s) \leq 1 \text{ for all } (i, s). \end{aligned} \tag{23d}$$

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*

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

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 we can ignore the incentive compatibility constraint of the store manager and pay 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 $\max_{\bar{I}} r\mathbf{E}[S|\bar{I}] - C(\bar{I})$, 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 $r \frac{d}{d\bar{I}} \mathbf{E}[S|\bar{I}] = C'(\bar{I})$.

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

$$\begin{aligned} W(\bar{I}) &\triangleq \min_w \mathbb{E}[w(I, S)|e_o^H, e_m^H] \\ \text{s.t.} \quad &\mathbb{E}[w(I, S)|e_o^H, e_m^H] - c(e_o^H, e_m^H) \geq \underline{U} \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}[w(I, S)|e_o^H, e_m^H] - \mathbb{E}[w(I, S)|e_o, e_m] &\geq c(e_o^H, e_m^H) - c(e_o, e_m) \text{ for all } (e_o, e_m) \\ 0 \leq w(i, s) &\leq \bar{w} \text{ for all } (i, s) \end{aligned}$$

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

PROPOSITION 12. (i) *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.*

(ii) *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(i) 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(ii) 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^* :

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

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 $I^* > I^{NV}$. In other words, the firm over-invests in inventory as compared to 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 examined incentive issues at the intersection of operations and marketing. We show 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 waitlist) can we construct a compensation plan that is both optimal and resolve 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 arms-length company cannot stop the

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. In addition, we hope 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 responsible for one of each task. 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 in making operational decisions, such as inventory stocking levels (see, e.g., Sen and Alp 2019). 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.

Acknowledgments

The authors are grateful to the Department Editor, Associate Editor, and three anonymous reviewers for the constructive review process. They also appreciate Anil Arya, Maqbool Dada, Nicole DeHoratius, Vishal Gaur, 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 the paper. The authors also benefited from interacting with session participants at the 2018 MSOM Conference (Dallas, Texas), 2019 INFORMS Marketing Science Conference (Rome, Italy), 2019 POMS International Conference (Brighton, UK), and INFORMS 2019 Annual Meeting (Seattle, Washington). The authors contributed equally, with names listed in alphabetical order; Rongzhu Ke is the corresponding author of the paper.

References

- Anupindi, R., M. Dada, S. Gupta. 1998. Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Sci.* **17**(4) 406–423.
- Arya, A., J. Glover, B. Mittendorf. 2007. Aligning incentives by capping bonuses. Antle, R., F. Gjesdal, P. J. Liang, eds. *Essays in Accounting Theory in Honour of Joel S. Demski* (New York, NY: Springer) 169–182.

- Atali, A., H. L. Lee, and Ö. Özer. 2009. If the inventory manager knew: Value of visibility and RFID under imperfect inventory information. *Stanford University Working Paper*.
- Baiman, S., S. Netessine, R. Saouma. 2010. Informativeness, incentive compensation, and the choice of inventory buffer. *Accounting Rev.* **85**(6) 1839–1860.
- Barvinok, A. 2002. *A Course in Convexity*. Graduate Studies in Mathematics **54**. Providence, RI: American Mathematical Society.
- Beck, A., C. Peacock. 2009. *New Loss Prevention: Redefining Shrinkage Management*. Basingstoke, United Kingdom: Palgrave Macmillan.
- Besbes, O., A. Muharremoglu. 2013. On implications of demand censoring in the newsvendor problem. *Management Sci.* **59**(6) 1407–1424.
- Beyer, A., I. Guttman, I. Marinovic. 2014. Optimal contracts with performance manipulation. *J. Accounting Res.* **52**(4) 817–847.
- Bond, P., A. Gomes. 2009. Multitask principal-agent problems: Optimal contracts, fragility, and effort misallocation. *J. Econom. Theory* **144**(1) 175–211.
- Chao, G. H., S. M. R. Irvani, R. C. Savaskan. 2009. Quality improvement incentives and product recall cost sharing contracts. *Management Sci.* **55**(7) 1122–1138.
- Chen, K.-Y., S. Li, Y. Rong. 2019. The behavioral promise and pitfalls in compensating store managers. *Management Sci.* Forthcoming.
- Chu, L. Y., G. Lai. 2013. Salesforce contracting under demand censorship. *Manufacturing Service Oper. Management* **15**(2) 320–334.
- Chu, L. Y., D. E. M. Sappington. 2009. Implementing high-powered contracts to motivate intertemporal effort supply. *RAND J. Econom.* **40**(2) 296–316.
- Conlon, C. T., J. H. Mortimer. 2013. Demand estimation under incomplete product availability. *Amer. Econom. J.: Microeconomics* **5**(4) 1–30.
- Crama, P., Reyck, B.D. and Degraeve, Z., 2008. Milestone payments or royalties? Contract design for R&D licensing. *Oper. Res.* **56**(6) 1539–1552.
- Dai, T., K. Jerath. 2013. Salesforce compensation with inventory considerations. *Management Sci.* **59**(11) 2490–2501.
- Dai, T., K. Jerath. 2016. Impact of inventory on quota-bonus contracts with rent sharing. *Oper. Res.* **64**(1) 94–98.
- Dai, T., K. Jerath. 2019. Salesforce contracting under uncertain demand and supply: Double moral hazard and optimality of smooth contracts. *Marketing Sci.* **38**(5) 852–870.
- DeHoratius, N., A. Raman. 2007. Store manager incentive design and retail performance: An exploratory investigation. *Manufacturing Service Oper. Management* **9**(4) 518–534.

- Dewatripont, M., I. Jewitt, J. Tirole. 1999. The economics of career concerns, Part II: Application to missions and accountability of government agencies. *Rev. Econom. Stud.* **66**(1) 199–217.
- Feiler, D. C., J. D. Tong, R. Larrick. 2013. Biased judgment in censored environments. *Management Sci.* **59**(3) 573–591.
- Feltham, G. A., J. Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *Accounting Rev.* **69**(3) 429–453.
- Grossman, S. J., O. D. Hart. 1983. An analysis of the principal-agent problem. *Econometrica* **51**(1) 7–45.
- Herweg, F., D. Muller, P. Weinschenk. 2010. Binary payment schemes: moral hazard and loss aversion. *Amer. Econom. Rev.* **100**(5) 2451–2477.
- Ho, T. H., C. S. Tang. 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., P. Milgrom. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *J. Law Econom. Organ.* **7**(Sp) 24–52.
- Huh, W. T., R. Levi, P. Rusmevichientong, J. B. Orlin. 2011. Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Oper. Res.* **59**(4) 929–941.
- Innes, R. D. 1990. Limited liability and incentive contracting with ex-ante action choices. *J. Econom. Theory* **52** 45–67.
- Jain, S. 2012. Self-control and incentives: An analysis of multiperiod quota plans. *Marketing Sci.* **31**(5) 855–869.
- Jain, A., N. Rudi, T. Wang. 2014. Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Oper. Res.* **63**(1) 134–150.
- Jerath, K., S. Netessine, Z. J. Zhang. 2007. Can we all get along? incentive contracts to bridge the marketing and operations divide. *University of Pennsylvania Working Paper*.
- Jerath, K., S.-H. Kim, R. Swinney. 2017. Product quality in a distribution channel with inventory risk. *Marketing Sci.* **36**(5) 747–761.
- Jewitt, I. 2008. Moral hazard with bounded payments. *J. Econom. Theory* **143**(1) 59–82.
- Ke, R., C. T. Ryan. 2018. Monotonicity of optimal contracts without the first-order approach. *Oper. Res.* **66**(4) 1101–1118.
- Krishnan, H., R. A. Winter. 2012. The economic foundations of supply chain contracting. *Foundations and Trends in Technology, Information and Operations Management* **5**(3–4) 147–309.
- Krishnan, J., M. Fisher. 2005. Store level execution at Wawa. Case study, University of Pennsylvania.
- Lacker, J. M., J. A. Weinberg. 1989. Optimal contracts under costly state falsification. *J. Political Econom.* **97**(6) 1345–1363.

- Laffont, J. J., D. Martimort. 2009. *The Theory of Incentives: The Principal-Agent Model*. Princeton, NJ: Princeton University Press.
- Lal, R., V. Srinivasan. 1993. Compensation plans for single-and multi-product salesforces: An application of the Holmstrom-Milgrom model. *Management Sci.* **39**(7) 777–793.
- Long, X, J. Nasiry. 2019. Wage transparency and social comparison in salesforce compensation. *Management Sci.* Forthcoming.
- Mirrlees, J. A. 1999. The theory of moral hazard and unobservable behaviour: Part I. *Rev. Econom. Stud.* **66**(1) 3–21.
- Misra, S., E. J. Pinker, R. A. Shumsky. 2004. Salesforce design with experience-based learning. *IIE Transactions* **36**(10) 941–952.
- Nikoofal, M. E., M. Gümüş. 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, J. S., V. Srinivasan. 1996. Quota-based compensation plans for multiterritory heterogeneous salesforces. *Management Sci.* **42**(10) 1454–1462.
- Rogerson, W. P. 1985. The first-order approach to principal-agent problems. *Econometrica* **53**(6) 1357–1367.
- Rudi, N., D. Drake. 2014. Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. *Management Sci.* **60**(5) 1334–1345.
- Sen, A., O. Alp. 2019. Delegation of stocking decisions under asymmetric demand information. *Manufacturing Service Oper. Management*. Forthcoming.
- Sethi, S. P., G. L. Thompson. 2000. *Optimal Control Theory: Applications to Management Science and Economics*. New York, NY: Springer.
- Shapiro, B. P. 1977. Can marketing and manufacturing coexist? *Harvard Bus. Rev.* **55**(5) 104–114.
- Ton, Z., A. Raman. 2004. The effect of product variety and inventory levels on misplaced products at retail stores: A longitudinal study. *Harvard Business School Working Paper*.

Online Appendix to “Incentive Design for Operations-Marketing Multitasking”

In this online appendix, we provide proofs of the results in the main body, additional technical details, and robustness checks.

OA.1. Proofs of Results Stated in the Paper

Proof of Theorem 1: We need the following two lemmas.

LEMMA OA1. *An optimal contract for (4) exists.*

Proof of Lemma OA1. Due to constraint (4d), we may assume w is a function in $L^\infty(\mathcal{X})$, the space of uniformly bounded functions on \mathcal{X} . Moreover, those same constraints ensure the feasible region of (4) is bounded in the norm on $L^\infty(\mathcal{X})$. Hence, by Alaoglu’s Theorem (see Theorem 5.105 in Aliprantis and Border 2006), the set $\{w : 0 \leq w(\vec{x}) \leq \bar{w}\}$ is compact in the weak topology $\sigma(L^\infty(\mathcal{X}), L^1(\mathcal{X}))$ (for a definition of this weak topology, see Section 5.14 of Aliprantis and Border (2006)). Because $f(\cdot|\vec{a})$ is in $L^1(\mathcal{X})$ for all $\vec{a} \in \mathcal{A}$, the constraints (4b) and (4c) are continuous in the $\sigma(L^\infty(\mathcal{X}), L^1(\mathcal{X}))$ topology, and so the feasible region is a closed subset of $\{w : 0 \leq w(\vec{x}) \leq \bar{w}\}$ and thus also compact in the $\sigma(L^\infty(\mathcal{X}), L^1(\mathcal{X}))$ topology. Moreover, the objective function $V(w, a)$ is continuous in the $\sigma(L^\infty(\mathcal{X}), L^1(\mathcal{X}))$ topology, and so, by Weierstrass’s Theorem (see Theorem 2.35 in Aliprantis and Border 2006), an optimal contract exists. \square

Let W denote the set of feasible contracts to (4). An *extremal contract* of W is a contract that cannot be written as the convex combination of two other feasible contracts. That is, $w \in W$ is an extremal contract if $w^1, w^2 \in W$, and $\lambda_1, \lambda_2 \in (0, 1)$ with $\lambda_1 + \lambda_2 = 1$ do not exist such that $w = \lambda_1 w^1 + \lambda_2 w^2$. The next result is a consequence of Barvinok (2002, Proposition III.5.3).

LEMMA OA2. *Every extremal feasible contract to (4) is a bang-bang contract.*

Proof of Lemma OA2. This argument can be adapted from Proposition III.5.3 in Barvinok (2002). Our problem (4) is a linear program in $L^\infty(\mathcal{X})$ with finitely many constraints. Proposition III.5.3 in Barvinok (2002) shows that in linear programs in $L^\infty[0, 1]$, extremal solutions have a bang-bang structure. This result can be adjusted to the multidimensional setting over the compact set \mathcal{X} using standard arguments. Details are omitted. \square

Bauer’s Maximum Principle (see Theorem 7.69 in Aliprantis and Border (2006)) states that every lower semicontinuous concave function has an extreme-point minimizer over a compact convex set. The feasible region W is convex because all constraints are linear. The compactness of W and the continuity of the objective of (4) were argued in the proof of Lemma OA1. Hence, by Bauer’s

Maximum Principle, an optimal extremal contract exists. Therefore, by [Lemma OA2](#), an optimal bang-bang contract exists. \square

Proof of Theorem 2. We use Barvinok (2002) Proposition IV.12.6, which is based on duality and complementary slackness, to characterize the structure of an optimal solution to our linear program over $L^\infty[0, 1]$. Adapted to our setting, this method involves setting a Lagrange multiplier μ for the IR constraint (4b) and Lagrange multipliers $\lambda_1, \lambda_2, \dots, \lambda_m$ for the constraints in (4c). Given a choice of nonnegative dual multipliers μ and λ_i , we define a function

$$p(\vec{x}) \triangleq \max\{0, -f(\vec{x}|\vec{a}^*) + \mu f(\vec{x}|\vec{a}^*) + \sum_{i=1}^m \lambda_i R_i(\vec{x}) f(\vec{x}|\vec{a}^*)\} \quad (\text{OA1})$$

and, by Barvinok (2002, Proposition IV.12.6), an optimal bang-bang contract has the form

$$w^*(\vec{x}) = \begin{cases} \bar{w} & \text{if } p(\vec{x}) > 0 \\ 0 & \text{if } p(\vec{x}) = 0 \end{cases}, \quad (\text{OA2})$$

assuming $\{\vec{x} : -f(\vec{x}|\vec{a}^*) + \mu f(\vec{x}|\vec{a}^*) + \sum_{i=1}^m \lambda_i R_i(\vec{x}) f(\vec{x}|\vec{a}^*) = 0\}$ has measure zero.

Observe that $p(\vec{x}) > 0$ if and only if

$$-1 + \mu + \sum_{i=1}^m \lambda_i R_i(\vec{x}) > 0, \quad (\text{OA3})$$

which can be rewritten as

$$\sum_{i=1}^m \omega_i R_i(\vec{x}) > \frac{1-\mu}{\sum_{i=1}^m \lambda_i},$$

where $\omega_i \triangleq \frac{\lambda_i}{\sum_{i=1}^m \lambda_i}$. In the first step, we divide both sides of the argument in the “max” in (OA1) by $f(\vec{x}|\vec{a}^*)$, which is positive for all $\vec{x} \in \mathcal{X}$. In the last step, we divide through by $\sum_{i=1}^m \lambda_i$ which we assume is nonzero. This assumption is without loss, because otherwise $\lambda_i = 0$ for all i , and so (OA3) either holds for all x or no x . Thus, if $\sum_{i=1}^m \lambda_i = 0$, the extremal contract is either constant at 0 or constant at \bar{w} . In either case, the contract is either not feasible (violates IR) or not optimal (pays out the maximum) and thus can be excluded from consideration. Observe that ω_i is nonnegative because λ_i is nonnegative for all i . Using this equivalence and defining

$$t \triangleq \frac{1-\mu}{\sum_{i=1}^m \lambda_i},$$

we may re-express the optimal bang-bang contract in (OA2) as¹⁶

$$w^*(\vec{x}) = \begin{cases} \bar{w} & \text{if } \sum_{i=1}^m \omega_i R_i(\vec{x}) \geq t. \\ 0 & \text{otherwise.} \end{cases}$$

¹⁶ Note we change the strict inequality in (OA2) to a weak inequality here. Because $p(\vec{x}) = 0$ is assumed to be a measure-zero event, this change can be made without loss.

This completes the proof. \square

Proof of Theorem 3. By Theorem 2, an optimal contract to (4) exists that is feasible to (8). Thus, the optimal value of (8) is at least the optimal value of (4). Moreover, because the feasible region of (8) is a restriction of the feasible region of (4) (i.e., it restricts to information-trigger contracts), the value of the former cannot exceed the value of latter. Together, this implies both problems have the same optimal value. An optimal solution (ω^*, t^*) of (8) yields the trigger contract

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

which is a feasible solution to (4). Moreover, w^* attains the optimal value of (4) because (ω^*, t^*) is optimal to (8) and the values of both problems are equal. Therefore, w^* is optimal to (4). \square

Proof of Lemma 1. When the context is clear, we lighten notation to $\Pr(I \leq i, S \leq s)$. First, note

$$\Pr(I \leq i, S \leq s) = \Pr(I \leq i, \min\{Q, I\} \leq s) = \Pr(I \leq i, Q < I, Q \leq s) + \Pr(I \leq s, Q \geq I)$$

because $s \leq i$. We can develop this derivation further by noting that

$$\begin{aligned} \Pr(I \leq i, Q < I, Q \leq s) &= \Pr(Q < I \leq i, Q \leq s) \\ &= \int_0^s \int_q^i f(j|e_o) g(q|e_m) dj dq \\ &= \int_0^s (F(i|e_o) - F(q|e_o)) g(q|e_m) dq \\ &= F(i|e_o) \int_0^s g(q|e_m) dq - \int_0^s F(q|e_o) g(q|e_m) dq \\ &= F(i|e_o) G(s|e_m) - \int_0^s F(q|e_o) g(q|e_m) dq. \end{aligned}$$

It is useful to further analyze this by integration by parts to conclude that

$$\Pr(I \leq i, Q < I, Q \leq s) = F(i|e_o) G(s|e_m) - F(s|e_o) G(s|e_m) + \int_0^s G(q|e_m) f(q|e_o) dq. \quad (\text{OA4})$$

Moreover,

$$\begin{aligned} \Pr(I \leq s, Q \geq I) &= \int_0^s \left(\int_j^{\bar{Q}} g(q|e_m) dq \right) f(j|e_o) dj \\ &= \int_0^s (1 - G(j|e_m)) f(j|e_o) dj \\ &= F(s|e_o) - \int_0^s G(j|e_m) f(j|e_o) dj. \end{aligned} \quad (\text{OA5})$$

From (OA4) and (OA5), we can conclude that

$$\Pr(I \leq i, S \leq s) = \Pr(I \leq i, Q < I, Q \leq s) + \Pr(I \leq s, Q \geq I)$$

$$\begin{aligned}
&= F(i|e_o)G(s|e_m) - F(s|e_o)G(s|e_m) + \int_0^s G(q|e_m)f(q|e_o)dq + F(s|e_o) - \int_0^s G(j|e_m)f(j|e_o)dj \\
&= F(s|e_o) + G(s|e_m)[F(i|e_o) - F(s|e_o)]
\end{aligned}$$

for $s < i$, and that if $s = i$,

$$\Pr(I \leq i, S \leq s) = F(i|e_o).$$

To conclude, we have derived the joint cumulative distribution function as

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

and complete the proof. \square

Proof of Proposition 1. This follows from implications of the MLRP assumptions stated in [Section 3](#).

To explore the implications of the MLRP assumption, it is useful to represent $R_{e_o, e_m}^{\text{NSO}}(i, s)$ and R_{e_o, e_m}^{SO} explicitly:

$$R_{e_o^H, e_m^L}^{\text{NSO}}(i, s) = 1 - \frac{g(s|e_m^L)}{g(s|e_m^H)}, \quad R_{e_o^L, e_m^H}^{\text{NSO}}(i, s) = 1 - \frac{f(i|e_o^L)}{f(i|e_o^H)}, \quad \text{and} \quad R_{e_o^L, e_m^L}^{\text{NSO}}(i, s) = 1 - \frac{f(i|e_o^L)g(s|e_m^L)}{f(i|e_o^H)g(s|e_m^H)}; \quad (\text{OA6})$$

$$R_{e_o^H, e_m^L}^{\text{SO}}(i) = 1 - \frac{1-G(i|e_m^L)}{1-G(i|e_m^H)}, \quad R_{e_o^L, e_m^H}^{\text{SO}}(i) = 1 - \frac{f(i|e_o^L)}{f(i|e_o^H)}, \quad \text{and} \quad R_{e_o^L, e_m^L}^{\text{SO}}(i) = 1 - \frac{f(i|e_o^L)(1-G(i|e_m^L))}{f(i|e_o^H)(1-G(i|e_m^H))}. \quad (\text{OA7})$$

The MLRP lends monotonicity to the ratios on the right-hand sides of (9) and (10) that allows us to reveal the structure of the bonus regions B^{NSO} and B^{SO} defined in (12) and (13), respectively.

We state these implications in the following lemma.

LEMMA OA3. *The following hold: (i) both $\frac{f(i|e_o^L)}{f(i|e_o^H)}$ and $\frac{1-G(i|e_m^L)}{1-G(i|e_m^H)}$ are nonincreasing and nonconstant in i ; and (ii) $\frac{g(s|e_m^L)}{g(s|e_m^H)}$ is nonincreasing and nonconstant in s .*

Finally, note $R_{e_o^H, e_m^L}^{\text{NSO}}(i, s)$ is constant in i and $R_{e_o^L, e_m^H}^{\text{NSO}}(i, s)$ is constant in s .

Returning to the proof of the proposition, let

$$s^*(i) \triangleq \min\{s : \varphi^{\text{NSO}}(i, s) = t\}, \quad (\text{OA8})$$

where

$$\varphi^{\text{NSO}}(i, s) \triangleq \sum_{e_o, e_m} \omega_{e_o, e_m} R^{\text{NSO}}(i, s). \quad (\text{OA9})$$

These objects are illustrated in [Figure OA.1](#). The critical inventory value i_s is the unique fixed point of s^* . That is, it is defined by

$$s^*(i_s) = i_s. \quad (\text{OA10})$$

We first note the domain of s^* need not be all $[0, \bar{I}]$, because for a given i , an s might not exist such that $\varphi^{\text{NSO}}(i, s) = t$. However, properties on the $R_{e_o, e_m}^{\text{NSO}}$ imply that once some \bar{i} and s exist such

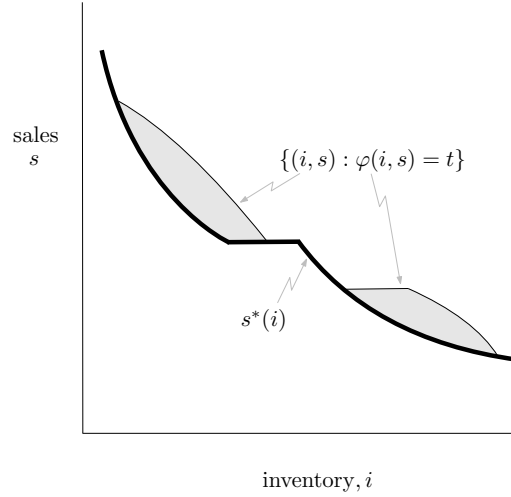


Figure OA.1 Illustration of the level set $\{(i, s) : \varphi^{\text{NSO}}(i, s) = t\}$ of the function φ^{NSO} (defined in (OA9)) and the function $s^*(i)$ (defined in (OA8)).

that $\varphi(\bar{i}, s) = t$ is nonempty, the same is true for any i larger than \bar{i} . That is, the domain of s^* is an interval of the form $[\bar{i}, \bar{I}]$.

Next, we show the mapping $s^*(i)$ is well behaved. Specifically, it is a nonincreasing, continuous, and almost everywhere differentiable function of i on its domain. The reasoning is as follows. As described in the paragraph above (OA8), each of the $R_{e_o, e_m}^{\text{NSO}}(i, s)$ are continuous, nonincreasing, and nonconstant in each of its coordinates. Hence, the same is true of the function $\varphi(i, s)$. Hence the level set $\{(i, s) : \varphi(i, s) = t\}$ has the structure illustrated in Figure OA.1. That is, the level set $\{(i, s) : \varphi(i, s) = t\}$ is the region between two nonincreasing and continuous functions. Observe that the graph of $s^*(i)$ in (i, s) -space is precisely the lower envelope of $\{(i, s) : \varphi(i, s) = t\}$. Thus, we can conclude s^* is a nonincreasing, continuous, and almost everywhere differentiable function of i .

Finally, we prove the existence of uniqueness of i_s . Because $s^*(i)$ is a nonincreasing and continuous function of i and we have assumed B^{NSO} has a positive measure, the arg min in (OA10) is nonempty (by Brouwer's Fixed Point Theorem (Corollary 17.56 in Aliprantis and Border (2006))) and is a singleton (because s^* is nonincreasing and the 45° line is strictly increasing). Hence, a unique choice exists for i_s . \square

Proof of Proposition 2. When B^{SO} has positive measure, we set

$$i_m = \min \{i : \varphi^{\text{SO}}(i) = t\},$$

where

$$\varphi^{\text{SO}}(i) \triangleq \sum_{e_o, e_m} \omega_{e_o, e_m} R_{e_o, e_m}^{\text{SO}}(i).$$

If B^{SO} has zero measure, we simply set $i_m = \bar{I}$.

To see that the set of i such that $\varphi^{\text{SO}}(i) = t$ is nonempty, observe that $\varphi^{\text{SO}}(\bar{I}) \geq t$ because (\bar{I}, \bar{I}) must be in every bonus region, and an i exists such that $\varphi^{\text{SO}}(i) < 0$. The latter follows because properties (OA6) and (OA7) imply $\int_{D^{\text{SO}}} R_{e_o, e_m}^{\text{SO}}(i, s) f(i|e_o) g(i|e_m) ds di \leq 0$ for all e_o, e_m , and so for all e_o, e_m , an i exists such that $R_{e_o, e_m}^{\text{SO}}(i) < 0$. Hence, by this continuity of $\varphi^{\text{SO}}(i)$, the set of i such that $\varphi^{\text{SO}}(i) = t$ is nonempty. Again, from the monotonicity properties of φ^{SO} , the bonus region in the stockout case is precisely as defined in Proposition 2. \square

Proof of Proposition 3. From the definition of $s^*(i)$, i_s , and i_m , the following holds for every valid choice of t and $(\omega_{e_o^H, e_m^L}, \omega_{e_o^L, e_m^L})$:

$$\begin{aligned} 1 - t &= \omega_{e_o^H, e_m^L} \frac{1 - G(i_m|e_m^L)}{1 - G(i_m|e_m^H)} + \omega_{e_o^L, e_m^H} \frac{f(i_m|e_o^L)}{f(i_m|e_o^H)} + \omega_{e_o^L, e_m^L} \frac{f(i_m|e_o^L)(1 - G(i_m|e_m^L))}{f(i_m|e_o^H)(1 - G(i_m|e_m^H))} \\ &= \omega_{e_o^H, e_m^L} \frac{g(s^*(i_s)|e_m^L)}{g(s^*(i_s)|e_m^H)} + \omega_{e_o^L, e_m^H} \frac{f(i_s|e_o^L)}{f(i_s|e_o^H)} + \omega_{e_o^L, e_m^L} \frac{f(i_s|e_o^L)g(s^*(i_s)|e_m^L)}{f(i_s|e_o^H)g(s^*(i_s)|e_m^H)} \\ &= \omega_{e_o^H, e_m^L} \frac{g(i_s|e_m^L)}{g(i_s|e_m^H)} + \omega_{e_o^L, e_m^H} \frac{f(i_s|e_o^L)}{f(i_s|e_o^H)} + \omega_{e_o^L, e_m^L} \frac{f(i_s|e_o^L)g(i_s|e_m^L)}{f(i_s|e_o^H)g(i_s|e_m^H)}, \end{aligned}$$

where the second equality follows because $s^*(i_s) = i_s$. MRLP distributions also have increasing failure rates, so we have $\frac{1 - G(\cdot|e_m^L)}{1 - G(\cdot|e_m^H)} < \frac{g(\cdot|e_m^L)}{g(\cdot|e_m^H)}$, which implies

$$\begin{aligned} &\omega_{e_o^H, e_m^L} \frac{g(i_s|e_m^L)}{g(i_s|e_m^H)} + \omega_{e_o^L, e_m^H} \frac{f(i_s|e_o^L)}{f(i_s|e_o^H)} + \omega_{e_o^L, e_m^L} \frac{f(i_s|e_o^L)g(i_s|e_m^L)}{f(i_s|e_o^H)g(i_s|e_m^H)} \\ &\leq \omega_{e_o^H, e_m^L} \frac{g(i_m|e_m^L)}{g(i_m|e_m^H)} + \omega_{e_o^L, e_m^H} \frac{f(i_m|e_o^L)}{f(i_m|e_o^H)} + \omega_{e_o^L, e_m^L} \frac{f(i_m|e_o^L)g(i_m|e_m^L)}{f(i_m|e_o^H)g(i_m|e_m^H)}. \quad \square \end{aligned}$$

Because f and g satisfy the MLRP, $\omega_{e_o^H, e_m^L} \frac{g(i|e_m^L)}{g(i|e_m^H)} + \omega_{e_o^L, e_m^H} \frac{f(i|e_o^L)}{f(i|e_o^H)} + \omega_{e_o^L, e_m^L} \frac{f(i|e_o^L)g(i|e_m^L)}{f(i|e_o^H)g(i|e_m^H)}$ is a decreasing function of i . As result, we can conclude $i_1^* \geq i_2^*$.

Proof of Proposition 5. We first prove (i). For $i < i_m$, notice that $w^*(s', i) = w^*(s'', i) = 0$ for all $(i, s'), (i, s'') \in D$ with $s' \leq s''$. For $i \in [i_m, i_s]$, observe that $w(s, i) = 0$ if $s < i$ and \bar{w} if $s = i$. Thus, $w^*(s', i) \leq w^*(s'', i) = 0$ for all $(i, s'), (i, s'') \in D$ with $s' \leq s''$. Finally, suppose $i > i_s$. By the structure of B^{NSO} and B^{SO} in Propositions 1 and 2, $w^*(s, i) = \bar{w}$ if $s^*(i) \leq s \leq i$, and 0 otherwise. Hence, again, $w^*(s', i) \leq w^*(s'', i) = 0$ for all $(i, s'), (i, s'') \in D$ with $s' \leq s''$. This implies w^* is monotone in s . We now show (ii). First, we show w^* is monotone in i . Suppose $s \geq i_s$. This implies $w(i, s) = \bar{w}$ for all $i \geq s$ (which is needed for $(i, s) \in D$) and so w^* is monotone in i in this region. Suppose otherwise that $s < i_s$. In this case, $w(i, s) = 0$ for $i \leq \min(s^*)^{-1}(s)$ and \bar{w} otherwise, which again yields monotonicity in i . Joint monotonicity now follows from monotonicity in both direction i and s . Turning now to (iii), if $i_m < i_s$, an $\epsilon > 0$ exists such that $w^*(i_m, i_m) = \bar{w}$ but $w^*(i_m + \epsilon, i_m) = 0$, and thus w^* is not monotone in i .

Finally, we show w^* also fails joint monotonicity when $i_m < i_s$. Consider the point $(i^\circ, s^\circ) \triangleq (i_s + \epsilon, s^*(i_s + \epsilon))$ on the graph of s^* . This means $w^*(i^\circ, s^\circ) = \bar{w}$ because the graph of s^* for $i \geq i_s$

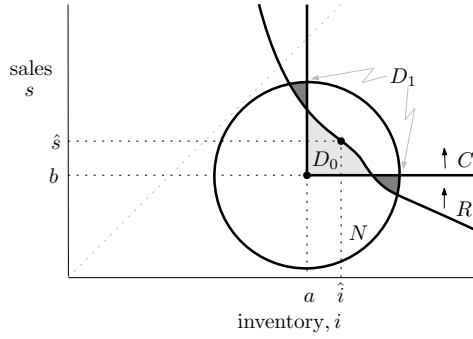


Figure OA.2 Illustration of the objects in the proof of **Proposition 6**.

lies in the bonus region of w^* . Choose any (i, s) in the open line segment between (i_m, i_m) and (i°, s°) with $i < i_s$. Such a choice is possible because $i_m < i_s$. See **Figure 3** for an illustration.

Note (i, s) does not lie in the bonus region. Indeed, $i < i_s$ (by construction) and $s^*(i) < i_s$ (because $s^\circ \leq i_s$ and s^* is a nonincreasing function). Hence, $w^*(i, s) = 0$. This yields a contradiction of monotonicity. Observe that $(i_m, i_m) < (i, s) < (i^\circ, s^\circ)$ but $w^*(i_m, i_m) = w^*(i^\circ, s^\circ) = \bar{w}$ and $w^*(i, s) = 0$.¹⁷ \square

Proof of Proposition 6. We prove both by noting that no optimal compensation plan can have a bonus region that has a “corner,” defined as follows. Let B denote the bonus region of an optimal contact, that is, where the store manager receives a positive bonus. Let (a, b) be a point in B and let $C \triangleq (a, b) + \mathbb{R}_+^n$ denote the (translated) cone of points that are pointwise no smaller than (\bar{i}, \bar{s}) . We say (a, b) is a corner point of B if (a, b) is an isolated extreme point of B . That is, a nonempty neighborhood N of (\bar{i}, \bar{s}) exists, such that $B \cap N = C \cap N$. This claim rules out corner compensation plans, but also any compensation plan with a “corner” as defined above.

We first prove the claim in the simplest case, in which on the no-jump constraint for (e_o^L, e_m^L) is tight, the other two no-jump constraints (for (e_o^H, e_m^L) and (e_o^L, e_m^H)) are slack. This case is the easiest to understand and will make clear what needs to be handled in the more challenging settings. Let w^* be an optimal compensation plan with a corner at (a, b) go to a new point $(\hat{i}, \hat{s}) = (a, b) + \delta(1, 1)$, where δ is chosen sufficiently small so that the no-jump constraints for (e_o^H, e_m^L) and for (e_o^L, e_m^H) remain slack. Now, define the set $R = \left\{ (i, s) : R_{e_o^L, e_m^L}^{\text{NSO}}(i, s) \geq R_{e_o^L, e_m^L}^{\text{NSO}}(\hat{i}, \hat{s}) \right\}$, which is the $R_{e_o^L, e_m^L}^{\text{NSO}}(\hat{i}, \hat{s})$ -superlevel sets of $R_{e_o^L, e_m^L}^{\text{NSO}}(i, s)$. Now, δ is also chosen sufficiently small so that both of the sets

$$D_0 \triangleq \{(i, s) : s < i\} \cap (C \setminus R)$$

$$D_1 \triangleq \{(i, s) : s < i\} \cap (R \cap N \setminus C)$$

¹⁷ Not this proof allows for the possibility that $s^*(i) = s^*(i_s)$ for all $i \geq i_s$, which cannot be ruled out under the MLRP assumptions.

have positive measure (where N is defined when we say (a, b) is a corner). We intersect both sets with $\{(i, s) : s < i\}$ so that we only deal with points in the interior of the domain of the optimal compensation plan w^* (which is only defined on $\{(i, s) : i \leq s\}$). Such a choice for δ is possible because the boundary of the set R expressed by the implicit function theorem by $r(i)$ as a function of i is a *strictly decreasing* function of i . The fact that $r(i)$ is both a function and is strictly decreasing is due to the assumption that f and g satisfy the strict MLRP. See Figure 4 for a visual representation of the sets D_0 and D_1 .

Now, consider the perturbation function h defined as follows:

$$h(s, i) = \begin{cases} -\epsilon_0 & \text{if } (i, s) \in D_0 \\ \epsilon_1 & \text{if } (i, s) \in D_1 \\ 0 & \text{otherwise,} \end{cases}$$

where $\epsilon_0, \epsilon_1 > 0$ are chosen so that $\mathbb{E}[h|e_o^H, e_m^H] = 0$. Such a choice is possible because D_0 and D_1 both have positive measure. Now consider the new optimal compensation plan $w' = w^* + h$. We claim w' is also an optimal compensation plan. Indeed,

$$\mathbb{E}[w'|e_o^H, e_m^H] = \mathbb{E}[w^*|e_o^H, e_m^H] + \mathbb{E}[h|e_o^H, e_m^H] = \mathbb{E}[w^*|e_o^H, e_m^H]$$

because $\mathbb{E}[h|e_o^H, e_m^H] = 0$. Thus, if we can show w' is feasible, it is optimal. Because we have assumed the IR constraint is not binding at w^* and ϵ_0 , and D_2 can be chosen sufficiently small so that the IR constraint is satisfied at w' (indeed, the condition that $\mathbb{E}[h|e_o^H, e_m^H] = 0$ is only a single linear constraint on ϵ_0 and D_2 , and thus a degree of freedom from that requirement allows us to drive ϵ_0 and D_2 arbitrarily small). Now consider the IC constraint. We claim that

$$\begin{aligned} \int_{s \leq i} R_{e_o^L, e_m^L}(i, s) w'(i, s) f(i|e_o^L) g(s|e_m^L) dids &> \int_{s \leq i} R_{e_o^L, e_m^L}(i, s) w^*(i, s) f(i|e_o^L) g(s|e_m^L) dids \quad (\text{OA11}) \\ &\geq c(e_o^H, e_m^H) - c(e_o^L, e_m^L) \end{aligned}$$

holds for all (e_o, e_m) . Notice the second inequality holds because w^* is feasible to the IC constraint (here we are taking the form of IC constraints from (4c)). Thus, it remains to show (OA11). This follows because

$$\begin{aligned} &\int_{s < i} R_{e_o^L, e_m^L}^{\text{NSO}}(i, s) h(s, i) f(i|e_o^L) g(s|e_m^L) dids \\ &= -\epsilon_0 \int_{D_0} R_{e_o^L, e_m^L}^{\text{NSO}}(i, s) f(i|e_o^L) g(s|e_m^L) dids + D_2 \int_{D_1} R_{e_o^L, e_m^L}^{\text{NSO}}(i, s) f(i|e_o^L) g(s|e_m^L) dids \\ &> R_{e_o^L, e_m^L}^{\text{NSO}}(a, b) \left[-\epsilon_0 \int_{D_0} f(i|e_o^L) g(s|e_m^L) dids \right] + R_{e_o^L, e_m^L}^{\text{NSO}}(\hat{i}, \hat{s}) \left[D_2 \int_{D_1} f(i|e_o^L) g(s|e_m^L) dids \right] \end{aligned}$$

$$> R(a, b) \left[-\epsilon_0 \int_{D_0} f(i|e_o^L)g(s|e_m^L) dids + D_2 \int_{D_1} f(i|e_o^L)g(s|e_m^L) dids \right] = 0,$$

where the key fact in each step is that $R_{e_o^L, e_m^L}^{\text{NSO}}(i, s)$ is coordinatewise strictly increasing. This implies (OA11). Indeed, observe that it suffices to integrate in the region $s < i$ to conclude the IC constraints bind for w' because $w'(i, s) = w^*(i, s)$ for $i = s$ (due to $h(i, s) = 0$ for $i = s$). We thus conclude w' is an optimal compensation plan.

In fact, we have shown something more in (OA11): the IC constraints are slack at optimal compensation plan w' . However, because problem (2) is linear, every optimal compensation plan must be on the boundary of the feasible region. This implies either w' must bind the IR constraint, which violates our assumption of positive rents, which is a contradiction. This completes the proof in the special case of the argument in which only the no-jump constraint for (e_o^L, e_m^L) was initially tight.

We now allow the possibility that at least one of the other two no-jump constraints is tight at (a, b) . In the case in which exactly one of the no-jump constraints for (e_o^H, e_m^L) or (e_o^L, e_m^H) is tight, an argument largely analogous to the previous one can be conducted. Here, the region D_1 will consist of only one “piece” (the D_1 in Figure OA.2 has two distinct pieces) because the no-jump constraints for (e_o^H, e_m^L) or (e_o^L, e_m^H) correspond to horizontal and vertical superlevel sets for $R_{e_o, e_m}^{\text{NSO}}(i, s)$ from the discussion following (OA7). The difficulty here is that when both of the no-jump constraints for (e_o^H, e_m^L) or (e_o^L, e_m^H) are tight, “room” remains to construct D_1 as in Figure OA.2. In this setting, we need to construct three regions, D_0 , D_1 , and D_2 , where a tradeoff exists between the value of the perturbation in regions D_1 and D_2 to lead to a strictly increasing covariance as we were able to show in (OA11).

To make this argument, we need the following definitions. We also move a distance δ along $(1, 1)$ to the point (\hat{i}, \hat{s}) . The superlevel set R and the region D_0 are defined exactly as before. To define D_1 and D_2 , we need the following concepts. At (a, b) , $R_{e_o^H, e_m^L}(i, s) \triangleq R_2(s)$ is constant in i at b with value $R_1(b)$ and so we consider the horizontal line through the point (a, b) . Where this line intersects R is denoted (i_1, b) . The region D_1 is chosen below the horizontal line to the right of (i_1, b) , above R , and inside N . Similarly, $R_{e_o^L, e_m^H}(i, s) \triangleq R_2(i)$ is constant in s at a (at the value $R_2(a)$). The region D_2 is chosen to the left of the vertical line above (a, s_2) and above the lower envelope of R . The specific sets D_1 and D_2 and δ are chosen so that an $\bar{\epsilon} > 0$ exists such that

$$\mathbb{E}[R_1(s)|D_1] - R_1(b) \leq \bar{\epsilon} \text{ and } \mathbb{E}[R_2(i)|D_2] - R_2(a) \leq \bar{\epsilon},$$

where (as used above) $\mathbb{E}[\cdot]$ is the expectation with respect to distributions with effort (e_o^H, e_m^H) . From the regions D_0 , D_1 , and D_2 , we define the perturbation

$$h(i, s) = \begin{cases} -\epsilon_0 & \text{if } (i, s) \in D_0 \\ \epsilon_1 & \text{if } (i, s) \in D_1 \\ \epsilon_2 & \text{if } (i, s) \in D_2 \\ 0 & \text{otherwise.} \end{cases}$$

Using a logic similar to the simpler case above, it suffices to show that $\mathbb{E}[h(S, I)] = 0$ and $\mathbb{E}[R(S, I)h(S, I)] > 0$. We now have three degrees of freedom, and so this is possible. For the condition $\mathbb{E}[h(S, I)] = 0$, this only requires

$$\epsilon_0 \mathbb{P}(D_0) = \epsilon_1 \mathbb{P}(D_1) + \epsilon_2 \mathbb{P}(D_2),$$

where $\mathbb{P}[\cdot]$ is the probability measure with respect to distributions with effort (e_o^H, e_m^H) . As for the covariance condition $\mathbb{E}[R(S, I)h(S, I)] > 0$, this analysis is more delicate. We first show how to find ϵ_1 and ϵ_2 such that

$$\begin{aligned} & \frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} \left(\epsilon_1 \int_{D_1} g(s)f(i)dsdi + \epsilon_2 \int_{D_2} g(s)f(i)dsdi \right) \\ & < \epsilon_1 \int_{D_1} R_1(s)g(s)f(i)dsdi + \epsilon_2 \int_{D_2} R_1(s)g(s)f(i)dsdi \end{aligned}$$

and

$$\begin{aligned} & \frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} \left(\epsilon_1 \int_{D_1} g(s)f(i)dsdi + \epsilon_2 \int_{D_2} g(s)f(i)dsdi \right) \\ & < \epsilon_1 \int_{D_1} R_2(i)g(s)f(i)dsdi + \epsilon_2 \int_{D_2} R_2(i)g(s)f(i)dsdi. \end{aligned}$$

To see that, we note the coefficient of ϵ_1 in the first inequality is

$$\int_{D_1} R_1(s)g(s)f(i)dsdi - \frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} \int_{D_1} g(s)f(i)dsdi > 0,$$

and the coefficient of ϵ_1 in the second inequality is

$$\int_{D_1} R_2(i)g(s)f(i)dsdi - \frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} \int_{D_1} g(s)f(i)dsdi < 0.$$

Therefore, the existence of such ϵ_1 and ϵ_2 depends on

$$\begin{aligned} & \frac{\frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} \int_{D_2} g(s)f(i)dsdi - \int_{D_2} R_2(i)g(s)f(i)dsdi}{\int_{D_1} R_2(i)g(s)f(i)dsdi - \frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} \int_{D_1} g(s)f(i)dsdi} \\ & > \frac{\epsilon_1}{\epsilon_2} > \frac{\frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} \int_{D_2} g(s)f(i)dsdi - \int_{D_2} R_1(s)g(s)f(i)dsdi}{\int_{D_1} R_1(s)g(s)f(i)dsdi - \frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} \int_{D_1} g(s)f(i)dsdi}. \end{aligned}$$

The sufficient condition is

$$\frac{\frac{\int_{D_2} R_2(i)g(s)f(i)dsdi}{\int_{D_2} g(s)f(i)dsdi} - \frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi}}{\frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} - \frac{\int_{D_1} R_2(i)g(s)f(i)dsdi}{\int_{D_1} g(s)f(i)dsdi}} > \frac{\frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} - \frac{\int_{D_2} R_1(s)g(s)f(i)dsdi}{\int_{D_2} g(s)f(i)dsdi}}{\frac{\int_{D_1} R_1(s)g(s)f(i)dsdi}{\int_{D_1} g(s)f(i)dsdi} - \frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi}}.$$

We can choose D_0 such that

$$\frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} - R_2(a) \leq \bar{\epsilon} \text{ and } \frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} - R_1(b) \leq \bar{\epsilon}.$$

Choose D_1 and D_2 to make $\frac{\int_{D_2} R_2(i)g(s)f(i)dsdi}{\int_{D_2} g(s)f(i)dsdi}$ as large as possible and $\frac{\int_{D_1} R_2(i)g(s)f(i)dsdi}{\int_{D_1} g(s)f(i)dsdi}$ as close to $R_2(a)$ as possible (indeed the distance is also controlled by $\bar{\epsilon}$), and $\frac{\int_{D_2} R_1(s)g(s)f(i)dsdi}{\int_{D_2} g(s)f(i)dsdi}$ as large as possible. Then, it suffices to have

$$\frac{\frac{\int_{D_2} R_2(i)g(s)f(i)dsdi}{\int_{D_2} g(s)f(i)dsdi} - \frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi}}{\frac{\int_{D_0} R_2(i)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} - R_2(a) + \bar{\epsilon}} > \frac{\frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi} - (R_1(\bar{s}) - \bar{\epsilon})}{\frac{\int_{D_1} R_1(s)g(s)f(i)dsdi}{\int_{D_1} g(s)f(i)dsdi} - \frac{\int_{D_0} R_1(s)g(s)f(i)dsdi}{\int_{D_0} g(s)f(i)dsdi}}.$$

Note both $\frac{\int_{D_2} R_2(i)g(s)f(i)dsdi}{\int_{D_2} g(s)f(i)dsdi}$ and $\frac{\int_{D_2} R_1(s)g(s)f(i)dsdi}{\int_{D_1} g(s)f(i)dsdi}$ can be sufficiently large. In the limit case, we can make

$$\frac{\frac{\int_{D_2} R_2(i)g(s)f(i)dsdi}{\int_{D_2} g(s)f(i)dsdi} - R_2(a)}{2\bar{\epsilon}} > \frac{2\bar{\epsilon}}{\frac{\int_{D_1} R_1(s)g(s)f(i)dsdi}{\int_{D_1} g(s)f(i)dsdi} - R_1(b)},$$

which is always possible. \square

Proof of Proposition 9: The proof of parts (i) and (ii) are similar to that of Proposition 1. Due to Theorem 2 there exists an optimal information trigger contract involving likelihood ratio functions $R_{e_o, e_m}(i, z)$ as described in (19) and (20).

For part (i), note that (*) equals

$$\frac{\int_i^{\bar{Q}} \frac{1}{q-i} \varphi\left(\frac{z-i}{q-i}\right) g(q|e_m^L) dq}{\int_i^{\bar{Q}} \frac{1}{q-i} \varphi\left(\frac{z-i}{q-i}\right) g(q|e_m^H) dq} = \frac{\int_i^{\bar{Q}} \frac{1}{q-i} \varphi\left(\frac{z-i}{q-i}\right) g(q|e_m^L) dq}{\int_i^{\bar{Q}} \frac{1}{q-i} \varphi\left(\frac{z-i}{q-i}\right) \frac{g(q|e_m^H)}{g(q|e_m^L)} g(q|e_m^L) dq} \leq \frac{1}{\frac{g(i|e_m^H)}{g(i|e_m^L)}},$$

which is bounded, where $\frac{g(q|e_m^H)}{g(q|e_m^L)} \geq \frac{g(i|e_m^H)}{g(i|e_m^L)}$ by MLRP. Now, as $\int_i^{\bar{Q}} \frac{1}{q-i} \varphi\left(\frac{z-i}{q-i}\right) g(q|e_m^L) dq \rightarrow \infty$, it must be that $\int_i^{\bar{Q}} \frac{1}{q-i} \varphi\left(\frac{z-i}{q-i}\right) \frac{g(q|e_m^H)}{g(q|e_m^L)} g(q|e_m^L) dq \rightarrow \infty$ as well. First, by changes of variables, $\varsigma = q - i$,

$$\frac{\int_0^{\bar{Q}-i} \frac{1}{\varsigma} \varphi\left(\frac{z-i}{\varsigma}\right) g(\varsigma + i|e_m^L) d\varsigma}{\int_0^{\bar{Q}} \frac{1}{\varsigma} \varphi\left(\frac{z-i}{\varsigma}\right) g(\varsigma + i|e_m^H) d\varsigma} = \lim_{\epsilon \rightarrow 0} \frac{\int_\epsilon^{\bar{Q}-i} \frac{1}{\varsigma} \varphi\left(\frac{z-i}{\varsigma}\right) g(\varsigma + i|e_m^L) d\varsigma}{\int_\epsilon^{\bar{Q}} \frac{1}{\varsigma} \varphi\left(\frac{z-i}{\varsigma}\right) g(\varsigma + i|e_m^H) d\varsigma}.$$

Therefore, by L'Hopital's rule, we have

$$\lim_{\epsilon \rightarrow 0} \frac{\int_\epsilon^{\bar{Q}-i} \frac{1}{\varsigma} \varphi\left(\frac{z-i}{\varsigma}\right) g(\varsigma + i|e_m^L) d\varsigma}{\int_\epsilon^{\bar{Q}} \frac{1}{\varsigma} \varphi\left(\frac{z-i}{\varsigma}\right) g(\varsigma + i|e_m^H) d\varsigma} = \lim_{\epsilon \rightarrow 0} \frac{\frac{1}{\epsilon} \varphi\left(\frac{z-i}{\epsilon}\right) g(\epsilon + i|e_m^L)}{\frac{1}{\epsilon} \varphi\left(\frac{z-i}{\epsilon}\right) g(\epsilon + i|e_m^H)} = \frac{g(i|e_m^L)}{g(i|e_m^H)},$$

which is independent of z . This establishes the result, because it also implies $R_{e_o, e_m}^{\text{LS}}(i, z) = R_{e_o, e_m}^{\text{NLS}}(i, z)$ at $z = i$, and so B^{NLS} has the structure expressed in the result.

For part (ii), we show $R^{\text{LS}}(i, z|e_o, e_m)$ is nondecreasing in z . This suffices to show an $i_\ell \in (0, \bar{I}]$ exists such that

$$B^{\text{LS}} = \{(i, z) : i \leq i_\ell \text{ and } \ell^*(i) \leq z \leq \bar{Q}\} \cup \{(i, z) : i \geq i_\ell \text{ and } i \leq z \leq \bar{Q}\},$$

using an argument analogous to that of [Proposition 1](#). The final part of the proof argues $i_\ell = i_s$, establishing the result.

To establish the first part of the argument, we first claim $(*)$ in [\(20\)](#) is nonincreasing in z under [Assumption 1](#). First, we can write

$$\int_i^{\bar{Q}} \gamma'(z|q, i) g(q|e_m) dq = \int_i^{\bar{Q}} \frac{\gamma'(z|q, i)}{\gamma(z|q, i)} \frac{g(q|e_m)}{g(q|e_m^*)} \gamma(z|q, i) g(q|e_m^*) dq.$$

Note $\frac{\gamma'(z|q, i)}{\gamma(z|q, i)}$ is nondecreasing in q by [Assumption 1](#), $\frac{g(q|e_m)}{g(q|e_m^*)}$ is nonincreasing in q . Therefore, the two functions $\frac{\gamma'(z|q, i)}{\gamma(z|q, i)}$ and $\frac{g(q|e_m)}{g(q|e_m^*)}$ are negatively correlated given any z . Therefore,

$$\frac{\int_i^{\bar{Q}} \frac{\gamma'(z|q, i)}{\gamma(z|q, i)} \frac{g(q|e_m)}{g(q|e_m^*)} \gamma(z|q, i) g(q|e_m^*) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq} \leq \frac{\int_i^{\bar{Q}} \frac{\gamma'(z|q, i)}{\gamma(z|q, i)} \gamma(z|q, i) g(q|e_m^*) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq} \frac{\int_i^{\bar{Q}} \frac{g(q|e_m)}{g(q|e_m^*)} \gamma(z|q, i) g(q|e_m^*) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq}.$$

By canceling out one of the $\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq$ in the denominator, we obtain

$$\begin{aligned} \int_i^{\bar{Q}} \frac{\gamma'(z|q, i)}{\gamma(z|q, i)} \frac{g(q|e_m)}{g(q|e_m^*)} \gamma(z|q, i) g(q|e_m^*) dq &\leq \int_i^{\bar{Q}} \frac{\gamma'(z|q, i)}{\gamma(z|q, i)} \gamma(z|q, i) g(q|e_m^*) dq \frac{\int_i^{\bar{Q}} \frac{g(q|e_m)}{g(q|e_m^*)} \gamma(z|q, i) g(q|e_m^*) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq} \\ &= \int_i^{\bar{Q}} \gamma'(z|q, i) g(q|e_m^*) dq \frac{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq}. \end{aligned}$$

It remains to show i_s and i_ℓ are equal. This requires that $R_{e_o, e_m}^{\text{LS}}(i, i) = R_{e_o, e_m}^{\text{NLS}}(i, i)$, which occurs if

$$\frac{\int_i^{\bar{Q}} \gamma(i|q, i) g(q|e_m) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq} = \frac{g(i|e_m)}{g(i|e_m^*)}.$$

But this holds because $\Lambda \in (0, 1)$, and so $\gamma(i|q, i) dq$ is the Dirac measure $\delta(i = q)$, thus yielding the result.

Proof of Proposition 10: It suffices to show $\frac{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq}$ is nonincreasing in i . We need

$$\frac{\partial}{\partial i} \left[\frac{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq} \right] \leq 0.$$

It is equivalent to show

$$\frac{-h(i|q, i) g(i|e_m) + \int_i^{\bar{Q}} \frac{\partial}{\partial i} \gamma(z|q, i) g(q|e_m) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m) dq} - \frac{-h(i|q, i) g(i|e_m^*) + \int_i^{\bar{Q}} \frac{\partial}{\partial i} \gamma(z|q, i) g(q|e_m^*) dq}{\int_i^{\bar{Q}} \gamma(z|q, i) g(q|e_m^*) dq} \leq 0.$$

The fact that

$$\frac{-h(i|q, i)g(i|e_m)}{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m)dq} + \frac{h(i|q, i)g(i|e_m^*)}{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m^*)dq} < 0$$

is due to **Assumption 1**:

$$\frac{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m^*)dq}{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m)dq} > \frac{g(i|e_m^*)}{g(i|e_m)}.$$

It remains to show

$$\frac{\int_i^{\bar{Q}} \frac{\partial}{\partial i} \gamma(z|q, i)g(q|e_m)dq}{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m)dq} \leq \frac{\int_i^{\bar{Q}} \frac{\partial}{\partial i} \gamma(z|q, i)g(q|e_m^*)dq}{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m^*)dq}.$$

By Young's Theorem, **Assumption 2** implies $\frac{\partial \gamma(z|q, i)}{\gamma(z|q, i)}$ is nondecreasing in q , and so we have

$$\begin{aligned} & \int_i^{\bar{Q}} \frac{\partial}{\partial i} \gamma(z|q, i)g(q|e_m)dq \\ &= \int_i^{\bar{Q}} \frac{\frac{\partial}{\partial i} \gamma(z|q, i)}{\gamma(z|q, i)} \frac{g(q|e_m)}{g(q|e_m^*)} \gamma(z|q, i)g(q|e_m^*)dq \\ &\leq \int_i^{\bar{Q}} \frac{\frac{\partial}{\partial i} \gamma(z|q, i)}{\gamma(z|q, i)} \gamma(z|q, i)g(q|e_m^*)dq \frac{\int_i^{\bar{Q}} \frac{g(q|e_m)}{g(q|e_m^*)} \gamma(z|q, i)g(q|e_m^*)dq}{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m^*)dq} \\ &= \int_i^{\bar{Q}} \frac{\partial}{\partial i} \gamma(z|q, i)g(q|e_m^*)dq \frac{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m)dq}{\int_i^{\bar{Q}} \gamma(z|q, i)g(q|e_m^*)dq}, \end{aligned}$$

which shows the desired inequality.

Proof of Proposition 12. First, note $f(i|e_o)$ may potentially involve \bar{I} , so we do some transformation. Let $\tau = i/\bar{I}$ be the ratio. Suppose $\tau \in [0, 1]$ has a distribution $\tilde{F}(\tau|e_o)$, which is independent of \bar{I} . Then, $F(i|e_o) = \tilde{F}(i/\bar{I}|e_o)$ and $f(i|e_o) = \frac{1}{\bar{I}} \tilde{f}(i/\bar{I}|e_o)$.

We denote $\tilde{s}(\mu, \lambda, \tau)$ as the cut-off solving $\mu + \sum_j \lambda_j (1 - \frac{g(s|\hat{e}_m^j) \tilde{f}(\tau|\hat{e}_o^j)}{g(s|e_m^*) \tilde{f}(\tau|e_o^*)}) = 1$ for $s < \tau \bar{I}$, where j is the index for the j -th IC constraint. And denote $\tau_s(\mu, \lambda)$ as the minimum solution for $\tilde{s}(\mu, \lambda, \tau) = \tau \bar{I}$, $\tau_m(\mu, \lambda)$ as the minimum solution for $\mu + \sum_j \lambda_j (1 - \frac{(1-G(\tau \bar{I}|\hat{e}_m^j)) \tilde{f}(\tau|\hat{e}_o^j)}{(1-G(\tau \bar{I}s|e_m^*)) \tilde{f}(\tau|e_o^*)}) = 1$.

By strong duality, and **Theorem 2**, we have

$$\begin{aligned} -W(\bar{I}) &= \min_{\mu, \lambda} \phi(\mu, \lambda, \bar{I}) \\ &= \min_{\mu, \lambda} \int_{\tau_s(\mu, \lambda)}^1 \int_{\tilde{s}(\mu, \lambda, \tau)}^{\tau \bar{I}} \left(-1 + \mu + \sum_j \lambda_j \left(1 - \frac{g(s|\hat{e}_m^j) \tilde{f}(\tau|\hat{e}_o^j)}{g(s|e_m^*) \tilde{f}(\tau|e_o^*)} \right) \right) g(s|e_m^*) \tilde{f}(\tau|e_o^*) ds d\tau \\ &\quad + \int_{\tau_m(\mu, \lambda)}^1 \left(-1 + \mu + \sum_j \lambda_j \left(1 - \frac{(1-G(\tau \bar{I}|\hat{e}_m^j)) \tilde{f}(\tau|\hat{e}_o^j)}{(1-G(\tau \bar{I}s|e_m^*)) \tilde{f}(\tau|e_o^*)} \right) \right) (1-G(\tau \bar{I}s|e_m^*)) \tilde{f}(\tau|e_o^*) d\tau \\ &\quad - \sum_j \lambda_j [c(e^*) - c(\hat{e}^j)] - \mu [c(e^*) + \underline{U}]. \end{aligned}$$

Note the dual is convex in (μ, λ) . Let (μ^*, λ^*) be the solution of the dual minimization. Then, by the envelope theorem, we have

$$\begin{aligned}
-\frac{dW(\bar{I})}{d\bar{I}} &= \frac{\partial}{\partial \bar{I}} \phi(\mu^*, \lambda^*, \bar{I}) \\
&= \int_{\tau_s(\mu^*, \lambda^*)}^1 \tau \left(-1 + \mu^* + \sum_j \lambda_j^* \left(1 - \frac{g(\tau \bar{I} | \hat{e}_m^j) \tilde{f}(\tau | \hat{e}_o^j)}{g(\tau \bar{I} | e_m^*) \tilde{f}(\tau | e_o^*)} \right) \right) g(\tau \bar{I} | e_m^*) \tilde{f}(\tau | e_o^*) d\tau \\
&\quad - \int_{\tau_m(\mu^*, \lambda^*)}^1 \tau \left((-1 + \mu^*) g(\tau \bar{I} | e_m^*) \tilde{f}(\tau | e_o^*) + \sum_j \lambda_j^* [g(\tau \bar{I} | e_m^*) \tilde{f}(\tau | e_o^*) - g(\tau \bar{I} | \hat{e}_m^j) \tilde{f}(\tau | \hat{e}_o^j)] \right) d\tau \\
&= - \int_{\tau_m(\mu^*, \lambda^*)}^{\tau_s(\mu^*, \lambda^*)} \tau \left(-1 + \mu^* + \sum_j \lambda_j^* \left(1 - \frac{g(\tau \bar{I} | \hat{e}_m^j) \tilde{f}(\tau | \hat{e}_o^j)}{g(\tau \bar{I} | e_m^*) \tilde{f}(\tau | e_o^*)} \right) \right) g(\tau \bar{I} | e_m^*) \tilde{f}(\tau | e_o^*) d\tau \\
&> - \left(-1 + \mu^* + \sum_j \lambda_j^* \left(1 - \frac{g(\tau_s(\mu^*, \lambda^*) \bar{I} | \hat{e}_m^j) \tilde{f}(\tau_s(\mu^*, \lambda^*) | \hat{e}_o^j)}{g(\tau_s(\mu^*, \lambda^*) \bar{I} | e_m^*) \tilde{f}(\tau_s(\mu^*, \lambda^*) | e_o^*)} \right) \right) \int_{\tau_m(\mu^*, \lambda^*)}^{\tau_s(\mu^*, \lambda^*)} \tau g(\tau \bar{I} | e_m^*) \tilde{f}(\tau | e_o^*) d\tau = 0,
\end{aligned}$$

where the inequality is by MLRP and $\tau_s(\mu^*, \lambda^*) \geq \tau_m(\mu^*, \lambda^*)$.

OA.2. Numerical Illustration of a Mast-and-Sail Compensation Plan

We now use a numerical example to illustrate an optimal bonus region of the “mast and sail” structure (as seen in [Figure 2\(a\)](#)) in the case of outcome distributions that satisfy the MLRP. This example shows the delicacy of numerical computation in this setting, which is typical of moral hazard problems. Our extensive numerical simulations use similar logic to that found in this example.

EXAMPLE OA.1. Consider the following instance of the multitasking store manager. The distribution functions of operating and marketing effort are $F(i|e_o) = i^{e_o}$ and $G(s|e_m) = s^{e_m}$, respectively, where $e_o \in \{e_o^L, e_o^H\}$ 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 $c(e_o^H, e_m^H) = 5$, $c(e_o^H, e_m^L) = c(e_o^L, e_m^H) = 4$ and $c(e_o^L, e_m^L) = 2$. The resource constraint for the firm has $\bar{w} = 10$.

For now, we suppose an optimal choice of ω (as guaranteed to exist by [Theorem 2](#)) has $\omega_{e_o^L, e_m^L} = 1$ and $\omega_{e_o^H, e_m^L} = \omega_{e_o^L, e_m^H} = 0$. We construct the associated trigger value t below, and also show the resulting compensation plan with these choices of parameters is indeed feasible to (4) and thus optimal.

The condition $R_{e_o, e_m}^{\text{NSO}}(i, s) \geq t$ can be expressed as

$$1 - t \geq \frac{f(i|e_o)g(s|e_m)}{f(i|e_o^H)g(s|e_m^H)} = \frac{e_o}{e_o^H} \frac{e_m}{e_m^H} i^{e_o - e_o^H} s^{e_m - e_m^H}$$

and the condition $R_{e_o, e_m}^{\text{SO}}(i) \geq t$ amounts to

$$1 - t \geq \frac{e_o i^{e_o - 1} (1 - i^{e_m})}{e_o^H i^{e_o^H - 1} (1 - i^{e_m^H})}.$$

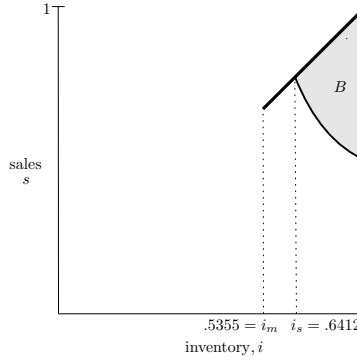


Figure OA.3 The bonus region of the optimal compensation plan for **Example OA.1**.

These two conditions are equivalent to (suppose $e_o < e_o^H$)

$$i \geq \begin{cases} \frac{1}{4s(1-t)} & \text{if } s < i \\ i_m(t) & \text{if } s = i \end{cases},$$

where $i_m(t) \triangleq \frac{-1 + \sqrt{1+2/(1-t)}}{2}$. Therefore, the bonus region B^{NSO} is

$$\{(i, s) : \frac{1}{4i(1-t)} \leq s < i \text{ and } i \geq i_s(t)\},$$

where $i_s(t) \triangleq \frac{1}{2\sqrt{1-t}}$.

The next step is to determine t . Because $\omega_{e_o^L, e_m^L} = 1$, the NJ constraint between (e_o^H, e_m^H) and (e_o^L, e_m^L) is tight. We can isolate for t in the resulting equality to determine t . The tight NJ constraint is

$$\int_{R(i, s|e) \geq t} R(i, s) f(i|e_o^*) g(s|e_m^*) ds di = (c(e_o^H, e_m^H) - c(e_o^L, e_m^L)) \cdot \bar{w}^{-1},$$

which can be rewritten as

$$\int_{i_s(t)}^1 \int_{\frac{1}{4i(1-t)}}^i (4si - 1) ds di + \int_{i_m(t)}^1 (2i(1 - i^2) - (1 - i)) di = 0.3.$$

Solving for t results in $t^* \simeq 0.3919$.

We check whether the trigger compensation plan w^* with $t^* \simeq 0.3919$, $\omega_{e_o^L, e_m^L} = 1$, and $\omega_{e_o^H, e_m^L} = \omega_{e_o^L, e_m^H} = 0$ is an optimal compensation plan. It suffices to check that the remaining no-jump constraints are feasible. We first check that no profitable deviation to $e = (e_o^H, e_m^L)$ exists. Under the trigger compensation plan, the marginal revenue of deviation to (e_o^H, e_m^L) is

$$\bar{w} \int_{i_s(t)}^1 \int_{\frac{1}{4i(1-t)}}^i (4si - 2i) ds di + \bar{w} \int_{i_m(t)}^1 (2i(1 - i^2) - 2i(1 - i)) di.$$

Plugging $t^* \simeq 0.3919$ into the above object yields $1.748 > 1 = c(e_o^H, e_m^H) - c(e_o^H, e_m^L)$. Therefore, the store manager will not deviate to (e_o^H, e_m^L) .

Similarly, the store manager's marginal revenue of deviation to (e_o^L, e_m^H) is

$$\bar{w} \int_{i_s(t)}^1 \int_{\frac{1}{4i(1-t)}}^i (4si - 2s) ds di + \bar{w} \int_{i_m(t)}^1 (2i(1 - i^2) - (1 - i^2)) di.$$

Plugging $t^* \simeq 0.3919$ into the above object yields $1.864 > 1 = c(e_o^H, e_m^H) - c(e_o^L, e_m^H)$. Therefore, the store manager will not deviate to (e_o^L, e_m^H) . The bonus region is illustrated in [Figure OA.3](#). Note $i_m(t^*) = 0.5355 < i_s(t^*) = 0.6412$, and so we have a mast-and-sail bonus region as plotted in [Figure OA.3](#). Under this optimal compensation plan, the probability of payout under the optimal mast-and-sail compensation plan is 51.96%.

OA.3. Numerical Illustrations of Corner Compensation Plans

To give a sense of how to compute the optimal corner compensation plan we give two concrete examples. These examples refer to some auxiliary results in the same online appendix that help us analyze corner compensation plans.

EXAMPLE OA.2 ([EXAMPLE OA.1](#) CONTINUED). Returning to the set up in [Example OA.1](#), the optimal corner compensation plan solves problem (14) and can be simply stated as:

$$\begin{aligned} \min_{a,b} \quad & (1 - b^2)(1 - a^2) \\ \text{s.t.} \quad & (1 - b^2)(1 - a^2) - (1 - b)(1 - a) \geq 0.3 \\ & (1 - b^2)(1 - a^2) - (1 - b^2)(1 - a) \geq 0.2 \\ & (1 - b^2)(1 - a^2) - (1 - b)(1 - a^2) \geq 0.2 \\ & b \leq a. \end{aligned}$$

Note we have formulated the objective to minimize the expected probability of paying out the bonus \bar{w} (following [Proposition 7](#)), which is equivalent to the objective function (14a).

Now, the optimality condition for the optimal corner compensation plan

$$\frac{\mathcal{H}^f(a|e_o^*)}{\mathcal{H}^g(b|e_m^*)} = \frac{\mathcal{H}^f(a|e_o)}{\mathcal{H}^g(b|e_m)}$$

from equation (15) implies

$$\frac{\frac{2a}{(1-a^2)}}{\frac{2b}{(1-b^2)}} = \frac{\frac{1}{(1-a)}}{\frac{1}{(1-b)}},$$

which yields $a = b$. Combined with the other case of [Proposition 8](#), we have $a = b$ in optimality. In this case, we can show the optimal choice of ω has $\omega_{e_o^L, e_m^L} = 1$ and $\omega_{e_o^H, e_m^L} = \omega_{e_o^L, e_m^H} = 0$. The optimal corner compensation plan can be obtained through solving $(1 - b^2)(1 - b^2) - (1 - b)^2 = 0.3$, which yields $b^* = 0.5234$. Under the optimal corner compensation plan, the probability of payout is 0.5271. By comparison, the probability of payout under the optimal mast-and-sail compensation plan is 0.5196. So the performance gap is $(0.5271 - 0.5196)/0.5196 = 1.44\%$.

EXAMPLE OA.3. Consider an instance with the same setting as in [Example OA.1](#) except that the cost functions are as follows: $c(e_o^H, e_m^H) = 3.4$, $c(e_o^H, e_m^L) = 1.5$, $c(e_o^L, e_m^H) = 1.8$, and $c(e_o^L, e_m^L) = 1$. In this case, we can show the optimal choice of ω has $\omega_{e_o^H, e_m^L} = 1$ and $\omega_{e_o^L, e_m^L} = \omega_{e_o^L, e_m^H} = 0$. Thus, we can solve for b^* by setting $(1 - b^2)(1 - b^2) - (1 - b^2)(1 - b) = 0.19$, which yields $b^* = 0.4896$, which corresponds to a probability of payout of 0.5781. By comparison, the optimal mast-and-sail compensation plan has a probability of payout of 0.5148. Thus, using a corner compensation plan leads to an efficiency loss of $(0.5781 - 0.5148)/0.5148 = 12.30\%$.

OA.4. Fully Observed Demand

A natural benchmark is to look at the scenario where demand is fully observed and not censored by inventory; that is, both the firm and store manager can observe Q . This situation is much simpler than the case of censored demand. The analysis in [Section 4](#) still applies and it can be shown that $R_{e_o, e_m}(i, q) = 1 - \frac{f(i|e_o)g(q|e_m)}{f(i|e_o^*)g(q|e_m^*)}$, which is precisely what we analyzed before as $R_{e_o, e_m}^{NSO}(i, s)$ in [\(9\)](#) for $s \leq i$. Similar reasoning thus yields the result:

PROPOSITION OA1. *A nonincreasing and continuous function q^f exists such that an optimal contract w^f to the fully observed demand exists with the form*

$$w^f(i, q) = \begin{cases} \bar{w} & \text{if } (i, q) \in B^f \\ 0 & \text{otherwise,} \end{cases}$$

where $B^f = \{(i, q) : q \geq q^f(i)\}$.

Recall, just as in [Proposition 1](#) for s^* , the domain of the function q^f need not be all of $[0, \bar{I}]$, and an \bar{i} may exist such that $B^f \subseteq [\bar{i}, \bar{I}] \times [0, \bar{Q}]$. Thus, the bonus region of the optimal contract in the fully observed demand case is an “expanded sail” that is not truncated by the 45° line.

Next, we use a numerical experiment to demonstrate the efficiency loss due to demand censoring, by comparing the firm’s expected additional payment to the store manager — compared to the first-best scenario — for inducing the same effort level. We illustrate our result in [Figure OA.4](#).

OA.5. Approximating Mast-and-Sail Contracts

In this section, we consider approximating mast-and-sail contracts when we allow these contracts to be nonmonotone in nature, in contrast to [Section 7.2](#), where attention was restricted to monotone approximations. The purpose of this exercise is to understand what aspects of the mast-and-sail structure are driving optimality.

We consider the following three nonmonotone approximations, in addition to weighted-sum threshold policies:

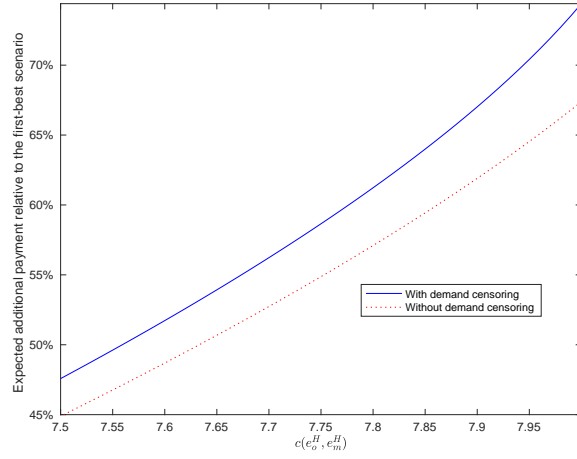


Figure OA.4 Expected additional payment relative to the first-best scenario with and without demand censoring. **Parameters:** $\bar{w} = 20$, $c(e_o^L, e_m^L) = 1$ and $c(e_o^H, e_m^L) = c(e_o^L, e_m^H) = 5$. We vary $c(e_o^H, e_m^H)$ between 7.5 and 8.

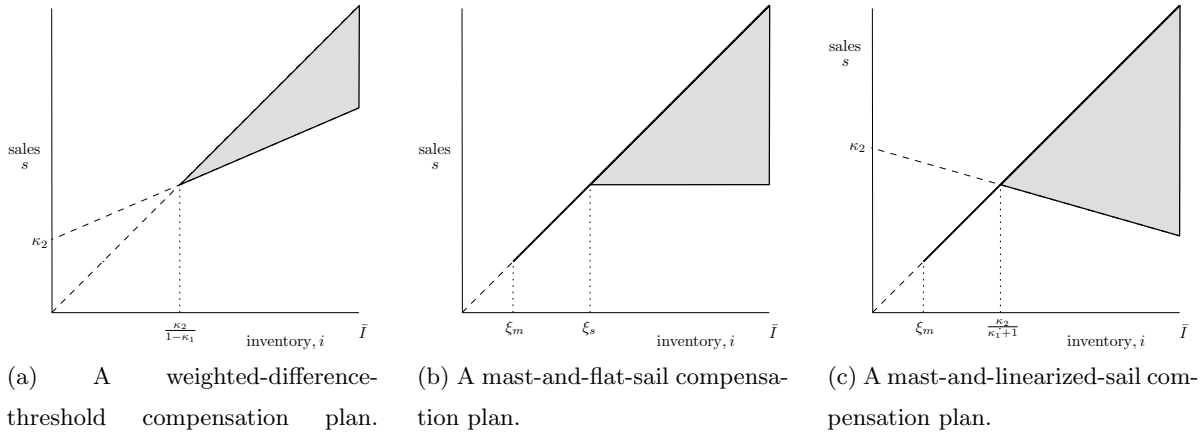


Figure OA.5 Nonmonotone simplifications of the mast-and-sail contracts.

- (a) *Weighted-difference threshold compensation plan*: This compensation plan is similar to a weighted-sum threshold compensation plan, except that the threshold for the bonus payout is a weighted *difference* of the sales quantity and inventory level. Specifically, the agent receives a bonus if the sales quantity s and inventory level i satisfy $s - \kappa_1 \cdot i \geq \kappa_2$, for some $\kappa_1, \kappa_2 \geq 0$.
- (b) *Mast-and-flat-sail compensation plan*: This compensation plan has both a “mast” and a “sail” except that the bottom of the sail is flattened. It has two parameters, ξ_m and ξ_s , where $0 < \xi_m \leq \xi_s < 1$, such that the agent receives a bonus under one of the following two conditions: (a) the sales quantity $s \geq \xi_m$ and all inventory is cleared and (b) the sales quantity $s \geq \xi_s$ and not all inventory is cleared.

- (c) *Mast-and-linearized-sail compensation plan*: This simple contract most closely mimics the structure of the mast-and-sail contract and has three parameters, ξ_m , κ_1 and κ_2 , all nonnegative, such that the agent receives a bonus under one of the following two conditions: (a) the sales quantity $s \geq \xi_m$ and all inventory is cleared or (b) the sales quantity s and realized inventory level i satisfy $s + \kappa_1 \cdot i \geq \kappa_2$ and not all inventory is cleared.

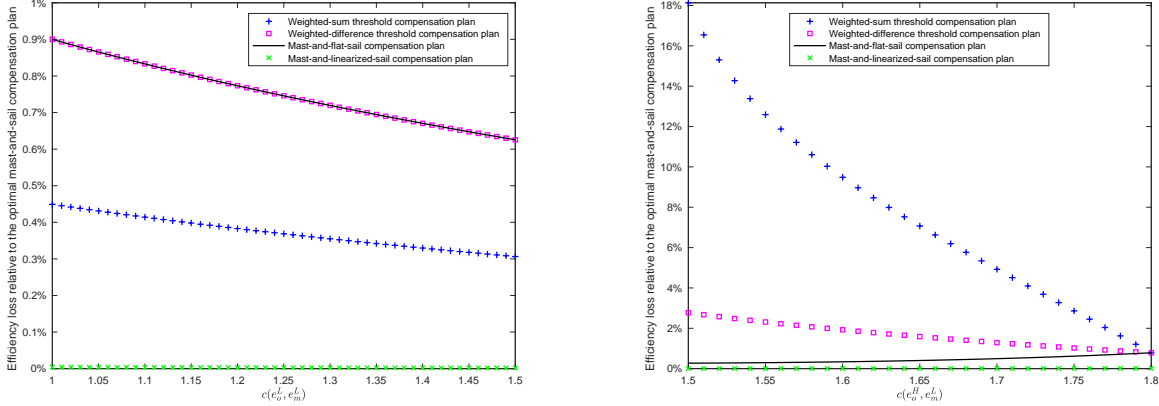
These contract types are illustrated visually in [Figures OA.5\(a\)](#) and [OA.5\(c\)](#). One can view the mast-and-flat-sail compensation plan as a “demand-censoring-aware” threshold policy because it can be viewed as a generalized sales-quota-bonus compensation plan with two thresholds, depending on whether inventory is cleared. Similarly, the mast-and-linearized-sail compensation plan is a “demand-censoring-aware” weighted-sum compensation plan. In this case, however, the linear portion is always downward sloping, because it mimics the nonincreasing function s^* .

We evaluate the performance of the above three compensation plans (along with weighted-sum threshold contracts) through extensive numerical experiments, and illustrate two representative scenarios in [Figure OA.6](#). We draw the following observations. First, the performance of the weighted-sum threshold compensation plan is similar to that of the corner (or sales-quota-bonus) compensation plan in that it is near optimal when the marketing and operational activities are sufficiently complementary (see [Figure OA.6\(a\)](#)), yet far from the optimal otherwise (see [Figure OA.6\(b\)](#)). Interestingly, the weighted-difference threshold compensation plan performs reasonably well in the latter case.

Second, the mast-and-flat-sail compensation plan, despite having both a “mast” and a “sail,” can be outperformed by a “triangular sail” of the weighted-sum threshold described in [Section 7.2](#) (see [Figure OA.6\(a\)](#)). However, the shape of the triangular sail is important here. Contracts with two “pieces” (a mast and something resembling a sail) can tailor incentives to the “cleared inventory” and “not cleared inventory” cases in isolation from each other. A single “piece” contract (like a weighted threshold compensation plan) needs to effectively “bridge” between these two cases by designing the transition from “cleared inventory” to “not cleared inventory.” A perfect “bridge” is not always possible to build, but our experiments show single-piece contracts with strong performance are possible if constructed correctly. The correct construction (whether a weighted-sum or weighted-difference threshold) is the one that most closely mimics the structure of the underlying optimal mast-and-sail contract. For instance, a weighted-difference contract mimics the case in which the bottom tip of the mast is lower than the bottom tip of the sail. Note, however, that weighted-difference compensation plans are also not monotone, again underscoring the fact that non-monotonicity is somehow endemic to the case of pure demand censoring explored here.

This leads to our final observation. The performance of the mast-and-linearized-sail compensation plan is consistently near optimal (or optimal). It treats the “cleared inventory” case and “not

cleared inventory” case individually and with relatively little restriction (except that the sail is triangular). Thus, it gains many of the benefits of the mast-and-sail compensation plan but remains simple and easier to compute. The drawback, of course, is that it remains non-monotone and thus susceptible to *ex post* manipulation of inventory just as in the mast-and-sail setting.



(a) Parameters: $\bar{w} = 10$, $c(e_o^H, e_m^L) = c(e_o^L, e_m^H) = 3.0$, and $c(e_o^H, e_m^H) = 3.5$. We vary $c(e_o^L, e_m^L)$ between 1 and 1.5. (b) Parameters: $\bar{w} = 10$, $c(e_o^L, e_m^L) = 1$, $c(e_o^L, e_m^H) = 1.8$, and $c(e_o^H, e_m^H) = 3.5$. We vary $c(e_o^H, e_m^L)$ between 1.5 and 1.8.

Figure OA.6 Performance of the optimal weighted-sum threshold , weighted-difference threshold, mast-and-flat-sail, and mast-and-linearized-sail compensation plans, relative to the optimal mast-and-sail compensation plan. We assume the same random distributions for I and S as in **Figure OA.5**.

OA.6. Single-tasking versus Multitasking

A natural question is to compare a multitasking manager to a single-tasking manager responsible for only one of these tasks. In this section, we ask about the optimal incentives for an inventory manager who finds himself in the same situation as the store manager in the main body of the paper, but cannot influence demand through his efforts (i.e., he can only undertake operational effort). Building on our analysis of the general multitasking setting in **Section 4** (in particular **Theorem 2**), one can show the optimal contract for the inventory manager problem is an inventory quota-bonus compensation plan.

PROPOSITION OA2. *There exist nonnegative multipliers ω and a “target” t such that an optimal solution to the single-tasking inventory manager problem of the following form exists:*

$$w^*(i, s) = \begin{cases} \bar{w} & \text{if } R(i, s) \geq t \\ 0 & \text{otherwise} \end{cases},$$

where now $R(i, s) = 1 - \frac{f(i|e^L)}{f(i|e^H)}$, and where the goal is to implement high inventory effort.

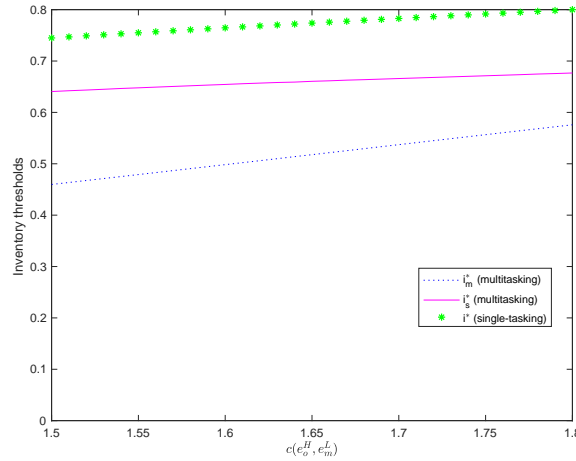


Figure OA.7 Comparison of inventory thresholds under single-tasking versus multitasking. All the parameters are the same as in **Figure 5(b)**

Under the MLRP assumption, it is straightforward to show the condition $R(i, s) \geq t$ translates into an inventory threshold i^* where

$$w^*(i, s) = \begin{cases} \bar{w} & \text{if } i \geq i^*. \\ 0 & \text{if } i < i^*. \end{cases} \quad (\text{OA12})$$

Given the optimality of a quota-bonus structure for the inventory management problem, an interesting question is whether the inventory threshold should be higher for the single-tasking agent than the multitasking agent. We explore this question numerically. Consider the same set of parameters as in **Figure 5(b)**. In the case of single-tasking, we fix the marketing effort e_m as e_m^H and look for an inventory threshold i^* specified in (OA12). We observe from **Figure OA.7** that under single-tasking, the firm consistently chooses an inventory threshold i^* that is higher than both inventory thresholds (i.e., i_s^* and i_m^*) derived from the multitasking setting. In particular, this result implies an inventory manager has a more stringent requirement for earning a bonus in terms of inventory than a multitasker who is responsible for inventory (among other things).

OA.7. Effect of Compensation Ceiling on Compensation Plans and Firm Profitability

We now numerically illustrate the effect of the compensation ceiling (\bar{w}) on the design of the optimal compensation plan and the firm's profitability. Consider a case in which all the parameters are the same as in **Figure 5(a)** except that (a) we fix $c(e_o^L, e_m^L) = 1.0$ and (b) we vary the value of \bar{w} from 9 to 21. We compute the optimal compensation plans and plot three scenarios in **Figure OA.8**, corresponding to the cases of $\bar{w} = 9, 15$, and 21, respectively. In addition, we compute the firm's

expected cost of compensating the store manager for each combination of parameters. We do not explicitly report the firm's expected profit, which also depends on its unit revenue; a lower compensation cost suggests a higher expected profit. Figure OA.9 shows the relationship between \bar{w} and the firm's expected cost of compensation.

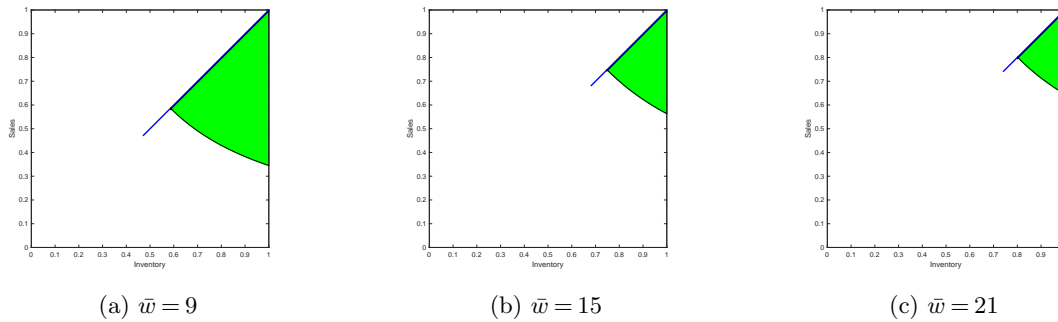


Figure OA.8 The effect of compensation ceiling on the optimal bonus region. In each panel, the blue line shows the “mast” part of the bonus region, whereas the green area shows the “sail” part.

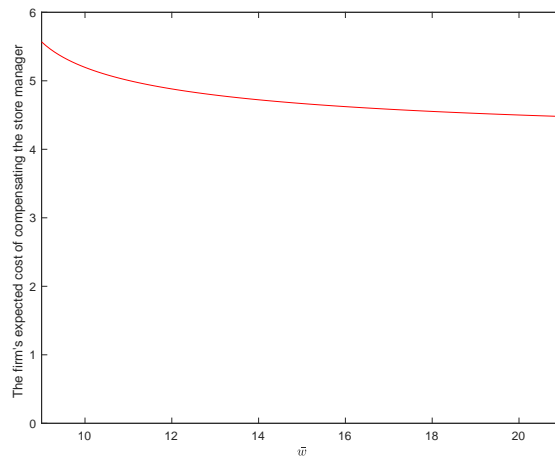


Figure OA.9 The effect of \bar{w} on the firm's expected cost of compensating the store manager.

OA.8. Microfounding the Compensation Ceiling

In this section, we attempt to microfound the compensation ceiling by examining the case in which the agent's probability of receiving the bonus cannot be below a pre-specified threshold (denoted by \mathcal{P}). Our analysis is motivated by the observation that in practice, firms aim to ensure each agent, when opting to exert effort, can receive the bonus with a reasonably high likelihood.

In an influential book on salesforce compensation, Zoltners et al. (2006, pp. 53–57) point out two key factors in the effect of contracts to motivate effort in salesforce staff — engagement (measured by an “engagement rate”) and excitement (measured by an “excitement index”). The engagement rate “measures what percentage of a sales force receives incentive pay with a plan,” whereas the excitement index measures “the rate at which salespeople earn their last incremental dollar” (in our case, the bonus). Clearly, a tension exists between the engagement rate and excitement index. Zoltners et al. (2006) states that too low an engagement rate can hurt the morale of the salesforce. Although we are examining a multitasking setting, we believe the engagement-excitement tradeoff characterized by Zoltners et al. (2006) is relevant to our setting because a higher compensation ceiling (\bar{w}) makes the optimal compensation plan more rewarding yet less achievable. Hence, the choice of \bar{w} illustrates this engagement-excitement tradeoff. Zoltners et al. (2006) argue that an “ideal engagement rate” exists and depends on a number of factors. We believe this notion provides practical justification for the notion that \bar{w} could be chosen to match this “ideal engagement rate” that balances the engagement-excitement tradeoff to best motivate the store manager to exert high effort.

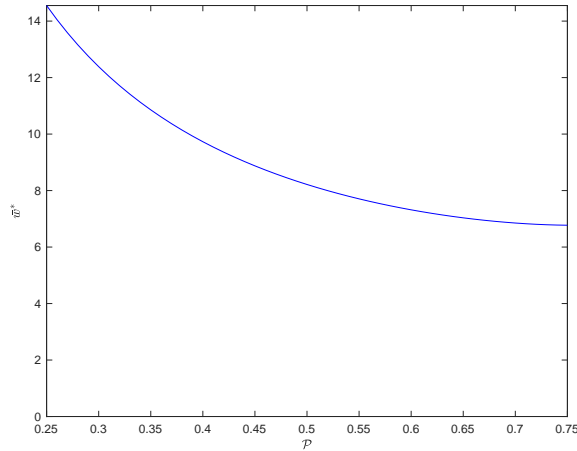


Figure OA.10 The maximum \bar{w} as a function of the agent’s minimum likelihood of receiving the bonus (\mathcal{P}), which varies between 25% and 75%. All the parameters are the same as in Figure 5(a) except that we fix $c(e_o^L, e_m^L) = 1.0$.

Note that for any given \bar{w} , we can characterize the optimal compensation plan as in Section 5.2, which gives the optimal solution specified by $t^*(\bar{w})$ such that the bonus region consists of two parts:

$$B^{\text{NSO}}(\bar{w}) \triangleq \left\{ (i, s) \in D^{\text{NSO}} : \sum_{e_o, e_m} \omega_{e_o, e_m} R_{e_o, e_m}^{\text{NSO}}(i, s) \geq t^*(\bar{w}) \right\}$$

and

$$B^{\text{SO}}(\bar{w}) \triangleq \left\{ (i, s) \in D^{\text{SO}} : \sum_{e_o, e_m} \omega_{e_o, e_m} R_{e_o, e_m}^{\text{SO}}(i, s) \geq t^*(\bar{w}) \right\}.$$

Thus, the problem is equivalent to finding the maximum \bar{w} that satisfies

$$\Pr \left(\sum_{e_o, e_m} \omega_{e_o, e_m} R_{e_o, e_m}^{\text{NSO}}(i, s) \geq t^*(\bar{w}) \right) + \Pr \left(\sum_{e_o, e_m} \omega_{e_o, e_m} R_{e_o, e_m}^{\text{SO}}(i, s) \geq t^*(\bar{w}) \right) \geq \mathcal{P}. \quad (\text{OA13})$$

Using (OA13) as an additional constraint and endogenously choosing \bar{w} , we conduct a numerical experiment and illustrate in Figure OA.10 a sensitivity analysis showing how \mathcal{P} restricts the range of \bar{w} in the incentive-design problem. Our experiment shows that as such a likelihood increases, as one would expect, the value of \bar{w} decreases. Thus, by incorporating the agent's likelihood of receiving a bonus, this extension can be viewed as a natural way of microfounding \bar{w} .

References

- Aliprantis, C.D., K.C. Border. 2006. *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Second Edition. New York, NY: Springer.
- Barvinok, A. 2002. *A Course in Convexity*. Graduate Studies in Mathematics **54**. Providence, RI: American Mathematical Society.
- Zoltners, A. A., P. Sinha, G. A. Zoltners. 2006. *The Complete Guide to Sales Force Incentive Compensation: How to Design and Implement Plans that Work*. AMACOM, New York, NY.