Optimal Salesforce Compensation with General Demand and Operational Considerations

Haotian Song
School of Management, Zhejiang University, Hangzhou, 310058, China, hasong@zju.edu.cn

Guoming Lai
McCombs School of Business, The University of Texas, Austin, Texas 78712, USA, guoming.lai@mccombs.utexas.edu

Wenqiang Xiao
Stern School of Business, New York University, New York, NY 10012, USA, wx2@stern.nyu.edu

Problem definition: We investigate the optimal sales compensation scheme in the context of private information and unobservable actions, while considering common operational factors encountered in practice including inventory costs, lost sales, and ordering delegation. **Methodology/results:** Based on an agency model with general demand and cost functions, we derive optimality conditions for implementable contracts that can achieve the second-best outcome in such scenarios. The contracts are in the forms of a menu with linear compensation for demand or sales, incorporating inventory costs. Moreover, the contracts feature adjustments in compensation corresponding to the ordering level if it is delegated. **Managerial implications:** Our study reveals that, under reasonably mild conditions, optimal sales contracts can still maintain relatively simple forms, even when confronted with common operational factors and generalized demand and cost functions. However, the contracts must be tailored to suit the operational settings, incorporating adjustments based on inventory costs, lost sales, and ordering delegation. Intriguingly, neither the loss of demand information nor the delegation of inventory decisions would compromise system efficiency at optimum.

Key words: optimal contract, sales and operations planning, supply/demand mismatch, demand censoring, inventory delegation

1. Introduction

Sales compensation design is vital for motivating sales teams and driving firm success. Its impact spans various dimensions, including gathering market information and stimulating demand. However, despite extensive research on the design, a gap exists in understanding how sales forces interact with operational factors, such as supply and demand mismatch, censored demand information due to lost sales, and ordering delegation, which can substantially impact a firm's profitability. Indeed, effectively managing the supply and demand mismatch poses a formidable challenge across diverse industries like groceries, fashion, electronics, and pharmaceuticals. These sectors grapple with product characteristics such as short life-cycles and lengthy lead time, leading to rapid inventory depreciation and high demand uncertainty. Pharmaceutical companies face excess inventories and shortages simultaneously (Ebel et al. 2013). Major retailers like Wal-Mart and Whole Foods have struggled with shortages, resulting in disappointed customers (Dudley 2014, Peterson 2018). These mismatches have led to financial losses amounting to billions of dollars.

The root causes of these losses are attributed not only to product characteristics but also to ineffective coordination between sales divisions or retail outlets and central offices. In practice, firms employ varied approaches to inventory management, with some delegating ordering decisions to sales divisions or retail outlets. The rationale for delegation is based on the belief that the sales divisions, being in close proximity to customers and the local market, possess valuable information and insights regarding evolving customer preferences, local competition dynamics, sales events, and the current financial conditions and purchasing intentions of major clients. This proximity could allow the sales divisions to gather real-time data and firsthand knowledge that can inform inventory decisions. However, the sales divisions may not always leverage their superior information in a manner that aligns with the firm's best interest. This misalignment becomes particularly pronounced when the sales divisions are incentivized through sales commissions or bonuses tied to meeting sales targets. These incentives encourage them to boost sales by maintaining ample available inventory, resulting in systematic over-ordering or inflated demand forecasts.

Caro and Gallien (2010) reported Zara's inventory ordering process, which granted store managers full control or direct influence over inventory decisions at centralized distribution centers through order placements. However, these managers, whose compensation was tied to total sales in their stores, were enticed to frequently order more than necessary, especially when facing potential rationing risk from distribution centers. Similarly, a global pharmaceutical company, as documented by Scheele et al. (2018), relied on sales forecasts for monthly inventory decisions but found a systematic over-forecasting, with an average inflation of 16.2%. This could be blamed on the company's incentive system, which included a sales bonus but lacked rewards or penalties based on forecast accuracy. Moreover, Van Donselaa et al. (2010) examined a European supermarket chain where store managers were empowered to freely adjust recommended inventory orders. Granting this autonomy resulted in an average inventory increase of 9.6% over the centralized replenishment system.

On the contrary, many prominent retail chains in the United States, such as Wal-Mart, Target and Whole Foods, have implemented centrally controlled inventory systems. The rationale behind this practice is grounded in the theory of efficiency. To prepare the goods at a distribution center according to store layouts and deliver them directly to the store shelves "just in time" based on the orders of the central system may achieve significant cost efficiency, while reducing the likelihood of local excess inventory. However, without effectively leveraging the superior information of the retail outlets, the centrally controlled system may result in substantial stock shortages, as reported in the previous news articles.

Besides the private information held by the sales divisions and retail outlets, their efforts play a critical role in influencing and improving demand through activities such as client visits, enhanced customer services, and product consultations (Dudley 2014, Peterson 2018). However, neither their private information nor their efforts might be revealed to the firm, particularly when the actual demand information is potentially censored by the inventory level due to lost sales. This prompts the question of how firms can leverage sales division insights, incentivize efforts, and manage inventory effectively. To address this gap, we develop an agency model encompassing hidden action (sales division effort) and hidden information (local market conditions). By placing minimal constraints on the demand and cost functions, our model subsumes a broad spectrum of models found in the existing literature, and our results remain robust across exceptionally diverse scenarios. We incorporate the supply and demand mismatch, the demand information censoring, and the inventory ordering role into the model, which leads to four setups: demand-based contracting with controlled inventory, sales-based contracting with controlled inventory, demand-based contracting with delegated inventory, and sales-based contracting with delegated inventory.

Our analysis not only highlights the subtle differences in these setups, which can be ascribed to operational factors, but also provides actionable mechanisms to address the central question mentioned above. Specifically, we demonstrate that linear commission contracts in the form of menus (MLC) are highly effective in achieving the second-best outcome under similarly mild conditions in all of these scenarios. This finding holds true for the assumed general demand and cost functions, ensuring the practical implementability of these mechanisms in real-world environments. Therefore, there is no need to undertake organizational restructuring solely for the purpose of aligning incentives, for instance, for the aforementioned firms that employ different strategies to manage their operations.

These commission contracts, however, differ for different setups. The optimal contracts in the benchmark setting would not properly incentivize the agent in any of those scenarios involving operational factors. In the setting closest to the benchmark, where the demand is contractible and the inventory decision is centrally controlled, the commission rates need to be adjusted lower compared to the benchmark case, taking into account the inventory costs, which in turn motivates a lower effort level. Conversely, the transfer payments need to be adjusted higher depending on the desired inventory level. In the setting where the contracts must be written based on sales due to censored demand information, the commission rates and the transfer payments need be adjusted, according not only to the inventory costs but also to the expected sales function. As a result of the loss of demand information, the commission rates in the menu may need to vary more in order to incentivize the agent to truthfully report her private information and exert the second-best effort.

When the inventory decision is delegated, it becomes crucial to incorporate incentives into the contracts to motivate the agent to make the second-best inventory decision. In the scenario where the demand is contractible, the commission rates at their optimum remain the same as in the case where the inventory decision is centralized. However, besides the standard terms, the transfer payments are structured with terms that depend on the chosen inventory level, mimicking a centralized newsvendor payoff function. Notably, the motivation for effort and information revelation is decoupled from the motivation for the inventory decision in this scenario. In contrast, in the more intricate case where the contracts must be based on sales, the commissions are contingent upon the inventory level selected by the agent. As a result, the terms in the transfer payments used to incentivize the inventory decision must be structured taking into account the commission rates.

1.1. Related Literature

Our work is related to the marketing and economics literature on the optimal agency compensation scheme. Basu et al. (1985) characterized the firm's optimal contract under a setting with pure hidden action (i.e., demand-enhancement efforts). Their findings emphasized the necessity of a convex increasing scheme, tailored to compensate agents with an escalating marginal cost of effort. Consequently, the firm's optimal contract exhibits a convex increase in sales. Laffont and Tirole (1986) and Rao (1990) explored scenarios featuring both hidden action and hidden information and demonstrated the optimality of linear contracts within the general contract space. Each linear contract comprises a fixed salary and a commission rate contingent on the agent's truthful report of the local market condition. This optimality result not only holds theoretical elegance but also sheds light on the widespread adoption of linear compensation schemes, which stand as one of the two most prevalent incentive structures in practice (see Li et al. 2020). However, the marketing and economic literature largely assumes unlimited supplies, implying that demand, irrespective of its magnitude, can always be met without incurring extra costs. This perspective overlooks a central operational issue, namely, the potential supply and demand mismatch. In contrast, our work complements this oversight by integrating supply and demand mismatch as well as delegation of inventory decisions. We show that the optimal scheme can still retain a linear structure but with a new compensation component contingent on operational metrics. Moreover, we find that neither demand information censoring nor delegation of inventory decisions results in efficiency loss.

Our work falls into the growing body of studies on salesforce compensation with considerations of operational factors. Chen (2005) compared a forecast-based compensation scheme with a menu of linear contracts in a model where the agent possesses private information of the market condition, which aids the firm's production and inventory planning. Sohoni et al. (2011) compared piece-wise convex compensation schemes and quota-based compensation schemes, taking the sales variance into account. Khanjari et al. (2014) examined the performance of a menu of linear contracts for a retailer- or manufacturer-employed sales agent. Unlike these papers that focus on specific contract forms, we consider the most general contract space within which we characterize an optimal compensation scheme.

Chu and Lai (2013) and Dai and Jerath (2013) studied the optimal contract problem under full information with censored demand and hidden action and showed the optimality of the simple and commonly-used sales-quota-based bonus scheme when the agent is protected by limited liability—a specific type of risk aversion. Our model differs from theirs in two aspects. First, we allow the agent to have private information of her local market condition, which is common in practice. Notably, the agent is risk-neutral in our study because the agency problem with hidden information and hidden action under risk aversion (or limited liability) is intractable in most scenarios, let alone the addition of inventory delegation. Second, inventory delegation was not considered in their models, where the firm decides the inventory after offering the contract to the agent. Dai et al. (2021) incorporated both demand censoring and agent multitasking into their agency problem, where the agent is responsible for enhancing demand through marketing effort and available inventory through operational effort. They unveiled the general optimality of contracts with a bang-bang structure under mild assumptions. Our work differs in two ways. First, Dai et al. (2021) did not consider the presence of private information, a pivotal component central to our model. Second, they treated the available inventory as a stochastic outcome resulting from the agent's operational effort, rather than a direct decision variable, as is the case in our study. Furthermore, our findings diverge from those presented in the three aforementioned papers. We demonstrate that neither demand censorship nor inventory delegation compromises the system efficiency. In particular, when the inventory decision is made by the sales division, the classical menu of linear contracts falls short in motivating the right order quantity. However, convenient adjustments to the linear contracts can swiftly realign it with the attainment of the second-best outcome. These findings hold practical relevance, particularly in scenarios where sales divisions are typically large and risk-neutral, and inventory decisions are entrusted to them.

Chen et al. (2016) considered an agent not only exerting sales effort but also acquiring information and found that the forecast-based contract can outperform the menu of linear contracts. Xiao and Xiao (2020) delineated the optimal contract with the consideration of supply and demand mismatch. They showed that the S-shaped scheme is optimal under deterministic demand and that a tailor-made concave (quadratic) scheme is optimal under stochastic demand. However, these studies either assumed that the firm decides the inventory or assumed that the inventory is exogenously provided.

The remainder of the paper is organized as follows. Section 2 describes the model. In Sections 3 and 4, we analyze the optimal compensation under inventory control and delegation, respectively. We discuss the insights in Section 5 and conclude in Section 6.

2. The Model

A risk-netural firm needs to design a contract for its risk-netural sales division (or retail store) in a local market (herein referred to as the "agent"), who is responsible for selling a product at a unit price p. Due to her closer proximity to local customers compared to the firm, the agent possesses superior information on market conditions, denoted by Θ. To be precise, we assume that the agent directly observes the actual condition $\Theta = \theta$, while the firm is only privy to the distribution of Θ, denoted by $F(\cdot)$ (with density $f(\cdot)$ and support

$[\underline{\theta}, \bar{\theta}]$). We refer to the agent that observes condition θ as the θ-type agent. The agent can invest effort, denoted by e, into promoting demand, subject to a convex and increasing cost function $C(e)$. The θ-type agent exerting e effort results in a deterministic output $R(e,\theta)$, which reflects the predictable portion of the actual sales volume. To account for market uncertainty, we represent the demand function as $d = d(R(e,\theta),\epsilon)$, where ϵ stands for a stochastic shock following a distribution G (with density $g(\cdot)$).

To ease our analysis, for any given pair of output level r and type θ, we conversely define an effort function $E(r,\theta)$ such that $R(E(r,\theta),\theta) = r$. Consequently, instead of the preceding model where the agent selects an effort level e entailing a cost of $C(e)$, we can amalgamate θ and e into r and contemplate an equivalent model, where the θ-type agent chooses an output level r incurring a cost of $K(r,\theta) \equiv C(E(r,\theta))$. This reformulated model will be our focus throughout the paper. As an example, if we consider the demand and cost functions of Chen (2005), $d = \theta + e + \epsilon$ and $C(e) = e^2/2$, the corresponding demand and cost functions in our equivalent model will be $d = r + \epsilon$ and $K(r,\theta) = (r-\theta)^2/2$.

We adhere to the standard assumption of increasing hazard rate, i.e., $H(\theta) \equiv (1 - F(\theta))/f(\theta)$ decreases in θ. In addition, we introduce the following regularity assumptions regarding the demand, effort, and cost functions (throughout the paper, the numbered subscripts applied to multivariate functions denote partial derivatives relative to the respective variables).

Assumption 1. *All functions are three times continuously differentiable. Moreover,*

(C) $C_1 > 0,\ C_{11} > 0,\ C_{111} \geq 0;$

(D) $d_1 > 0,\ d_2 > 0,\ d_{11} \leq 0;$

(E) $E_1 > 0,\ E_2 < 0,\ E_{11} \geq 0,\ E_{12} \leq 0,\ E_{22} \geq 0,\ E_{112} \leq 0,\ E_{122} \geq 0.$

These regularity assumptions are commonly employed in agency theory, and their interpretations follow conventional norms. For instance, the cost for effort is convex and accelerating. The demand increases in the type, effort and shock, but may exhibit a diminishing pattern. For the implicit effort function, $E_1 > 0$ indicates that a higher output requires a higher effort, while $E_2 < 0$ implies that a higher-type agent needs less effort to reach a certain output level; $E_{12} \leq 0$ suggests that the additional effort required for a unit increase in output is less for a higher-type agent (i.e., E_1 decreases in θ); $E_{122} \leq 0$ indicates that this counterbalancing interplay between the output level and the agent's type on effort is less pronounced for a higher-type agent (i.e., E_{12} decreases in θ), and so forth. Our comprehensive model subsumes a multitude of classical models found in the agency literature (see, e.g., Laffont and Tirole 1986, Rao 1990, Chen 2005, Chen et al. 2016), such as $d=\theta+e+\epsilon$, $d=\theta e\epsilon$ and $d=\theta e+\epsilon$, and also encompasses more intricate demand forms, such as $d=(\theta+e)\epsilon+\ln(1+\theta+e+\epsilon)$ and $d=\theta e\epsilon+\sqrt{1+\theta e+\epsilon}$.

The determination of the order quantity (or equivalently, the inventory level), denoted by Q, can either be made by the firm or the agent. We refer to the former scenario as the "control" setting and the latter as the "delegation" setting. We introduce an auxiliary quantity function $l(r,Q)$, established by the equation $d(r,l(r,Q))=Q$. For instance, if $d=r+\epsilon$, then $l(r,Q)=Q-r$. As per Assumption (D), it is evident that $l_1<0$ and $l_2>0$. In line with the conventional inventory literature, we assume that the firm's production cost is c per unit and the agent's hassle cost is h per unit. The results derived in this paper can be extended to any scenario with convexly increasing production and hassle costs. The realized sales, denoted by s, signifies the demand censored by the inventory level, expressed as $s(r,Q,\epsilon)=\min\{d(r,\epsilon),Q\}$. For ease of notation, we denote $D(r)\triangleq\mathbb{E}_\epsilon[d(r,\epsilon)]$ and $S(r,Q)\triangleq\mathbb{E}_\epsilon[s(r,Q,\epsilon)]$ as the expected demand and realized sales, respectively.

Hinging on the contracting scenarios, the firm may receive accurate demand information or only observe realized sales, thereby having the option of contracting either on demand (i.e., uncensored) or on sales (i.e., censored). We look at all four settings: demand-based control environment (CD), sales-based control environment (CS), demand-based delegation environment (DD), and sales-based delegation environment (DS). Our focus in the main paper lies on menus of linear commission contracts (MLC) for the purpose of implementability.

3. The Control Setting

This section focuses on the control setting where the firm decides the inventory level based on the market condition θ reported by the agent. By the revelation principle, we can, without loss of optimality, restrict our attention to the class of direct mechanisms that comprise a payment and a quantity, i.e., $\{T(d,\theta),Q(\theta)\}$ or $\{T(s,\theta),Q(\theta)\}$, depending on the availability of demand or sales information. The firm incurs the production cost $cQ(\theta)$ and the agent incurs the hassle cost $hQ(\theta)$. When the selling season commences, the θ-type agent selects an output level r, incurring the cost, $K(r,\theta)$, and then the demand $d(r,\epsilon)$ is materialized, resulting in sales $s = \min\{d, Q(\theta)\}$. Any excess demand is lost, or the leftover inventory is salvaged at zero value. Finally, the firm collects the sales revenue ps, and the agent receives the compensation $T(d,\theta)$ or $T(s,\theta)$.

In the following, we present the formulation for the demand-based contracting setting. To ensure the agent's truthtelling, the firm needs to impose the incentive-compatibility constraint (IC). Specifically, when the θ-type agent chooses the $\hat{\theta}$-type contract, she earns:

$$\pi(\theta,\hat{\theta}) = \max_{r\geq 0} \mathbb{E}_\epsilon[T(d(r,\epsilon),\hat{\theta})] - hQ(\hat{\theta}) - K(r,\theta).$$

Let $r(\theta,\hat{\theta})$ be the maximizer (if it does not exist then $\pi(\theta,\hat{\theta}) = \infty$, which is not what the firm wants) and denote $r(\theta) = r(\theta,\theta)$ and $\pi(\theta) = \pi(\theta,\theta)$. The IC constraint follows:

$$\pi(\theta) \geq \pi(\theta,\hat{\theta}), \quad \forall \theta,\hat{\theta} \in [\underline{\theta},\bar{\theta}]. \tag{IC}$$

The individual-rationality constraint (IR) to ensure the agent's participation follows:

$$\pi(\theta) \geq 0, \quad \forall \theta \in [\underline{\theta},\bar{\theta}]. \tag{IR}$$

Under these two conditions, the firm's profit follows:

$$\begin{aligned}\Pi(T,Q) &= \mathbb{E}_{\Theta,\epsilon}[p\min\{d(r(\Theta),\epsilon),Q(\Theta)\} - cQ(\Theta) - T(d(r(\Theta),\epsilon),\Theta)] \\ &= \mathbb{E}_{\Theta,\epsilon}[pS(r(\Theta),Q(\Theta)) - cQ(\Theta) - T(d(r(\Theta),\epsilon),\Theta)].\end{aligned}$$

Therefore, the firm's mechanism design problem is:

$$\begin{aligned}\max_{T,Q} \quad & \Pi(T,Q) \\ \text{s.t.} \quad & \text{(IC) and (IR)}.\end{aligned} \tag{$\mathrm{P_{CD}}$}$$

In the CS setting, the firm's mechanism design problem, denoted as ($\mathrm{P_{CS}}$), can be formulated by replacing demand with sales. To solve these problems in the contract space, our approach first derives an upper bound and then establishes sufficient conditions under which MLC can achieve the upper bound, resulting in the second-best outcome.

Proposition 1. *(P_{CD}) and (P_{CS}) have a common upper bound, where the outcome $(r^*(\theta), Q^*(\theta))$ solves the following system of equations:*

$$\begin{cases} pS_1(r,Q) - K_1(r,\theta) + K_{12}(r,\theta)H(\theta) = 0, \\ pS_2(r,Q) - c - h = 0. \end{cases} \tag{1}$$

The first branch of Eq. (1) is the balance equation from the sales aspect, where $pS_1(r,Q)$ is the marginal return of the output, while $K_1(r,\theta)-K_{12}(r,\theta)H(\theta)$ is the overall marginal cost, comprising the marginal cost associated with the output and the one attributed to the information rent. The second branch is the balance equation from the operational aspect, where $pS_2(r,Q)$ is the marginal return from increasing inventory, while $c+h$ is the overall marginal cost, comprising the production cost and the hassle cost. An ideal outcome must have marginal returns equal to marginal costs on both fronts. We can verify that $S_1(r,Q)=\int_{-\infty}^{l(r,Q)} d_1(r,\epsilon)\mathrm{d}G(\epsilon)$ and $S_2(r,Q)=1-G(l(r,Q))$. Moreover, by definition, $d(r,l(r,Q))=Q$, i.e., $l(r,Q)$ represents the quantity to match the output level with the inventory level. Hence, we can rewrite Eq. (1) as follows for clearer insight:

$$\begin{cases} p\int_{-\infty}^{G^{-1}(\frac{p-c-h}{p})} d_1(r,\epsilon)\mathrm{d}G(\epsilon)-K_1(r,\theta)+K_{12}(r,\theta)H(\theta)=0, \\ l(r,Q)=G^{-1}(\frac{p-c-h}{p}). \end{cases} \tag{2}$$

$G^{-1}(\frac{p-c-h}{p})$ is the safety stock level that optimally balances the mismatch cost arising from the random noise ϵ. The ratio $\frac{p-c-h}{p}$ corresponds to the firm's targeted fill rate and aligns with the critical fractile in the classical newsvendor literature.

PROPOSITION 2. *The second-best outcome $(r^*(\theta),Q^*(\theta))$ can be implemented by MLC if*

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))}\geq 0, \qquad \forall\theta\in[\underline{\theta},\bar{\theta}] \tag{D}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{K_1(r^*(\theta),\theta)}{S_1(r^*(\theta),Q^*(\theta))}\geq 0, \qquad \forall\theta\in[\underline{\theta},\bar{\theta}], \tag{S}$$

depending on the availability of demand or sales information, respectively.

The above proposition establishes sufficient conditions under which MLC achieves the upper bound and hence results in the second-best outcome for both settings. Commission contracts provide the agent with payments linearly increasing in demand, based on commission rates. To induce the second-best effort, commission rates must cover the agent's marginal cost for each unit increase in contractible output due to her effort. Moreover, to induce truthful reporting, an agent who reports a higher market condition should be offered a commission contract with a higher marginal return on the realized contractible output, ensuring that agents observing inferior market conditions have no incentive to mimic. As a result, the commission rates must increase with the agent's type, which is captured by Conditions (D) and (S). When these conditions are met, we can characterize the optimal MLCs to achieve the second-best outcome, as shown below.

PROPOSITION 3. *The optimal MLCs, $T_{CD}(d,\theta) = \alpha_{CD}(\theta)d + \beta_{CD}(\theta)$ and $T_{CS}(s,\theta) = \alpha_{CS}(\theta)s + \beta_{CS}(\theta)$, for ($P_{CD}$) and ($P_{CS}$) can be characterized respectively by:*

$$\begin{cases} \alpha_{CD}(\theta) & = \frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))}, \\ \beta_{CD}(\theta) & = -\alpha_{CD}(\theta)D(r^*(\theta)) + K(r^*(\theta),\theta) + hQ^*(\theta) - \int_{\underline{\theta}}^{\theta} K_2(r^*(z),z)\,dz; \end{cases}$$

and

$$\begin{cases} \alpha_{CS}(\theta) & = \frac{K_1(r^*(\theta),\theta)}{S_1(r^*(\theta),Q^*(\theta))}, \\ \beta_{CS}(\theta) & = -\alpha_{CS}(\theta)S(r^*(\theta),Q^*(\theta)) + K(r^*(\theta),\theta) + hQ^*(\theta) - \int_{\underline{\theta}}^{\theta} K_2(r^*(z),z)\,dz. \end{cases}$$

The difference between Conditions (D) and (S), as well as the optimal contracts, arises from the fact that in the CD setting, the commission is paid based on realized demand, whereas in the CS setting, it is paid based on realized sales. In the former (latter), the commission rates must cover the agent's marginal cost for each unit increase in demand (sales) due to her effort. Conditions (D) and (S) ensure the monotonicity of the commission rates with respect to the agent's type and enable the proposed MLCs to achieve effective screening. It is evident that $S_1(r^*(\theta),Q^*(\theta)) < D_1(r^*(\theta))$, which implies that $\alpha_{CS}(\theta) > \alpha_{CD}(\theta)$. When contracting based on sales, the presence of demand information censoring naturally dissuades the agent from generating output (compared to contracting based on demand), as any realized demand beyond the inventory level does not contribute to her compensation. Consequently, to induce the same ideal output level, the firm must offer a higher commission rate. On the other hand, the base payments of the optimal MLCs guarantee that the agent's payoff precisely matches her information rent in both settings. The establishment of these conditions and MLCs poses significant technical challenges. This is the first characterization of such results in functional forms in the existing literature.

4. The Delegation Setting

In this section, we study the scenarios where the inventory ordering decision is delegated to the agent. The sequence of events is modified as follows. First, before the selling season, the firm offers the agent a direct mechanism $T(d,Q,\theta)$ or $T(s,Q,\theta)$. Second, the agent observes the realized value of market condition $\Theta = \theta$ and truthfully reports θ given that it is in her best interest to do so. The agent requests inventory Q, fulfilled by the firm, for which the firm and the agent incur the inventory cost cQ and hQ, respectively. Third, the selling season starts, the agent selects an output level r, incurring the cost, $K(r,\theta)$, and then the demand $d(r,\epsilon)$ is materialized, resulting in sales $s = \min\{d, Q\}$. Any excess demand is lost, or the leftover inventory is salvaged at zero value. Finally, the firm collects the sales revenue ps, and the agent receives the compensation $T(d,Q,\theta)$ or $T(s,Q,\theta)$.

In the demand-based contracting setting, the agent's profit can be written as:

$$\pi(\theta,\hat{\theta}) = \max_{(r,Q)\geq 0} \mathbb{E}_\epsilon[T(d(r,\epsilon),Q,\hat{\theta})] - hQ - K(r,\theta).$$

Denote the maximizers of $\pi(\theta,\hat{\theta})$ by $r(\theta,\hat{\theta})$ and $Q(\theta,\hat{\theta})$, and let $r(\theta) = r(\theta,\theta)$, $Q(\theta) = Q(\theta,\theta)$, and $\pi(\theta) = \pi(\theta,\theta)$. When ensuring truthtelling, the firm earns:

$$\Pi(T) = \mathbb{E}_{\Theta,\epsilon}[pS(r(\Theta),Q(\Theta)) - cQ(\Theta) - T(d(r(\Theta),\epsilon),\Theta)].$$

The firm's problem, $(\mathrm{P}_{\mathrm{DD}})$, is to maximize $\Pi(T)$ subject to the agent's (IC) and (IR) constraints. We can similarly formulate the firm's problem in the sales-based contracting setting, $(\mathrm{P}_{\mathrm{DS}})$, by replacing the demand function with sales.

Despite the inventory decision being delegated to the agent, Proposition 4 demonstrates that the second-best outcome and the sufficient conditions for MLCs to achieve the second best, as characterized for the inventory-controlled settings, continue to apply.

Proposition 4. *The second-best outcome characterized in Proposition 1 and the sufficient Conditions (D) and (S) characterized in Proposition 2 for MLCs to achieve the second-best outcome remain the same for (P_{DD}) and (P_{DS}), respectively.*

When the inventory decision is delegated to the agent, the firm needs to incentivize the agent not only to truthfully report the market condition and exert sales effort but also to choose an optimal inventory level. As such, the firm faces more restrictions than in inventory-controlled settings, and thus the upper bound characterized there remains to upper bound the firm's problems in inventory-delegated settings. More intriguingly, Proposition 4 asserts that these two incentivization tasks can be coordinated, and the upper bound, and therefore the second-best outcome, is achievable with MLCs under the same sufficient conditions. Because the sufficient conditions only apply to the commission rates, this result indicates that the incentivization of the inventory decision does not affect the incentivization of information revelation and sales effort. Apparently, if the firm were to simply offer the MLCs characterized in the previous section to the agent, the incentives would be misaligned. The inventory level would only appear in the hassle cost term, i.e., hQ, in the agent's profit function, incentivizing her to opt for no inventory. Conversely, if the agent were exempted from inventory costs, she would inflate the inventory request, as discussed in the news articles mentioned in the introduction. It is worth noting that sales quota-based contracts could also lead to similar outcomes, as the marginal compensation becomes infinitely large when sales approach the quota. These deficiencies arise due to their inadequate consideration of the cost and performance implications of inventory.

In the following, we present the optimal MLCs that can motivate the agent to request the right inventory level, induce truthful reporting, and achieve the most desired output.

PROPOSITION 5. *The optimal MLCs, $T_{DD}(d,Q,\theta)=\alpha_{DD}(\theta)d+\beta_{DD}(\theta)+\gamma_{DD}(\theta)Q-Q^2$ and $T_{DS}(s,Q,\theta)=\alpha_{DS}(\theta)s+\beta_{DS}(\theta)-\gamma_{DS}(\theta)Q$, for ($P_{DD}$) and ($P_{DS}$) can be characterized respectively by:*

$$\begin{cases} \alpha_{DD}(\theta) & = \frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))}, \\ \beta_{DD}(\theta) & = -\alpha_{DD}(\theta)D(r^*(\theta)) + K(r^*(\theta),\theta) + hQ^*(\theta) - \int_{\underline{\theta}}^{\theta} K_2(r^*(z),z)\,dz - \gamma_{DD}(\theta)Q^*(\theta) + Q^*(\theta)^2, \\ \gamma_{DD}(\theta) & = 2Q^*(\theta) + h; \end{cases}$$

and

$$\begin{cases} \alpha_{DS}(\theta) & = \frac{K_1(r^*(\theta),\theta)}{S_1(r^*(\theta),Q^*(\theta))}, \\ \beta_{DS}(\theta) & = -\alpha_{DS}(\theta)S(r^*(\theta),Q^*(\theta)) + K(r^*(\theta),\theta) + hQ^*(\theta) - \int_{\underline{\theta}}^{\theta} K_2(r^*(z),z)\,dz + \gamma_{DS}(\theta)Q^*(\theta), \\ \gamma_{DS}(\theta) & = \alpha_{DS}(\theta)\frac{c+h}{p} - h. \end{cases}$$

From Proposition 5, we observe that the commission rates of the optimal MLCs remain the same as those in the inventory-controlled settings. However, the base payments now depend on the requested inventory level. The two cases regarding whether demand or sales are contractible represent subtle differences.

In the DD setting where contracts are contingent on the realized demand, this metric is unaffected by the inventory level, and vice versa. Consequently, the firm's goal of stimulating the agent to generate the most desired output is not entangled with its goal of motivating the agent to order the right quantity. The firm can employ the same commission rate used in $T_{CD}(d,\theta)$ to effectively induce $r^*(\theta)$, and a separate term included in the base payment, $\gamma_{DD}(\theta)Q - Q^2$, to induce the right inventory request. Note that this inventory term satisfies the first-order condition at the optimal inventory level $Q^*(\theta)$, and it is unrelated to the commission rate. The agent would be penalized for any deviation from requesting $Q^*(\theta)$. In contrast, in the DS setting where contracts are contingent on the realized sales, this metric is influenced not only by the demand but also by the inventory level. As such, despite similar problem formulations, the inducements of the output and the inventory decision become intertwined. Although the commission rate, set to elicit the ideal output level $r^*(\theta)$, is unaffected by the inventory level, the term to induce the optimal inventory request now depends on the commission rate. Specifically, the instrument, $\gamma_{DS}(\theta)$, consists of two parts. The constant term, $-h$, is introduced to compensate the agent for her hassle cost. The type-dependent term, $\alpha_{DS}(\theta)\frac{c+h}{p}$, where the ratio $\frac{c+h}{p}$ is derived from $1 - G(l(r^*(\theta), Q^*(\theta)))$, is designed to neutralize the spillover effect from the sales commission to the inventory decision. The overall base payments align the agent's payoff with her information rent in both cases.

5. Managerial Discussion

Throughout Sections 3 and 4, our analysis has demonstrated that the same pair of ideal output and inventory levels can be implemented by MLCs, as long as two sufficient conditions are met. Hence, we arrive at one of the key revelations from this paper: under these conditions, the fundamental trade-offs regarding the benefits and costs of demand enhancement and demand fulfillment remain intact irrespective of demand information censoring or inventory delegation. That is, neither demand information censoring nor the delegation of inventory decisions impedes the incentivization of truth-telling or compromises the overall efficiency of the system.

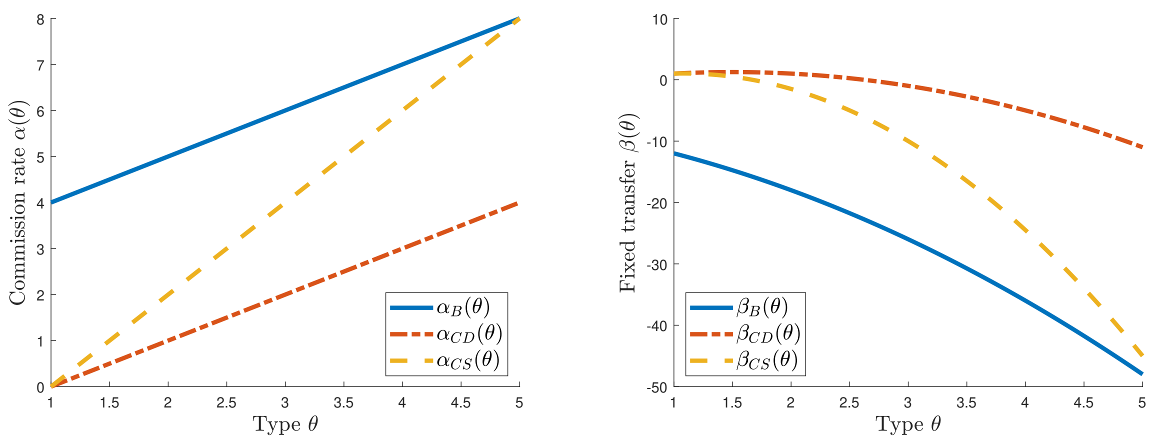
Proposition 6. *When Conditions* (D) *and* (S) *are satisfied,* $\Pi^{CD} = \Pi^{CS} = \Pi^{DD} = \Pi^{DS}$.

This theoretical finding is not only intriguing but also offers valuable insights for managerial decision-making. It asserts the possibility of maintaining the second-best outcome regardless of whether the inventory decision is centralized or delegated, and whether the demand information is censored or uncensored. As discussed in the introduction, certain firms (such as Wal-Mart and Whole Foods) favor centralized inventory decisions, while others (such as Zara, a global pharmaceutical company, and a European supermarket chain mentioned in news articles) typically delegate these decisions to their salesforce or local stores. Despite various factors influencing these preferences, our study demonstrates that, under reasonably mild conditions, appropriately designed simple commission contracts can align incentives, gather information, and effectively motivate effort and inventory decisions in each case, eliminating the need for organizational restructuring solely to align incentives.

It is important to note that the contracts employed in these scenarios differ in their characteristics and value. First, different from the conventional results in prior literature (e.g. Laffont and Tirole 1986, Rao 1990), the ideal effort level to induce now must take into account the inventory costs, while the commission rates need to incorporate the inventory costs that the agent incurs to supply effort. Second, if the demand information is censored, the commission rates need to be structured based on the expected sales function that involves the inventory level. Third, if the inventory decision is delegated, the commission contracts need to include an incentive term contingent on the requested inventory that penalizes any deviation from the optimal inventory level. To illustrate these distinctions, we present a specific example in the following.

Consider an additive demand function, given by $d = \theta + e + \epsilon$, and a quadratic effort cost function, represented by $C(e) = e^2/2$. It can be verified that both Conditions (D) and (S) hold. The second-best effort level follows: $e^*(\theta) = p - c - h - H(\theta)$, while the second-best inventory decision satisfies: $Q^*(\theta) = p - c - h + \theta - H(\theta) + G^{-1}(\frac{p-c-h}{p})$. These optimal decisions can be effectively executed through contracts in both controlled and delegated inventory scenarios, regardless of whether or not the demand information is censored.

Figure 1 illustrates the comparisons between the contracts used in controlled inventory scenarios and the benchmark, which represents a situation with abundant free inventory. In the case where the demand information is not censored, the commission rates, depicted by the dash-dotted curve in the left subplot, shift downward in parallel with respect to the benchmark rates represented by the solid curve. This shift reflects the reduction in effort after considering the unit inventory cost $c + h$. However, when the demand information is censored by the inventory level, the commission rates depicted by the dashed curve, exhibit a steeper slope. This steeper slope is necessary to motivate the agent to reveal information and exert the second-best effort. The reason for this is that a higher θ implies a greater likelihood of the actual demand information being censored, leading to a “reduced” compensation. Therefore, the commission rates in the menu need to vary more when sales are used for contracting, compared to the case where the demand is contractible. Regarding the transfer payment, the comparisons are reversed (note that the agent's reservation profit is zero). Firstly, after accounting for the inventory cost, a higher transfer payment is required for the agent, represented by the dash-dotted curve in the right subplot, compared to the benchmark case represented by the solid curve. Secondly, the transfer payment under censored demand information, depicted by the dashed curve, decreases at a faster rate to negative values than that without demand information censoring. This is because the agent receives a higher commission under information censoring as θ increases, resulting in a larger decrease in the transfer payment to extract rent.



(a) Commission rates as functions of θ. (b) Fixed transfers as functions of θ.

Figure 1 **Parameters:** $p = 8$, $c = 3$, $h = 1$, $\Theta \sim U[1,5]$ **and** $\epsilon \sim U[-1,1]$.

When the inventory decision is delegated to the agent and the second-best inventory level is implemented, both the commission rate and the transfer payment align with their counterparts under controlled inventory decisions. Therefore, we do not provide a separate depiction. Instead, Figure 2 showcases the agent's profit functions in relation to Q under the contracts outlined in the propositions. Recall that the inventory level affects the commission when the information is censored, whereas it does not in the other case. In this figure, the agent's profit function with censored demand information, depicted by the dashed curve, exhibits a smaller curvature compared to the counterpart with complete demand information represented by the solid curve. Nevertheless, the agent makes the same best profit when she selects the second-best inventory level.

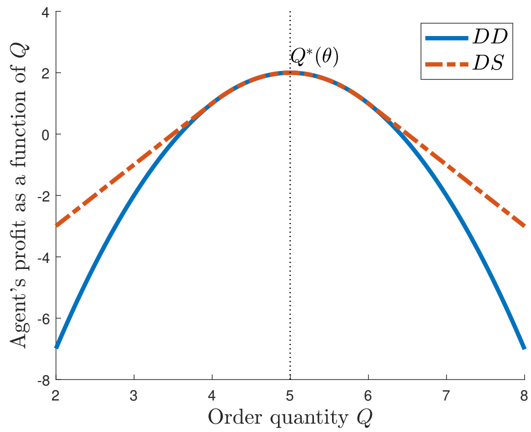


Figure 2 **Agent's profits as functions of** Q **in the delegation setting. Parameters:** $p=8$**,** $c=3$**,** $h=1$**,** $\Theta \sim U[1,5]$**,** $\epsilon \sim U[-1,1]$**, and** $\theta=3$**.**

The above findings offer several important managerial takeaways. First, simplicity reigns supreme: when the requisite condition is met, the optimal outcome can be attained by convenient compensation schemes, even when confronted with intricate demand and cost functions. Second, adaptability is key: compensation schemes need be tailored to accommodate the nuances of different operational aspects, ranging from the presence or absence of demand information to the delegation of inventory decision to the sales division. Finally, a rather unexpected revelation emerges from our research: under these optimality conditions, neither demand information censoring nor inventory delegation poses a threat to system efficiency. This underscores the resilience of well-designed compensation schemes in achieving desired outcomes.

6. Conclusion

In this study, we delved into the intricate realm of designing compensation schemes for a firm's sales division, where the challenges of hidden information and hidden actions abound. Our analytical approach was firmly grounded in the classical agency model, allowing us to explore scenarios where both sales and operational factors come into play, including supply and demand mismatch, demand information censoring, and inventory delegation. Notably, our analysis considered highly generalized demand and cost functions. As such, our model subsumes a broad spectrum of models found in the existing literature, and our results remain robust across exceptionally diverse scenarios.

Our investigation commenced with a benchmark model stripped of operational features, focusing on the fundamental question of how to achieve the second-best outcome.

Intriguingly, we established that the implementation of this optimal outcome hinges on the fulfillment of a specific condition, rooted in the inherent characteristics of demand and cost functions. This insight offers a clear and elegant path for designing compensation schemes under diverse circumstances. Venturing further, we probed four distinct models, each representing a unique operational setup. Remarkably, we unearthed similar optimality conditions that pave the way for attaining the second-best outcome through the exclusive use of implement-friendly linear commission contracts or by complementing them with quantity-related components. Our study not only offers valuable insights into the intricate task of compensation scheme design but also highlights the robustness and adaptability of these schemes in addressing complex real-world challenges. It provides a foundation for firms to navigate the intricate interplay between hidden information, hidden actions, and operational realities in their quest for optimal performance and efficiency.

We conclude by discussing a couple of extensions to our study. Firstly, while we establish the conditions that ensure the optimality of MLC, it is worth exploring mechanisms when these conditions are not satisfied. In Appendix B, we investigate menus of quadratic commission contracts (MQC), of which MLC is a special case. Our investigation uncovers interesting technical findings. When these conditions are not met, MQC can still achieve the second-best outcome and be optimal if the demand is contractible and subject to mild restrictions on its derivatives. The analysis also reveals that within such environments, the conditions are both sufficient and necessary for the optimality of MLC. However, if the contracts must rely on sales, the validity of these results is compromised. That is, with censored demand information, the conditions are no longer necessary conditions for the optimality of MLC; moreover, if the conditions are not met, MQC may lose its optimality, and characterizing the optimal mechanism becomes challenging. Differently, inventory delegation does not affect these conditions. Hence, our investigation shows that in a more complex environment, different operational factors such as inventory delegation versus lost sales can have distinct effects on the parties' profits, in addition to influencing compensation design. Secondly, our study assumes that the agent is risk-neutral, which is reasonable for relatively large sales divisions. However, when the agent is risk-averse, including limited liability, the mechanism design problem involving hidden information, hidden action and operational factors becomes significantly more challenging. Addressing these aspects would require making valuable technical advancements in future research.

References

Bolandifar E, Feng T, Zhang F (2018) Simple Contracts to Assure Supply under Noncontractible Capacity and Asymmetric Cost Information. *Manufacturing & Service Operations Management* 20(2) 217-231.

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

Caro F, Gallien J (2010) Inventory management of a fast-fashion retail network. *Operations Research* 58(2):257-273.

Chen F (2005) Salesforce incentives, market information, and production/inventory planning. *Management Science* 51(1):60-75.

Chen F, Lai G, Xiao W (2016) Provision of incentives for information acquisition: Forecast-based contracts versus menus of linear contracts. *Management Science* 62(7):1899-1914.

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

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

Dai T, Ke R, Ryan CT (2021) Incentive Design for Operations-Marketing Multitasking. *Management Science* 67(4):2211-2230.

Dudley R (2014) Wal-Mart sees $3 billion opportunity filling empty shelves. *Bloomberg News.*

Ebel T, Larsen E, Shah K (2013) Strengthening health care's supply chain: A five-step plan. *McKinsey Quarterly* 1-6.

Gonik J (1978) Tie salesmen's bonuses to their forecasts. *Harvard Business Review* 56(3):116-123.

Grocery Manufacturers Association (2015) Solving the out-of-stock problem: A FMI/GMA trading partner alliance report.

Khanjari N, Iravani S, Shin H (2014) The impact of the manufacturer-hired sales agent on a supply chain with information asymmetry. *Manufacturing & Service Operations Management* 16(1):76–88.

Laffont JJ, Tirole J (1986) Using Cost observation to regulate firms. *Journal of Political Economy* 94(3):614-641.

Li S, Chen KY, Rong Y (2020) The behavioral promise and pitfalls in compensating store managers. *Management Science* 66(10):4899-4919.

Peterson H (2018) 'Seeing someone cry at work is becoming normal': Employees say Whole Foods is using 'scorecards' to punish them. *Business Insider.*

Picard P (1987) On the design of incentive schemes under moral hazard and adverse selection. *Journal of Public Economics.* 33(3):305-331.

Rao R (1990) Compensating heterogeneous salesforces: some explicit solutions. *Marketing Science* 9(4):319-341.

Scheele LM, Thonemann UW, Slikker M (2018) Designing incentive systems for truthful forecast information sharing within a firm. *Management Science* 64(8):3690–3713.

Sohoni MG, Chopra S, Mohan U, Sendil N (2011) Threshold incentives and sales variance. *Production and Operations Management* 20(4):571–586.

Turner R, Lasserre C, Beauchet P (2007) Innovation in field force bonuses: Enhancing motivation through a structured process-based approach. *Journal of Medical Marketing* 7(2):126-135.

van Donselaa KH, Gaur V, van Woensel T, Broekmeulen RA, Fransoo JC (2010) Ordering behavior in retail stores and implications for automated replenishment. *Management Science* 56(5):766–784.

Xiao B, Xiao W (2020) Technical note: Optimal salesforce compensation with supply/demand mismatch costs. *Production and Operations Management* 29(1):62-71.

Appendix A: Proofs

The following lemmas are useful. For notational convenience, let $\bar{F}=1-F$ and $\bar{G}=1-G$.

LEMMA 1. *The cost function $K(r,\theta)$ satisfies: $K_1>0$, $K_2<0$, $K_{11}>0$, $K_{12}<0$, $K_{112}\leq 0$, and $K_{122}\geq 0$.*

Proof of Lemma 1. The monotonic properties of $K(r,\theta)\equiv C(E(r,\theta))$ follows from:

$$
\begin{aligned}
K_1(r,\theta) &= \underbrace{C_1(E(r,\theta))}_{>0}\underbrace{E_1(r,\theta)}_{>0}>0\\
K_2(r,\theta) &= \underbrace{C_1(E(r,\theta))}_{>0}\underbrace{E_2(r,\theta)}_{<0}<0\\
K_{11}(r,\theta) &= \underbrace{C_{11}(E(r,\theta))}_{>0}\underbrace{E_1(r,\theta)^2}_{>0}+\underbrace{C_1(E(r,\theta))}_{>0}\underbrace{E_{11}(r,\theta)}_{\geq 0}>0\\
K_{12}(r,\theta) &= \underbrace{C_{11}(E(r,\theta))}_{>0}\underbrace{E_1(r,\theta)}_{>0}\underbrace{E_2(r,\theta)}_{<0}+\underbrace{C_1(E(r,\theta))}_{>0}\underbrace{E_{12}(r,\theta)}_{\leq 0}<0\\
K_{112}(r,\theta) &= \underbrace{C_{111}(E(r,\theta))}_{\geq 0}\underbrace{E_2(r,\theta)}_{<0}\underbrace{E_1(r,\theta)^2}_{\geq 0}+2\underbrace{C_{11}(E(r,\theta))}_{>0}\underbrace{E_1(r,\theta)}_{>0}\underbrace{E_{12}(r,\theta)}_{\leq 0}\\
&\quad+\underbrace{C_{11}(E(r,\theta))}_{>0}\underbrace{E_2(r,\theta)}_{<0}\underbrace{E_{11}(r,\theta)}_{\geq 0}+\underbrace{C_1(E(r,\theta))}_{>0}\underbrace{E_{112}(r,\theta)}_{\leq 0}\leq 0\\
K_{122}(r,\theta) &= \underbrace{C_{111}(E(r,\theta))}_{\geq 0}\underbrace{E_1(r,\theta)}_{>0}\underbrace{E_2(r,\theta)^2}_{>0}+\underbrace{C_{11}(E(r,\theta))}_{>0}(\underbrace{E_{12}(r,\theta)}_{\leq 0}\underbrace{E_2(r,\theta)}_{<0}+\underbrace{E_1(r,\theta)}_{>0}\underbrace{E_{22}(r,\theta)}_{\geq 0})\\
&\quad+\underbrace{C_{11}(E(r,\theta))}_{>0}\underbrace{E_2(r,\theta)}_{<0}\underbrace{E_{12}(r,\theta)}_{\leq 0}+\underbrace{C_1(E(r,\theta))}_{>0}\underbrace{E_{122}(r,\theta)}_{\geq 0}\geq 0. \quad \square
\end{aligned}
$$

LEMMA 2. *In* ($\mathrm{P_{CD}}$), ($\mathrm{P_{CS}}$), ($\mathrm{P_{DD}}$), *and* ($\mathrm{P_{DS}}$), *we all have $\pi'(\theta)=-K_2(r(\theta),\theta)$ and $\mathbb{E}_\Theta[\pi(\Theta)]=\pi(\underline{\theta})-\mathbb{E}_\Theta[K_2(r(\Theta),\Theta)H(\Theta)]$. Consequently, it is without loss of optimality for the firm to restrict to compensation schemes that induce $\pi(\underline{\theta})=0$.*

Proof. The proofs for all four problems are essentially identical. We take the one for ($\mathrm{P_{CS}}$) for illustration. We can rewrite $\pi(\theta)$ as

$$\pi(\theta)=\max_{\hat{\theta}\in[\underline{\theta},\bar{\theta}],r\geq 0}\mathbb{E}_\epsilon[T(\min\{d(r,\epsilon),Q(\hat{\theta})\},\hat{\theta})]-hQ(\hat{\theta})-K(r,\theta).$$

Note that θ only appears in $K(r,\theta)$. By the envelope theorem, we have

$$\pi'(\theta)\equiv\frac{\mathrm{d}\pi(\theta)}{\mathrm{d}\theta}=\frac{\partial}{\partial\theta}\left[\mathbb{E}_\epsilon[T(\min\{d(r,\epsilon),Q(\hat{\theta})\},\hat{\theta})]-hQ(\hat{\theta})-K(r,\theta)\right]\Bigg|_{\substack{\hat{\theta}=\theta\\ r=r(\theta)}}=-K_2(r(\theta),\theta)$$

and this yields

$$
\begin{aligned}
\mathbb{E}_\Theta[\pi(\Theta)] &= -\int_{\underline{\theta}}^{\bar{\theta}}\pi(\theta)\mathrm{d}\bar{F}(\theta)=\underbrace{-\pi(\theta)\bar{F}(\theta)|_{\underline{\theta}}^{\bar{\theta}}}_{=\pi(\underline{\theta})}+\underbrace{\int_{\underline{\theta}}^{\bar{\theta}}\bar{F}(\theta)\mathrm{d}\pi(\theta)}_{=\int_{\underline{\theta}}^{\bar{\theta}}\pi'(\theta)\frac{\bar{F}(\theta)}{f(\theta)}\mathrm{d}F(\theta)}\\
&= \pi(\underline{\theta})+\mathbb{E}_\Theta[\pi'(\Theta)H(\Theta)]=\pi(\underline{\theta})-\mathbb{E}_\Theta[K_2(r(\Theta),\Theta)H(\Theta)]
\end{aligned}
$$

Since, on the other hand,

$$\mathbb{E}_\Theta[\pi(\Theta)]=\mathbb{E}_{\Theta,\epsilon}[T(\min\{d(r(\Theta),\epsilon),Q(\Theta)\})-hQ(\Theta)-K(r(\Theta),\Theta)],$$

we have

$$\begin{aligned}\Pi(T) &= \mathbb{E}_\Theta[pS(r(\Theta),Q(\Theta)) - \pi(\Theta) - (c+h)Q(\Theta) - K(r(\Theta),\Theta)] \\ &= \mathbb{E}_\Theta[pS(r(\Theta),Q(\Theta)) + K_2(r(\Theta),\Theta)H(\Theta) - (c+h)Q(\Theta) - K(r(\Theta),\Theta)] - \pi(\underline{\theta}),\end{aligned}$$

and thus it is optimal for the firm to have $\pi(\underline{\theta}) = 0$. $\square$

LEMMA 3. *The expected realized sales S is concave in (r,Q).*

Proof. Recall that $Q = d(r, l(r,Q))$. We have

$$\begin{aligned}0 &= \frac{\partial Q}{\partial r} = \frac{\partial}{\partial r} d(r,l(r,Q)) = d_1(r,l(r,Q)) + d_2(r,l(r,Q))l_1(r,Q) \\ 1 &= \frac{\partial Q}{\partial Q} = \frac{\partial}{\partial Q} d(r,l(r,Q)) = d_2(r,l(r,Q))l_2(r,Q)\end{aligned}$$

which implies $d_1(r,l(r,Q))l_2(r,Q) = -l_1(r,Q)$. Then we have

$$\begin{aligned}S(r,Q) &= \int_{-\infty}^{l(r,Q)} d(r,\epsilon)g(\epsilon)\mathrm{d}\epsilon + Q\bar{G}(l(r,Q)) \\ S_1 &= \underbrace{d(r,l(r,Q))}_{=Q} g(l(r,Q))l_1(r,Q) + \int_{-\infty}^{l(r,Q)} d_1(r,\epsilon)g(\epsilon)\mathrm{d}\epsilon - Qg(l(r,Q))l_1(r,Q) = \int_{-\infty}^{l(r,Q)} d_1(r,\epsilon)g(\epsilon)\mathrm{d}\epsilon \\ S_2 &= \underbrace{d(r,l(r,Q))}_{=Q} g(l(r,Q))l_2(r,Q) + \bar{G}(l(r,Q)) - Qg(l(r,Q))l_2(r,Q) = \bar{G}(l(r,Q)) \\ S_{11} &= \int_{-\infty}^{l(r,Q)} \underbrace{d_{11}(r,\epsilon)}_{\le 0} g(\epsilon)\mathrm{d}\epsilon + \underbrace{d_1(r,l(r,Q))}_{>0} g(l(r,Q))l_1(r,Q) < 0 \\ S_{12} &= -g(l(r,Q))l_1(r,Q) > 0 \\ S_{22} &= -g(l(r,Q))l_2(r,Q) < 0 \\ \det(S) &= S_{11}S_{22} - S_{12}^2 \ge -d_1(r,l(r,Q))g(l(r,Q))l_1(r,Q)g(l(r,Q))l_2(r,Q) - [g(l(r,Q))l_1(r,Q)]^2 = 0\end{aligned}$$

which implies the Hessian of S is negative semidefinite and thus S is concave in (r,Q). $\square$

Proof of Proposition 1. Per Lemma 2, in either ($\mathrm{P_{CD}}$) or ($\mathrm{P_{CS}}$), the firm's objective is to maximize $\mathbb{E}_\Theta[M(r(\Theta),Q(\Theta),\Theta)]$, where

$$M(r,Q,\theta) \equiv pS(r,Q) - (c+h)Q - K(r,\theta) + K_2(r,\theta)H(\theta).$$

We can derive the second-order condition (SOC) of $M(r,Q,\theta)$ and use Lemmas 1 and 3 to show its strict concavity in (r,Q) given θ fixed. Then, its system of first-order conditions (FOC) uniquely determines the optimal pair of decisions (r^*,Q^*):

$$\begin{cases} M_1(r,Q,\theta) \equiv pS_1(r,Q) - K_1(r,\theta) + K_{12}(r,\theta)H(\theta) = 0, \\ M_2(r,Q,\theta) \equiv pS_2(r,Q) - c - h = 0. \end{cases}$$

We further have the following monotone properties. The second equation above does not contain θ and thus directly gives a relation between r and Q. Define $Q^\star(r) = \arg\max_Q pS(r,Q) - cQ - hQ$ (if $Q^\star(r)$ is not unique, we can pick any branch, but by strict concavity it has to be unique at $r = r^*(\theta)$). By its definition,

$$\max_{r,Q} M(r,Q,\theta) = \max_r M(r,Q^\star(r),\theta) \implies r^*(\theta) = \arg\max_r M(r,Q^\star(r),\theta).$$

The cross partial derivative of $M(r,Q^\star(r),\theta)$ w.r.t. (r,θ) is

$$\frac{\partial}{\partial r}\left(\frac{\partial M(r,Q^\star(r),\theta)}{\partial\theta}\right)=\frac{\partial^2}{\partial r\partial\theta}[-K_1(r,\theta)+K_2(r,\theta)H(\theta)]=-\underbrace{K_{12}}_{<0}+\underbrace{K_{122}}_{\geq 0}\underbrace{H}_{\geq 0}+\underbrace{K_{12}}_{<0}\underbrace{H'}_{\leq 0}>0,$$

which means $M(r,Q^\star(r),\theta)$ is supermodular in (r,θ), and thus by Topkis' Theorem, $r^*(\theta)$ is strictly increasing in θ. Take the derivative w.r.t. r on both sides of $M_2(r,Q^\star(r),\theta)=0$ and we have

$$-p\underbrace{g(l(r,Q^\star))}_{\geq 0}\left(\underbrace{l_1(r,Q^\star)}_{<0}+\underbrace{l_2(r,Q^\star)}_{>0}\frac{\mathrm{d}Q^\star}{\mathrm{d}r}\right)=0$$

which implies $\frac{\mathrm{d}Q^\star}{\mathrm{d}r}\geq 0$ and thus $Q^*(\theta)=Q^\star(r^*(\theta))$ is increasing in θ. $\square$

We merge the proofs of Propositions 2 and 3 together.

Proofs of Propositions 2 and 3. For ease of notation, we omit the subscripts.

We first consider (P_{CD}) and $T_{CD}(d,\theta)$. By the definition of $\alpha(\theta)$ and by Condition (D), we have $\alpha'(\theta)\geq 0$. Under T_{CD}, we have

$$\begin{aligned}\pi(\theta,\hat\theta)&=\max_{r\geq 0}\ \mathbb{E}_\epsilon[\alpha(\hat\theta)d(r,\epsilon)+\beta(\hat\theta)]-hQ(\hat\theta)-K(r,\theta)\\&=\max_{r\geq 0}\ \alpha(\hat\theta)D(r)+\beta(\hat\theta)-hQ(\hat\theta)-K(r,\theta).\end{aligned}$$

The cross partial derivative of the maximand w.r.t. (r,θ) is $-K_{12}(r,\theta)>0$, which implies strict supermodularity, and thus $r(\theta,\hat\theta)$ is strictly increasing in θ.

Under truthtelling, we have

$$\pi(\theta)=\max_{r\geq 0}\ \alpha(\theta)D(r)+\beta(\theta)-hQ(\theta)-K(r,\theta).$$

The FOC of the maximand w.r.t. r is $\alpha(\theta)D_1(r)-K_1(r,\theta)=0$, and the second-order condition (SOC), $\alpha(\theta)D_{11}(r)-K_{11}(r,\theta)<0$, implies the strict concavity of maximand and the uniqueness of the maximizer. It can be verified that given the proposed $\alpha(\theta)$, $r=r^*(\theta)$ is the root, i.e., $r(\theta)=\arg\max_{r\geq 0}\pi(\theta)=r^*(\theta)$. Therefore, the firm's ideal output $r^*(\theta)$ is induced, which leads to the firm's ideal profit. The proposed instrument $\beta(\theta)$ is set to guarantee that $\pi(\theta)=-\int_{\underline{\theta}}^{\theta}K_2(r^*(z),z)\mathrm{d}z$.

To show that T_{CD} satisfies (IC), we proceed as follows. By the envelope theorem, we have

$$\begin{aligned}-K_2(r(\theta),\theta)=\pi'(\theta)&=\frac{\partial}{\partial\theta}\Big[\alpha(\theta)D(r)+\beta(\theta)-hQ(\theta)-K(r,\theta)\Big]\Big|_{r=r(\theta)}\\&=\alpha'(\theta)D(r(\theta))+\beta'(\theta)-hQ'(\theta)-K_2(r(\theta),\theta)\\\implies\beta'(\theta)&=-\alpha'(\theta)D(r(\theta))+hQ'(\theta).\end{aligned}$$

By the envelope theorem again, we have

$$\begin{aligned}\frac{\partial\pi(\theta,\hat\theta)}{\partial\hat\theta}&=\frac{\partial}{\partial\hat\theta}\Big[\alpha(\hat\theta)D(r)+\beta(\hat\theta)-hQ(\hat\theta)-K(r,\theta)\Big]\Big|_{r=r(\theta,\hat\theta)}\\&=\alpha'(\hat\theta)D(r(\theta,\hat\theta))+\beta'(\hat\theta)-hQ'(\hat\theta)\\&=\alpha'(\hat\theta)(D(r(\theta,\hat\theta))-D(r(\hat\theta))).\end{aligned}$$

Recall that $\alpha'(\hat\theta)\geq 0$ and that $r(\theta,\hat\theta)$ is strictly increasing in θ. If $\hat\theta<\theta$, then $d(r(\theta,\hat\theta),\epsilon)>d(r(\hat\theta),\epsilon)$ and thus $\frac{\partial\pi(\theta,\hat\theta)}{\partial\hat\theta}>0$; if $\hat\theta>\theta$, then $d(r(\theta,\hat\theta),\epsilon)<d(r(\hat\theta),\epsilon)$ and thus $\frac{\partial\pi(\theta,\hat\theta)}{\partial\hat\theta}<0$; and if $\hat\theta=\theta$, then $d(r(\theta,\hat\theta),\epsilon)=d(r(\hat\theta),\epsilon)$. This implies $\theta=\arg\max_{\hat\theta}\pi(\theta,\hat\theta)$.

We then consider ($\mathrm{P_{CS}}$) and $T_{CS}(s,\theta)$. By the definition of $\alpha(\theta)$ and by Condition (S), we have $\alpha'(\theta)\geq 0$. Under the proposed contract $T_{CS}(s,\theta)$, we have

$$\pi(\theta,\hat{\theta})=\max_{r\geq 0}\ \alpha(\hat{\theta})S(r,Q(\hat{\theta}))+\beta(\hat{\theta})-hQ(\hat{\theta})-K(r,\theta).$$

Since the cross partial derivative of the maximand w.r.t. (r,θ) is $-K_{12}(r,\theta)>0$, which implies strict supermodularity, $r(\theta,\hat{\theta})$ is strictly increasing in θ.

Under truthtelling, we have

$$\pi(\theta)=\max_{r\geq 0}\ \alpha(\theta)S(r,Q(\theta))+\beta(\theta)-hQ(\theta)-K(r,\theta).$$

The FOC of the maximand w.r.t. r is $\alpha(\theta)S_1(r,Q(\theta))-K_1(r,\theta)=0$, and the SOC, $\alpha(\theta)S_{11}(r,Q(\theta))-K_{11}(r,\theta)<0$, implies that the maximizer is unique. It can be verified that given the proposed $\alpha(\theta)$, $r=r^*(\theta)$ is the root. The proposed instrument $\beta(\theta)$ is set to guarantee that $\pi(\theta)=-\int_{\underline{\theta}}^{\theta}K_2(r^*(z),z)\mathrm{d}z$.

To show that T_{CS} satisfies (IC), we proceed as follows. By the envelope theorem, we have

$$\begin{aligned}-K_2(r(\theta),\theta)=\pi'(\theta)&=\frac{\partial}{\partial\theta}\Big[\alpha(\theta)S(r,Q(\theta))+\beta(\theta)-hQ(\theta)-K(r,\theta)\Big]\Big|_{r=r(\theta)}\\&=\alpha'(\theta)S(r(\theta),Q(\theta))+\alpha(\theta)S_2(r(\theta),Q(\theta))Q'(\theta)+\beta'(\theta)-hQ'(\theta)-K_2(r(\theta),\theta)\\\Longrightarrow\ \beta'(\theta)&=-\alpha'(\theta)S(r(\theta),Q(\theta))-\alpha(\theta)Q'(\theta)S_2(r(\theta),Q(\theta))+hQ'(\theta)\end{aligned}$$

By the envelope theorem again, we have

$$\begin{aligned}\frac{\partial\pi(\theta,\hat{\theta})}{\partial\hat{\theta}}&=\frac{\partial}{\partial\hat{\theta}}\Big[\alpha(\hat{\theta})S(r,Q(\hat{\theta}))+\beta(\hat{\theta})-hQ(\hat{\theta})-K(r,\theta)\Big]\Big|_{r=r(\theta,\hat{\theta})}\\&=\underbrace{\alpha'(\hat{\theta})}_{\geq 0}[S(r(\theta,\hat{\theta}),Q(\hat{\theta}))-S(r(\hat{\theta}),Q(\hat{\theta}))]+\alpha(\hat{\theta})\underbrace{Q'(\hat{\theta})}_{\geq 0}[S_2(r(\theta,\hat{\theta}),Q(\hat{\theta}))-S_2(r(\hat{\theta}),Q(\hat{\theta}))].\end{aligned}$$

Note that both $S(r,Q)$ and $S_2(r,Q)$ are increasing in r. Following the same reasoning as before, we have $\theta=\arg\max_{\hat{\theta}}\pi(\theta,\hat{\theta})$. $\square$

Proof of Proposition 4. The analysis procedure is identical to that in the proof of Proposition 3, as we have the same expression of $M(r,Q,\theta)$, so that we can derive the exact same upper-bound pair of decisions, (r^*,Q^*). To show that this upper bound is attainable under Condition (D) in ($\mathrm{P_{DD}}$) or under Condition (S) in ($\mathrm{P_{DS}}$), we proceed to the proof of the subsequent proposition. $\square$

Proof of Proposition 5. For ease of notation, we omit the subscripts.

We first consider ($\mathrm{P_{DD}}$) and $T_{DD}(d,Q,\theta)$. By the definition of $\alpha(\theta)$ and by Condition (D), we have $\alpha'(\theta)\geq 0$. Under the proposed contract T_{DD}, we have $\pi(\theta,\hat{\theta})=\max_{r,Q\geq 0}N(r,Q,\theta,\hat{\theta})$, where

$$N(r,Q,\theta,\hat{\theta})=\alpha(\hat{\theta})D(r)+\beta(\hat{\theta})+\gamma(\hat{\theta})Q-Q^2-hQ-K(r,\theta).$$

Note that in the maximand, the term involving Q is $\gamma(\hat{\theta})Q-Q^2-hQ$, which by the FOC leads to $Q(\theta,\hat{\theta})=Q(\hat{\theta})=\frac{\gamma(\hat{\theta})-h}{2}$, $\forall\theta\in[\underline{\theta},\bar{\theta}]$. Provided $Q(\hat{\theta})$, the cross partial derivative of the maximand w.r.t. (r,θ) is $-K_{12}(r,\theta)>0$, so $r(\theta,\hat{\theta})$ is strictly increasing in θ. Moreover, $r(\theta,\hat{\theta})$ has to satisfy the FOC

$$\alpha(\hat{\theta})D_1(r)-K_1(r,\theta)=0.$$

It can be verified that under truthtelling (i.e., $\theta = \hat{\theta}$) and given the proposed instruments $\{\alpha(\theta), \gamma(\theta)\}$, the pair (r^*, Q^*) is the unique root. The proposed $\beta(\theta)$ is set to guarantee that $\pi(\theta) = -\int_{\underline{\theta}}^{\theta} K_2(r^*(z), z)\mathrm{d}z$.

To show that T_{DD} satisfies (IC), we proceed as follows. By the envelope theorem,

$$\begin{aligned}
-K_2(r(\theta),\theta) = \pi'(\theta) &= \frac{\partial}{\partial\theta}\Big[\alpha(\theta)D(r) + \beta(\theta) + \gamma(\theta)Q - Q^2 - hQ - K(r,\theta)\Big]\Big|_{\substack{r=r(\theta)\\ Q=Q(\theta)}} \\
&= \alpha'(\theta)D(r(\theta)) + \beta'(\theta) + \gamma'(\theta)Q(\theta) - K_2(r(\theta),\theta) \\
\implies \beta'(\theta) &= -\alpha'(\theta)D(r(\theta)) - \gamma'(\theta)Q(\theta)
\end{aligned}$$

By the envelope theorem again, we have

$$\begin{aligned}
\frac{\partial\pi(\theta,\hat{\theta})}{\partial\hat{\theta}} &= \frac{\partial}{\partial\hat{\theta}}\Big[\alpha(\hat{\theta})D(r) + \beta(\hat{\theta}) + \gamma(\hat{\theta})Q - Q^2 - hQ - K(r,\theta)\Big]\Big|_{\substack{r=r(\theta,\hat{\theta})\\ Q=Q(\theta,\hat{\theta})}} \\
&= \alpha'(\hat{\theta})D(r(\theta,\hat{\theta})) + \beta'(\hat{\theta}) + \gamma'(\hat{\theta})Q(\theta,\hat{\theta}) \\
&= \underbrace{\alpha'(\hat{\theta})}_{\geq 0}(D(r(\theta,\hat{\theta})) - D(r(\hat{\theta}))) + \gamma'(\hat{\theta})(\underbrace{Q(\theta,\hat{\theta}) - Q(\hat{\theta})}_{=0}).
\end{aligned}$$

The rest of the analysis is the same as that in the proof of Proposition 3, and we have $\theta = \arg\max_{\hat{\theta}} \pi(\theta, \hat{\theta})$.

<u>We then consider ($\mathrm{P_{DS}}$) and $T_{DS}(s,Q,\theta)$.</u> By the definition of $\alpha(\theta)$ and by Condition (D), we have $\alpha'(\theta) \geq 0$. Under the proposed contract T_{DS}, we have $\pi(\theta,\hat{\theta}) = \max_{r,Q\geq 0} N(r,Q,\theta,\hat{\theta})$, where

$$N(r,Q,\theta,\hat{\theta}) = \alpha(\hat{\theta})D(r) + \beta(\hat{\theta}) + \gamma(\hat{\theta})Q - Q^2 - hQ - K(r,\theta).$$

Note that in the maximand, the term involving Q is $\gamma(\hat{\theta})Q - Q^2 - hQ$, which by the FOC leads to $Q(\theta,\hat{\theta}) = Q(\hat{\theta}) = \frac{\gamma(\hat{\theta})-h}{2}$, $\forall\theta \in [\underline{\theta}, \bar{\theta}]$. Provided $Q(\hat{\theta})$, the cross partial derivative of the maximand w.r.t. (r,θ) is $-K_{12}(r,\theta) > 0$, so $r(\theta,\hat{\theta})$ is strictly increasing in θ. Moreover, $r(\theta,\hat{\theta})$ has to satisfy the FOC

$$\alpha(\hat{\theta})D_1(r) - K_1(r,\theta) = 0.$$

It can be verified that under truthtelling (i.e., $\theta = \hat{\theta}$) and given the proposed instruments $\{\alpha(\theta), \gamma(\theta)\}$, the pair (r^*, Q^*) is the unique root. The proposed $\beta(\theta)$ is set to guarantee that $\pi(\theta) = -\int_{\underline{\theta}}^{\theta} K_2(r^*(z), z)\mathrm{d}z$.

To show that T_{DS} satisfies (IC), we proceed as follows. By the envelope theorem,

$$\begin{aligned}
-K_2(r(\theta),\theta) = \pi'(\theta) &= \frac{\partial}{\partial\theta}\Big[\alpha(\theta)D(r) + \beta(\theta) + \gamma(\theta)Q - Q^2 - hQ - K(r,\theta)\Big]\Big|_{\substack{r=r(\theta)\\ Q=Q(\theta)}} \\
&= \alpha'(\theta)D(r(\theta)) + \beta'(\theta) + \gamma'(\theta)Q(\theta) - K_2(r(\theta),\theta) \\
\implies \beta'(\theta) &= -\alpha'(\theta)D(r(\theta)) - \gamma'(\theta)Q(\theta)
\end{aligned}$$

By the envelope theorem again, we have

$$\begin{aligned}
\frac{\partial\pi(\theta,\hat{\theta})}{\partial\hat{\theta}} &= \frac{\partial}{\partial\hat{\theta}}\Big[\alpha(\hat{\theta})D(r) + \beta(\hat{\theta}) + \gamma(\hat{\theta})Q - Q^2 - hQ - K(r,\theta)\Big]\Big|_{\substack{r=r(\theta,\hat{\theta})\\ Q=Q(\theta,\hat{\theta})}} \\
&= \alpha'(\hat{\theta})D(r(\theta,\hat{\theta})) + \beta'(\hat{\theta}) + \gamma'(\hat{\theta})Q(\theta,\hat{\theta}) \\
&= \underbrace{\alpha'(\hat{\theta})}_{\geq 0}(D(r(\theta,\hat{\theta})) - D(r(\hat{\theta}))) + \gamma'(\hat{\theta})(\underbrace{Q(\theta,\hat{\theta}) - Q(\hat{\theta})}_{=0}).
\end{aligned}$$

The rest of the analysis is the same as that in the proof of Proposition 3, and we have $\theta = \arg\max_{\hat{\theta}} \pi(\theta, \hat{\theta})$. □

Proof of Proposition 6. Because the upper-bound pair (r^*, Q^*) is attainable under Condition (D) in ($\mathrm{P_{CD}}$) and ($\mathrm{P_{DD}}$) and under Condition (S) in ($\mathrm{P_{CS}}$) and ($\mathrm{P_{DS}}$), the firm's resulting optimal profits in all four settings become identical.

Appendix B: The optimality of MQC under demand-based contracting

In this appendix, by additionally assuming $d_{12} \geq 0$ and $d_{112} \geq 0$, we show that in the demand-based contracting problems, i.e., ($\mathrm{P_{CD}}$), and ($\mathrm{P_{DD}}$), the upper-bound outcome (r^*, Q^*) can be consistently implemented by a menu of quadratic contracts (MQC), regardless of Condition (D). In contrast, proposing a quadratic compensation scheme in the sales-based contracting problems, ($\mathrm{P_{CS}}$), and ($\mathrm{P_{DS}}$), results in an ill-behaved agent's profit maximization problem, wherein the uniqueness of $r^*(\theta)$ is no longer verifiable. This contrast sheds light on how the operation factor, which is the censorship of demand information here, can fundamentally alter the optimal contact.

For notational convenience, define $\Phi(r) = \mathbb{E}_\epsilon[d(r,\epsilon)^2]$. We need the following lemmas.

Lemma 4. *Let Z be a random variable, $A(z)$ and $B(z)$ be two monotonic functions of z. If one is increasing and the other is decreasing, then $\mathbb{E}[A(Z)B(Z)] \leq \mathbb{E}[A(Z)]\mathbb{E}[B(Z)]$. If both are increasing or decreasing, then $\mathbb{E}[A(Z)B(Z)] \geq \mathbb{E}[A(Z)]\mathbb{E}[B(Z)]$.*

Proof. Let Z_1 and Z_2 be two independent identical copies of the random variable Z. If one of $A(z)$ and $B(z)$ is increasing and the other is decreasing, then we have

$$\begin{aligned}
&\mathbb{E}[(A(Z_1)-A(Z_2))(B(Z_1)-B(Z_2))] \leq 0 \\
\Longleftrightarrow\ &\underbrace{\mathbb{E}[A(Z_1)B(Z_1)]}_{=\mathbb{E}[A(Z)B(Z)]} + \underbrace{\mathbb{E}[A(Z_2)B(Z_2)]}_{=\mathbb{E}[A(Z)B(Z)]} \leq \underbrace{\mathbb{E}[A(Z_1)B(Z_2)]}_{=\mathbb{E}[A(Z)]\mathbb{E}[B(Z)]} + \underbrace{\mathbb{E}[A(Z_2)B(Z_1)]}_{=\mathbb{E}[A(Z)]\mathbb{E}[B(Z)]} \\
\Longleftrightarrow\ &\mathbb{E}[A(Z)B(Z)] \leq \mathbb{E}[A(Z)]\mathbb{E}[B(Z)].
\end{aligned}$$

The same trick can be used to prove the statement when both functions are increasing or decreasing. □

Lemma 5. *We have*

$$\frac{\mathrm{d}}{\mathrm{d}r}\frac{\Phi_1(r)}{D_1(r)} > 0.$$

Proof. First we have

$$\frac{\mathrm{d}}{\mathrm{d}r}\frac{\Phi_1(r)}{2D_1(r)} = \frac{\mathrm{d}}{\mathrm{d}r}\frac{\mathbb{E}_\epsilon[d(r,\epsilon)d_1(r,\epsilon)]}{\mathbb{E}_\epsilon[d_1(r,\epsilon)]} = \frac{\mathbb{E}_\epsilon[d(r,\epsilon)d_{11}(r,\epsilon)+d_1(r,\epsilon)^2]\mathbb{E}_\epsilon[d_1(r,\epsilon)] - \mathbb{E}_\epsilon[d(r,\epsilon)d_1(r,\epsilon)]\mathbb{E}_\epsilon[d_{11}(r,\epsilon)]}{\mathbb{E}_\epsilon[d_1(r,\epsilon)]^2}.$$

Since $d(r,\epsilon)$, $d_1(r,\epsilon)$ and $d_{11}(r,\epsilon)$ are all increasing in ϵ, per Lemma 4, the numerator is

$$\begin{aligned}
&\mathbb{E}_\epsilon[d(r,\epsilon)d_{11}(r,\epsilon)+d_1(r,\epsilon)^2]\mathbb{E}_\epsilon[\underbrace{d_1(r,\epsilon)}_{>0}] - \mathbb{E}_\epsilon[d(r,\epsilon)d_1(r,\epsilon)]\mathbb{E}_\epsilon[\underbrace{d_{11}(r,\epsilon)}_{<0}] \\
\geq &(\mathbb{E}_\epsilon[d(r,\epsilon)]\mathbb{E}_\epsilon[d_{11}(r,\epsilon)] + \mathbb{E}_\epsilon[d_1(r,\epsilon)^2])\mathbb{E}_\epsilon[d_1(r,\epsilon)] - \mathbb{E}_\epsilon[d(r,\epsilon)]\mathbb{E}_\epsilon[d_1(r,\epsilon)]\mathbb{E}_\epsilon[d_{11}(r,\epsilon)] \\
= &\mathbb{E}_\epsilon[d_1(r,\epsilon)^2]\mathbb{E}_\epsilon[d_1(r,\epsilon)] > 0.
\end{aligned}$$

Proposition 7. *For ($\mathrm{P_{CD}}$), the contract $T_{CD}(d,\theta) = \delta_{CD}d^2 + \alpha_{CD}(\theta)d + \beta_{CD}(\theta)$ is optimal, where*

$$\begin{aligned}
\delta_{CD} &= \min\left\{0, \min_{\theta\in[\underline{\theta},\bar{\theta}]} \frac{\frac{d}{d\theta}\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))}}{\frac{d}{d\theta}\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))}}\right\} \\
\alpha_{CD}(\theta) &= \frac{K_1(r^*(\theta),\theta) - \delta_{CD}\Phi_1(r^*(\theta))}{D_1(r^*(\theta))} \\
\beta_{CD}(\theta) &= -[\delta_{CD}\Phi(r^*(\theta)) + \alpha_{CD}(\theta)D(r^*(\theta))] + K(r^*(\theta),\theta) + hQ^*(\theta) - \int_{\underline{\theta}}^{\theta} K_2(r^*(z),z)dz.
\end{aligned}$$

Proof. We show that the proposed MQC implements (r^*, Q^*). For ease of notation, we omit the subscripts. Because $r^*(\theta)$ strictly increases in θ, so does $\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))}$ by Lemma 5, and thus δ is well-defined. Essentially, δ is set to guarantee that $\alpha'(\theta) \geq 0$. To see this, notice that by definition, $\delta \leq 0$ and

$$\delta \leq \min_{\theta \in [\underline{\theta}, \bar{\theta}]} \frac{\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))}}{\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))}} \leq \frac{\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))}}{\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))}}.$$

Therefore,

$$\begin{aligned}
\alpha'(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))} - \delta_{CD}\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))} \underbrace{-\delta_{CD}}_{\geq 0} \underbrace{\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))}}_{>0} \\
&\geq \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))} - \frac{\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))}}{\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))}} \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))} = 0.
\end{aligned}$$

Under the proposed contract, we have

$$\begin{aligned}
\pi(\theta, \hat{\theta}) &= \max_{r \geq 0} \ \mathbb{E}_\epsilon[\delta d(r,\epsilon)^2 + \alpha(\hat{\theta})d(r,\epsilon) + \beta(\hat{\theta})] - hQ(\hat{\theta}) - K(r,\theta) \\
&= \max_{r \geq 0} \ \delta\Phi(r) + \alpha(\hat{\theta})D(r) + \beta(\hat{\theta}) - hQ(\hat{\theta}) - K(r,\theta).
\end{aligned}$$

Because $\lim_{r \to -\infty} \pi(\theta, \hat{\theta}) = -\infty$, $r(\theta, \hat{\theta}) = \arg\max_{r \geq 0} \pi(\theta, \hat{\theta})$ is well-define. To show that the above maximand has a unique maximizer, we will use the following fact: if a continuously differentiable function is strictly concave at every stationary point, then its stationary point is unique and is the maximizer. The FOC gives

$$\mathbb{E}_\epsilon[(2\delta d(r,\epsilon) + \alpha(\hat{\theta}))d_1(r,\epsilon)] - K_1(r,\theta) = 0,$$

and to validate the SOC

$$2\underbrace{\delta\mathbb{E}_\epsilon[d_1(r(\theta,\hat{\theta}),\epsilon)^2]}_{\leq 0} + \mathbb{E}_\epsilon[(\alpha(\hat{\theta}) + 2\delta d(r(\theta,\hat{\theta}),\epsilon))d_{11}(r(\theta,\hat{\theta}),\epsilon)] - \underbrace{K_{11}(r(\theta,\hat{\theta}),\theta)}_{>0} < 0$$

locally at $r = (\theta, \hat{\theta})$, it suffices to show that the middle term is negative. In the FOC, since $\alpha(\hat{\theta}) + 2\delta d(r(\theta,\hat{\theta}),\epsilon)$ decreases in ϵ and $d_1(r(\theta,\hat{\theta}),\epsilon)$ increases in ϵ, per Lemma 4, we have

$$\begin{aligned}
0 < \underbrace{K_1(r(\theta,\hat{\theta}),\theta) = \mathbb{E}_\epsilon[(2\delta d(r(\theta,\hat{\theta}),\epsilon) + \alpha(\hat{\theta}))d_1(r(\theta,\hat{\theta}),\epsilon)]}_{\text{FOC}} &\leq \mathbb{E}_\epsilon[2\delta d(r(\theta,\hat{\theta}),\epsilon) + \alpha(\hat{\theta})]\mathbb{E}_\epsilon[\underbrace{d_1(r(\theta,\hat{\theta}),\epsilon)}_{>0}] \\
\Longrightarrow \mathbb{E}_\epsilon[2\delta d(r(\theta,\hat{\theta}),\epsilon) + \alpha(\hat{\theta})] &\geq \frac{K_1(r(\theta,\hat{\theta}),\theta)}{\mathbb{E}_\epsilon[d_1(r(\theta,\hat{\theta}),\epsilon)]} > 0.
\end{aligned}$$

Since $d_{11}(r(\theta,\hat{\theta}),\epsilon)$ increases in ϵ, per Lemma 4 again, we have

$$\mathbb{E}_\epsilon[(\alpha(\hat{\theta}) + 2\delta d(r(\theta,\hat{\theta}),\epsilon))d_{11}(r(\theta,\hat{\theta}),\epsilon)] \leq \underbrace{\mathbb{E}_\epsilon[\alpha(\hat{\theta}) + 2\delta d(r(\theta,\hat{\theta}),\epsilon)]}_{>0}\mathbb{E}_\epsilon[\underbrace{d_{11}(r(\theta,\hat{\theta}),\epsilon)}_{\leq 0}] \leq 0,$$

which verifies the SOC. Consequently, $r(\theta, \hat{\theta})$ is unique.

Since θ only appears in $K(r,\theta)$ in the maximand, the cross partial derivative of the maximand is simply $-K_{12} > 0$, which implies the strict supermodularity, and thus $r(\theta, \hat{\theta})$ is strictly increasing in θ.

Under truthtelling, we have

$$\pi(\theta) = \max_{r \geq 0} \ \delta\Phi(r) + \alpha(\theta)D(r) + \beta(\theta) - hQ(\theta) - K(r,\theta).$$

The FOC of the maximand w.r.t. r is

$$\delta\Phi_1(r) + \alpha(\theta)D_1(r) - K_1(r,\theta) = 0,$$

and it can be easily checked that given the proposed instruments δ and $\alpha(\theta)$, $r = r^*(\theta)$ is the root, i.e., the firm's ideal output $r^*(\theta)$ is induced so that its ideal profit is attained. We remark that the proposed instrument $\beta(\theta)$ is set to guarantee that $\pi(\theta) = -\int_{\underline{\theta}}^{\theta} K_2(r^*(z), z)\mathrm{d}z$.

We show that this MQC satisfies (IC). With a little abuse of notations, we simply write $r(\theta) = r^*(\theta)$ and $Q(\theta) = Q^*(\theta)$. By the envelope theorem, we have

$$\begin{aligned}
-K_2(r(\theta),\theta) = \pi'(\theta) &= \frac{\partial}{\partial\theta}\Big[\delta\Phi(r) + \alpha(\theta)D(r) + \beta(\theta) - hQ(\theta) - K(r,\theta)\Big]\Big|_{r=r(\theta)} \\
&= \alpha'(\theta)D(r(\theta)) + \beta'(\theta) - hQ'(\theta) - K_2(r(\theta),\theta) \\
\implies \beta'(\theta) &= -\alpha'(\theta)D(r(\theta)) + hQ'(\theta).
\end{aligned}$$

By the envelope theorem again, we have

$$\begin{aligned}
\frac{\partial\pi(\theta,\hat\theta)}{\partial\hat\theta} &= \frac{\partial}{\partial\hat\theta}\Big[\delta\Phi(r) + \alpha(\hat\theta)D(r) + \beta(\hat\theta) - hQ(\hat\theta) - K(r,\theta)\Big]\Big|_{r=r(\theta,\hat\theta)} \\
&= \alpha'(\hat\theta)D(r(\theta,\hat\theta)) + \beta'(\hat\theta) - hQ'(\hat\theta) \\
&= \alpha'(\hat\theta)[D(r(\theta,\hat\theta)) - D(r(\hat\theta))].
\end{aligned}$$

The rest of the analysis is the same as that in the proof of Proposition 3, and we have $\theta = \arg\max_{\hat\theta}\pi(\theta,\hat\theta)$. $\square$

PROPOSITION 8. *For* ($\mathrm{P_{DD}}$)*, the contract* $T_{DD}(d,\theta) = \delta_{DD}d^2 + \alpha_{DD}(\theta)d + \beta_{DD}(\theta) + \gamma_D D(\theta)Q - Q^2$ *is optimal, where*

$$\begin{aligned}
\delta_{DD} &= \min\left\{0, \min_{\theta\in[\underline{\theta},\bar\theta]} \frac{\frac{d}{d\theta}\frac{K_1(r^*(\theta),\theta)}{D_1(r^*(\theta))}}{\frac{d}{d\theta}\frac{\Phi_1(r^*(\theta))}{D_1(r^*(\theta))}}\right\} \\
\alpha_{DD}(\theta) &= \frac{K_1(r^*(\theta),\theta) - \delta_{DD}\Phi_1(r^*(\theta))}{D_1(r^*(\theta))} \\
\gamma_{DD}(\theta) &= 2Q^*(\theta) + h \\
\beta_{DD}(\theta) &= -[\delta_{DD}\Phi(r^*(\theta)) + \alpha_{DD}(\theta)D(r^*(\theta)) + \gamma_{DD}Q^*(\theta) - Q^*(\theta)^2] \\
&\quad + K(r^*(\theta),\theta) + hQ^*(\theta) - \int_{\underline{\theta}}^{\theta} K_2(r^*(z),z)dz.
\end{aligned}$$

Proof. We show that the proposed MQC implements (r^*, Q^*). Note that $\delta_{DD} = \delta_{CD}$ and $\alpha_{DD}(\theta) = \alpha_{CD}(\theta)$, so δ_{DD} is well-defined and $\alpha'_{DD}(\theta) \geq 0$. For ease of notation, we omit the subscripts. We have

$$\pi(\theta,\hat\theta) = \max_{r,Q\geq 0} \delta\Phi(r) + \alpha(\hat\theta)D(r) + \beta(\hat\theta) + \gamma(\hat\theta)Q - Q^2 - hQ - K(r,\theta).$$

Note that in the maximand, the term involving Q is $\gamma(\hat\theta)Q - Q^2 - hQ$, which by the FOC leads to $Q(\theta,\hat\theta) = Q(\hat\theta) = \frac{\gamma(\hat\theta)-h}{2}$, $\forall\theta\in[\underline{\theta},\bar\theta]$. Provided $Q(\hat\theta)$, the cross partial derivative of the maximand w.r.t. (r,θ) is $-K_{12}(r,\theta) > 0$, so $r(\theta,\hat\theta)$ is strictly increasing in θ. The term involving r is $\delta\Phi(r) + \alpha(\hat\theta)D(r) - K(r,\theta)$, and with the same argument as in Proposition 7 (because the two FOC's have identical expressions), $r(\theta,\hat{\theta}) = \arg\max_{r\geq 0}\pi(\theta,\hat{\theta})$ is well-defined, unique, and increasing in θ. It can be verified that under truthtelling (i.e., $\theta=\hat{\theta}$) and given the proposed instruments $\{\alpha(\theta),\gamma(\theta)\}$, the pair (r^*,Q^*) is the root. The proposed $\beta(\theta)$ is set to guarantee that $\pi(\theta) = -\int_{\underline{\theta}}^{\theta} K_2(r^*(z),z)\mathrm{d}z$.

We show that this MQC satisfies (IC). By the envelope theorem,

$$\begin{aligned}
-K_2(r(\theta),\theta) = \pi'(\theta) &= \frac{\partial}{\partial\theta}\Big[\delta\Phi(r)+\alpha(\theta)D(r)+\beta(\theta)+\gamma(\theta)Q-Q^2-hQ-K(r,\theta)\Big]\Big|_{\substack{r=r(\theta)\\ Q=Q(\theta)}}\\
&= \alpha'(\theta)D(r(\theta))+\beta'(\theta)+\gamma'(\theta)Q(\theta)-K_2(r(\theta),\theta)\\
\implies \beta'(\theta) &= -\alpha'(\theta)D(r(\theta))-\gamma'(\theta)Q(\theta)
\end{aligned}$$

By the envelope theorem again, we have

$$\begin{aligned}
\frac{\partial\pi(\theta,\hat{\theta})}{\partial\hat{\theta}} &= \frac{\partial}{\partial\hat{\theta}}\Big[\delta\Phi(r)+\alpha(\hat{\theta})D(r)+\beta(\hat{\theta})+\gamma(\hat{\theta})Q-Q^2-hQ-K(r,\theta)\Big]\Big|_{\substack{r=r(\theta,\hat{\theta})\\ Q=Q(\theta,\hat{\theta})}}\\
&= \alpha'(\hat{\theta})D(r(\theta,\hat{\theta}))+\beta'(\hat{\theta})+\gamma'(\hat{\theta})Q(\theta,\hat{\theta})\\
&= \underbrace{\alpha'(\hat{\theta})}_{\geq 0}[D(r(\theta,\hat{\theta}))-D(r(\hat{\theta}))]+\gamma'(\hat{\theta})(\underbrace{Q(\theta,\hat{\theta})-Q(\hat{\theta})}_{=0}).
\end{aligned}$$

The rest of the analysis is the same as that in the proof of Proposition 3, and we have $\theta = \arg\max_{\hat{\theta}}\pi(\theta,\hat{\theta})$. $\square$