Title
The Role of Market Evolution in Channel Contracting

Journal
Management Science, 65(5)

ISSN
0025-1909

Authors
Gao, Long
Mishra, Birendra K

Publication Date
2019

DOI
10.1287/mnsc.2018.3057

Peer reviewed

The Role of Market Evolution in Channel Contracting

Long Gao
A. Gary Anderson Graduate School of Management, University of California, Riverside, CA 92521 long.gao@ucr.edu

Birendra K. Mishra
A. Gary Anderson Graduate School of Management, University of California, Riverside, CA 92521 barry.mishra@ucr.edu

Real markets evolve over time. They often exhibit complex behaviors, such as autocorrelation, continuity, and nonstationarity. How do these behaviors affect channel contracting? We study the problem in a bilateral channel where the retailer has private information on evolving market conditions. We characterize the optimal contract under arbitrary market evolution. The central notion is market inertia: it prices retailer's information advantage, dictates price and quantity response over time, and determines the contract complexity. Using market inertia, we identify a general property—stochastic linearity—that justifies the use of simple contracts for a much larger class of channel conditions. For practitioners, we offer refined guidance: (i) when the market has linear dynamics, simple contracts are sufficient; (ii) when the market is continuous, the quantity distortion should be pervasive; (iii) when the market is nonstationary, the distortion can vanish, intensify, stay constant, or even go non-monotonic over time. By highlighting the central role of realistic market behaviors, this paper advances our understanding of channel theory and practice.

Key words: market evolution; contracting; information sharing; information asymmetry; distribution channel

---

1. Introduction

Many distribution channels operate in volatile markets (He et al. 2008, 2017). To improve forecasting, downstream retailers often closely track market conditions for better demand information. If properly shared and used, this information can greatly reduce supply-demand mismatch, improve production, and enhance channel efficiency (Gao et al. 2012).

To ensure truthful information sharing, manufacturers need to write a long-term contract. The problem is complicated by three factors. First, the long-term contract should govern repeated transactions over multiple periods, during which market conditions may evolve (Dekimpe and Hanssens 2003). Therefore, the contract should account for *market evolution*. Second, each period the retailer may learn new consumer information, develop local market expertise, and gain fresh information advantage over the manufacturer; he may mislead the manufacturer to secure better price. Therefore, the contract should account for *information asymmetry*, providing incentives for truthful information sharing (Guo and Iyer 2010). Third, both parties are forward-looking: current actions may affect both parties' future learning and reactions, and the prospect of future transactions may also shape the actions taken now (Chintagunta et al. 2006). Therefore, the contract should account for such *forward-looking behavior* (Bernstein and Martínez-de Albéniz 2016).

The channel literature is largely silent on how to write such a contract. So far it has primarily focused on contracting under the *simplistic* market assumptions, e.g., static condition, binary uncertainty, and IID dynamics. Yet real markets rarely behave this way. They often exhibit *complex* behaviors that defy simplistic characterizations. For example, the carryover effect, the decay advertising pattern (Tellis 2006), the new product diffusion (Bass 1995), none of them fits the simplistic market assumptions. In general, real markets evolve stochastically, continuously, and nonstationarily (Pauwels et al. 2004); the simplistic market assumptions (e.g., static, binary, and IID) become increasingly untenable.[1] The gap between modeling and reality greatly limits the applicability of many existing channel results (Bronnenberg et al. 2005). Indeed, practitioners are often suspicious of the policies derived from the simplistic assumptions—they are "*insensitive to how real world works*" (Reiss and Wolak 2007). To improve policy recommendations, we need to model more realistic market behaviors (Wittink 2005).[2]

In this paper, we put realistic market behaviors on the center stage. We seek to understand how they drive contract response. We take the manufacturer's perspective and address four questions: (i) How should a channel adapt to an evolving market? (ii) How do realistic market behaviors affect the existing results? (iii) What determines the contract complexity? (iv) What are the new policy recommendations for channel managers?

To address these questions, we consider a dyadic channel where the market condition evolves stochastically over time. The manufacturer sells a perishable product through a retailer over multiple periods. Due to proximity and expertise, the retailer has the private knowledge of the market condition, but he may lack the incentive to share it with the manufacturer. The manufacturer has the dominant bargaining power and can commit to a long-term contract at the outset. In each period, the retailer privately learns the market condition, sets the retail price, orders production, pays the manufacturer, and sells the products in the end market; the manufacturer receives the order, produces the product, and gets paid accordingly. The transaction then repeats. Both parties are strategic, forward-looking, and profit-maximizing.

We make three contributions to the channel literature. The first is modeling. We develop a general framework that captures market evolution, information asymmetry, and forward-looking behavior. It uses time-series approach to model three market behaviors—autocorrelation, continuity, and nonstationarity. The framework greatly enhances our ability to analyze channel contracts:

[1] Few markets assume only two conditions: price and quantity change continuously. Markets in different time behave differently: IID dynamics is the exception, not the norm (Hamilton 1994, Pauwels et al. 2004). Indeed, 60% of marketing performance variables, and 78% of sales variables, are not stable, bur rather evolve over time (Dekimpe and Hanssens 1995).

[2] Wittink (2005): "As long as researchers do not capture . . . how other relevant parties may change behavior as a function of changes in market conditions, the models will fail to make correct predictions of marketplace outcomes."

we are able to unify the existing results, tackle new channel issues, and propose more credible policy recommendations.

The second contribution is theoretical. Because of arbitrary demand process and more complicated constraints, our model is significantly more difficult to solve. To overcome the technical challenge, we first lift the decision space into a larger but simpler space, and then nail down the solutions that are optimal in both spaces. Using this relaxation technique, we are able to fully characterize the optimal contract under arbitrary demand process—the most general case in the channel literature.

The main conceptual challenge is how to *price* retailer's information advantage. In the dynamic setting, the higher type retailer enjoys better sales potential now, and he is also more likely to enjoy it in the future. Unless he is paid an information rent, he would manipulate for a better profit. The manipulation could have short-term, long-term, and cumulative effects. To measure them, we develop the notion of market inertia, which represents how the market change at one point affects future market conditions. The notion connects the time-series data with the decision model. Using this notion, we can express the information rent as the sum of weighted sales-potential advantage in all future periods, where the weight is the market inertia. Hence, the price of retailer's information advantage is precisely the option value of his manipulation potential over time.

Market inertia is central to the optimal contract. Besides controlling price and quantity over time, it also serves as a "sufficient statistic" for the contract complexity. Indeed, market inertia compresses all the relevant market information into a single term; it then enters the contract in a simple factorization form. As such, the complexity of contracts boils down to the complexity of market inertia. Under the lens of market inertia, many of seemingly complex evolutions are in fact quite simple. For example, continuous nonstationary processes can have history-independent expressions of the market inertia. Leveraging this fact, we identify the general property—stochastic linearity—that guarantees the simplicity of the optimal contract. Unlike binary or IID demand, this property is common in practice, enjoying broad empirical support (Pauwels et al. 2004). As such, we justify the use of simple contracts for a much larger class of channel conditions.

Our third contribution is managerial. For channel managers, we provide refined policy guidance for realistic market behaviors. (i) When the market is stochastically linear, managers need only simple contracts, without tracking past market conditions. (ii) When the market is continuous, the quantity distortion should be pervasive. The intuition is that, continuous evolution increases retailer heterogeneity and thus the complexity of incentive provision. To ensure truthful information sharing, the distortion is necessarily more refined and pervasive. (iii) When the market condition is nonstationary, the distortion can vanish, stay constant, intensify, or go nonmonotonic over time. The intuition is that, market evolution can either dampen or heighten retailer's information advantage over time; in response, the distortion should follow in lockstep, either vanishing or intensifying over time.

2. Relation to the Literature

Our work connects two paradigms in the channel literature: analytical modeling and time-series econometrics. The analytical paradigm focuses on how to coordinate the channel and improve efficiency. Two culprits of inefficiency are double marginalization and information asymmetry (Klibanoff and Morduch 1995).

Early studies focus on *symmetric information* (Cachon 2003). They find wholesale prices alone cannot coordinate the channel (Kolay and Shaffer 2013). To eliminate double marginalization (Tirole 1988), researchers have proposed several coordination instruments: quantity discount (Jeuland and Shugan 1983), two-part tariff (Moorthy 1987), franchise agreements (Desai and Srinivasan 1995), product returns (Padmanabhan and Png 1997), bargaining power (Iyer and Villas-Boas 2003), and retail price maintenance (Iyer 1998).[3] In essence, these instruments follow the same nonlinear pricing scheme, which can achieve the first best under static, symmetric information. This insight has been extended to dynamic settings.[4] Yet the symmetric-information assumption is not innocuous. By assuming away incentive compatibility constraints, this literature does not address incentive issues in information sharing—a paramount concern in practice.

Recent studies address incentive issues under *asymmetric information*; see, e.g., Chu (1992), Desai and Srinivasan (1995), Iyer (1998), Villas-Boas (1998), Desai (2000), Mishra and Prasad (2005), He et al. (2008), Sudhir and Datta (2008), Gal-Or et al. (2008), Guo (2009), Guo and Iyer (2010), Dukes et al. (2011), Mittendorf et al. (2013), Gao et al. (2014), Gao (2015), Jiang et al. (2016), Gümüş (2017). They use either signaling or screening models, in which incentive compatibility is central. Technically, the off-equilibrium analysis is important for signaling, but unnecessary for screening because of the revelation principle (Fudenberg and Tirole 1991). Two main insights from static screening are: (i) the quantity for the low type is always downward distorted; (ii) channel coordination is unattainable. These insights are for static settings; how they fare in dynamic environments is unclear. Our contribution is three-fold. We develop a unified framework with arbitrary demand, delineate the applicability of the existing results, and offer refined policy recommendations.

[3] The behaviorial literature finds that behaviorial factors—such as risk preference, bounded rationality (Ho and Zhang 2008), trust (Özer et al. 2011), fairness (Cui et al. 2007, Katok et al. 2014)—may also serve the coordination purpose.

[4] See, e.g., Shugan (1985), Jørgensen (1986), Eliashberg and Jeuland (1986), Chintagunta and Jain (1992), Jørgensen and Zaccour (2003), Chiang (2012).

In the literature, the simple contract puzzle—why simple contracts can arise from the *complex* world—has attracted many explanations; e.g., product non-specificity and bargaining power (Iyer and Villas-Boas 2003), arbitrage (Tirole 1988), antitrust laws (McAfee 2009), contractual and communication complexities (Mookherjee 2006), and behavioral factors such as trust and fairness concerns (Cui et al. 2007, Katok et al. 2014, Özer et al. 2011). The issue of market dynamics, however, has not been well studied.

On this issue, Battaglini (2005) and Lobel and Xiao (2017) are the most relevant studies. They use the mechanism design approach (see, e.g., Courty and Hao 2000, Pavan et al. 2014).[5] They find that simple contracts can arise from dynamic settings, when the dynamics is either binary or IID. Building on this literature, we address the puzzle under more complex market behaviors (autocorrelation, continuity, and nonstationarity), and explicate their policy implications for channel management. Moreover, we pinpoint the general property—stochastic linearity—that guarantees the contract simplicity. This result legitimizes the use of simple contracts in much broader channel situations. As such, our work enriches the methods and insights of the channel literature.

3. Formulation

Consider a distribution channel. The upstream manufacturer sells a perishable product through the downstream retailer over T periods. The channel operates in a make-to-order fashion, carrying no stock over time. In each period the consumer (inverse) demand $P_t = z_t - q_t$ is determined by the *market condition* (sales potential) z_t and quantity q_t. The market condition $z_t \in \mathcal{Z} \equiv [\ell, h]$ evolves over time, and only the retailer observes z_t. Let $z^t \equiv (z_0, z_1, \ldots, z_t)$ be a history (path) at time t, and $z_s^t = (z_s, z_{s+1}, \ldots, z_t)$ the history from time s to t. Both firms are forward-looking, risk neutral, and profit maximizing; they share the same discount factor $\delta \in [0, 1]$.

The sequential game proceeds as follows. (1) In period 0, the manufacturer offers the retailer a long-term contract: $(p_t, q_t)_t \equiv \big\{(p_t(z^t), q_t(z^t)) : \ z^t \in \mathcal{Z}^t, t \le T\big\}$, where $p_t(z^t)$ is the *total payment* for ordering quantity $q_t(z^t)$. (2) In each period $t \ge 0$, the retailer first observes z_t, then orders $q_t(\hat{z}^t)$, where $\hat{z}^t = (\hat{z}_0, \ldots, \hat{z}_t)$. (3) The manufacturer produces at marginal cost c, delivers the product, and gets paid $p_t(\hat{z}^t)$. (4) The retailer sells the product for revenue $R(q_t(\hat{z}^t), z_t) = \big(z_t - q_t(\hat{z}^t)\big) q_t(\hat{z}^t)$. (5) The market condition evolves to z_{t+1}, and the stage game of (2)–(5) repeats.

From the manufacturer's perspective, the contract should control two strategic maneuvers of the retailer. First, the market condition z_t is critical for channel efficiency (matching supply with demand); but the better informed retailer may misreport, if the cost saving of doing so outweighs

[5] The mechanism design literature is vast. The most relevant papers include Courty and Hao (2000), Fernandes and Phelan (2000), Battaglini (2005), Eső and Szentes (2007), Bergemann and Välimäki (2010), Pavan et al. (2014), Deb and Said (2015), Lobel and Xiao (2017), Kuribko et al. (2018). For comprehensive reviews, see Bergemann and Said (2011) and Vohra (2012). For book-length treatments, see Laffont and Martimort (2001) and Bolton and Dewatripont (2005).

its revenue reduction. Therefore, the contract should offer a *truthtelling incentive* for credible information sharing. Second, the retailer may walk away, if taking his outside option is more profitable than the channel relationship. Therefore, the contract should ensure the *participation constraint*. To write such a contract, the manufacturer must account for evolving market conditions, and price retailer's information advantages properly. Both tasks hinge on an accurate market evolution model.

3.1. Time-Series Model for Market Evolution

It has long been recognized that models useful for policy recommendations should be good descriptive models first (Franses 2005, Kuzu et al. 2018). For market evolution, the empirical literature documents three key behaviors: autocorrelation, continuity, and nonstationarity (Dekimpe and Hanssens 1995). For example, market conditions are usually serially correlated, because of agents' habit persistence, forward-looking behavior, and dynamic responses to exogenous variables (Chintagunta et al. 2006). They may vary continuously within a stable range, e.g., in the mature phase of a product life cycle. They may fluctuate drastically around an upward trend, e.g., in the initial phase of new product sales (Bass 1995).

We use a general time-series model to capture these behaviors. The equation of motion is $z_{t+1} = G_{t+1}(z_t, \epsilon_{t+1})$, where ϵ_{t+1} is the random shock, and function G_{t+1} increases in (z_t, ϵ_{t+1}). The evolution induces (cumulative) transition probability $\Lambda_{t+1}(z_{t+1}|z_t)$ with density $\lambda_{t+1}(z_{t+1}|z_t)$. This is the workhorse model in time-series econometrics (Hamilton 1994). The function G_{t+1} represents specifications such as constant, IID, autoregressive, and generalized linear models. It can be readily estimated from demand data. As such, the model directly links to channel practice.[6]

This model captures several realistic market behaviors. (i) The state space $\mathcal{Z} \subset \mathbb{R}_+$ allows *continuous* market changes. (ii) The random shock ϵ_{t+1} captures *stochastic* market factors, e.g., competitive moves, consumer preference shifts (Villas-Boas 1999), and macro business cycles. (iii) The time dependence of function G_{t+1} allows *nonstationary* evolution—market conditions can follow different distributions in different periods (Gao et al. 2018). This is to capture the trend and seasonality of the time series; e.g., decay advertising effects in the Koyck model, and diffusion pattern of new product sales in the Bass model (Tellis 2006). Indeed, it is one of the most general time-series models, allowing both parametric and nonparametric specifications. (iv) The monotonicity of G_{t+1} ensures the first order stochastic dominance of market evolution. This is to capture the reality that similar market conditions tend to persist for a while before changing substantially to another one; e.g., the carryover effect. (v) Finally, the structure of G_{t+1} specifies the autocorrelation (intertemporal correlation)—the defining feature of a time series (Hamilton 1994).

[6] For example, the Bass diffusion model follows $N_{t+1} = G(N_t, \epsilon_{t+1}) = N_t + p(m - N_t) + q\frac{N_t}{m}(m - N_t) + \epsilon_{t+1}$, where N_t is the sales till period t, m market size, p innovation parameter, and q imitation parameter (Dekimpe et al. 2008).

Next, we develop the notion of market inertia. It serves three purposes later: to price retailer's information advantage, to control price and quantity over time, and to determine the contract complexity. The pricing rationale goes as follows. Observing private information z_t, the retailer can manipulate it by Δz_t for a better deal. The manipulation Δz_t has short-term, long-term, and cumulative effects. To assess all three effects, the notion should be able to measure how market changes are intertemporally linked, i.e., how the change Δz_t at one point in time affects future conditions. Conceptually, the notion is akin to the expected *marginal effect* of infinitesimal change Δz_t on $\{z_\tau\}_{\tau>t}$, holding all else equal. We call it *market inertia*.

Formally, we define market inertia as follows. First, we define the one-period market inertia by $\xi(z_t, z_{t+1}) \equiv \mathbb{E}_\epsilon\left[\partial_{z_t} G_{t+1}(z_t, \epsilon) \,\middle|\, z_{t+1} = G_{t+1}(z_t, \epsilon)\right]$, conditioning on the shocks that generate the path (z_t, z_{t+1}). By the chain rule, we then define the t-period inertia as $\xi(z^t) \equiv \prod_{\tau=0}^{t-1} \xi(z_\tau, z_{\tau+1})$. Market inertia is easy to estimate from demand data. For many commonly used time-series models, it also has a clear statistical interpretation: e.g., the autoregressive process $z_{t+1} = \alpha z_t + \epsilon_{t+1}$ has market inertia $\xi(z_t, z_{t+1}) = \alpha$, and $\xi(z^t) = \alpha^t$, which coincides with the *autocorrelation function* (ACF) $\rho_t \equiv \frac{\mathrm{cov}(z_t, z_0)}{\sqrt{\mathrm{var}(z_t)\cdot\mathrm{var}(z_0)}} = \alpha^t$ (Box et al. 2011).[7] In this case, the market inertia is *independent of history details* z^t, a key property for channel contracting (cf. §5.2).

Market inertia measures all three effects of retailer manipulation. (i) The one-period inertia $\xi(z_t, z_{t+1})$ is a short-term elasticity, which measures the expected effect of a change in z_t on z_{t+1}, holding constant the shocks ϵ with $z_{t+1} = G_{t+1}(z_t, \epsilon)$. (ii) The t-period inertia $\xi(z^t)$ is a long-term elasticity, which measures how the change Δz_0 propagates along path z^t to affect future z_t. (iii) Finally, the weighted sum of inertia $\xi(z^t)$ measures the accumulative effect. The key takeaway is that, the slower the market evolution, the longer the effects of manipulation, the higher the market inertia.

3.2. Decision Model for Contracting

We build on the mechanism design literature (Vohra 2012). By the *revelation principle* (Myerson 1986), we can focus on the direct revelation mechanisms. Formally, we frame the manufacturer's contracting problem as a dynamic program, while casting the retailer's strategic reactions as its constraints. By incentive compatibility, truth telling is the retailer's equilibrium strategy—he cannot gain from deviation.[8] Under the *truth telling* strategy, in each period t, the retailer reports truthfully $\hat{z}_t(z^t) = z_t$, and obtains his *equilibrium continuation payoff*

$$U_t(z^{t-1}, z_t) = R(q_t(z^t), z_t) \;-\; p_t(z^t) \;+\; \delta \mathbb{E}_{z_{t+1}}[U_{t+1}(z^{t-1}, z_t, z_{t+1}) | z_t]. \tag{1}$$

[7] The formal derivation is in Lemma 1. Market inertia $\xi(z^t)$ and ACF ρ_t are conceptually different objects: the former measures the *marginal* effects; the latter measures the linear relationship (correlation). So the agreement here is coincidental, driven by the linear dynamics.

[8] In the adverse selection context, the *revelation principle* guarantees that the off-equilibrium behavior is irrelevant for finding optimal contracts: one only needs to consider direct revelation mechanisms (Fudenberg and Tirole 1991).

By the one-step-deviation principle (Mailath and Samuelson 2006), we can focus on the *one-step-deviation* strategies, in which the retailer can misreport in the current period, but returns to the truth-telling strategy from $t+1$ onward. For history z^{t-1}, current state z_t, and report $\hat{z}_t$, he has continuation payoff $\tilde{U}_t(z^{t-1}, \hat{z}_t; z_t) = R(q_t(z^{t-1}, \hat{z}_t), z_t) - p_t(z^{t-1}, \hat{z}_t) + \delta \mathbb{E}_{z_{t+1}}[U_{t+1}(z^{t-1}, \hat{z}_t, z_{t+1})|z_t]$. Hence, $U_t(z^t) = U_t(z^{t-1}, z_t) = \tilde{U}_t(z^{t-1}, z_t; z_t)$.

The retailer's strategic reactions entail two classes of constraints. The first is the *participation constraints*: $IR_0(z_0) \equiv U_0(z_0) \geq 0$, which ensure the retailer's participation, by offering him higher payoff than his outside option (normalized to zero). The second class is the *incentive compatibility constraints*: $IC_t(z^t) \equiv U_t(z^t) - \max_{\hat{z}_t} \tilde{U}_t(z^{t-1}, \hat{z}_t; z_t) \geq 0$, which enforce truth telling, by guaranteeing that reporting truthfully is the retailer's best choice: $U_t(z^t) = \max_{\hat{z}_t} \tilde{U}_t(z^{t-1}, \hat{z}_t; z_t)$. As such, the retailer would not entertain off-equilibrium actions—they would never do better. The manufacturer then solves

$$V_0 = \max \left\{ \mathbb{E}\left[\textstyle\sum_{t \geq 0} \delta^t \Big(p_t(z^t) - c \cdot q_t(z^t) \Big) \right] : \quad IR_0(z_0) \geq 0, \quad IC_t(z^t) \geq 0, \forall z^t, \forall t \right\}. \qquad (\mathcal{P}_0)$$

This formulation $(\mathcal{P}_0)$ is fairly general, explicitly modeling arbitrary demand process and intertemporal interactions. It is also more difficult to solve. We tackle it with the relaxation technique. First, we construct a new problem $(\mathcal{P}_1)$ with larger but simpler decision space $\mathcal{F}_1$. Second, we characterize the optimal solution set $\mathcal{F}_1^*$ of $(\mathcal{P}_1)$. Third, we pinpoint solutions in $\mathcal{F}_1^*$ that are also feasible and hence optimal for $(\mathcal{P}_0)$. We execute this idea and detail the solution procedure in the Appendix.

4. How Should the Channel Adapt to Changing Market Conditions?

We first characterize the optimal contract. The key issue is how to set information rent and quantity distortion to ensure credible information sharing. The central notion is market inertia.

THEOREM 1. *For problem $(\mathcal{P}_0)$, the optimal quantity and price are*

$$q_t(z^t) = \underbrace{\frac{z_t - c}{2}}_{\text{efficient quantity}} - \underbrace{\frac{1}{2} \cdot \eta(z_0) \cdot \xi_t(z^t)}_{\text{distortion } \mathcal{D}(z^t)}, \qquad p_t(z^t) = \underbrace{[z_t - q_t(z^t)] \cdot q_t(z^t)}_{\text{sales revenue}} - \underbrace{\Big(U_t(z^t) - \delta \mathbb{E}_{z_{t+1}}[U_{t+1}(z^{t+1})|z_t] \Big)}_{\text{per period rent}}.$$

The information rent for retailer-z^t is $U_t(z^t) = U_t(z^{t-1}, \ell) + \displaystyle\int_{\ell}^{z_t} \mathrm{d}\hat{z}_t \cdot \mathbb{E}\Big[\textstyle\sum_{\tau \geq 0} \delta^\tau \cdot \xi(z_t^{t+\tau}) \cdot q_{t+\tau}(z^{t+\tau}) \; |\hat{z}_t \Big]$.

Dynamic information rent: The manufacturer should pay the dynamic information rents U_t for credible information sharing.[9] The rent arises from information asymmetry and strategic interaction. Consider a high type retailer z_0. He enjoys the advantage of Δz_0-better sales potential over his lower peers $z_0' = z_0 - \Delta z_0$. So for the same quantity q_0, he can command a higher retail price

[9] Morton Kamien: "Economists, unlike lawyers, believe you have to pay people to tell the truth" (Vohra 2011).

and greater profit; e.g., price at $P_0 = (z_0 - q_0) > (z_0' - q_0) = P_0'$, and gain $q_0 \Delta z_0$ more. Unless he is paid a premium, retailer-z_0 would manipulate—act as if he were retailer-z_0'—to pocket in the difference $q_0 \Delta z_0$. To prevent such manipulation, the manufacturer must pay the premium—the information rent U_0.

This requires pricing retailer's advantage properly. Intuitively, the more heterogeneous the market, the severe the information asymmetry, the stronger the manipulation incentive, the larger the information rent. The market (type) heterogeneity is measured by the (inverse) hazard rate $\eta(z_0) \equiv \frac{1-\Lambda_0(z_0)}{\lambda_0(z_0)}$. This is the central idea of *static screening* (Mussa and Rosen 1978). Indeed, when the future does not count ($\delta = 0$), we recover the static case $\mathbb{E}U_0(z_0) = U_0(\ell) + \mathbb{E}[\eta(z_0)q_0(z_0)]$, in which hazard rate η alone is sufficient for pricing retailer's advantage.

In the evolving market, this is no longer the case. The high type retailer-z^t now enjoys both short- and long-term advantages over the lower type (z^{t-1}, z_t') with $z_t' = z_t - \Delta z_t$. The *short-term advantage* is familiar—each period the high type has better sales potential and can gain $q_t \Delta z_t$ more revenue.[10] The *long-term advantage* is new—the high type is $\partial_{z_t} \Lambda_{t+1}(z_{t+1}|z_t) \cdot \Delta z_t$ more likely to have higher sales potential in the next period and beyond. To price these two advantages, the rent U_t must account for *all* future market evolution and retailer's manipulation potentials.

Theorem 1 shows how. Managerially, the rent U_t is the present value of retailer's option to manipulate over time. Technically, the rent U_t is the sum of weighted short-term advantage $q_{t+\tau} \Delta z_{t+\tau}$ in all future periods, where the weight $\xi(z_t^{t+\tau})$ is the market inertia that measures retailer's residual long-term advantage τ periods into the future.

Quantity Distortion: The manufacturer should reduce quantity q_t from the first-best level $(z_t - c)/2$ by distorting downward $\mathcal{D}(z^t) \equiv \eta(z_0)\xi(z^t)/2$. The first-best level fluctuates with the current market condition z_t. The distortion is more involved: it is necessary for *almost all* retailers,[11] and may depend on the entire history $z^t = (z_0, \ldots, z_t)$. Its purpose is to reduce the information rent: by distorting production downward by a small amount Δq from the first best, only a second-order loss in direct profit arises, but a first-order gain in strategic rent reduction is secured.[12]

To get the intuition of *how* the market inertia drives distortion, we now dissect the first order condition for $q_1(z^1)$ (cf. Lemma 3's proof):

$$\underbrace{\left[\delta \cdot \mathbb{P}(z^1) \cdot \left(z_1 - 2q_1(z^1)\right) - c\right]}_{\text{direct revenue effect}} - \underbrace{\delta \cdot [1 - \Lambda_0(z_0)] \cdot [\Lambda_1(z_1'|z_0') - \Lambda_1(z_1'|z_0)]}_{\text{strategic rent effect}} = 0. \tag{2}$$

[10] The gain comes from $\partial_{z_t} R(q_t, z_t) \cdot \Delta z_t = \partial_{z_t}[z_t q_t - (q_t)^2] \cdot \Delta z_t = q_t \Delta z_t$.

[11] We speak of *almost all* in the following sense: Since $\eta(h) = 0$, the retailers in set $H_0 = \{z^t : z_0 = h, t \geq 0\}$ will not distort; they order the first-best instead. But given the continuous nature of market conditions, the set of those retailers is of probability zero; i.e., $\mathcal{Z} = [\ell, h] \subset \mathbb{R}$ implies $\mathbb{P}\{H_0\} = 0$.

[12] The optimal quantity q_t is determined by two trade-offs: (i) the revenue benefit and production cost in channel operation (Li and Gao 2008); (ii) the efficiency loss and rent extraction in strategic interaction.

Intuitively, increasing $q_1(z^1)$ by infinitesimal Δq has two countervailing effects. The first is a direct revenue effect of efficiency gain $[(z_t - 2q_t) - c] \cdot \Delta q$ at z^1 (because of the reduced distortion),which translates into the additional revenue gain at time 0:

$$\delta \cdot \underbrace{\lambda_1(z_1|z_0)\lambda_0(z_0)\Delta z}_{\approx \mathbb{P}(z_0, z_1)} \cdot [(z_1 - 2q_1) - c] \cdot \Delta q. \tag{GAIN}$$

The second effect is the rent increase for ensuring the truth telling of higher type retailers $z_0' > z_0$. This effect is more involved. As the retailer z^1 raises quantity by Δq, to ensure IC_1, his higher counterpart (z_0, z_1') with $z_1' > z_1$ must be paid $\partial_{q_1}[\partial_{z_1} R(q_1, z_1)] \cdot \Delta q = \Delta q$ more for truth telling; but then their predecessor z_0' with $z_0' > z_0$ must be paid $[\Lambda_0(z_1'|z_0) - \Lambda_t(z_1'|z_0')] \cdot \Delta q$ more for truth telling, because retailer-z_0' is $[\Lambda_0(z_1'|z_0) - \Lambda_t(z_1'|z_0')]$ more likely than retailer-z_0 to have rosier future condition $z_1' > z_1$ (by the first order stochastic dominance). Viewed at time 0, there are $\mathbb{P}(z_0' > z_0) = [1 - \Lambda_0(z_0)]$ such retailers. Hence, the total loss is

$$\delta \cdot [1 - \Lambda_0(z_0)] \cdot [\Lambda_1(z_1'|z_0) - \Lambda_1(z_1'|z_0')] \cdot \Delta q. \tag{LOSS}$$

Balancing these two countervailing effects, (GAIN)=(LOSS), we obtain:

$$\frac{1}{2} \cdot \underbrace{[(z_1 - 2q_1(z^1)) - c]}_{\text{marginal channel profit}} = \frac{1}{2} \cdot \underbrace{\frac{1 - \Lambda_0(z_0)}{\lambda_0(z_0)}}_{\text{inv. hazard rate } \eta(z_0)} \cdot \underbrace{\frac{[\Lambda(z_1'|z_0) - \Lambda(z_1'|z_0')]/\Delta z}{\lambda_1(z_1|z_0)}}_{\text{market inertia } \xi(z_0, z_1)} = \frac{1}{2} \cdot \eta(z_0) \cdot \xi(z^1), \tag{3}$$

where the distortion $\frac{1}{2} \cdot \eta(z_0) \cdot \xi(z^1)$ is driven by the product of hazard rate $\eta(z_0)$ and market inertia $\xi(z^1)$.

Intuitively, the distortion is to resolve the tension between revenue gain and rent extraction. In static screening, the tension is across retailer type; so the hazard rate alone is sufficient to determine the distortion $\frac{1}{2} \cdot \eta(z_0)$. In dynamic screening, the tension is far more complicated: it spreads across both type and time. Resolving it requires one to determine the size and *timing* of the distortion. This would entail complex tracking of entire history $(z_0, \ldots, z_t)$ and intertemporal calculation—a daunting task.

Yet Theorem 1 shows this task can get done easily. All one needs is a simple multiplication of market inertia $\xi(z^t)$. This greatly simplifies the contract computation. The simplification is akin to sufficient statistics and factorization theorem: the market inertia compresses the entire history $(z_0, \ldots, z_t)$ into a single term $\xi(z^t)$, and that term enters the contract in the factorization form. In this sense, the market inertia is a "*sufficient statistic*": it summarizes all the relevant market information for contracting.

COROLLARY 1. *For problem ($\mathcal{P}_0$), the complexity of market inertia determines the complexity of the optimal contract.*

5. Policy Implications

Researchers have derived many theoretical insights for channel management. These insights are mainly based on simplistic market behaviors; e.g., static setting, IID demand, and binary type. Yet the further the models are abstracted away from reality, the weaker the credibility of the policy recommendations (Reiss 2011). Indeed, practitioners often question: Can these insights be applied to more realistic contexts?[13]

5.1. The Limitations of Static Screening

The static screening framework has two main predictions (Mussa and Rosen 1978): (a) the low type is always downward distorted; (b) under information asymmetry, channel coordination—the first best—is unattainable. Neither holds in the dynamic settings. The main reason is that, market inertia can change distortion behavior over time, and hence the predictions.

For prediction (a), consider two dynamic scenarios: (i) retailers with $z_0 = h$ (hence $\eta(h) = 0$); (ii) independent market conditions (hence $\xi(z^t) = 0$). In either scenario, the distortion vanishes for $t > 0$, and the retailer-z^t orders the first best $q_t = (z_t - c)/2$, regardless of his current type z_t. Hence prediction (a) fails in dynamic settings.

For prediction (b), the channel cannot coordinate in static settings, because the static distortion is always positive; $\eta(z_0)/2 > 0$ for all $z_0 \in [\ell, h)$. This is no longer the case for dynamic screening. When long-run market inertia converges to zero, the dynamic distortion vanishes—*the channel can achieve coordination* in the long run; see (–▲–) in Figure 1. For example, the linear dynamics $z_{t+1} = \gamma + \alpha_{t+1} z_t + \epsilon_{t+1}$ with $\alpha_{t+1} \in [0, 1)$, has $\lim_t \xi(z^t) = \lim_t(\alpha_1 \alpha_2 \ldots \alpha_t) = 0$, which implies zero distortion eventually $\lim_t \eta(z_0)\xi(z^t)/2 = 0$, and hence the channel coordinates. This overturns prediction (b).

5.2. The "Simple Contract" Puzzle

A contract is *simple* if it is independent of history details (Bolton and Dewatripont 2005, Plambeck and Taylor 2006). The prevalence of simple contracts has long puzzled researchers. The world is complex. Theory predicts that optimal contracts should depend on all the relevant information, hence complex; but channel managers often use simple contracts. When and why can simple contracts be optimal in the complex world?

Lobel and Xiao (2017) make great progress on this issue. For an inventory system, they show that the optimal contact is simple under the dynamic, IID demand. In our context, the IID demand implies: (i) all future periods are *stochastically* similar to the initial period; (ii) retailer's information

[13] These insights still hold if one can recover their premises in practice. For example, for linear dynamics $z_t = \alpha_t z_{t-1} + \epsilon_t$, (i) the static condition arises, when $\alpha_t = 1$ and $\mathbb{E}\epsilon_t = \mathrm{var}(\epsilon_t) = 0$, $\forall t$; (ii) the IID condition prevails, when $\alpha_t = 0$, $\forall t$; (iii) the stationary condition $0 \le \xi_t < 1$ is equivalent to $0 \le \min \alpha_t \le \max \alpha_t < 1$.

advantage (knowing z_0) lasts for only one period. After the initial period, both parties are equally informed of the future—they both forecast future conditions with the same prior Λ_0. As such, distortion is necessary only in the initial period. Hence, the optimal contract is simple.

We model the complexity of the world in the sense of arbitrary demand process. This allows us to offer a new explanation for the "simple contract" puzzle. The central argument is that contracts are firms' best response to market dynamics. Intuitively, the dynamics seems only to complicate contracts, because the time dimension greatly expands contingencies. But this intuition tells only half the story: the dynamics also allows additional flexibility of arranging activities across time, whose effects are fully captured by the market inertia. Thus, the dynamics per se does not determine the contract complexity; it is the market inertia that determines the complexity. As such, simple contracts can and do arise, as the best response to changing environments, even in the world of complex dynamics (autocorrelation). We formalize this explanation below.

PROPOSITION 1. *For problem ($\mathcal{P}_0$), the optimal contract is simple, if the market is* stochastic-linear*: for* $t \geq 0$, $z_{t+1} = \gamma + \alpha_{t+1} z_t + \epsilon_{t+1}$, *where* ϵ_{t+1} *is the random shock.*

Proposition 1 identifies the general statistical property that guarantees the contract simplicity. Indeed, the linear dynamics has market inertia $\xi(z^t) = \alpha_1 \cdots \alpha_t$; the resulting quantity $q_t(z^t) = (z_t - c)/2 - (\alpha_1 \cdots \alpha_t)\eta(z_0)/2$ is independent of history detail $(z_1, z_2, \ldots, z_{t-1})$, and hence simple. With zero slope $\alpha_{t+1} = 0$, the IID demand is a special case of stochastic linearity.

The intuition for Proposition 1 is as follows. The market inertia plays dual roles: empirically, it measures the *marginal effect* of the dynamics (in the time-series model); theoretically, it is also a *sufficient statistic* that determines the contract complexity (in the contract decision model). So, for the optimal contract to be simple, $\xi(z^t)$ should be independent of history details. Technically, the history independence requires that, except the the first order effect, there should be no higher order influences. Hence, the dynamics is necessarily linear.

Proposition 1 is reassuring. For managers, as long as their markets are stochastic-linear, they can focus on simple contracts and safely ignore past observations, without worrying about efficiency loss. More importantly, linear dynamics are common in reality (Dekimpe et al. 2008). As such, we justify the use of simple contracts for a much broader class of channel conditions.

5.3. Beyond Simplistic Market Behaviors

For a channel manager, a key question is *how to adapt* to realistic market conditions. Our model provides such a decision support: from demand data, one can estimate the market inertia, then classify market conditions, and finally craft policy response accordingly.

Market inertia is easy to estimate. It is straightforward if we have estimated the dynamics G_{t+1} from time series analysis. In case the evolution is properly parameterized, we can further reduce data requirement (Gao et al. 2017). For example, if the market condition follows a Markov process (Luo et al. 2016, Chen et al. 2017), we can back out G_{t+1} directly. To illustrate how to classify evolution and formulate policy, we consider the linear dynamics with market inertia $\xi_t = \alpha_1 \cdot \alpha_2 \ldots \alpha_t$.

Case $0 \le \xi_t < 1$: Past shocks have only short-term effect; their long-term effect diminishes over time. Hence, the market inertia vanishes eventually, $\lim_t \xi_t = 0$, resulting in *stable* market condition with a fixed mean and finite variance. Intuitively, this case characterizes business-as-usual, the mature and decline stages of new product sales. Technically, as $t \to \infty$, distortion $\eta(z_0)\xi_t/2 \downarrow 0$, quantity $q_t \uparrow (z_t - c)/2$ (see –☆– in Figure 1), and retail price $P_t \downarrow (z_t + c)/2$. The key policy response is therefore *vanishing distortion*; see (–▲–) in Figure 1.

This case has two implications. (i) Managers should phase out quantity distortion and information rent gradually, because the impact of manipulation diminishes over time. (ii) As distortion fades away, the retailer should reduce retail price, resulting in *decreasing price path*. The literature has explained the deceasing price path by learning curve effect, skim pricing strategy, and competition (Rao 2009). We show it may also emerge endogenously from vanishing distortion and improving efficiency.

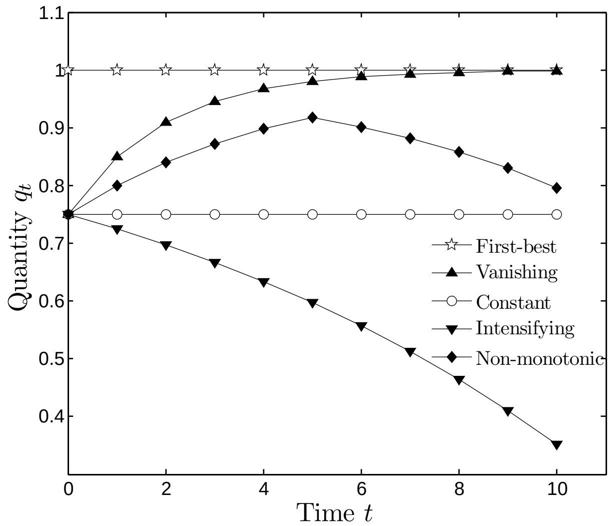
Case $\xi_t = 1$: Each shock has a permanent effect on the future. Hence, the market inertia persists, $\lim_t \xi_t = 1$. Market condition z_t has no fixed mean; its variance increases with time. This case characterizes random walk markets, e.g., $(z_{t+1} - z_t) = \epsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$. The key policy response is *constant distortion*: although quantity $q_t(z^t) = (z_t - c)/2 - \eta(z_0)/2$ still responds to current market condition z_t through the first best level $(z_t - c)/2$, its distortion $\eta(z_0)/2$ should remain constant; see (–○–) in Figure 1.

Case $\xi_t > 1$: Past shocks become increasingly more important for the future. This case characterizes exponential growth conditions, such as viral market phenomena and rapid growth stage of new product sales. They can be driven by the word-of-mouth contagion, network effects, and hysteresis effect. The key policy response is *intensifying distortion*. Intuitively, because market inertia $\xi_t > 1$, the impact of early manipulation—retailer's advantage—amplifies over time. To ensure truth telling in *early* periods, the manufacturer must neutralize the amplified advantage by intensifying future distortion $\eta(z_0)\xi_t/2$; see (–▼–) in Figure 1.

Our policy recommendations complement existing ones. For example, Battaglini (2005) studies the binary model with Markov demand. He establishes two main principles for the stable type process (i.e., $0 \le \xi_t < 1$): a generalized no distortion at the top (GNDT), and the vanishing distortion at the bottom (VDB). Two policy recommendations are: (i) the distortion should be limited to consistent-low types only, $\ell^t = (\ell, \ell, \ldots, \ell)$; (ii) the distortion should diminish over time.

Figure 1 **Expected order quantity $\mathbb{E}q_t$ for linear dynamics:** $c = 0.5$, $\eta(z_0) = 0.5$, $\epsilon_t \sim \mathcal{N}(0, 0.3^2)$, $z_t = 2$, $\forall t$

Note: Vanishing distortion $\alpha_t = 0.6$; Constant distortion (second best) $\alpha_t = 1$; Intensifying distortion $\alpha_t = 1.1$; Non-monotonic distortion $\alpha_t = 0.8$ for $t < 5$, and $\alpha_t = 1.2$ for $t \geq 5$.



Our work differs from the binary model in two substantive ways. The first is applicability. Real markets are often complex: market phenomena, such as viral products and random walk, are continuous and nonstationary. The binary model cannot capture these complexities of real markets. In contrast, our model builds on arbitrary demand and thus forges the direct link to reality. This difference is critical to channel managers: given the continuous nature of market conditions $z_t \in [\ell, h] \subset \mathbb{R}_+$, the very concept of high or low type is often ill-defined, let alone implementation.

The second substantive difference is in policy recommendations. (i) The binary model prescribes the first best quantity for the majority of retailers, and limits the distortion to types ℓ^t only. Our model prescribes the opposite: in a continuous market, almost all retailers should distort quantities, and only retailers with $z_0 = h$ order the first best.[14] This reverse is driven by the increased complexity to ensure IC constraints. In the binary model, the manufacturer faces only two types; to ensure IC, distorting the ℓ-type is sufficient. In the continuous model, however, the manufacturer faces multiple types in $[\ell, h]$. To ensure IC, he must ensure truthtelling from *all* types above ℓ, i.e., $z_0 \in (\ell, h]$. This requires far more refined responses, resulting in pervasive distortion.

(ii) The binary model prescribes *vanishing distortion* over time. As we have shown, this is the right policy for the stable market conditions ($0 \leq \xi_t < 1$). However, when markets follow random walk ($\xi_t = 1$) or grow exponentially over time ($\xi_t > 1$), the distortion need not vanish. In these markets, to neutralize the escalated advantage of the retailer, the manufacturer should respond in kind, *intensifying distortion* over time. Indeed, the distortion should move in lockstep with the

[14] Given the continuous space $\mathcal{Z} = [\ell, h]$, the set of retailers who order the first best is of probability zero: $\mathbb{P}\{z^t : z_0 = h, t \geq 0\} = 0$.

market inertia: when the market is expecting *regime changes* (Hamilton 2010), the distortion can even follow non-monotonic patterns; see (–♦–) in Figure 1.

In summary, our results call for caution when applying the channel insights based on simplistic market assumptions. Indeed, real market conditions are often complex, requiring far more subtle responses.[15]

6. Concluding Remarks

Real markets evolve over time. Their behaviors are central to channel management. A key issue is how these behaviors affect channel contracting. The existing theoretical studies mainly focus on simplistic market behaviors, which often limit their applicability. This paper puts realistic market behaviors—autocorrelation, continuity, and nonstationarity—on the center stage. We develop a general framework that captures market evolution, information asymmetry, and strategic interactions over time. To overcome the technical challenge, we transform the problem into one defined on a higher dimension. Using this relaxation technique, we are able to fully characterize the optimal contract under arbitrary market evolution—the most general case in the channel literature.

We develop the notion of market inertia that links time-series data to contracting. Market inertia plays three roles: price retailer's information advantage, set contract response to evolving market conditions, and determine the contract complexity. Using market inertia, we identify a general property—stochastic linearity—that justifies the use of simple contracts in a much wider range of channel conditions.

For practitioners, we offer three refined prescriptions: linear market dynamics requires only simple contracts; continuous markets entail pervasive quantity distortion; nonstationary markets call for subtler responses (e.g., the distortion can be vanishing, constant, intensifying, or even non-monotonic over time). By highlighting the central role of realistic market behaviors, we take a step closer to a better understanding of channel theory and practice.

Acknowledgements

We thank Miguel Villas-Boas (the department editor) and the review team for their insightful comments and suggestions, which have significantly improved the paper. We have also benefited from the discussions with Elodie Adida, Anil Arya, Volodymyr Babich, Saif Benjaafar, Xiuli Chao, Li Chen, Ying-Ju Chen, Leon Chu, Charles Corbett, Tony Haitao Cui, Lingxiu Dong, Fernando Bernstein, Liang Guo, Mehmet Gumus, Ozalp Ozer, Ashutosh Prasad, Jeannette Song, Brian Tomlin, Rajeev Tyagi, and Shuya Yin. All remaining errors are our own.

[15] Albert Einstein: "Everything should be made as simple as possible, *but not simpler*."

References

Bass, Frank M. 1995. Empirical generalizations and marketing science: A personal view. *Marketing Science* **14**(3_supplement) G6–G19.

Battaglini, M. 2005. Long-term contracting with markovian consumers. *American Economic Review* 637–658.

Bergemann, D., M. Said. 2011. Dynamic auctions. *Wiley Encyclopedia of Operations Research and Management Science* .

Bergemann, Dirk, Juuso Välimäki. 2010. The dynamic pivot mechanism. *Econometrica* **78**(2) 771–789.

Bernstein, Fernando, Victor Martínez-de Albéniz. 2016. Dynamic product rotation in the presence of strategic customers. *Management Science* .

Bolton, P., M. Dewatripont. 2005. *Contract theory*. MIT press.

Box, George EP, Gwilym M Jenkins, Gregory C Reinsel. 2011. *Time series analysis: forecasting and control*, vol. 734. Wiley.

Bronnenberg, Bart J, Peter E Rossi, Naufel J Vilcassim. 2005. Structural modeling and policy simulation. *Journal of Marketing Research* **42**(1) 22–26.

Cachon, G. P. 2003. Supply chain coordination with contracts. *Handbooks in operations research and management science* **11** 227–339.

Chen, Li, Jing-Sheng Song, Yue Zhang. 2017. Serial inventory systems with markov-modulated demand: Derivative bounds, asymptotic analysis, and insights. *Operations Research* **65**(5) 1231–1249.

Chiang, Wei-yu Kevin. 2012. Supply chain dynamics and channel efficiency in durable product pricing and distribution. *Manufacturing & Service Operations Management* **14**(2) 327–343.

Chintagunta, Pradeep, Tülin Erdem, Peter E Rossi, Michel Wedel. 2006. Structural modeling in marketing: Review and assessment. *Marketing Science* **25**(6) 604–616.

Chintagunta, Pradeep K, Dipak Jain. 1992. A dynamic model of channel member strategies for marketing expenditures. *Marketing Science* **11**(2) 168–188.

Chu, Wujin. 1992. Demand signalling and screening in channels of distribution. *Marketing Science* **11**(4) 327–347.

Courty, Pascal, Li Hao. 2000. Sequential screening. *The Review of Economic Studies* **67**(4) 697–717.

Cui, Tony Haitao, Jagmohan S Raju, Z John Zhang. 2007. Fairness and channel coordination. *Management Science* 1303–1314.

Deb, Rahul, Maher Said. 2015. Dynamic screening with limited commitment. *Journal of Economic Theory* .

Dekimpe, Marnik, DM Hanssens. 1995. Empirical generalizations about market evolution. *24th Annual EMAC Conference*. Ecole Supérieure des Sciences Economiques et Commerciales (ESSEC).

Dekimpe, Marnik G, Philip Hans Franses, Dominique M Hanssens, Prasad A Naik. 2008. Time-series models in marketing. *Handbook of marketing decision models*. Springer, 373–398.

Dekimpe, Marnik G, Dominique M Hanssens. 2003. Persistence modeling for assessing marketing strategy performance .

Desai, Preyas S. 2000. Multiple messages to retain retailers: Signaling new product demand. *Marketing Science* **19**(4) 381–389.

Desai, Preyas S, Kannan Srinivasan. 1995. Demand signalling under unobservable effort in franchising: Linear and nonlinear price contracts. *Management Science* **41**(10) 1608–1623.

Dukes, Anthony, Esther Gal-Or, Tansev Geylani. 2011. Who benefits from bilateral information exchange in a retail channel? *Economics Letters* **112**(2) 210–212.

Eliashberg, Jehoshua, Abel P Jeuland. 1986. The impact of competitive entry in a developing market upon dynamic pricing strategies. *Marketing Science* **5**(1) 20–36.

Eső, Péter, Balazs Szentes. 2007. Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies* **74**(3) 705–731.

Fernandes, Ana, Christopher Phelan. 2000. A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory* **91**(2) 223–247.

Franses, Philip Hans. 2005. On the use of econometric models for policy simulation in marketing. *Journal of Marketing Research* **42**(1) 4–14.

Fudenberg, Drew, Jean Tirole. 1991. *Game theory*. MIT Press.

Gal-Or, Esther, Tansev Geylani, Anthony J Dukes. 2008. Information sharing in a channel with partially informed retailers. *Marketing Science* **27**(4) 642–658.

Gao, Long. 2015. Long-term contracting: The role of private information in dynamic supply risk management. *Production and Operations Management* **24**(10) 1570–1579.

Gao, Long, Zhaolin Li, Biying Shou. 2014. Information acquisition and voluntary disclosure in an export-processing system. *Production and Operations Management* **23**(5) 802–816.

Gao, Long, Junmin Shi, Michael Gorman, Ting Luo. 2018. Business analytics for intermodal capacity management. *UCR working paper* .

Gao, Long, Susan H Xu, Michael O Ball. 2012. Managing an available-to-promise assembly system with dynamic short-term pseudo-order forecast. *Management Science* **58**(4) 770–790.

Gao, Long, Nan Yang, Renyu Zhang, Ting Luo. 2017. Dynamic supply risk management with signal-based forecast, multi-sourcing, and discretionary selling. *Production and Operations Management* **26**(7) 1299–1415.

Gümüş, Mehmet. 2017. A tale of two information asymmetries in competitive supply chains. *Handbook of Information Exchange in Supply Chain Management*. Springer, 189–213.

Guo, Liang. 2009. The benefits of downstream information acquisition. *Marketing Science* **28**(3) 457–471.

Guo, Liang, Ganesh Iyer. 2010. Information acquisition and sharing in a vertical relationship. *Marketing Science* **29**(3) 483–506.

Hamilton, James D. 2010. Regime switching models. *Macroeconometrics and time series analysis*. Springer, 202–209.

Hamilton, James Douglas. 1994. *Time series analysis*, vol. 2. Princeton university press Princeton.

He, Chuan, Johan Marklund, Thomas Vossen. 2008. Research note-vertical information sharing in a volatile market. *Marketing Science* **27**(3) 513–530.

He, Long, Ying Rong, Zuo-Jun Max Shen. 2017. Product geographical distribution under recall risk .

Ho, Teck-Hua, Juanjuan Zhang. 2008. Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Management Science* **54**(4) 686–700.

Iyer, Ganesh. 1998. Coordinating channels under price and nonprice competition. *Marketing Science* **17**(4) 338–355.

Iyer, Ganesh, J Miguel Villas-Boas. 2003. A bargaining theory of distribution channels. *Journal of Marketing Research* 80–100.

Jeuland, Abel P, Steven M Shugan. 1983. Managing channel profits. *Marketing science* **2**(3) 239–272.

Jiang, Baojun, Lin Tian, Yifan Xu, Fuqiang Zhang. 2016. To share or not to share: Demand forecast sharing in a distribution channel. *Marketing Science* **35**(5) 800–809.

Jørgensen, Steffen. 1986. Optimal production, purchasing and pricing: A differential game approach. *European Journal of Operational Research* **24**(1) 64–76.

Jørgensen, Steffen, Georges Zaccour. 2003. Channel coordination over time: Incentive equilibria and credibility. *Journal of Economic Dynamics and Control* **27**(5) 801–822.

Katok, Elena, Tava Olsen, Valery Pavlov. 2014. Wholesale pricing under mild and privately known concerns for fairness. *Production and Operations Management* **23**(2) 285–302.

Klibanoff, Peter, Jonathan Morduch. 1995. Decentralization, externalities, and efficiency. *The Review of Economic Studies* **62**(2) 223–247.

Kolay, Sreya, Greg Shaffer. 2013. Contract design with a dominant retailer and a competitive fringe. *Management Science* .

Kuribko, Nataliya, Tracy Lewis, Fang Liu, Jing-Sheng Song. 2018. Long term partnership for achieving efficient capacity allocation. *Operations Research,* forthcoming .

Kuzu, Kaan, Long Gao, Susan H. Xu. 2018. To wait or not to wait: The theory and practice of ticket queues. *Manufacturing & Service Operations Management,* forthcoming .

Laffont, J.J., D. Martimort. 2001. *The theory of incentives: the principal-agent model*. Princeton University Press.

Li, Z., L. Gao. 2008. The effects of sharing upstream information on product rollover. *Production and Operations Management* **17**(5) 522–531.

Lobel, Ilan, Wenqiang Xiao. 2017. Optimal long-term supply contracts with asymmetric demand information. *Operations Research* .

Luo, Ting, Long Gao, Yalçın Akçay. 2016. Revenue management for intermodal transportation: the role of dynamic forecasting. *Production and Operations Management* **25**(10) 1658–1672.

Mailath, George J, Larry Samuelson. 2006. *Repeated games and reputations: long-run relationships*. Oxford University Press, USA.

McAfee, R Preston. 2009. *Competitive solutions: the strategist's toolkit*. Princeton University Press.

Mishra, Birendra K, Ashutosh Prasad. 2005. Delegating pricing decisions in competitive markets with symmetric and asymmetric information. *Marketing Science* **24**(3) 490–497.

Mittendorf, Brian, Jiwoong Shin, Dae-Hee Yoon. 2013. Manufacturer marketing initiatives and retailer information sharing. *Quantitative Marketing and Economics* **11**(2) 263–287.

Mookherjee, Dilip. 2006. Decentralization, hierarchies, and incentives: A mechanism design perspective. *Journal of Economic Literature* 367–390.

Moorthy, K Sridhar. 1987. Managing channel profits: Comment. *Marketing Science* **6**(4) 375–379.

Mussa, Michael, Sherwin Rosen. 1978. Monopoly and product quality. *Journal of Economic theory* **18**(2) 301–317.

Myerson, Roger B. 1986. Multistage games with communication. *Econometrica: Journal of the Econometric Society* 323–358.

Özer, Ö., Y. Zheng, K.Y. Chen. 2011. Trust in forecast information sharing. *Management Science* **57**(6) 1111–1137.

Padmanabhan, V, Ivan PL Png. 1997. Manufacturer's return policies and retail competition. *Marketing Science* **16**(1) 81–94.

Pauwels, Koen, Imran Currim, Marnik G Dekimpe, Dominique M Hanssens, Natalie Mizik, Eric Ghysels, Prasad Naik. 2004. Modeling marketing dynamics by time series econometrics. *Marketing Letters* **15**(4) 167–183.

Pavan, Alessandro, Ilya Segal, Juuso Toikka. 2014. Dynamic mechanism design: A myersonian approach. *Econometrica* **82**(2) 601–653.

Plambeck, Erica L, Terry A Taylor. 2006. Partnership in a dynamic production system with unobservable actions and noncontractible output. *Management Science* **52**(10) 1509–1527.

Rao, Vithala R. 2009. *Handbook of pricing research in marketing*. Edward Elgar Publishing.

Reiss, Peter C. 2011. Structural workshop paper—descriptive, structural, and experimental empirical methods in marketing research. *Marketing Science* **30**(6) 950–964.

Reiss, Peter C, Frank A Wolak. 2007. Structural econometric modeling: Rationales and examples from industrial organization. *Handbook of econometrics* **6** 4277–4415.

Shugan, Steven M. 1985. Implicit understandings in channels of distribution. *Management Science* **31**(4) 435–460.

Sudhir, K, Sumon Datta. 2008. Pricing in marketing channels. *Handbook of Pricing (V. Rao Ed.), Edward Elgar Publishers* .

Tellis, Gerard J. 2006. Modeling marketing mix. *Handbook of marketing research* 506–522.

Tirole, Jean. 1988. *The Theory of Industrial Organization: Jean Tirole*. MIT press.

Villas-Boas, J Miguel. 1998. Product line design for a distribution channel. *Marketing Science* **17**(2) 156–169.

Villas-Boas, J Miguel. 1999. Dynamic competition with customer recognition. *The Rand Journal of Economics* 604–631.

Vohra, Rakesh V. 2011. *Mechanism Design: a linear programming approach*, vol. 47. Cambridge University Press.

Vohra, R.V. 2012. Dynamic mechanism design. *Surveys in Operations Research and Management Science* **17**(1) 60–68.

Wittink, Dick R. 2005. Econometric models for marketing decisions. *Journal of Marketing Research* **42**(1) 1–3.

Supplemental File

Table 1 Notation

c	unit production cost
δ	discount factor, $\delta \in (0,1)$
z_t	market condition in period t, $z_t \in \mathcal{Z}$, where $\mathcal{Z} \equiv [\ell, h]$
z^t	the history from 0 to t; $z^t \equiv (z_0, \ldots, z_t) \in \mathcal{Z}^t$, where $\mathcal{Z}^t \equiv \times_{s=0}^t \mathcal{Z}$
z_s^t	the history from s to t; $z_s^t = (z_s, z_{s+1}, \ldots, z_t)$
$\Lambda_0(z_0)$	the prior (cdf) of z_0, with pdf λ_0
$\eta(z_0)$	the inverse hazard rate of z_0, $\eta(z_0) = \frac{1-\Lambda_0(z_0)}{\lambda_0(z_0)}$
$[f]^{-1}$	inverse of function $f \in \mathcal{Y}^{\mathcal{X}}$; $[f]^{-1}(y) \equiv \{ x \in \mathcal{X} : f(x) = y \}$, $\forall y \in \mathcal{Y}$
$(q_t)_t$	sequence $\{q_t : 0 \le t \le T\}$
$\partial_x f(x,y)$	the derivative of function f with respect to x
$\mathbb{E}_{z_t^s}[f(z_t^s) \vert z_t]$	conditional expectation with respect to random vector $z_t^s = (z_t, z_{t+1}, \ldots, z_s)$ given z_t
$\Lambda_{t+1}(z'_{t+1} \vert z'_t)$	the cumulative transition probability $\mathbb{P}(z_{t+1} \le z'_{t+1} \vert z_t = z'_t)$, with density $\Lambda_{t+1}(z'_{t+1} \vert z'_t)$

Lemma 1 and Its Proof

The following technical lemma is critical for characterizing the optimal contract.

LEMMA 1. *(i) For the process $z_{t+1} = G_{t+1}(z_t, \epsilon_{t+1})$, with differentiable transition probability $\Lambda_{t+1}(z_{t+1}|z_t)$ and density $\lambda_{t+1}(z_{t+1}|z_t)$, we have*

$$\partial_{z_t} G_{t+1}(z_t, \epsilon) = \partial_{z_t}\left([\Lambda_{t+1}]^{-1}(\epsilon|z_t)\right) = -\frac{\partial_{z_t}\Lambda_{t+1}(z_{t+1}|z_t)}{\lambda_{t+1}(z_{t+1}|z_t)}. \tag{4}$$

(ii) For $\alpha_{t+1} \ge 0$, the process $z_{t+1} = \alpha_{t+1} z_t + \epsilon_{t+1}$, has market inertia $\xi(z_t, z_{t+1}) = \alpha_{t+1}$, $\xi_t(z^t) = \prod_{s \le t} \alpha_s$. It satisfies first order stochastic dominance, with $\partial_{z_t}\Lambda(z_{t+1}|z_t) = -\alpha_{t+1} \cdot \lambda(z_{t+1}|z_t)$.

Proof: **Part (i).** To ease notation, we suppress time index from Λ_{t+1} and λ_{t+1}, i.e., $\Lambda = \Lambda_t$, and $\lambda = \lambda_t$, $\forall t$.

For two consecutive states (z_t, z_{t+1}), let $\Lambda(z_{t+1}|z_t) = \epsilon \in [0,1]$. By definition, $\Lambda(\cdot|z_t)$ is increasing in z_{t+1}. Hence, its inverse function is well defined, and

$$[\Lambda]^{-1}(\epsilon|z_t) = z_{t+1} = G(z_t, \epsilon). \tag{5}$$

Substituting it into $\Lambda(z_{t+1}|z_t) = \epsilon$, we have

$$\Lambda\Big([\Lambda]^{-1}(\epsilon|z_t)\Big|z_t\Big) = \epsilon.$$

This implies that $\phi(z_t) \equiv \Lambda\Big([\Lambda]^{-1}(\epsilon|z_t)\Big|z_t\Big) = \epsilon$ is independent of z_t. Taking total derivative of $\phi(z_t) = \epsilon$, by the chain rule, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z_t}\phi(z_t) &= \lambda\Big([\Lambda]^{-1}(\epsilon|z_t)\Big|z_t\Big) \cdot \partial_{z_t}[\Lambda]^{-1}(\epsilon|z_t) \ + \ \partial_{z_t}\Lambda\Big([\Lambda]^{-1}(\epsilon|z_t)\Big|z_t\Big) \\ &= \lambda(z_{t+1}\big|z_t) \cdot \partial_{z_t}[\Lambda]^{-1}(\epsilon|z_t) \ + \ \partial_{z_t}\Lambda(z_{t+1}|z_t) \\ &= \partial_{z_t}\epsilon = 0,\end{aligned}$$

which implies:

$$\partial_{z_t} G(z_t, \epsilon) = \partial_{z_t}\left([\Lambda]^{-1}(\epsilon|z_t)\right) = -\frac{\partial_{z_t}\Lambda(z_{t+1}|z_t)}{\lambda(z_{t+1}|z_t)}.$$

Part (ii). By part (i), we have

$$\frac{\partial_{z_t}\Lambda(z_{t+1}|z_t)}{\lambda(z_{t+1}|z_t)} = -\partial_{z_t}\left([\Lambda]^{-1}(\epsilon|z_t)\right) = -\partial_{z_t}\left(z_{t+1}\right) = -\partial_{z_t}\left(\alpha_{t+1}z_t + \epsilon\right) = -\alpha_{t+1}.$$

By Eq. (6), $\partial_{z_t}\Lambda(z_{t+1}|z_t) \le 0$, hence the first order stochastic dominance follows. ■

Proof of Theorem 1

We use the relaxation technique to solve $(\mathcal{P}_0)$. First, we construct a relaxed problem $(\mathcal{P}_1)$ with a simpler feasible set $\mathcal{F}_1 \supset \mathcal{F}_0$. Second, we characterize the optimal solution set $\mathcal{F}_1^*$ of $(\mathcal{P}_1)$. Third, we verify a solution in $\mathcal{F}_1^*$ is also feasible for $(\mathcal{P}_0)$, hence it must be optimal for $(\mathcal{P}_0)$.

Formally, for problem $(\mathcal{P}_i): V_i = \max_{\mathcal{F}_i} \tilde{V}$, $i = 0, 1$, let $\mathcal{F}_i$ and $\mathcal{F}_i^*$ be its *feasible* and *optimal* sets. Clearly, $\mathcal{F}_0 \equiv \{(p_t, q_t)_t : IR_t, IC_t, \forall t\}$. Then $\mathcal{F}_0 \subset \mathcal{F}_1$ implies $V_0 \leq V_1$ and

$$\mathcal{F}_1^* \cap \mathcal{F}_0 \subset \mathcal{F}_0^*.$$

We execute this idea in three lemmas: relaxation Lemma 2, Characterization Lemma 3, and Verification Lemma 4. First, we construct a simpler set $\mathcal{F}_1$ by replacing set $\mathcal{F}_0$'s IC_t constraints with *necessary* derivative condition:

LEMMA 2. **(Relaxation)** *If the contract* $(p_t, q_t)_t$ *is sequentially incentive compatible, then*
(i) $U_t(\hat{z}^{t-1}, z_t)$ *increases in* z_t.
(ii) $U_t(\hat{z}^{t-1}, z_t)$ *is differentiable in* z_t *almost everywhere, with*

$$\partial_{z_t} U_t(\hat{z}^{t-1}, z_t) = \textstyle\sum_{\tau \geq 0} \mathbb{E}_{z_{t+1}^{t+\tau}} \left[\delta^\tau \cdot \xi\big(z_t^{t+\tau}\big) \cdot q_{t+\tau}(\hat{z}^{t-1}, z_t, z_{t+1}^{t+\tau}) | z_t \right]. \tag{6}$$

Let $\mathcal{F}_1 \equiv \{(p_t, q_t)_t : IR_0,\ Eq.(6), \forall t\}$. Hence, $\mathcal{F}_0 \subset \mathcal{F}_1$. This relaxation is critical, because after changing decision variables from $(p_t, q_t)_t$ to $(U_t, q_t)_t$, we can use the derivative condition (6) to express $(\mathcal{P}_1)$ by quantities $(q_t)_t$ alone. This allows a simple characterization:

LEMMA 3. **(Characterization)** *For* $(p_t, q_t)_t \in \mathcal{F}_1^*$, *we have* $q_t(z^t)$ *increases in each* z_s, $s = 0, \ldots, t$, *and*

$$q_t(z^t) = \frac{z_t - c}{2} - \frac{1}{2} \cdot \eta(z_0) \cdot \xi\big(z^t\big), \tag{7}$$

$$p_t(z^t) = [z_t - q_t(z^t)] \cdot q_t(z^t) - \Big(U_t(z^t) - \delta \mathbb{E}[U_{t+1}(z^{t+1}) | z_t]\Big), \tag{8}$$

where

$$U_t(z^t) = U_t(z^{t-1}, \ell) + \int_\ell^{z_t} \Big\{ \textstyle\sum_{\tau \geq 0} \mathbb{E}_{z_{t+1}^{t+\tau}} \left[\delta^\tau \cdot \xi(\hat{z}_t, z_{t+1}^{t+\tau}) \cdot q_{t+\tau}(z^{t-1}, \hat{z}_t, z_{t+1}^{t+\tau}) \big| \hat{z}_t \right] \Big\} \cdot \mathrm{d}\hat{z}_t. \tag{9}$$

Finally, we argue that $(q_t, q_t)_t \in \mathcal{F}_1^*$ is indeed incentive compatible and hence optimal.

LEMMA 4. **(Verification)** *The solution* $(p_t, q_t)_t$ *in* (7)–(9) *is feasible and hence optimal for* $(\mathcal{P}_0)$.

The results then follow immediately. The proofs of these lemmas are next. ■

Proof of Lemma 2

We proceed in two steps. To ease notation, we suppress time index from Λ_t and λ_t.

STEP 1: We prove parts (i) and (ii) together by backward induction.

For the last period T, we have a canonical, static screening problem (Fudenberg and Tirole 1991), with the incentive compatibility constraint $U_T(\hat{z}^{T-1}, z_T) = \max_{\hat{z}_T} \tilde{U}_T(\hat{z}^{T-1}, \hat{z}_T; z_T)$, where

$$\tilde{U}_T(\hat{z}^{T-1}, \hat{z}_T; z_T) = \big(z_T - q_T(\hat{z}^T)\big) q_T(\hat{z}^T) - p_T(\hat{z}^T).$$

Clearly, $\tilde{U}_T(\hat{z}^T; z_T)$ increases in z_T, and hence differentiable almost everywhere (a.e.) in z_T, with $\partial_{z_T}\tilde{U}_T(\hat{z}^T; z_T) = q_T(\hat{z}^T)$. By the Envelope Theorem, $U_T(z^{T-1}, z_T)$ is also differentiable in z_T almost everywhere. Thus, parts (i) and (ii) hold for period T.

Suppose parts (i) and (ii) hold for $t+1$. We show they also hold for period t. Since U_{t+1} is differentiable a.e., by Fubini's Theorem, we have

$$\mathbb{E}[U_{t+1}(\hat{z}^t, z_{t+1})|z_t] = \int_{\mathcal{Z}} \partial_{z_{t+1}} U_{t+1}(\hat{z}^t, z_{t+1})[1-\Lambda(z_{t+1}|z_t)]\,\mathrm{d}z_{t+1}, \tag{10}$$

which follows from

$$\begin{aligned}
\mathbb{E}[U_{t+1}(\hat{z}^t, z_{t+1})|z_t] &= \int_{\mathcal{Z}} \Lambda(\,\mathrm{d}w|z_t)\cdot U_{t+1}(\hat{z}^t, w) = \int_\ell^h \lambda(w|z_t)\,\mathrm{d}w \int_\ell^w \partial_v U_{t+1}(\hat{z}^t, v)\,\mathrm{d}v \\
&= \int_\ell^h\int_\ell^w \partial_v U_{t+1}(\hat{z}^t, v)\lambda(w|z_t)\,\mathrm{d}v\,\mathrm{d}w = \int_\ell^h\int_v^h \partial_v U_{t+1}(\hat{z}^t, v)\lambda(w|z_t)\,\mathrm{d}w\,\mathrm{d}v \quad \text{Fubini's theorem} \\
&= \int_\ell^h \partial_v U_{t+1}(\hat{z}^t, v)\,\mathrm{d}v \int_v^h \lambda(w|z_t)\,\mathrm{d}w = \int_\ell^h \partial_v U_{t+1}(\hat{z}^t, v)[1-\Lambda(v|z_t)]\,\mathrm{d}v.
\end{aligned}$$

Hence,

$$\tilde{U}_t(\hat{z}^t; z_t) = [z_t - q_t(\hat{z}^t)]q_t(\hat{z}^t) - p_t(\hat{z}^t) + \delta\int_{\mathcal{Z}} \partial_{z_{t+1}} U_{t+1}(\hat{z}^t, z_{t+1})\cdot[1-\Lambda(z_{t+1}|z_t)]\cdot\mathrm{d}z_{t+1}. \tag{11}$$

Incentive compatibility implies that, for $z_t'' > z_t'$,

$$\begin{aligned}
&U_t(\hat{z}^{t-1}, z_t'') - U_t(\hat{z}^{t-1}, z_t') \\
\geq\ & \tilde{U}_t(\hat{z}^{t-1}, z_t'; z_t'') - \tilde{U}_t(\hat{z}^{t-1}, z_t'; z_t') \quad \text{by incentive compatibility} \\
=\ & \underbrace{[z_t'' - z_t']\cdot q_t(\hat{z}^{t-1}, z_t')}_{\geq 0} + \delta\int_{\mathcal{Z}} \underbrace{\partial_{z_{t+1}} U_{t+1}(\hat{z}^{t-1}, z_t', z_{t+1})}_{\geq 0}\cdot\underbrace{[\Lambda(z_{t+1}|z_t') - \Lambda(z_{t+1}|z_t'')]}_{\geq 0}\cdot\mathrm{d}z_{t+1} \quad \text{by Eq.(11)} \\
\geq\ & 0,
\end{aligned}$$

where the last inequality holds because $z_t'' > z_t'$, U_{t+1} increases in z_{t+1}, and $\Lambda(\cdot|z_t)$ is stochastic increasing in z_t. Hence, $U_t(\hat{z}^{t-1}, z_t)$ increases in z_t. This completes the induction step for part (i).

For part (ii), at any differentiable point z_t, by hypothesis and the definition of $\tilde{U}_t$, we have

$$\begin{aligned}
&\partial_{z_t}\tilde{U}_t(\hat{z}^t; z_t) \\
=\ & \partial_{z_t}\Big(\big[(z_t - q_t(\hat{z}^t))q_t(\hat{z}^t)\big] - p_t(\hat{z}^t)\Big) + \delta\partial_{z_t}\int_{\mathcal{Z}} \Lambda(\,\mathrm{d}w|z_t)\cdot U_{t+1}(\hat{z}^t, w) \\
=\ & q_t(\hat{z}^t) + \delta\partial_{z_t}\int_\ell^h \partial_w U_{t+1}(\hat{z}^t, w)[1-\Lambda(w|z_t)]\,\mathrm{d}w \qquad \text{by Fubini's Theorem, Eq. (10)} \\
=\ & q_t(\hat{z}^t) + \delta\int_\ell^h \partial_{z_t}\left\{\partial_w U_{t+1}(\hat{z}^t, w)[1-\Lambda(w|z_t)]\right\}\mathrm{d}w \quad \text{Lebesgue's Dominated Convergence Theorem} \\
=\ & q_t(\hat{z}^t) + \delta\int_\ell^h \partial_w U_{t+1}(\hat{z}^t, w)\cdot(-\partial_{z_t}\Lambda(w|z_t))\cdot\mathrm{d}w \\
=\ & q_t(\hat{z}^t) + \delta\int_\ell^h \partial_w U_{t+1}(\hat{z}^t, w)\cdot\left(\frac{-\partial_{z_t}\Lambda(w|z_t)}{\lambda(w|z_t)}\right)\cdot\lambda(w|z_t)\,\mathrm{d}w \\
=\ & q_t(\hat{z}^t) + \delta\int_\ell^h \partial_w U_{t+1}(\hat{z}^t, w)\cdot\xi(z_t, w)\cdot\lambda(w|z_t)\,\mathrm{d}w \qquad \text{by Lemma 1} \\
=\ & q_t(\hat{z}^t) + \delta\int_{\mathcal{Z}} \xi(z_t^{t+1})\cdot\partial_{z_{t+1}} U_{t+1}(\hat{z}^t, z_{t+1})\cdot\Lambda(\,\mathrm{d}z_{t+1}|z_t).
\end{aligned}$$

Hence, we have

$$\partial_{z_t}\tilde{U}_t(\hat{z}^t;z_t) = q_t(\hat{z}^t) + \delta\mathbb{E}_{z_{t+1}}[\xi(z_t^{t+1})\cdot\partial_{z_{t+1}}U_{t+1}(\hat{z}^t,z_{t+1})|z_t]. \tag{12}$$

Applying Eq. (12) inductively, we have

$$\begin{aligned}
\partial_{z_t}\tilde{U}_t(\hat{z}^t;z_t) &= q_t(\hat{z}^t) + \delta\mathbb{E}[\xi(z_t^{t+1})\cdot\partial_{z_{t+1}}U_{t+1}(\hat{z}^t,z_{t+1})|z_t] \\
&= q_t(\hat{z}^t) + \delta\mathbb{E}\left[\xi(z_t^{t+1})\cdot\Big(q_{t+1}(\hat{z}^t,z_{t+1}) + \delta\mathbb{E}[\xi(z_{t+1}^{t+2})\cdot\partial_{z_{t+2}}U_{t+2}(\hat{z}^t,z_{t+1},z_{t+2})|z_{t+1}]\Big)\Big|z_t\right] \quad \text{by (12)} \\
&= q_t(\hat{z}^t) + \delta\cdot\mathbb{E}[\xi(z_t^{t+1})\cdot q_{t+1}(\hat{z}^t,z_{t+1})|z_t] + \delta^2\cdot\mathbb{E}[\xi(z_t^{t+1})\cdot\xi(z_{t+1}^{t+2})\cdot\partial_{z_{t+2}}U_{t+2}(\hat{z}^t,z_{t+1}^{t+2})|z_t] \\
&\;\;\vdots \\
&= q_t(\hat{z}^t) + \textstyle\sum_{\tau\geq 1}\delta^\tau\cdot\mathbb{E}\left[\prod_{s=t}^{t+\tau-1}\xi(z_s^{s+1})\cdot q_{t+\tau}(\hat{z}^t,z_{t+1}^{t+\tau})\Big|z_t\right].
\end{aligned}$$

By the chain rule, we have $\xi(z_t^{t+\tau}) = \prod_{s=t}^{t+\tau-1}\xi(z_s^{s+1})$, and $\xi(z_t^s)=1$ for $t>s$. Hence,

$$\partial_{z_t}\tilde{U}_t(\hat{z}^t;z_t) = \textstyle\sum_{\tau\geq 0}\delta^\tau\cdot\mathbb{E}_{z_{t+1}^{t+\tau}}\left[\xi(z_t^{t+\tau})\cdot q_{t+\tau}(\hat{z}^t,z_{t+1}^{t+\tau})|z_t\right]. \tag{13}$$

Finally, Eq. (6) follows from the incentive compatibility $U_t(z^t) = \max_{\hat{z}_t}\tilde{U}_t(z^{t-1},\hat{z}_t;z_t)$, and the Envelope Theorem: $\partial_{z_t}U_t(z^t) = \partial_{z_t}\tilde{U}_t(z^{t-1},\hat{z}_t;z_t)|_{\hat{z}_t=z_t}$. ■

Proof of Lemma 3

First, Eq. (9) follows from Lemma 2 and $U_t(z^t) = U_t(z^{t-1},\ell) + \int_\ell^{z_t}\partial_{z_t}U_t(z^{t-1},\hat{z}_t)\,\mathrm{d}\hat{z}_t$.

Recall $\eta(z_0) = \frac{1-\Pi_0(z_0}{\pi_0(z_0)}$ is the inverse hazard rate of z_0. For the retailer-z_0, his optimal payoff is

$$U_0(z_0) = \int_\ell^{z_0}\partial_{z_0}U_0(\hat{z}_0)\,\mathrm{d}\hat{z}_0 = \int_\ell^{z_0}\mathrm{d}\hat{z}_0\cdot\left[\textstyle\sum_{\tau\geq 0}\mathbb{E}[\delta^\tau\cdot\xi(z^\tau)\cdot q_\tau(z^\tau)|\hat{z}_0]\right].$$

Hence,

$$\begin{aligned}
\mathbb{E}[U_0(z_0)] &= \int_{\mathcal{Z}}U_0(z)\,\mathrm{d}\Pi_0(z) = -\int_{\mathcal{Z}}U_0(z)\,\mathrm{d}[1-\Pi_0(z)] \\
&= -U_0(z)[1-\Pi_0(z)|_{z=\ell}^h + \int_{\mathcal{Z}}\partial_{z_0}U_0(z)[1-\Pi_0(z)]\,\mathrm{d}z \\
&= U_0(\ell) + \int_{\mathcal{Z}}\partial_{z_0}U_0(z)\frac{1-\Pi_0(z)}{\pi_0(z)}\pi_0(z)\,\mathrm{d}z \\
&= U_0(\ell) + \mathbb{E}[\partial_{z_0}U_0(z_0)\eta(z_0)] \\
&= U_0(\ell) + \mathbb{E}\left[\textstyle\sum_{t\geq 0}\delta^t\cdot\eta(z_0)\cdot\xi(z^t)\cdot q_t(z^t)\right]
\end{aligned} \tag{C}$$

Under the optimal contract $(p_t,q_t)_{t\geq 0}\in\mathcal{F}_1^*$, the manufacturer obtains the expected profit, which is the total revenue minus the production cost and the retailer's expected payoff $\mathbb{E}U_0(z_0)$. The objective of $(\mathcal{P}_1)$ becomes

$$\begin{aligned}
\tilde{V}\Big((p_t,q_t)_{t\geq 0}\Big) &= \mathbb{E}\left[\sum_{t\geq 0}\delta^t\big(R(q_t(z^t),z_t) - cq_t(z^t)\big)\right] - \mathbb{E}[U_0(z_0)] \\
&= \mathbb{E}\left[\sum_{t\geq 0}\delta^t\cdot\Big\{[z_t - q_t(z^t)]q_t(z^t) - cq_t(z^t) - \eta(z_0)\cdot\xi(z^t)\cdot q_t(z^t)\Big\}\right] - \mathbb{E}U_0(\ell), \text{by Eq. (C)} \\
&= \mathbb{E}\left[\sum_{t\geq 0}\delta^t\cdot\Big\{\Big(z_t - c - \eta(z_0)\cdot\xi(z^t)\Big)\cdot q_t(z^t) - \big(q_t(z^t)\big)^2\Big\}\right] - \mathbb{E}U_0(\ell),
\end{aligned}$$

which is a function of quantities $(q_t)_t$ alone. We can use point-wise optimization to derive q_t. The first order condition of the integrand with respect to q_t yields Eq. (7):

$$q_t(z^t) = \frac{z_t - c}{2} - \frac{1}{2} \cdot \eta(z_0) \cdot \xi(z^t).$$

Finally, Eq. (8) follows from the definition of $U_t(z^t) = R\big(q_t(z^t), z_t\big) - p_t(z^t) + \delta\mathbb{E}[U_{t+1}(z^{t+1})|z_t]$. ■

Proof of Lemma 4

It suffices to show q_t is IC_t, i.e., $U_t(\hat{z}^{t-1}, z_t) - \tilde{U}(\hat{z}^{t-1}, z_t'; z_t) \geq 0$. First, suppose $z_t > z_t'$.

$$\begin{aligned}
&U_t(\hat{z}^{t-1}, z_t) - \tilde{U}(\hat{z}^{t-1}, z_t'; z_t) \\
&= \tilde{U}_t(\hat{z}^{t-1}, z_t; z_t) - \tilde{U}(\hat{z}^{t-1}, z_t'; z_t) \quad \text{by the definitions of } U_t \text{ and } \tilde{U}_t \\
&= [\tilde{U}_t(\hat{z}^{t-1}, z_t; z_t) - \tilde{U}(\hat{z}^{t-1}, z_t'; z_t')] + [\tilde{U}_t(\hat{z}^{t-1}, z_t'; z_t') - \tilde{U}(\hat{z}^{t-1}, z_t'; z_t)] \\
&= [U_t(\hat{z}^{t-1}, z_t) - U_t(\hat{z}^{t-1}, z_t')] \; - \; [\tilde{U}(\hat{z}^{t-1}, z_t'; z_t) - \tilde{U}_t(\hat{z}^{t-1}, z_t'; z_t')] \quad \text{by the definitions of } U_t \text{ and } \tilde{U}_t \\
&= \int_{z_t'}^{z_t} \partial_{z_t} U_t(\hat{z}^{t-1}, \hat{z}_t)\, \mathrm{d}\hat{z}_t \;\; - \int_{z_t'}^{z_t} \partial_{z_t} \tilde{U}_t(\hat{z}^{t-1}, z_t'; \hat{z}_t)\, \mathrm{d}\hat{z}_t \quad \text{by Lemma 2} \\
&= \int_{z_t'}^{z_t} \Big\{ \partial_{z_t} U_t(\hat{z}^{t-1}, \hat{z}_t) - \partial_{z_t} \tilde{U}_t(\hat{z}^{t-1}, z_t'; \hat{z}_t) \Big\}\, \mathrm{d}\hat{z}_t \\
&= \int_{z_t'}^{z_t} \left\{ \sum_{\tau \geq 0} \mathbb{E}_{z_{t+1}^{t+\tau}} \Big[\delta^\tau \cdot \xi(z_t^{t+\tau}) \cdot \Big(q_{t+\tau}(\hat{z}^{t-1}, \hat{z}_t, z_{t+1}^{t+\tau}) - q_{t+\tau}(\hat{z}^{t-1}, z_t', z_{t+1}^{t+\tau}) \Big) \Big| \hat{z}_t \Big] \right\} \mathrm{d}\hat{z}_t, \text{Lemma 2, Eq. (13)} \\
&\geq 0. \quad \text{since } q_t(z^t) \text{ is nondecreasing in every argument.}
\end{aligned}$$

The case for $z_t \leq z_t'$ is similar. ■

Proof of Proposition 1

For the optimal contract to be simple, $\xi(z^t)$ should be independent of history details. The history independence requires that, except the the first order effect, there should be no higher order influences. For the Taylor expansion $z_{t+1} = G(z_t, \epsilon_{t+1}) = \gamma + \alpha_1 z_t + O(z_t^2) + \epsilon_{t+1}$, we take its derivative relative to z_t and obtain:

$$\partial_{z_t} z_{t+1} = \partial_{z_t} G(z_t, \epsilon_{t+1}) = \alpha_1 + O'(z_t^2).$$

If higher order coefficients of $O(z_t^2)$ were nonzero, the observation z_t would influence current z_{t+1} through $O(z_t^2)$, i.e., depending on history z_t. Hence, the dynamics is linear. ■