



Salesforce Contracting Under Uncertain Demand and Supply: Double Moral Hazard and Optimality of Smooth Contracts

 Tinglong Dai,^a Kinshuk Jerath^b
^aCarey Business School, Johns Hopkins University, Baltimore, Maryland 21202; ^bColumbia Business School, Columbia University, New York, New York 10027

 Contact: dai@jhu.edu,  <https://orcid.org/0000-0001-9248-5153> (TD); jerath@columbia.edu,  <https://orcid.org/0000-0003-0732-5863> (KJ)

Received: March 8, 2018

Revised: December 5, 2018; March 13, 2019

Accepted: April 5, 2019

 Published Online in Articles in Advance:
 August 28, 2019

<https://doi.org/10.1287/mksc.2019.1171>

Copyright: © 2019 INFORMS

Abstract. We consider the compensation design problem of a firm that hires a salesperson to exert effort to increase demand. We assume both demand and supply to be uncertain with sales being the smaller of demand and supply and assume that, if demand exceeds supply, then unmet demand is unobservable (demand censoring). Under single moral hazard (i.e., when the salesperson's effort is unobservable to the firm), we show that the optimal contract has an extreme convex form in which a bonus is provided only for achieving the highest sales outcome even if low realized sales are due to low realized supply (on which the salesperson has no influence). Under double moral hazard (i.e., when the firm can also take supply-related actions that are unobservable to the salesperson), we show that the optimal contract is smoother as it involves positive compensation for intermediate sales outcomes to assure the salesperson that the firm does not have an incentive to deviate to an action that hurts the agent; in fact, under certain conditions, the contract is concave in sales. We also determine conditions under which, if possible, the firm should postpone contracting until after supply is realized.

History: Ganesh Iyer served as the senior editor and Dmitri Kuksov served as associate editor for this article.

Keywords: salesforce compensation • yield uncertainty • demand censoring • double moral hazard • quota-bonus contract • early versus late contracting

1. Introduction

Firms engage salespersons to increase demand for their products. Salesforce compensation is a major expense for firms, especially in business-to-business (B2B) settings, and the total spend of U.S. firms on salesforce compensation is approximately three times their spend on advertising (Zoltners et al. 2008). There is a large literature on salesforce compensation contracts rooted in agency theory (e.g., Holmstrom 1979, Basu et al. 1985, Holmstrom and Milgrom 1987, Lal and Srinivasan 1993, Park 1995, Raju and Srinivasan 1996, Kim 1997, Oyer 2000, Herweg et al. 2010, and Simester and Zhang 2014). This literature assumes that the agent's sales effort is unobservable and there is demand uncertainty, which makes it difficult to infer salesforce effort from observing realized demand, leading to the issue of moral hazard.

Realized demand can only be fulfilled if there is sufficient supply, and virtually all of the work on salesforce compensation has assumed that the supply is unbounded and always assured (typically, supply-related assumptions are not even explicitly stated). This, however, may not always be the case as firms may only stock a limited amount of inventory to meet short-run demand. Indeed, how much inventory to stock

under uncertain demand, often called the “newsvendor problem,” is a primary focus of study of the field of operations management (Porteus 2002, Cachon and Terweisch 2012). Recent work has considered the importance of supply in determining salesforce compensation contracts. For example, Dai and Jerath (2013, 2016) assume limited supply and show that, counterintuitively, this leads to higher powered contracts (in which “higher powered” means that the bonus is larger).

Furthermore, in practice, there are many situations in which a firm's inventory level may not only be limited, but may also not be fully predictable, that is, supply may be random. (Note that we use the terms “supply,” “inventory,” and “yield” interchangeably throughout the paper.) For example, in the case of wine production, “Vineyards are variable. Growers have known this for as long as they have been growing grapes” (Bramley and Hamilton 2004, p. 32). The uncertain yield has implications for demand fulfillment. Random yield is, indeed, a widely observed phenomenon in myriad industries and scenarios, including electronic fabrication and assembly (Lee and Yano 1988), mining (Kamrad and Ernst 2001), semiconductor manufacturing (Stapper and Rosner 1995), agriculture

(Kazaz 2004), refining and chemical manufacturing (Rajaram and Karmarkar 2002), vaccine and drug manufacturing (Dai et al. 2016), and multistage custom production processes (Wein 1992). Yield uncertainty may also play a significant role in the case of procuring from unreliable suppliers (Dada et al. 2007). Furthermore, Yano and Lee (1995) argue that random yield is prevalent even outside of the aforementioned industries/scenarios because of “many traditional discrete parts manufacturing processes that experience random yields.” For example, a wide variety of consumer products (e.g., smartphones) are assembled from multiple parts with uncertain yields. Another factor that may lead to supply uncertainty is inventory shrinkage, for example, unforeseen loss of stocked inventory resulting from theft, mismanagement (e.g., damage in the handling and storage of the product), expiration, etc., which is a common problem in warehouses and retail stores (Raman et al. 2001, Liu et al. 2010). In most, if not all, of these situations, the firms hire a salesforce to sell the products to other firms (B2B selling) or consumers (business-to-consumer or B2C selling), and it is important to understand the impact of supply uncertainty on salesforce compensation contracts. However, to our knowledge, the impact of supply uncertainty on salesforce compensation has not been studied, and this is a gap in the literature that we make an effort to start to fill.¹

An important consideration with random demand and limited (deterministic or random) supply is that, with positive probability, demand and supply do not match, and sales is the minimum of the two. We consider a setting in which the firm can only observe sales, which implies that, in the case when demand exceeds supply, the firm cannot observe the demand in excess of the supply. This is because it is typically not possible to keep track of demand that was or could have been realized but was not fulfilled because of a lack of inventory, especially if customers choose not to order or to postpone their purchase rather than backorder the product. This is a widely observed phenomenon commonly referred to as *demand censoring*. In recognition of its real-world importance, a growing economics, marketing, and operations literature has studied the managerial implications of demand censoring (Braden and Freimer 1991; Anupindi et al. 1998; Downs et al. 2001; Ding et al. 2002; Chen and Plambeck 2008; Lu et al. 2008; Besbes and Muharremoglu 2012; Conlon and Mortimer 2013; Dai and Jerath 2013, 2016; Rudi and Drake 2014; Chen et al. 2017). Demand censoring may be viewed as a specific form of information censoring and, in our setting, because of it, the firm cannot use the realized demand as the basis for determining how much to pay a salesperson. In fact, the firm has to work with realized sales, which is a worse signal than realized

demand (because of truncation at the inventory level) of the salesperson’s effort.

The following is an example of a B2B situation in which all the aspects that we highlight—namely salesforce, demand uncertainty, supply uncertainty, double moral hazard, and demand censoring—are operative. Consider a company that sells office products, such as electronic equipment (e.g., projectors), furniture, and stationery. The firm typically orders products and stocks them in a local warehouse with a lead time of several weeks or even several months; there is reasonable uncertainty about exactly how many of and when the ordered units will reach because of random supply disruptions and how many of the supplied units will actually be available (for reasons such as shrinkage), but the firm can take certain costly actions to increase the probability of high yield (i.e., inventory yield is uncertain, and to improve this yield, the firm makes an inventory decision/action unobserved to the sales agent). On the other hand, sales agents go out in the field to describe these products to prospective customers in the hope that they will be convinced of their benefits and will order these products from the company if and when the need for these arises for the customers (i.e., unobservable sales effort by the agent increases the level of demand, and realized demand is uncertain) with a promise that, after an order is placed, delivery will be done in a few days (i.e., if supply is not available in the short term, then realized demand cannot be met in the short term). When the customer actually wants to order, the customer may use a website on which the products are displayed along with whether they are available for immediate delivery or not, and sales from a geographical area are tied to the salesperson serving that area. In this case, if a prospective customer sees that a product is not available for immediate delivery, the customer may not even place the order, or possibly, an unavailable product is not even listed on the website, in which case, again, the consumers cannot order it or indicate that they wanted it (i.e., lost demand is not observed). It may also be possible that the customer calls the firm’s salesperson to place an order and is told that the particular product the customer wanted is not available; the salesperson may not record the missed order, and even if the firm did claim to record orders, the firm may not fully believe the lost demand number because the salesperson would have an incentive to inflate this number to claim that the salesperson generated high demand that was not fulfilled because of inventory issues (i.e., again, lost demand is not observed). Such a situation, which has all the essential components that we study, occurs in many B2B settings.

Another B2B situation in which this problem arises is in the media ad sales context. We directly quote

Robert Dillon, who was vice president of North America sales strategy and operations at Yahoo! (Sales Leadership Forum 2010, pp. 27–28):

Yahoo! is in the media business. We create inventory for people looking at web pages. We call that supply. Our sales representatives are out there generating demand from our advertisers. Supply can move dramatically up and down and it can move dramatically in various verticals. And, back to our market model, it's difficult to predict those shifts and difficult to explain those shifts when they happen. So a sales rep may have done everything right on the demand side, but because of some strange shift in inventory that's difficult to explain, they haven't hit their number. That's the challenge that we work through and try to compensate for and plan for.

Clearly, in this case, there is supply and demand variability, the sales agent can take actions to increase overall demand, the firm can take actions to increase overall supply, and unmet demand may go unexpressed by the buyer or unrecorded by the sales agent or the number may not be believed by the firm.

Such examples can also be readily provided for B2C settings, for example, for the smartphone division of a firm, such as Samsung. Briefly, the firm makes an inventory decision and can take actions to promote a high yield, but short-term yield is uncertain (this problem is especially acute in the smartphone industry²), and short-term demand can only be met with what is available. Customers go to off-line stores, such as a Samsung Experience Shop or a Samsung store-within-a-store in BestBuy, to obtain information about the smartphones, where in-store associates exert unobservable effort to convince customers of the benefits of the product(s) offered. Customers later go online to order the product they want, where they see availability status; typically, if they see a product as unavailable they are unable to place an order for it although sometimes they might not even see it listed on the website, and so lost demand is not observed. On the other hand, if a customer orders a product in-store and is told that it is out of stock, this fact may not be recorded (in which case lost sales are not recorded), and/or the firm may not believe these numbers as the in-store associate would have the incentive to artificially inflate the missed sales.

To study such situations, we construct a stylized principal–agent model of a firm that hires a salesperson to market a product with uncertainty in both supply and demand. To keep the model simple while conveying the main insights, we assume both supply and demand to have discrete distributions with the same support. The firm takes an inventory-related action to influence the supply distribution (but does not influence the demand distribution); the salesperson's effort influences the demand distribution

(but does not influence the supply distribution). We assume that the firm contracts with the salesperson before yield uncertainty is resolved. We characterize the firm's optimal contracting decision under the standard assumption in the contract theory literature that salesforce effort boosts demand in a way such that the demand distribution satisfies a monotone likelihood ratio property (MLRP) (which essentially implies that a higher demand level is a more reliable indicator that the salesperson has exerted effort); likewise, we model how the firm's inventory-related action influences supply by assuming the inventory distribution satisfies MLRP. Because of demand censoring, the firm can only observe the realized sales and has to contract on this as the outcome metric.

In a benchmark in which we fix the firm's inventory-related action, we show that the optimal compensation contract provides a bonus only if observed sales are the highest possible (even if sales were limited by a low inventory realization on which the salesperson has no control). This is a simple contract that is similar to the optimal salesforce compensation contract without supply limitations when the demand distribution satisfies MLRP, in which the salesperson is rewarded a bonus only when the most desirable demand outcome is achieved (see, e.g., Laffont and Martimort 2001).

The aforementioned contract provides a bonus only if observed sales are the highest possible even if sales were more limited by low inventory realization than by low demand realization. One may think of this as an overly extreme contract, especially because the salesperson's effort does not influence the supply distribution. Echoing this point, we analyze our focal scenario in which the firm's inventory-related action is endogenous, but there is randomness in the final supply that is available to meet demand. For instance, the firm may choose a high or low intensity of auditing inventory, proactively addressing upstream supply issues or controlling inventory shrinkage; all these activities influence the inventory available at the time of meeting demand. The salesperson observes the final inventory but does not observe the firm's inventory-related action. Just as demand uncertainty and effort unobservability imply a demand-related moral hazard problem for the firm (i.e., the firm cannot verify the salesperson's effort from the realized demand), if the firm's inventory-related action is unobservable to the salesperson, yield uncertainty implies a supply-related moral hazard problem for the salesperson (i.e., the salesperson cannot verify the firm's original inventory action from the final available inventory), and a *double moral hazard* problem arises.

In this case, in which the salesperson does not have transparency regarding the firm's supply-related action, we show that the firm optimally offers the salesperson a smoother contract, that is, positive bonus

is provided for intermediate sales outcomes as well. This is because the firm is tempted to take a less effective inventory-related action (without the salesperson's knowledge) if this helps the firm to reduce the expected compensation for the salesperson, and to motivate the salesperson under this concern, the firm may have to reward the salesperson even when the most desirable sales outcome is not achieved. In other words, under double moral hazard, when the firm's inventory-related action is not observable, the contract is lower powered; in fact, the contract may even be concave in realized sales. This is an interesting result because it shows that supply-side moral hazard can lead to concave contracts even when the agent is risk neutral; this is different from extant literature that argues that risk-neutral agents are offered convex contracts (Laffont and Martimort 2001, Dai and Jerath 2013), and concavity in a contract is typically driven by risk aversion of the agent (Basu et al. 1985, Rubel and Prasad 2016). We note that Zoltners et al. (2006) report that concave compensation plans (which they call "regressive" plans as opposed to convex compensation plans, which they call "progressive" plans) are widely used by firms; in this context, we show that regressive plans are possible even if the salesperson is not risk averse and because there may be double moral hazard.

Finally, we examine the optimal timing of contracting, that is, if it is possible to contract with the salesperson after yield uncertainty is resolved, should the firm do so?³ When the firm contracts with the salesperson after yield uncertainty is resolved, the inventory outcome is known, and the salesforce compensation is only contingent on the demand outcome. However, a trade-off arises from the reduced uncertainty (Dai and Jerath 2013). On the one hand, when the yield is low, the sales outcome is constrained by the low inventory level rather than a low demand outcome, and it is not worthwhile to engage the salesperson. In this case, the firm can avoid wasteful salesforce expenses (by not hiring the salesperson) in view of the inventory information. On the other hand, when the yield is relatively high but less than the highest possible realization of demand, the sales quantity is bounded above by the available inventory, and the firm faces the issue of demand censoring. Because of its limited observability of the sales outcome, the firm has to share a higher rent with the salesperson to induce the same demand inducing effort. Jointly, these two effects drive the optimal timing of salesforce contracting under yield uncertainty. We characterize the firm's optimal contracting decision, which allows us to show a number of counter-intuitive results. For instance, we find that, as the probability of high inventory outcome increases (i.e., there is a lower chance that the inventory outcome is

low), the firm might be more inclined to wait to contract with the salesperson after observing the inventory outcome.

Our paper contributes to the literature on double moral hazard. To the best of our knowledge, the prior double moral hazard literature with risk-neutral principals and agents (including, e.g., Cooper and Ross 1985, Romano 1994, Bhattacharyya and Lafontaine 1995, Kim and Wang 1998, and Roels et al. 2010) assumes that both parties have unlimited liability, whereas we assume that the agent has limited liability. The feature that the agent has limited liability, albeit unique to the double moral hazard literature, is a standard assumption in the salesforce compensation literature. This distinguishing feature in our model implies that, even under single moral hazard, the moral hazard problem cannot be solved by "selling the firm," and rent sharing is necessary. Double moral hazard further implies that the firm has to share more rent and settle with a less efficient contract that is smoother.

In addition to the literatures mentioned earlier, our paper contributes to the nascent literature on jointly modeling incentive and operational issues. Chen (2005) focuses on designing sales compensation contracts such that inventory can be managed more effectively through smoothing demand and eliciting more market information. Plambeck and Zenios (2003) derive an optimal incentive contract for a production manager for a specific production process. Dai and Jerath (2013, 2016) study salesforce compensation contracts under limited inventory but do not allow for uncertainty in inventory. This is the novel angle that we add. We show that, if the inventory decisions are exogenous or observable to the salesperson, then the contract form is similar to that without inventory considerations (an "extreme" contract that is convex in sales wherein a bonus is rewarded only if the highest possible sales outcome is achieved), but if inventory decisions are unobservable to the salesperson, then the contract is a smoother one (and may be concave in sales). Furthermore, supply uncertainty leads to the question of the timing of contracting and, for the class of contracts that we consider, we derive conditions for contracting before or after supply uncertainty is resolved.

The rest of the paper is organized as follows. In Section 2, we describe our model. In Section 3, we analyze the firm's optimal salesforce contract in a preliminary case with an exogenous inventory-related action. In Section 4, we consider the focal case of double moral hazard with endogenous, unobservable, inventory-related action by the firm followed by a random supply shock. In Section 5, we consider the firm's optimal timing of offering the incentive contract. In Section 6, we conclude with a discussion.

2. Model

We model a firm that manufactures/stocks and sells a product. The demand for the product is uncertain. The firm employs a salesperson to exert sales effort to increase the demand. We assume that the demand, denoted by D , can be high (H), medium (M), or low (L), $H > M > L > 0$. The salesperson's effort, denoted by e , can be high (e_H) or low (e_L) and influences the demand according to the following probabilities:

$$\Pr(D = \xi|e) = \begin{cases} p_\xi & \text{if } e = e_H \\ q_\xi & \text{if } e = e_L \end{cases} \text{ for } \xi \in \{H, M, L\}. \quad (1)$$

This states that, if the salesperson exerts high effort, the probabilities of demand being H, M , and L are p_H, p_M , and p_L , respectively, and if the salesperson exerts low effort, these probabilities are q_H, q_M , and q_L , respectively. Consistent with the principal–agent literature, we assume the MLRP such that

$$\frac{p_H}{q_H} > \frac{p_M}{q_M} > \frac{p_L}{q_L} > 0. \quad (2)$$

The MLRP property essentially states that a higher demand outcome is a more reliable indicator that the salesperson exerted effort. Note that the MLRP directly implies that $p_H > q_H$ and $p_L < q_L$, but p_M and q_M can have any relationship. We denote by $\psi > 0$ the salesperson's disutility of effort when the salesperson exerts high sales effort (i.e., $e = e_H$) and normalize the salesperson's disutility of effort to zero when the salesperson exerts low sales effort (i.e., $e = e_L$).

We assume that the firm has limited inventory to sell. This inventory level, denoted by I , is uncertain, and can be high (H), medium (M), or low (L). The firm can choose an *inventory-related action*, denoted by a , that influences the inventory according to the following probabilities:

$$\Pr(I = \xi|a) = \begin{cases} r_\xi & \text{if } a = a_H \\ s_\xi & \text{if } a = a_L \end{cases} \text{ for } \xi \in \{H, M, L\}. \quad (3)$$

When $a = a_H$, the firm takes a highly effective action to ensure ample inventory. Such an action may entail, for example, activities preventing inventory shrinkage, damage, and spoilage. When $a = a_L$, the firm takes a less effective inventory-related action. We assume that the inventory-related action is costless; this assumption is noncritical and helps us to understand the impact of inventory uncertainty in a clearer and simpler manner. Similar to the demand side, we assume MLRP for the supply side such that⁴

$$\frac{r_H}{s_H} > \frac{r_M}{s_M} > \frac{r_L}{s_L} > 0, \quad (4)$$

which implies $r_H > s_H$ and $r_L < s_L$.

We assume that both the firm and the salesperson are risk neutral. Unlike the firm, however, the salesperson has limited liability, implying that the salesperson must be protected from downside risk. Specifically, we normalize the salesperson's limited liability to zero, that is, the salesperson's salary must be non-negative under any outcome of demand. Limited liability is a widely observed feature of salesforce contracts in the industry, and this assumption is a standard one in the literature (cf. Laffont and Martimort 2001; examples in the salesforce literature include Sappington 1983; Park 1995; Kim 1997; Oyer 2000; and Dai and Jerath 2013, 2016). In our setting, the limited liability assumption implies that the firm provides a nonnegative fixed wage to the salesperson, which is aligned with industry practice. This directly implies that the firm cannot use a profit-sharing mechanism to achieve the first-best outcome. We also normalize the salesperson's reservation utility to zero without loss of generality.

The firm's revenue comes from matching supply with demand such that the actual selling quantity is $\min\{D, I\}$. Each unit of sales generates a revenue of $\rho > 0$. We assume that this per-unit price is exogenous and the salesperson does not adjust this price, an assumption that is uniformly made in the salesforce compensation literature and has real-world support (Chung et al. 2014). We assume that the effort cost is low relative to the unit revenue such that it is worthwhile for the firm to induce a high sales effort except in the case in which the firm contracts with the salesperson after the inventory is realized to be low. We also assume that inventory is costless (note that we have already assumed that the inventory-related action is costless). These assumptions of no inventory-side costs, though nonstandard, allow us to focus sharply on the effects of the uncertainty in inventory rather than on the costs related to inventory. Including the inventory-side costs will not lead to any qualitative change in the insights that our analysis provides.

Throughout the paper, we consider the setting in which the firm cannot verify the portion of realized demand D that is in excess of the stocked inventory level. For example, when the demand is H yet the inventory level is M , the realized sales are M , and the firm only knows the demand is no less than M as it cannot observe the actual demand. This phenomenon is referred to as demand censoring. As discussed in the introduction, a large number of papers in the literature document this important phenomenon and study its implications (Braden and Freimer 1991; Anupindi et al. 1998; Downs et al. 2001; Ding et al. 2002; Chen and Plambeck 2008; Lu et al. 2008; Besbes and Muharremoglu 2012; Conlon and Mortimer 2013; Dai and Jerath 2013, 2016; Rudi and Drake 2014;

Chen et al. 2017). Because of demand censoring, the firm has to use the sales quantity rather than the actual demand as the basis for determining the salesforce compensation plan.

We summarize the notation that we use in Table 1.

The timeline of the game, as illustrated in Figure 1, is as follows. First, the firm offers the salesperson a take-it-or-leave-it compensation contract, which the salesperson accepts or rejects. Second, the firm takes an inventory-related action to influence inventory distribution. Third, if the salesperson accepts the contract, the salesperson exerts effort to boost demand. Fourth, the inventory is realized as H, M , or L . Fifth, the demand is realized as H, M , or L , and sales are determined as the minimum of demand and inventory. Note that, in this formulation, the fourth and fifth stages can be merged into one stage.

3. Benchmark: Exogenous Inventory

In this section, we derive initial insights related to the salesperson’s compensation contract under random

yield. For this purpose, we “switch off” the part of the model in which the firm has an option to choose an inventory-related action. Instead, we fix the firm’s inventory-related action at $a = a_H$ such that inventory is random but its distribution is exogenous, which implies that, in this model, moral hazard exists only on the side of the salesperson. The insights obtained in this section enable us to better understand the forces at play in the main model with double moral hazard in the next section. We note that the setup and the results in this section are, nevertheless, interesting as well as novel to the literature for the following reasons. First, this section allows for uncertainty in *both* supply and demand; as discussed earlier, previous literature on salesperson compensation either assumes unlimited supply or limited but deterministic supply as in Dai and Jerath (2013, 2016). Second, we show that, if MLRP holds on the demand side and supply is uncertain, then a property similar to MLRP holds for realized sales (Lemma 1). Under this property, counterintuitively, it is optimal for the firm to provide a bonus for only the highest sales outcome even when low realized sales are due to low realized supply, which the salesperson cannot influence. Third, despite the presence of two sources of uncertainty (demand and supply), we show the firm incurs the same expected cost of compensating the salesperson as in the case with only demand uncertainty.

There are three possible sales outcomes, that is, H, M , and L . After the demand D and the inventory level I are realized, the actual sales, denoted by the random variable Y , are equal to $\min\{D, I\}$. Suppose that the salesperson exerts an effort level of e_H . Because we fix the inventory-related action at a_H , the probability of each possible sales outcome is

$$\Pr\{Y = \xi | e = e_H\} = \begin{cases} r_H p_H & \text{if } \xi = H \\ r_H p_M + r_M p_M + r_M p_H & \text{if } \xi = M \\ r_L + p_L - r_L p_L & \text{if } \xi = L. \end{cases}$$

Likewise, when the salesperson exerts an effort level of e_L , the probability of each possible sales outcome can be represented as

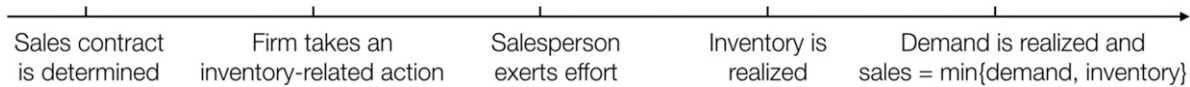
$$\Pr\{Y = \xi | e = e_L\} = \begin{cases} r_H q_H & \text{if } \xi = H \\ r_H q_M + r_M q_M + r_M q_H & \text{if } \xi = M \\ r_L + q_L - r_L q_L & \text{if } \xi = L. \end{cases}$$

Define $p_\xi^Y = \Pr\{Y = \xi | e = e_H\}$, $\xi \in \{H, M, L\}$; that is, p_ξ^Y is the probability that sales is equal to ξ under high effort given the probability distribution of inventory realization. Similarly, define $q_\xi^Y = \Pr\{Y = \xi | e = e_L\}$, $\xi \in \{H, M, L\}$. We can think of these as the parameters of the sales distributions with high and low effort

Table 1. Notation

D	Demand, which is subject to uncertainty and can be H (high), M (medium), or L (low)
I	Inventory, which is subject to uncertainty and can be H (high), M (medium), or L (low)
e	The salesperson’s effort, which can be high (e_H) or low (e_L)
ψ	The salesperson’s disutility from exerting a high effort level (i.e., $e = e_H$)
$\hat{\psi}$	The salesperson’s disutility from exerting a high effort level (i.e., $e = e_H$) if the effort is exerted after inventory is realized
p_ξ	The probability that demand is $\xi \in \{H, M, L\}$ when the salesperson exerts high effort (i.e., $e = e_H$)
q_ξ	The probability that demand is $\xi \in \{H, M, L\}$ when the salesperson exerts low effort (i.e., $e = e_L$)
a	The firm’s inventory-related action, which can be either highly effective (a_H) or less effective (a_L)
r_ξ	The probability that inventory is $\xi \in \{H, M, L\}$ when the firm takes a highly effective inventory-related action (i.e., $a = a_H$)
s_ξ	The probability that inventory is $\xi \in \{H, M, L\}$ when the firm takes a less effective inventory-related action (i.e., $a = a_L$)
ρ	Unit revenue
Y	Sales quantity, which is the minimum of demand and inventory (i.e., $Y = \min\{D, I\}$)
p_ξ^Y	The probability that sales quantity is equal to $\xi \in \{H, M, L\}$ under a high sales effort (i.e., $e = e_H$) and a highly effective inventory-related action (i.e., $a = a_H$)
q_ξ^Y	The probability that sales quantity is equal to $\xi \in \{H, M, L\}$ under a low sales effort (i.e., $e = e_L$) and a highly effective inventory-related action (i.e., $a = a_H$)
ΔS	The firm’s loss in its expected sales quantity under a high sales effort (i.e., $e = e_H$) when it switches from a highly effective inventory-related action ($a = a_H$) to a less effective one (i.e., $a = a_L$)
B_ξ	The firm’s bonus for the salesperson when the sales outcome $Y = \xi$ for $\xi \in \{H, M, L\}$

Figure 1. Timeline of Contracting Under Supply Uncertainty



given the parameters of the demand distributions with high and low effort and the parameters of the inventory distribution. In the lemma, we establish an important property.

Lemma 1. *If $\frac{p_H}{q_H} > \frac{p_M}{q_M} > \frac{p_L}{q_L}$ holds, then $\frac{p_H^Y}{q_H^Y} > \max\left\{\frac{p_L^Y}{q_L^Y}, \frac{p_M^Y}{q_M^Y}\right\}$ holds.*

In this scenario, the compensation contract of the salesperson should be specified for every possible realization of sales, accounting for every possible combination of demand and inventory realization that can lead to that particular realization of sales. However, the result of Lemma 1 simplifies the analysis of the scenario under consideration. The following proposition shows that the optimal salesforce compensation contract actually takes a very simple form. (We assume that the unit revenue is high enough such that it is always worthwhile to induce the salesperson to exert a high effort.)

Proposition 1. *The optimal compensation plan is to pay the salesperson a bonus of $\frac{\psi}{r_H(p_H - q_H)}$ if the sales are H units (i.e., if $I = D = H$) and zero otherwise. The firm’s expected payment for motivating a high salesforce effort is $\frac{\psi}{1 - q_H/p_H}$.*

This proposition states that, in the case of contracting before inventory realization, without knowledge of the actual inventory outcome, it is optimal for the firm to use an extreme, convex reward structure. Under this structure, the firm rewards the salesperson with a bonus only when the highest possible sales level is achieved and nothing for sales lower than this level (even if low sales were due to low yield realization over which the salesperson has no control).⁵ The intuition of the result is that, because of demand censoring, the firm cannot observe the true demand outcome and has to determine the salesperson’s effort level by observing the sales outcome. The distribution of the sales outcome is endogenous with the salesperson’s effort functioning as a key parameter. Given three possible sales outcomes (H , M , and L), the firm seeks the outcome that is most indicative of the fact that the salesperson has exerted a high effort. Mathematically, this problem corresponds to finding the outcome with the maximum likelihood ratio which, according to Lemma 1, is H .

Note that the only assumption needed for Proposition 1 to hold is MLRP in terms of demand distribution, which is a standard assumption in the contract-theory literature; no additional assumptions about

yield uncertainty are needed. Furthermore, Proposition 1 shows that, in the case of contracting before inventory realization, the amount of the bonus, given by $\psi/[r_H(p_H - q_H)]$, depends on r_H : the lower r_H is, the higher the bonus is. This makes intuitive sense; for instance, if the probability that the salesperson obtains the bonus is low because the probability of high inventory realization is low, then the salesperson should receive a larger bonus when the salesperson actually receives it. However, the firm’s expected payment to the salesperson, $\psi/(1 - q_H/p_H)$, is independent of r_H , $\xi \in \{H, M, L\}$. In other words, by contracting before inventory realization, the firm can fully address the risk resulting from yield uncertainty and incurs the same expected cost of salesforce compensation as in the case without yield uncertainty.

4. Endogenous Inventory-Related Action: Double Moral Hazard

In the previous section, we analyze a benchmark in which the firm’s inventory-related action is fixed at $a = a_H$. In this section, we analyze our main model, allowing the firm to endogenously determine its inventory-related action to influence the distribution of inventory. As discussed in Section 2, we assume that the firm’s inventory-related action is costless, which helps us concisely and crisply characterize the impact of inventory uncertainty; incorporating the cost of the inventory-related action will not qualitatively alter our findings. The firm’s inventory-related action may or may not be observable to the salesperson, and we analyze both of these cases. Here we consider the case in which the firm’s inventory-related action is endogenous and not observed by the agent (and, at the end of this section, we consider the case in which the firm’s inventory-related action is endogenous and observed by the agent).

The analysis until now shows that a contract that awards the salesperson a bonus only for the highest sales outcome is optimal. This is an extreme “bang bang” contract although, in reality, smoother contracts that reward a salesperson even for lower realized sales outcome are seen. In this section, we show that, if the firm’s inventory-related action is not observable to the agent, such a “smoother” or “lower powered” contract emerges (without adding any additional assumptions such as risk aversion, etc.). In this case, in the timeline in Figure 1, the second and third stages can be merged into one stage.

Note that this is the case of double moral hazard (see, e.g., Bhattacharyya and Lafontaine 1995), and the previous analysis does not hold. To see this, note that, under some conditions, once the salesperson has accepted the contract in Proposition 1, the firm might be tempted to deviate to an inventory-related action of a_L , under which the probability for the salesperson to earn a bonus decreases from $r_H p_H$ to $s_H p_H$ and the firm's expected payment decreases from $\frac{\psi}{1-q_H/p_H}$ to $\frac{s_H}{r_H} \cdot \frac{\psi}{1-q_H/p_H}$. (In an extreme case in which $s_H = 0$, that is, a less effective inventory-related action leads to impossibility of achieving a high inventory outcome, by switching from a_H to a_L , the firm effectively voids the salesperson's likelihood of receiving a bonus.) Thus, by switching from a_H to a_L , the firm's expected savings from salesforce compensation is

$$\left(1 - \frac{s_H}{r_H}\right) \cdot \frac{\psi}{1 - q_H/p_H}.$$

Switching from a_H to a_L , however, means a lower expected sales quantity and results in a loss of the firm's expected revenue. We define

$$\begin{aligned} \Delta S &= r_H p_H H + (r_H p_M + r_M p_M + r_M p_H) M \\ &\quad + (r_L + p_L - r_L p_L) L - [s_H p_H H \\ &\quad + (s_H p_M + s_M p_M + s_M p_H) M \\ &\quad + (s_L + p_L - s_L p_L) L] \\ &= (r_H - s_H) p_H H + [(r_H - s_H) p_M \\ &\quad + (r_M - s_M)(p_H + p_M)] M + (r_L - s_L)(1 - p_L) L \end{aligned}$$

as the absolute value of the firm's expected loss of sales by switching from a_H to a_L given that the salesperson chooses $e = e_H$. The magnitude of the firm's expected revenue loss is, thus, $\rho \Delta S$.

For ease of exposition, we define the following two constants:

$$\tau_1 \triangleq (p_H - q_H) r_M + (p_M - q_M)(r_H + r_M), \quad (5)$$

$$\tau_2 \triangleq (s_M - r_M)(p_H + p_M) - (r_H - s_H) p_M. \quad (6)$$

We focus on the case in which $\tau_2 > 0$ (i.e., $s_M - r_M$ is positive and sufficiently large, which is satisfied under the condition specified in endnote 4). We assume that $B_H^* \geq B_M^*$, which gives the parametric condition $\rho \Delta S \geq \left(1 - \frac{s_H + s_M}{r_H + r_M}\right) \cdot \frac{\psi}{1 - \frac{q_H + q_M}{p_H + p_M}}$, which also immediately implies that the firm chooses an inventory-related action of a_H . The following proposition provides the optimal compensation contract and shows that this is a smoother contract; that is, under some conditions, a bonus is awarded even for non-maximum sales outcomes. Note that the threshold $\frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$ in the proposition comes from the comparison between $\left(1 - \frac{s_H}{r_H}\right) \cdot \frac{\psi}{1 - q_H/p_H}$ and $\rho \Delta S$.

Proposition 2. *If the salesperson cannot verify the firm's inventory-related action, the optimal compensation contract is to pay the salesperson a bonus of B_i^* for a sales outcome $i \in \{H, M, L\}$ such that*

$$\text{i. If } \rho \geq \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}, B_H^* = \frac{\psi}{r_H(p_H - q_H)}, \text{ and } B_M^* = B_L^* = 0.$$

$$\text{ii. If } \rho < \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S},$$

$$B_H^* = \frac{\tau_1 \rho \Delta S + \tau_2 \psi}{\tau_1(r_H - s_H)p_H + \tau_2(p_H - q_H)r_H}, \quad (7)$$

$$B_M^* = \frac{(r_H - s_H)p_H \psi - (p_H - q_H)r_H \rho \Delta S}{\tau_1(r_H - s_H)p_H + \tau_2(p_H - q_H)r_H}, \text{ and} \quad (8)$$

$$B_L^* = 0. \quad (9)$$

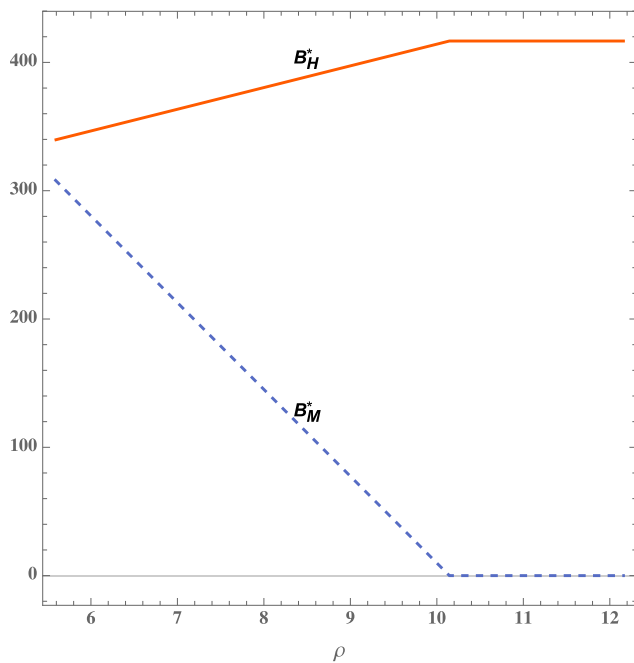
In both cases, the firm chooses a supply-side action of a_H .

Recall that, without double moral hazard, the optimal contract is independent of the per-unit revenue ρ . With double moral hazard, the optimal contract depends on ρ . When ρ is large enough (the case of Proposition 2(i)), the contract stays the same as in Proposition 1. However, when ρ is small enough (the case of Proposition 2(ii)), the firm pays bonuses for sales equal to M and H even though the MLRP is assumed to hold (as one would expect, $B_M^* \leq B_H^*$). Note that this is a fundamental change in the contractual form compared with the case of single-sided moral hazard, in which the firm paid a bonus just for sales equal to H . Specifically, the contract is smoother and not as high powered. A positive bonus for medium sales outcome is there to protect the salesperson from receiving no compensation at all should the firm, unobservable to the agent, choose an inventory-related action of a_L . Under MLRP, any positive compensation for an outcome that is not the most desirable causes a deadweight loss to the system, which is the inefficiency introduced into the system because of the unobservability of the firm's inventory-related action.

For the values of ρ in the case of Proposition 2(ii), as ρ increases, the firm reduces B_M^* , the bonus paid to the salesperson for a medium sales outcome, but increases B_H^* , the bonus paid to the salesperson for a high sales outcome; at $\rho = \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, the contract parameters $\{B_H^*, B_M^*\}$ are the same in cases (i) and (ii) of the proposition. This is illustrated in Figure 2. Intuitively, as ρ increases, everything else being the same, the firm has a lower incentive to switch its inventory-related action from a_H to a_L , making it less necessary to use B_M^* to protect the salesperson.

Note that, if $B_H^* - B_M^* > B_M^* - B_L^*$, then the optimal compensation plan is convex in sales, otherwise it is (weakly) concave. For $\rho \geq \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, the optimal

Figure 2. (Color online) Effect of ρ on B_H^* and B_M^*



Note. Parameters are $\{p_H, p_M, p_L\} = \{0.7, 0.2, 0.1\}$, $\{q_H, q_M, q_L\} = \{0.5, 0.2, 0.3\}$, $\{r_H, r_M, r_s\} = \{0.6, 0.15, 0.25\}$, $\{s_H, s_M, s_L\} = \{0.1, 0.4, 0.5\}$, $\{H, M, L\} = \{100, 75, 50\}$, and $\psi = 50$.

compensation plan is clearly convex. For $\rho < \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, the answer is more nuanced. Define

$$\hat{\rho} = \frac{(H - L)(r_H - s_H)p_H - (M - L)\tau_2}{(H - L)(p_H - q_H)r_H + (M - L)\tau_1} \cdot \frac{\psi}{\Delta S}.$$

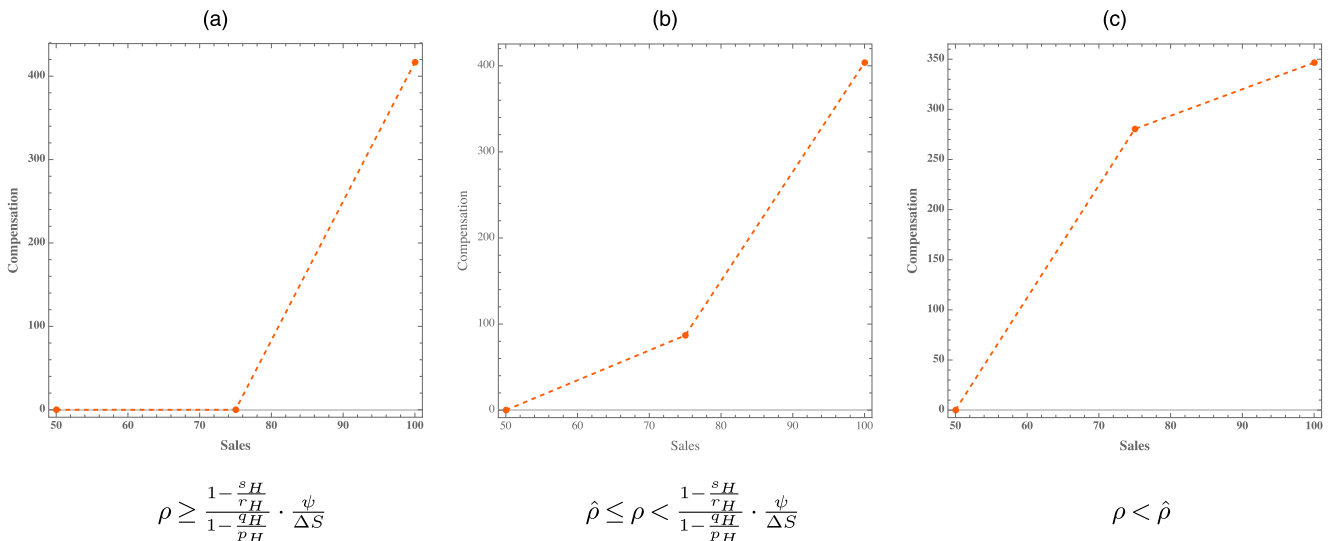
Then we have the following corollary that is immediate from Proposition 2(ii).

Corollary 1. If $\rho < \hat{\rho}$, the optimal compensation plan (B_L^*, B_M^*, B_H^*) is concave in the sales outcome; otherwise, it is convex.

The plots in Figure 3 illustrate the different types of contracts that are possible (in the plots, the dashed lines are for illustration only). In Figure 3(a), where $\rho \geq \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, the contract takes an extreme convex form with bonus awarded only when the realized sales are equal to H . In Figure 3(b), where $\hat{\rho} \leq \rho < \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, the contract takes a smoother yet convex form with a relatively small bonus awarded when the realized sales are equal to M and a large bonus awarded when the realized sales are equal to H . In Figure 3(c), where $\rho < \hat{\rho}$, the contract takes a smoother, concave form with a relatively large bonus awarded when the realized sales are equal to M and a not much larger bonus awarded when the realized sales are equal to H .

Corollary 1 states that the optimal compensation plan may be concave when the per-unit revenue ρ is small enough. The intuition behind this result is that, when ρ is sufficiently small, under the bang bang contract as characterized in Proposition 1—a convex compensation plan—the firm is tempted to choose a less effective inventory-related action to reduce its salesforce compensation without significant impact on its expected revenue. Anticipating this possibility, the salesperson would not accept this bang bang contract. Therefore, the firm must offer a contract with a compensation level for a medium sales outcome. Furthermore, under certain conditions, this compensation level should be sufficiently close to that

Figure 3. (Color online) Different Contractual Forms Under Double Moral Hazard



Notes. In panels (a)–(c), the value of ρ is 12, 9, and 6, respectively. All the other parameters are the same as in Figure 2.

for a high sales outcome, leading to a concave compensation plan.

Basu et al. (1985) show that it is optimal for the firm to choose a concave compensation plan (i.e., decreasing commission rates with outcome) only when the salesperson is risk averse (as characterized by, e.g., a constant risk-aversion utility function). Rubel and Prasad (2016) also show that, in their dynamic setting, concavity in the compensation plan arises from risk aversion. On the other hand, papers that assume a risk-neutral salesperson with limited liability show that extreme, nonlinear, quota-bonus plans that concentrate all variable payment at one outcome of sales are optimal (Park 1995, Kim 1997, Oyer 2000). Recent work shows that adding inventory considerations to these settings maintains the extreme form of the optimal contract (Dai and Jerath 2013, 2016) though it has implications for the reward amount and the inventory level.

In our case, the salesperson is risk neutral with limited liability. Nevertheless, we show that a smoother (even concave) contract can be optimal when there is unobservability in the inventory-related actions. In other words, we identify a force different from risk aversion—namely supply-related moral hazard—that can lead to a fundamentally different form of the incentive compensation plan.⁶ One can indeed expect supply-side moral hazard to be operative in reality because of unobservability of the firm's inventory-related action to a salesperson whose focus is on visiting clients in the field to increase demand rather than closely monitoring the firm's inventory-related action that is often undertaken by a different silo in the company.

We briefly discuss what happens if we assume the salesperson to be risk averse (while maintaining the assumption of limited liability). Oyer (2000) shows that, under single moral hazard, assuming limited liability with risk neutrality leads to a quota-bonus contract, that is, all marginal compensation is concentrated at one point, and adding risk aversion to that leads to marginal compensation being concentrated on a range of critical points, that is, it leads to a smoother contract. If we add risk aversion to our setting, then there are two forces leading to the contract getting smoother, namely double moral hazard and risk aversion of the salesperson, but the key insight that double moral hazard leads to a smoother contract is still operative.

We also note that our “discrete” model construction with three states for demand and inventory level and two effort levels, though certainly stylized, is not a limiting model setup when compared with a “continuous” model construction with continuous demand and effort. To see this, note that previous

research has already shown that, under limited liability and single moral hazard, the result that a quota-bonus contract is optimal is obtained for both discrete (Laffont and Martimort 2001, Dai and Jerath 2013) and continuous models (Park 1995, Kim 1997, Oyer 2000).⁷ Our main insight here is that, when there is double moral hazard, that is, the agent cannot observe a relevant action of the firm, the firm must make this contract smoother to assure the agent that the agent will still obtain some marginal compensation for lower demand outcomes. It is straightforward to see that this force should be operative equally in both discrete and continuous models.

4.1. Observable Inventory-Related Action

For completeness, we now briefly discuss the case in which the firm's inventory-related action as a_H or a_L is observable to the salesperson. We obtain the following proposition regarding the firm's optimal inventory-related action.⁸

Proposition 3. *If the firm's inventory-related action is observable, then the firm always sets this action as a_H . In addition, the firm chooses the same compensation contract as that in Proposition 1.*

The reason for this result is that, compared with an inventory-related action of a_H , an inventory-related action of a_L implies the same expected payment to the salesperson but a lower expected revenue (see proof for details). Given that the inventory-related action is a_H , the analysis in Section 3 applies, and the contract in Proposition 1 is optimal.

5. Timing of Contracting

So far, we have assumed that effort exertion by the salesperson must happen before yield is realized, which implies that the firm must contract with the salesforce before yield uncertainty is resolved. In some situations, however, it may be possible to exert sales effort after yield is realized, and in these situations, it may be beneficial for the firm to wait and contract after yield certainty is resolved. In this section, we analyze the case of offering a compensation contract after yield uncertainty is resolved and compare it with the case of offering a compensation contract before yield uncertainty is resolved (we call these cases “late contracting” and “early contracting,” respectively). This provides insights related to the optimal timing of contracting. (We assume that advanced contracts, such as those that include menus of contracts, cannot be used because of practical contracting frictions. We discuss this further at the end of this section.) We show that either early or late contracting may be optimal, depending on the interactions among yield uncertainty, demand censoring,

and moral hazard. The key trade-off is between wasteful salesforce compensation expenses in early contracting and overcompensation because of demand censoring in late contracting.

Before we proceed, we briefly note that there is a literature on early versus late contracting in labor and product markets that studies matching between market participants (Roth and Xing 1994, Priest 2010). A critical difference between our work and the early-contracting literature is that we consider a contracting problem in a single principal/single agent environment, in which the issue of matching between market participants is not relevant.

We assume the same supply-and-demand environment as specified in Section 2. In addition, we assume that the cost of effort exertion is higher if effort is exerted after inventory is realized. This assumption is coming from the idea that, to effect the same change in the demand distribution, that is, change it from (q_L, q_M, q_H) to (p_L, p_M, p_H) , in a shorter amount of time, the effort cost must be higher.⁹ We define the salesperson's costs of effort as ψ and $\hat{\psi}$, respectively, when effort exertion occurs before and after inventory realization, where $\hat{\psi} > \psi > 0$.

If it is possible to delay effort exertion, then, just as the firm has the flexibility of choosing the timing of contracting, in early contracting, the salesperson has the flexibility of choosing the timing of exerting effort. In other words, in early contracting, the salesperson may choose to delay effort exertion until after the inventory is realized. To see the incentive behind this, recall from Section 3 that the optimal contract awards a bonus to the salesperson only if sales equal H . The implication is that, if the inventory is M or L , the expected sales outcome is always less than H and the bonus is not awarded. The salesperson, by delaying effort exertion, may obtain inventory information and not necessarily choose a high effort level. Therefore, the firm would not offer this contract in equilibrium unless the firm expects the salesperson will conform with the firm's contracting choice by not delaying effort exertion. The following lemma specifies the condition for this to happen.

Lemma 2. *In the case of contracting before inventory realization, the salesperson chooses to exert effort before inventory realization if $\hat{\psi} \geq \psi/r_H$.*

Lemma 2 indicates that, if $\hat{\psi} \geq \psi/r_H$, early contracting is viable because the salesperson is better off

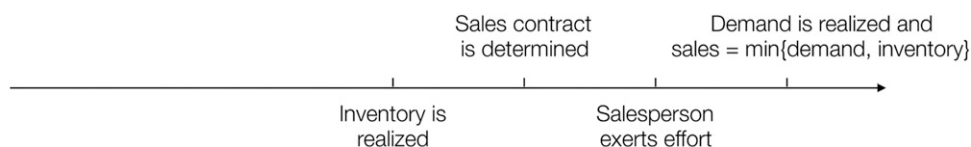
exerting effort before the inventory is realized. In the rest of this section, we focus on the interesting case in which $\hat{\psi} \geq \psi/r_H$.

5.1. Contracting After Inventory Realization

In this section, we consider the case of late contracting in which the firm contracts with the salesperson after the inventory is realized. The timeline of the game is as follows. First, the firm and the salesperson observe the realized inventory level as H, M , or L . Second, the firm offers the salesperson a take-it-or-leave-it compensation contract that the salesperson accepts or rejects. Third, the salesperson determines the salesperson's optimal effort level based on both the compensation plan and the inventory level. Fourth, the demand is realized as H, M , or L , and sales are determined as the minimum of demand and inventory. Figure 4 illustrates the timeline.

When inventory is realized, the firm observes this realized level. There is the possibility that the firm strategically does not disclose this inventory level to the salesperson. (We assume that the firm does not lie about the inventory level; that is, if it reports an inventory level to the agent, it reports truthfully. This is because, for any inventory level that the firm misreports, there is always a positive probability that the firm's lie is identified, in which case it can be taken to court. However, the firm can choose to not disclose the realized inventory level.) We note that the firm, in fact, *always truthfully discloses* the realized inventory level. To see this, consider that the realized inventory level is H . Then the firm discloses this information so that the agent puts in the agent's best effort. Next, assume that the realized inventory level is M . If the firm does not disclose this, then the agent knows the inventory level must not be H (from the preceding argument); that is, it is either M or lower (i.e., L). Clearly, this only implies that the agent has reduced incentive to work hard than if the agent knows that the inventory level is M ; therefore, once again, the firm discloses the inventory level. Next, assume that the realized inventory level is L . In this case, it does not matter to the firm whether it discloses or not as minimum demand is always L ; therefore, the firm again discloses the inventory level (and, if it does not disclose, then the agent can infer that it must not be H or M ; that is, it must be L). In other words, the firm always truthfully discloses the inventory level, and we simply assume that,¹⁰ in the first stage of the game,

Figure 4. Timeline for Contracting After Inventory Realization



both the firm and the salesperson observe the realized inventory level.¹¹

Under late contracting, the optimal compensation plan for the salesperson is different based on the realized inventory level. The analysis of this case is on the lines of the analysis in Dai and Jerath (2013), so we only provide a brief outline.

Case (i). $I = H$: Because of MLRP, the salesperson is paid a bonus only if the sales are H units, and this bonus is equal to

$$\frac{\hat{\psi}}{p_H - q_H}.$$

The firm's expected payment in this case is

$$\frac{\hat{\psi}}{1 - \frac{q_H}{p_H}}.$$

Case (ii). $I = M$: In this case, the maximum sales outcome is capped by the inventory level, leading to demand censoring. Because of the MLRP assumption, the firm chooses to reward the salesperson when the sales are M units. Under the optimal contract, the salesperson receives a positive bonus if the observed sales are M units and zero otherwise. The optimal bonus is

$$\frac{\hat{\psi}}{p_H + p_M - (q_H + q_M)}.$$

The firm's expected payment in this case is

$$\frac{\hat{\psi}}{1 - \frac{q_H + q_M}{p_H + p_M}}.$$

which, by MLRP, is higher than $\hat{\psi}/(1 - q_H/p_H)$ (the firm's expected payment in case (i)). Thus, because of demand censoring, the firm has to provide a higher expected payment despite obtaining a lower expected sales amount (because $\mathbb{E}[\min\{D, M\}] < \mathbb{E}[\min\{D, H\}]$). This scenario captures the drawback of contracting after inventory is realized: when the inventory is medium, both the firm and the salesperson understand that the sales outcome can never be high. Thus, the firm has to provide a bonus for a medium sales outcome which, according to Lemma 1, is not the outcome most indicative of the salesperson's effort choice, leading to overcompensation of the salesforce.

Case (iii). $I = L$: Because the inventory level is low, the sales would always be L units regardless of the demand. The firm, thus, chooses not to induce any salesforce effort, which is equivalent to offering a salesforce contract with a bonus of zero. Indeed, the firm can avoid unnecessary salesforce expenses in this low-yield scenario.

The following lemma holds for the case of late contracting.

Lemma 3. *In the case in which the firm contracts with the salesperson late, that is, after the inventory is realized, the following holds:*

i. *If $I = H$, the salesperson is offered a contract that pays a bonus only if the sales are H units, and this bonus is equal to $\frac{\hat{\psi}}{p_H - q_H}$.*

ii. *If $I = M$, the salesperson is offered a contract that pays a bonus only if the observed sales are M units, and this bonus is equal to $\frac{\hat{\psi}}{p_H + p_M - (q_H + q_M)}$.*

iii. *If $I = L$, the salesperson is not offered a contract of employment.*

The firm's expected payment for the salesperson is

$$r_H \cdot \frac{\hat{\psi}}{1 - \frac{q_H}{p_H}} + r_M \cdot \frac{\hat{\psi}}{1 - \frac{q_H + q_M}{p_H + p_M}}. \quad (10)$$

This analysis reveals that the key benefit of contracting after observing the inventory position is that the firm may avoid wasteful salesforce expenses when the inventory outcome is low. However, in the case in which the inventory outcome is medium, because of demand censoring, the firm has to pay a premium to induce a high salesforce effort.

5.2. Optimal Timing of Contracting

We now determine the optimal timing of contracting for the firm. Note that the firm has the same expected revenue in both early and late contracting (because, in both scenarios, in equilibrium, effort is exerted whenever effort exertion can increase sales). Therefore, it is sufficient to compare the firm's expected salesforce payment under the two scenarios (given in Proposition 2 for early contracting and in Equation (10) for late contracting). The following proposition presents the result for the case in which the per-unit revenue is sufficiently high (i.e., $\rho \geq \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$):

Proposition 4. *In the case of $\rho \geq \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, if*

$$\frac{1 - r_H}{r_M} \leq \frac{\hat{\psi}}{\psi} \cdot \frac{1 - \frac{q_H}{p_H}}{1 - \frac{q_H + q_M}{p_H + p_M}}, \quad (11)$$

then it is optimal for the firm to contract with the salesperson before the inventory is realized (and the contract is as per Proposition 2(i)). Otherwise, it is optimal for the firm to contract with the salesperson after the inventory is realized (and the contract is as per Lemma 3).

Note that the left-hand side of (11) has parameters related to yield uncertainty, and the right-hand side of (11) has parameters related to demand uncertainty and effort cost. Intuitively, the firm's optimal contract

timing depends on the trade-off between yield uncertainty and demand censoring. On the one hand, a high likelihood of the low inventory outcome makes it more appealing for the firm to observe the realized inventory level first before committing to the contracting decision with the salesperson. On the other hand, early commitment to the salesforce compensation plan helps to mitigate the effect of demand censoring because, without knowing the realized inventory level *ex ante*, the firm can hedge and specify a positive bonus awarded to the salesperson only when the sales are high. This is essentially transferring the burden of inventory uncertainty to the salesperson, and if the cost of delayed effort exertion is high enough, then the salesperson accepts this burden to be able to exert effort early. By saying that the firm may be better off by contracting before the inventory is realized, we have the following interesting implication: when choosing to contract before the inventory outcome is realized, the firm may benefit from its lack of inventory information because the bonus is paid to the salesperson only if the sales outcome is *H*.

We now generate several insights regarding the effect of supply-and-demand uncertainty on the firm’s optimal timing of contracting. First, consider the role of the inventory distribution parameters. Specifically, consider r_H , which is the probability of the high inventory outcome. The following corollary shows the impact of r_H on the firm’s optimal contracting timing.

Corollary 2. *In the case of $\rho \geq \frac{1-s_H}{1-p_H} \cdot \frac{\psi}{\Delta S}$, as r_H increases,*

- i. *Holding r_M constant, the firm is more inclined to contract with the salesperson before the inventory is realized.*
- ii. *Holding r_L constant, the firm is more inclined to contract with the salesperson after the inventory is realized.*

As the probability of a high inventory outcome increases, the probability of yield loss (i.e., the realized inventory is less than *H*) correspondingly decreases. As a result, inventory becomes a less restrictive factor, and one might intuit that the firm would be more inclined to contract with the salesperson before observing the realized inventory level, which is what Corollary 2(i) states. However, Corollary 2(ii) states that the opposite could be true under certain conditions. To understand this, note that holding r_L constant, as r_H increases, r_M decreases, reducing the issues that arise from demand censoring. Therefore, the advantages of late contracting dominate. More specifically, holding r_L constant and decreasing r_M lead to an increased ratio of r_L over r_M . Note that the left-hand side of (11) can be rewritten as $1 + r_L/r_M$, which captures the effect of yield uncertainty, whereas the right-hand side of (11) captures the effect of

demand censoring. The first effect occurs because of the possibility that the supply can fall at the lower bound of the support of the demand, whereas the second effect occurs because of the possibility that the supply is inadequate for fulfilling all the demand. The tension between these two effects drives the firm’s optimal timing of contracting. As r_L/r_M increases, in the case of yield loss, the firm is more likely to face a low inventory scenario than a medium inventory scenario. Recall from our previous analysis that contracting after inventory realization helps the firm avoid unnecessary salesforce efforts when faced with a low inventory outcome but might introduce the effect of demand censoring when faced with a medium inventory outcome. Therefore, as the probability of low inventory increases relative to the probability of medium inventory, the firm is more inclined to contract after observing the inventory outcome.

Second, we can consider the role of the demand distribution parameters. We have the following lemma regarding the monotonicity of the right-hand side of (11).

Lemma 4. *In the case of $\rho \geq \frac{1-s_H}{1-p_H} \cdot \frac{\psi}{\Delta S}$, if $p_M \leq q_M$ or $p_M > q_M$ and $\frac{p_H}{q_H} \geq \frac{p_M}{q_M} + \sqrt{\frac{p_M}{q_M} \cdot (p_M - 1) \left(1 + \frac{q_M}{q_H}\right)}$, the right-hand side of (11) decreases in p_H .*

This lemma immediately gives the following corollary.

Corollary 3. *In the case of $\rho \geq \frac{1-s_H}{1-p_H} \cdot \frac{\psi}{\Delta S}$, if p_H is higher, that is, the salesperson’s effort is more effective, the firm is more likely to contract with the salesperson after observing the realized inventory level.*

If the salesperson’s promoting effort is more effective, all else equal, the firm has a better indicator of effort exertion of the salesperson, reducing the issues resulting from demand censoring. Therefore, the advantages of late contracting dominate. In other words, for a higher p_H , the agency cost would be lower, and the firm would share a smaller rent with the salesperson to motivate a high effort level. This would make it more desirable to contract before inventory realization. On the other hand, given the same yield uncertainty, if p_H is higher, although the phenomenon of demand censoring becomes more salient (because, given a medium inventory level, the demand is more likely to exceed the inventory level, leading to unobserved demand), in the case of demand censoring, the firm can expect to share a lower rent to induce salesforce effort as shown in the second part of (10) ($r_M \cdot \psi / (1 - (q_H + q_M) / (p_H + p_M))$). This makes it more desirable to contract before inventory realization. Corollary 3 states that the first effect may

dominate the second one; that is, the reduction in the cost associated with demand censoring may not be as high as the reduction in agency costs without accounting for the effect of inventory.

Our results in this section suggest that the firm may prefer late contracting to early contracting under certain conditions. The reason is that, under late contracting, there are scenarios in which the salesperson does not exert effort; by comparison, under early contracting, the salesperson always exerts effort. This is a novel and nonobvious trade-off between yield uncertainty and demand censoring that our analysis reveals. Our comparison was between early contracting and late contracting, and we do not consider more elaborate contracts in the form of menus or hybrids of early and late contracting because, in practice, such contracts may be difficult to specify and execute. That said, we anticipate our result that late contracting may be sometimes preferred to hold qualitatively; whenever early contracting turns out to be not optimal, in the optimal contract, there must be some contracting elements that allow the agent not to exert effort contingent on inventory information.

In contrast to Proposition 4, the next proposition provides the optimal timing of contracting for the case in which the per-unit revenue is low (i.e., $\rho < \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$):

Proposition 5. *In the case of $\rho < \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, if*

$$p_H r_H B_H^* + (p_M r_H + p_H r_M + p_M r_M) B_M^* \leq r_H \cdot \frac{\hat{\psi}}{1 - \frac{q_H}{p_H}} + r_M \cdot \frac{\hat{\psi}}{1 - \frac{q_H + q_M}{p_H + p_M}} \quad (12)$$

then it is optimal for the firm to contract with the salesperson before the inventory is realized (and the contract is as per Proposition 2(ii)). Otherwise, it is optimal for the firm to contract with the salesperson after the inventory is realized (and the contract is as per Lemma 3).

When the per-unit revenue is sufficiently low, the firm has to offer a positive bonus for an intermediate sales quantity (i.e., $B_M^* > 0$). As a result, the firm has to incur a higher expected salesforce payment than the case in which the per-unit revenue is sufficiently high. Thus, compared with the case with a high per-unit revenue, the firm is less likely to contract with the salesperson before the inventory is realized.

One notable difference between Propositions 4 and 5 is that, in the former, the condition for early contracting to be optimal does not depend on the per-unit revenue, whereas, in the latter, it does (because both B_H^* and B_M^* depend on ρ). The following corollary follows from Proposition 5:

Corollary 4. *In the case of $\rho < \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, as ρ decreases, the firm is less likely to contract with the salesperson before the inventory is realized.*

The intuition behind Corollary 4 is that, in the case of a low per-unit revenue, as the per-unit revenue decreases, the firm faces the pressure to choose a less efficient contract to mitigate double moral hazard. Hence, early contracting becomes a less desirable option.

6. Conclusions and Discussion

We study salesforce compensation incentives in a setting in which both demand and supply are stochastic, and realized supply may be lower than realized demand (and unmet demand is not observable). Our research is the first to study salesforce compensation under supply uncertainty, which is an important real-world issue in many situations. Under moral hazard (i.e., when the salesperson's effort is unobservable to the firm), we characterize the optimal contract and show that it has an extreme convex form in which a bonus is provided only for achieving the highest sales outcome even if low realized sales are due to low realized supply over which the salesperson has no influence (this result is driven by Lemma 1, which we consider as one of our key results). However, when the inventory decision is endogenous but unobservable by the salesperson, that is, there is supply-related moral hazard, double moral hazard arises. We characterize the optimal contract and show that it may be smoother as it may involve positive compensation for intermediate sales outcomes; in fact, under certain conditions, the contract is concave in sales (Corollary 1). This is an important finding because, although previous literature argues that it is optimal to provide a risk-neutral agent an extreme convex contract, this result shows that adding supply-related moral hazard to the mix can lead to the optimal contract being a concave contract (which Zoltners et al. 2006 report is used widely by firms). Our findings, therefore, shed light on how operational considerations may drive the design of incentive contracts, especially when the salesperson does not have complete transparency regarding the inventory-related actions of the firm. In addition, we study whether the firm should contract with the salesperson before or after inventory realization (assuming it has the latter option at all) by characterizing the novel trade-off between avoiding unnecessary marketing expenses because of supply uncertainty (in the case of low yield) and overcompensation because of demand censoring (in the case of intermediate yield).

Our results are of relevance to managers in a number of ways. First, we show that, if there is no moral hazard on the firm's side, that is, the agent has full

transparency into the firm’s inventory-related actions, then, even if yield is random and even though the salesperson’s effort has no impact on inventory yield, the firm can use performance-based compensation contracts based on realized sales that are convex in shape. Second, we show that, if there is moral hazard on the firm’s side as well, that is, the agent does not have full transparency into the firm’s inventory-related actions, then the firm has to make the contract smoother and, under certain conditions, may even have to make this contract concave; we specify these conditions. Third, we show that, under random yield, it may sometimes be beneficial for the firm to wait until yield uncertainty is resolved to decide the compensation contract; however, it may not always be beneficial to do so, and we specify these conditions.

Our research can be extended in a number of directions. For instance, in the current model, we have assumed that the salesperson obtains inventory information exogenously. However, if the firm endogenously reveals inventory information, then it can determine when and how much information to release, and its revelation strategy itself may signal the inventory level. We have also assumed that the price of the product is exogenous. One alternative scenario (that is beyond the scope of this study) is to make this price endogenous and empower the salesperson with some ability to determine this price (Simester and Zhang 2014). In such an effort, it would be important to carefully determine the timing of the pricing decision as before or after yield uncertainty is resolved. Another interesting direction would be to study the case of asymmetric information, in which the firm or the salesperson may have better information than the other party on inventory or demand. A stream of literature on the interfaces of operations and marketing (e.g., Taylor 2006, Iyer et al. 2007, and Bialogorsky and Koenigsberg 2010) has examined the cases in which the downstream player in a supply chain may own the inventory; although it might not be reasonable in our case to assume that the salesperson rather than the firm can own the inventory, it is possible that the firm and the salesperson have different information about the inventory yield distribution. It would also be interesting to analyze how the firm may want to alleviate some of these problems; for example, should the firm invest in reducing the variability of supply to reduce the firm’s moral hazard issue or should it invest in improving the observability of lost demand? Finally, in many industries, firms may purchase insurance products (e.g., crop insurance in the agricultural setting) to hedge yield-related risks. An interesting research problem would be to study how yield-related insurance options impact promotional effort and salesforce compensation.

Along these lines, future research may incorporate financial hedging strategies (e.g., futures and options contracts) that are commonly employed in the semiconductor and related industries.

Acknowledgments

The authors are grateful to the senior editor, the associate editor, and two anonymous reviewers for the constructive review process. The authors thank Charles Angelucci, Dilip Mookherjee, and Raphael Thomadsen; as well as seminar participants at City University of Hong Kong, City University of New York, Columbia University, Duke University, Hong Kong University of Science and Technology, Washington University in St. Louis, University of Chicago, University of Southern California, Yale University, the 37th ISMS Marketing Science Conference (2015), INFORMS International Conference (2016), MSOM Conference (2016), and Zero-Decade Marketing Theory Symposium (2019) for their helpful comments.

Appendix. Proofs

Proof of Lemma 1. We first prove that $\frac{p_M^Y}{q_M^Y} < \frac{p_H^Y}{q_H^Y}$. By MLRP, we have

$$\frac{p_H}{q_H} > \frac{p_M}{q_M} > \frac{p_L}{q_L}. \tag{A.1}$$

Hence, we have

$$\begin{aligned} \frac{p_M^Y}{q_M^Y} &= \frac{r_H p_M + r_M p_M + r_M p_H}{r_H q_M + r_M q_M + r_M q_H} \\ &= \frac{(r_H + r_M) \cdot p_M + r_M p_H}{(r_H + r_M) \cdot q_M + r_M q_H} \\ &= \frac{\frac{p_M}{q_M} + \frac{r_M}{r_H + r_M} \cdot \frac{p_H}{q_M}}{1 + \frac{r_M}{r_H + r_M} \cdot \frac{q_H}{q_M}} \\ &< \frac{\frac{p_H}{q_H} + \frac{r_M}{r_H + r_M} \cdot \frac{p_H}{q_M}}{1 + \frac{r_M}{r_H + r_M} \cdot \frac{q_H}{q_M}} \text{ (by MLRP)} \\ &= \frac{\frac{p_H}{q_H} \cdot \left(1 + \frac{r_M}{r_H + r_M} \cdot \frac{q_H}{q_M}\right)}{1 + \frac{r_M}{r_H + r_M} \cdot \frac{q_H}{q_M}} \\ &= \frac{p_H}{q_H} = \frac{r_H p_H}{r_H q_H} = \frac{p_H^Y}{q_H^Y}. \end{aligned}$$

Next, we prove that $\frac{p_L^Y}{q_L^Y} < \frac{p_H^Y}{q_H^Y}$. Note from (A.1) that $p_H > q_H$ and $p_L < q_L$,¹² which gives $p_L/q_L < 1 < p_H/q_H$. Therefore,

$$\begin{aligned} \frac{p_L^Y}{q_L^Y} &= \frac{r_L + (1 - r_L)p_L}{r_L + (1 - r_L)q_L} \\ &< 1 \text{ (by } p_L < q_L) \\ &< \frac{p_H}{q_H} \text{ (by } p_H/p_L > 1) \\ &= \frac{r_H p_H}{r_H q_H} = \frac{p_H^Y}{q_H^Y}. \end{aligned}$$

Therefore, we have $\max\left\{\frac{p_L^Y}{q_L^Y}, \frac{p_M^Y}{q_M^Y}\right\} < \frac{p_H^Y}{q_H^Y}$. Q.E.D.

Proof of Proposition 1. Suppose that the firm pays the salesperson a bonus, denoted by B_H , when the sales outcome

is high and zero otherwise. The firm’s problem can be written as the following program:

$$\begin{aligned} \max_{B_H} \quad & \rho \cdot [r_H p_H H + (r_H p_M + r_M p_M + r_M p_H)M \\ & + (r_L + p_L - r_L p_L)L] - r_H p_H B_H \quad (\text{A.2}) \\ \text{s.t.} \quad & r_H p_H B_H - \psi \geq r_H q_H B_H \quad (\text{IC}) \\ & r_H p_H B_H - \psi \geq 0. \quad (\text{IR}) \end{aligned}$$

Clearly, the individual rationality (IR) constraint follows from the incentive compatibility (IC) constraint and is, thus, redundant. By solving this problem, we have $B_H^* = \frac{\psi}{r_H(p_H - q_H)}$. The firm’s expected payment to the salesperson is, thus, $r_H p_H B_H^* = \frac{\psi}{1 - q_H/p_H}$.

Next, we show that the firm only provides the salesperson a bonus when the sales outcome is H . Suppose the firm uses a different compensation scheme in which the salesperson is paid a bonus B_i when the sales outcome $i \in \{H, M, L\}$, and at least one of $\{B_M, B_L\}$ is positive. The firm’s problem can be written as the following program:

$$\begin{aligned} \max_{B_H} \quad & r_H p_H (\rho H - B_H) + (r_H p_M + r_M p_M + r_M p_H) \\ & \cdot (\rho M - B_M) + (r_L + p_L - r_L p_L)(\rho L - B_L) \quad (\text{A.3}) \\ \text{s.t.} \quad & r_H p_H B_H + (r_H p_M + r_M p_M + r_M p_H)B_M \\ & + (r_L + p_L - r_L p_L)B_L - \psi \\ & \geq r_H q_H B_H + (r_H q_M + r_M q_M + r_M q_H)B_M \\ & + (r_L + q_L - r_L q_L)B_L \quad (\text{IC}) \\ & r_H p_H B_H + (r_H p_M + r_M p_M + r_M p_H)B_M \\ & + (r_L + p_L - r_L p_L)B_L - \psi \geq 0. \quad (\text{IR}) \end{aligned}$$

Again, the IR constraint follows from the IC constraint and is, thus, redundant. The IC constraint gives $B_H \geq (\psi - \epsilon_M - \epsilon_L) / (r_H p_H - r_H q_H)$, where $\epsilon_M = B_M[(r_H p_M + r_M p_M + r_M p_H) - (r_H q_M + r_M q_M + r_M q_H)]$ and $\epsilon_L = B_L[(r_L + p_L - r_L p_L) - (r_L + q_L - r_L q_L)]$. Therefore, the firm’s expected payment to the salesperson is

$$\begin{aligned} & r_H p_H B_H + (r_H p_M + r_M p_M + r_M p_H)B_M \\ & + (r_L + p_L - r_L p_L)B_L \\ & \geq r_H p_H \cdot \frac{\psi - \epsilon_M - \epsilon_L}{r_H p_H - r_H q_H} \\ & + (r_H p_M + r_M p_M + r_M p_H) \\ & \cdot \frac{\epsilon_M}{(r_H p_M + r_M p_M + r_M p_H) - (r_H q_M + r_M q_M + r_M q_H)} \\ & + (r_L + p_L - r_L p_L) \\ & \cdot \frac{\epsilon_L}{(r_L + p_L - r_L p_L) - (r_L + q_L - r_L q_L)} \quad (\text{A.4}) \end{aligned}$$

$$= \frac{\psi - \epsilon_M - \epsilon_L}{1 - \frac{r_H q_H}{r_H p_H}} + \frac{\epsilon_M}{1 - \frac{r_H q_M + r_M q_M + r_M q_H}{r_H p_M + r_M p_M + r_M p_H}} + \frac{\epsilon_L}{1 - \frac{r_L + q_L - r_L q_L}{r_L + p_L - r_L p_L}} \quad (\text{A.5})$$

$$> \frac{\psi - \epsilon_M - \epsilon_L}{1 - \frac{r_H q_H}{r_H p_H}} + \frac{\epsilon_M}{1 - \frac{r_H q_H}{r_H p_H}} + \frac{\epsilon_L}{1 - \frac{r_H q_H}{r_H p_H}} \quad (\text{by Lemma 1}) \quad (\text{A.6})$$

$$= \frac{\psi}{1 - \frac{q_H}{p_H}}. \quad (\text{A.7})$$

Therefore, the firm is better off by paying the salesperson a positive bonus only when the sales outcome is H . Q.E.D.

Proof of Proposition 2. When $\rho \geq \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, we have

$$\rho \Delta S \geq \frac{\psi}{1 - \frac{q_H}{p_H}} - \frac{s_H}{r_H} \cdot \frac{\psi}{1 - \frac{q_H}{p_H}}. \quad (\text{A.8})$$

Note that the left-hand side of (A.8) denotes the absolute value of the firm’s expected revenue loss by switching its inventory-related action from a_H to a_L , whereas the right-hand side of (A.8) denotes the absolute value of the firm’s expected saving from its compensation to the salesperson for motivating an effort of e_H . In other words, by choosing a less effective inventory-related action (a_L), the firm’s revenue loss outweighs its savings from salesforce compensation. Thus, under the optimal salesforce compensation contract as characterized in Section 3, the firm does not have any incentive to change its inventory-related action from a_H to a_L as doing so would reduce its expected profit.

When $\rho < \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$, the optimal contract characterized in Section 3 cannot sustain because, if the salesperson chooses a high effort level, the firm is better off choosing an inventory-related action of a_L , which results in an expected reduction from salesforce force compensation, represented by $\frac{\psi}{1 - \frac{q_H}{p_H}} - \frac{s_H}{r_H} \cdot \frac{\psi}{1 - \frac{q_H}{p_H}}$ that is greater than the expected revenue loss ($\Delta S \cdot \rho$).

Thus, the firm must provide the salesperson with a positive bonus B_M when the sales outcome is M . The firm’s objective can be formulated as

$$\begin{aligned} & [p_H r_H H + (p_M r_H + p_H r_M + p_M r_M)M \\ & + (p_L + r_L - p_L r_L)L] \cdot \rho - p_H r_H B_H \\ & - (p_M r_H + p_H r_M + p_M r_M)B_M. \quad (\text{A.9}) \end{aligned}$$

The bonus B_M must be large enough such that the firm does not have any incentive to choose an inventory-related action of a_L (instead of a_H):

$$\begin{aligned} \Delta S \cdot \rho \geq & p_H r_H B_H + (p_H r_M + p_M r_H + p_M r_M)B_M \\ & - [p_H s_H B_H + (p_H s_M + p_M s_H + p_M s_M)B_M] \\ & = (r_H - s_H)p_H B_H \\ & - [(s_M - r_M)(p_H + p_M) - (r_H - s_H)p_M]B_M. \quad (\text{A.10}) \end{aligned}$$

In addition, the individual rationality constraint applies to ensure that the salesperson finds it optimal to exert high effort:

$$\begin{aligned} & p_H r_H B_H + (p_H r_M + p_M r_H + p_M r_M)B_M - \psi \\ & \geq q_H r_H B_H + (q_H r_M + q_M r_H + q_M r_M)B_M. \quad (\text{A.11}) \end{aligned}$$

The optimal contract parameters follow from maximizing (A.9) subject to (A.10) and (A.11). Q.E.D.

Proof of Corollary 1. The compensation plan (B_H^*, B_M^*, B_L^*) is concave (convex) if and only if $(B_H^* - B_M^*) / (H - M)$ is less (greater) than $(B_M^* - B_L^*) / (M - L)$.

If we view B_H^* and B_M^* as functions of ρ , that is,

$$B_H^*(\rho) = \frac{\tau_1 \rho \Delta S + \tau_2 \psi}{\tau_1 (r_H - s_H) p_H + \tau_2 (p_H - q_H) r_H} \text{ and}$$

$$B_M^*(\rho) = \frac{(r_H - s_H) p_H \psi - (p_H - q_H) r_H \rho \Delta S}{\tau_1 (r_H - s_H) p_H + \tau_2 (p_H - q_H) r_H},$$

then

$$\begin{aligned} \frac{dB_H^*(\rho)}{d\rho} &= \frac{\tau_1 \Delta S}{\tau_1(r_H - s_H)p_H + \tau_2(p_H - q_H)r_H} > 0, \\ \frac{dB_M^*(\rho)}{d\rho} &= \frac{-(p_H - q_H)r_H \Delta S}{\tau_1(r_H - s_H)p_H + \tau_2(p_H - q_H)r_H} < 0. \end{aligned}$$

If we view $(B_H^* - B_M^*)/(H - M) - (B_M^* - B_L^*)/(M - L)$ as a function of ρ and denote it by $\Delta B(\rho)$, then

$$\frac{\Delta B(\rho)}{\rho} = \frac{1}{H - M} \cdot \frac{dB_H^*(\rho)}{d\rho} - \left(\frac{1}{H - M} + \frac{1}{M - L} \right) \cdot \frac{dB_M^*(\rho)}{d\rho} > 0. \tag{A.12}$$

In other words, $\Delta B(\rho)$ increases in ρ . Setting $\Delta B(\rho) = 0$ gives

$$\hat{\rho} = \frac{(H - L)(r_H - s_H)p_H - (M - L)\tau_2}{(H - L)(p_H - q_H)r_H + (M - L)\tau_1} \cdot \frac{\psi}{\Delta S}. \tag{A.13}$$

By (A.12), we have $\Delta B(\rho) < 0$ if $\rho < \hat{\rho}$, and $\Delta B(\rho) \geq 0$ otherwise. Furthermore, we have from (A.13) that

$$\hat{\rho} < \frac{(H - L)(r_H - s_H)p_H}{(H - L)(p_H - q_H)r_H} \cdot \frac{\psi}{\Delta S} = \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S},$$

which completes the proof. Q.E.D.

Proof of Proposition 3. First, we show that an inventory-related action of a_L is a dominated option. Note that, when the firm chooses an inventory-related action of a_L , because such an action is observable to the salesperson, using an argument similar to those in the proof of Proposition 1, in the optimal contract, the firm pays the salesperson a bonus only for achieving sales equal to H , and the value of the bonus is $\frac{\psi}{s_H(p_H - q_H)}$. The firm’s expected payment to the salesperson is given by

$$s_H p_H \cdot \frac{\psi}{s_H(p_H - q_H)} = \frac{\psi}{1 - \frac{q_H}{p_H}},$$

which is equal to the firm’s expected payment to the salesperson when choosing an inventory-related action of a_H . In addition, we have from $\Delta S > 0$ that the firm’s expected revenue from an inventory-related action of a_L is lower than that from an inventory-related action of a_H . Therefore, the firm is better off choosing an inventory-related action of a_H instead of a_L .

Next, given that the inventory-related action is a_H , the optimal contract follows from Proposition 1. Q.E.D.

Proof of Lemma 2. Depending on the size of ρ , we have two cases:

i. $\rho \geq \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$. Consider the case in which the firm offers a contract before inventory realization, according to which a bonus of B is provided when $Y = H$. Under the contract, if the salesperson chooses to exert effort before inventory realization, the salesperson’s expected utility is $r_H p_H B - \psi$. If the salesperson waits until observing the realized inventory, then the salesperson would only exert effort when the inventory $I = H$, incurring a cost of $\hat{\psi} > \psi$. Thus, the salesperson’s expected utility is $r_H(p_H B - \hat{\psi})$. The condition for the salesperson to exert effort before inventory realization is $r_H p_H B - \psi \geq r_H(p_H B - \hat{\psi})$, which is equivalent to $\hat{\psi} \geq \psi/r_H$.

ii. $\rho < \frac{1 - \frac{s_H}{r_H}}{1 - \frac{q_H}{p_H}} \cdot \frac{\psi}{\Delta S}$. Note from Proposition 2(ii) that, in this case, the salesperson may receive a positive bonus when the sales outcome is M . Given a compensation contract with $B_H^*, B_M^* > 0$, by exerting effort before inventory realization, the salesperson’s expected utility is

$$r_H p_H B_H^* + (r_H p_M + r_M p_M + r_M p_H) B_M^* - \psi. \tag{A.14}$$

If the salesperson waits until observing the realized inventory and the realized inventory is M , the salesperson may or may not exert effort depending on the system parameter values. We discuss both cases one by one.

a. If the salesperson is willing to exert effort even if the realized inventory is M , the salesperson’s expected utility from waiting is

$$r_H(p_H B_H^* + p_M B_M^* - \hat{\psi}) + r_M[(p_H + p_M) B_M^* - \hat{\psi}]. \tag{A.15}$$

By comparing (A.14) with (A.15), we find the condition for the salesperson to exert effort before inventory realization is $\hat{\psi} \geq \psi/(r_H + r_M)$, which follows from $\hat{\psi} \geq \psi/r_H$.

b. If the salesperson is willing to exert effort only if the realized inventory is H , the salesperson’s expected utility from waiting is

$$r_H(p_H B_H^* + p_M B_M^* - \hat{\psi}). \tag{A.16}$$

By comparing (A.14) with (A.16), we identify the condition for the salesperson to exert effort before inventory realization: $\hat{\psi} > \frac{\psi}{r_H} - \frac{r_M(p_H + p_M)}{r_H} \cdot B_M^*$, which, again, follows from $\hat{\psi} \geq \psi/r_H$. Q.E.D.

Proof of Lemma 3. Similar to the argument in section 4 of Dai and Jerath (2013). Q.E.D.

Proof of Proposition 4. By comparing the firm’s expected salesforce compensation in the case of contracting after inventory is realized [see Equation (10)] against that in the case of contracting before inventory is realized [see Proposition 2(i)]. Q.E.D.

Proof of Corollary 2. Note that the left-hand side of (11), using $r_H + r_M + r_L = 1$, can be rewritten as $1 + r_L/r_M$, which increases in r_L and decreases in r_M ; the right-hand side of (11) is independent of $r_i, i \in \{H, M, L\}$. The corollary thus follows from Proposition 4. Q.E.D.

Proof of Lemma 4. Let us write the right-hand side of (11) as a function of p_H , that is, $f(p_H) = \hat{\psi}/\psi \cdot (1 - q_H/p_H)/(1 - (q_H + q_M)/(p_H + p_M))$. Its first-order derivative

$$f'(p_H) = \frac{\hat{\psi}}{\psi} \cdot \frac{p_M q_H (p_M - q_M) - (p_M q_H^2 + p_H^2 q_M - 2 p_H p_M q_H)}{p_H^2 (p_H + p_M - q_H - q_M)^2}$$

has a positive denominator. If $p_M \leq q_M$, its numerator

$$\begin{aligned} & p_M q_H (p_M - q_M) - (p_M q_H^2 + p_H^2 q_M - 2 p_H p_M q_H) \\ & \leq 0 - (p_M q_H^2 + p_H^2 q_M - 2 p_H p_M q_H) \\ & = -p_M (q_H^2 + p_H^2 - 2 p_H q_H) \\ & = -p_M (p_H - q_H)^2 < 0. \end{aligned}$$

If $p_M > q_M$, however, solving $p_M q_H (p_M - q_M) - (p_M q_H^2 + p_H^2 q_M - 2 p_H p_M q_H) < 0$ yields $\frac{p_H}{q_H} \geq \frac{p_M}{q_M} + \sqrt{\frac{p_M}{q_M} \cdot \left(\frac{p_M}{q_M} - 1 \right) \left(1 + \frac{q_M}{q_H} \right)}$. Q.E.D.

Proof of Corollary 3. Follows from Lemma 4. Q.E.D.

Proof of Proposition 5. By comparing the firm’s expected salesforce compensation in the case of contracting after inventory is realized [see Equation (10)] against that in the case of contracting before inventory is realized [see Proposition 2(ii)]. Q.E.D.

Proof of Corollary 4. It suffices to prove that the left-hand side of (12), that is,

$$p_H r_H B_H^* + (p_M r_H + p_H r_M + p_M r_M) B_M^* \quad (\text{A.17})$$

$$= \frac{p_H r_H (\tau_1 \rho \Delta S + \tau_2 \psi) + (p_M r_H + p_H r_M + p_M r_M) [(r_H - s_H) p_H \psi - (p_H - q_H) r_H \rho \Delta S]}{\tau_1 (r_H - s_H) p_H + \tau_2 (p_H - q_H) r_H} \quad (\text{A.18})$$

decreases in ρ .

Note that the denominator of (A.18) is independent of ρ . The first-order derivative of the numerator of (A.18) in terms of ρ is

$$[p_H r_H \tau_1 - (p_M r_H + p_H r_M + p_M r_M)(p_H - q_H) r_H] \Delta S,$$

which can be reorganized as

$$r_H q_H q_M [r_H + r_M] \cdot \left(\frac{p_M}{q_M} - \frac{p_H}{q_H} \right) \Delta S.$$

This quantity is negative because of MLRP. Hence, the proof is complete. Q.E.D.

Endnotes

¹ There is a recent literature on robust contracts in which the principal evaluates possible contracts by their worst-case performance over unknown actions that the agent may take (Antić 2014, Carroll 2015, Carroll and Meng 2016, Yu and Kong 2017). Our work here is of a different flavor as it can be thought of as the traditional principal-agent problem of Holmstrom (1979) but with uncertain supply that can sometimes be less than demand.

² See <https://bit.ly/2BIFWfB>.

³ We assume for this extension that advanced contracts, such as those that include menus of contracts, cannot be used because of practical contracting frictions. Otherwise, a sufficiently complex early contract can achieve any outcome that a late contract can achieve.

⁴ Note that MLRP for the supply side is neither necessary nor sufficient for our results to hold. We have assumed it as a parallel to the assumption of MLRP on the demand side, which makes it a palatable assumption. A sufficient condition for our analysis in Section 4 to hold is $r_H > s_H$ and $r_M < s_M - \frac{p_M}{p_H + p_M} \cdot (r_H - s_H)$ and $r_L < s_L$.

⁵ It is well known that, in a situation without inventory considerations, the MLRP on the demand distribution implies that the firm only needs to reward the salesperson when the highest demand outcome is achieved (see, e.g., Laffont and Martimort 2001).

⁶ Recent work has shown that linear contracts can be optimal in a setting with risk neutrality with limited liability of the agent. For instance, Kräkel and Schöttner (2016) and Jerath and Long (2018) show this in a dynamic setting, and Antić (2014), Carroll (2015), and Yu and Kong (2017) show this under a robust contracting paradigm with uncertainty about the agent’s technology, the agent’s action set, and the agent’s effectiveness, respectively. In our case, the optimal contract may be linear (though a linear contract is not always optimal).

⁷ Note that, in addition to MLRP, the continuous formulation in Park (1995) and Kim (1997) needs the convexity of the distribution function condition to hold although Oyer (2000) needs that the participation

constraint of the agent should not be binding; in this sense, the continuous formulations may be considered more restrictive than the discrete formulations. Furthermore, there are certain well-known technical challenges associated with the continuous modeling framework (most notably, regarding the first-order approach typically used under the assumptions of risk neutrality and limited liability of the salesperson). As Laffont and Martimort (2001, pp. 200–201) point out, “The first-order approach has been one of the most debated issues in contract theory” because the validity of the approach has not been well established, and “when the first-order approach is not valid, using it can be very misleading.” As a result, “most of the applied moral hazard literature” adopts a discrete formulation.

⁸ For conciseness of analysis, we consider a parametric space in which the firm prefers the salesperson to choose an effort of e_H regardless of whether the firm takes an inventory-related action of a_H or a_L .

⁹ This is easy to generate using a convex cost function of within-day effort. Assume that the salesperson can effect the change from (q_L, q_M, q_H) to (p_L, p_M, p_H) by working for a total of D hours. Also assume that to work h hours in a day, the salesperson’s cost is h^2 , that is, the marginal cost of effort of every additional hour worked in a day is higher. Lets say the salesperson has 10 days to work D hours, so the salesperson works $D/10$ hours per day at a total cost of $10(D/10)^2 = D^2/10$. Now suppose the salesperson must work D hours in five days, that is, the salesperson needs to work $D/5$ hours per day at a total cost of $5(D/5)^2 = D^2/5$, which is higher.

¹⁰ We note that this holds even for a demand distribution with more than three points of support.

¹¹ A variation to the timeline in Figure 4 is that the inventory is realized after the sales contract is determined but before the salesperson exerts effort. In this case, the firm again has the choice of whether it should disclose the inventory level to the salesperson (under the requirement that, if there is disclosure, it must be truthful). Following the same arguments as before, we can see that the firm discloses the inventory level.

¹² We can show this by contradiction. Suppose $p_H < q_H$; then (A.1) implies both $p_M < q_M$ and $p_L < q_L$, which is impossible to hold because $p_H + p_M + p_L = q_H + q_M + q_L = 1$. Likewise, suppose $p_L > q_L$; then (A.1) implies both $p_M > q_M$ and $p_H > p_H$, which is impossible to hold because $p_H + p_M + p_L = q_H + q_M + q_L = 1$.

References

- Anupindi R, Dada M, Gupta S (1998) Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Sci.* 17(4):406–423.
- Antić N 2014. Contracting with unknown technologies. Working paper, Northwestern University, Evanston, IL.
- Basu AK, Lal R, Srinivasan V, Staelin R (1985) Salesforce compensation plans: An agency theoretic perspective. *Marketing Sci.* 4(4): 267–291.
- Besbes O, Muharremoglu A (2012) On implications of demand censoring in the newsvendor problem. *Management Sci.* 59(6):1407–1424.
- Bhattacharyya S, Lafontaine F (1995) Double-sided moral hazard and the nature of share contracts. *RAND J. Econom.* 26(4):761–781.
- Biyalogorsky E, Koenigsberg O (2010) Ownership coordination in a channel: Incentives, returns, and negotiations. *Quant. Marketing Econom.* 8(4):461–490.
- Braden DJ, Freimer M (1991) Informational dynamics of censored observations. *Management Sci.* 37(11):1390–1404.
- Bramley RGV, Hamilton RP (2004) Understanding variability in wine-grape production systems. *Australian J. Grape Wine Res.* 10(1):32–45.
- Cachon GP, Terweisch C (2012) *Matching Supply with Demand: An Introduction to Operations Management*, 3rd ed. (McGraw-Hill Education, New York).
- Carroll G (2015) Robustness and linear contracts. *Amer. Econom. Rev.* 105(2):536–563.

- Carroll G, Meng D (2016) Locally robust contracts for moral hazard. *J. Math. Econom.* 62(1):36–51.
- Chen F (2005) Salesforce incentives, market information, and production/inventory planning. *Management Sci.* 51(1):60–75.
- Chen L, Plambeck EL (2008) Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing Service Oper. Management* 10(2):236–256.
- Chen L, Mersereau AJ, Wang Z (2017) Optimal merchandise testing with limited inventory. *Oper. Res.* 65(4):968–991.
- Chung DJ, Steenburgh T, Sudhir K (2014) Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans. *Marketing Sci.* 33(2):165–187.
- Conlon CT, Mortimer JH (2013) Demand estimation under incomplete product availability. *Amer. Econom. J. Microeconom.* 5(4):1–30.
- Cooper R, Ross TW (1985) Product warranties and double moral hazard. *RAND J. Econom.* 16(1):103–113.
- Dada M, Petrucci NC, Schwarz LB (2007) A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing Service Oper. Management* 9(1):9–32.
- Dai T, Cho S-H, Zhang F (2016) Contracting for on-time delivery in the US influenza vaccine supply chain. *Manufacturing Service Oper. Management* 18(3):332–346.
- Dai T, Jerath K (2013) Salesforce compensation with inventory considerations. *Management Sci.* 59(11):2490–2501.
- Dai T, Jerath K (2016) Impact of inventory on quota-bonus contracts with rent sharing. *Oper. Res.* 64(1):94–98.
- Ding X, Puterman ML, Bisi A (2002) The censored newsvendor and the optimal acquisition of information. *Oper. Res.* 50(3):517–527.
- Downs B, Metters R, Semple J (2001) Managing inventory with multiple products, lags in delivery, resource constraints, and lost sales: A mathematical programming approach. *Management Sci.* 47(3):464–479.
- Herweg F, Muller D, Weinschenk P (2010) Binary payment schemes: Moral hazard and loss aversion. *Amer. Econom. Rev.* 100(5):2451–2477.
- Holmstrom B (1979) Moral hazard and observability. *Bell J. Econom.* 10(1):74–91.
- Holmstrom B, Milgrom P (1987) Aggregation and linearity in the provision of intertemporal incentives. *Econometrica* 55(2):303–328.
- Iyer G, Narasimhan C, Niraj R (2007) Information and inventory in distribution channels. *Management Sci.* 53(10):1551–1561.
- Jerath K, Long F (2018). Multi-period incentive contracts. Working paper, Columbia University, New York.
- Kamrad B, Ernst R (2001) An economic model for evaluating mining and manufacturing ventures with output yield uncertainty. *Oper. Res.* 49(5):690–699.
- Kazaz B (2004) Production planning under yield and demand uncertainty with yield-dependent cost and price. *Manufacturing Service Oper. Management* 6(3):209–224.
- Kim SK (1997) Limited liability and bonus contracts. *J. Econom. Management Strategy* 6(4):899–913.
- Kim SK, Wang S (1998) Linear contracts and the double moral-hazard. *J. Econom. Theory* 82(2):342–378.
- Kräkel M, Schöttner A (2016) Optimal sales force compensation. *J. Econom. Behav. Organ.* 126(A):179–195.
- Laffont JJ, Martimort D (2001) *The Theory of Incentives: The Principal-Agent Model* (Princeton University Press, Princeton, NJ).
- Lal R, Srinivasan V (1993) Compensation plans for single-and multi-product salesforces: An application of the Holmstrom-Milgrom model. *Management Sci.* 39(7):777–793.
- Lee HL, Yano CA (1988) Production control in multistage systems with variable yield losses. *Oper. Res.* 36(2):269–278.
- Liu S, So KC, Zhang F (2010) Effect of supply reliability in a retail setting with joint marketing and inventory decisions. *Manufacturing Service Oper. Management* 12(1):19–32.
- Lu X, Song JS, Zhu K (2008) Analysis of perishable-inventory systems with censored demand data. *Oper. Res.* 56(4):1034–1038.
- Oyer P (2000) A theory of sales quotas with limited liability and rent sharing. *J. Labor Econom.* 18(3):405–426.
- Park E-S (1995) Incentive contracting under limited liability. *J. Econom. Management Strategy* 4(3):477–490.
- Plambeck E, Zenios S (2003) Incentive efficient control of a make-to-stock production system. *Oper. Res.* 51(3):371–386.
- Porteus EL (2002) *Foundations of Stochastic Inventory Theory* (Stanford University Press, Redwood City, CA).
- Priest GL (2010) Timing “disturbances” in labor market contracting: Roth's findings and the effects of labor market monopsony. *J. Labor Econom.* 28(2):447–472.
- Rajaram K, Karmarkar US (2002) Product cycling with uncertain yields: Analysis and application to the process industry. *Oper. Res.* 50(4):680–691.
- Raju JS, Srinivasan V (1996) Quota-based compensation plans for multiterritory heterogeneous salesforces. *Management Sci.* 42(10):1454–1462.
- Raman A, DeHoratius N, Ton Z (2001) Execution: The missing link in retail operations. *Calif. Management Rev.* 43(3):136–152.
- Roels G, Karmarkar US, Carr S (2010) Contracting for collaborative services. *Management Sci.* 56(5):849–863.
- Romano RE (1994) Double moral hazard and resale price maintenance. *RAND J. Econom.* 25(3):455–466.
- Roth AE, Xing X (1994) Jumping the gun: Imperfections and institutions related to the timing of market transactions. *Amer. Econom. Rev.* 84(4):992–1044.
- Rubel O, Prasad A (2016) Dynamic incentives in sales force compensation. *Marketing Sci.* 35(4):676–689.
- Rudi N, Drake D (2014) Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. *Management Sci.* 60(5):1334–1345.
- Sales Leadership Forum (2010) *The Art & Science of Quota Setting* (SalesGlobe, Atlanta).
- Sappington D (1983) Limited liability contracts between principal and agent. *J. Econom. Theory* 29(1):1–21.
- Simester D, Zhang J (2014) Why do salespeople spend so much time lobbying for low prices? *Marketing Sci.* 33(6):796–808.
- Stapper CH, Rosner RJ (1995) Integrated circuit yield management and yield analysis: Development and implementation. *IEEE Trans. Semiconductor Manufacturing* 8(2):95–102.
- Taylor TA (2006) Sale timing in a supply chain: When to sell to the retailer. *Manufacturing Service Oper. Management* 8(1):23–42.
- Wein AS (1992) Random yield, rework and scrap in a multistage batch manufacturing environment. *Oper. Res.* 40(3):551–563.
- Yano CA, Lee HL (1995) Lot sizing with random yields: A review. *Oper. Res.* 43(2):311–334.
- Yu Y, Kong X (2017) Robust contract designs: Linear contracts for sales force compensation. Working paper, City University of Hong Kong, Hong Kong.
- Zoltners AA, Sinha P, Lorimer SE (2008) Sales force effectiveness: A framework for researchers and practitioners. *J. Personal Selling Sales Management* 28(2):115–131.
- Zoltners AA, Sinha P, Zoltners GA (2006) *The Complete Guide to Sales Force Incentive Compensation: How to Design and Implement Plans that Work* (AMACOM, New York).