Discussion Paper Series – CRC TR 224

Discussion Paper No. 528
Project C 03

Supply Chain Frictions

Ying-Ju Chen[1]
Zhengqing Gui[2]
Ernst-Ludwig von Thadden[3]
Xiaojian Zhao[4]

April 2024

[1] School of Business and Management (ISOM), Hong Kong University of Science and Technology, Email: imchen@ust.hk
[2] Risk Management Institute, National University of Singapore, Email: zgui@nus.edu.sg
[3] University of Mannheim and CEPR, Email: vthadden@uni-mannheim.de
[4] Department of Economics, Monash Business School, Monash University, Email: xjzhao81@gmail.com

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Supply Chain Frictions*

Ying-Ju Chen † Zhengqing Gui ‡ Ernst-Ludwig von Thadden §
Xiaojian Zhao ¶

April 2024

Abstract

A central problem in supply chains is to coordinate the mismatch between supply and demand along the chain. This paper studies a problem of contracting between a manufacturer and a retailer who privately observes the retail demand materialized after the contracting stage. Under quite general assumptions, we show that the optimal contract must be either a wholesale contract or a buyback contract, depending on the retailer's ex-ante liquidity and bargaining power. In a buyback contract, the manufacturer requests an upfront payment from the retailer and buys back the unsold inventory at a previously agreed price. Depending on downstream liquidity and bargaining power this price may be constant or demand-dependent. Since return shipments are inefficient, retail supply and price will be lower than the first-best level. The optimal contracts are robust to several extensions including multiple retailers.

Keywords: Supply chains, informational frictions, buyback contracts, incentive compatibility, limited liability, ironing.

JEL Classification: D82, D86, L42, L60.

---

*This paper replaces an earlier paper entitled "Optimal Retail Contracts with Return Policies" with early work on some of the research presented here. We thank Kim-Sau Chung, Simon Loertscher, Volker Nocke, Harry Pei, Patrick Rey, Nicolas Schutz, as well as seminar and conference participants at the 2022 PKU-NUS Annual International Conference on Quantitative Finance and Economics, the 2023 Markets, Contracts, and Organizations Conference (Canberra), the 2023 European Association for Research in Industrial Economics (EARIE) Annual Conference (Rome), and the 2023 Asia-Pacific Industrial Organization Conference (APIOC, Hong Kong) for stimulating discussions. Von Thadden gratefully acknowledges financial support from the German Research Foundation (DFG) through CRC TR 224, C03.

†School of Business and Management (ISOM), Hong Kong University of Science and Technology. Email: imchen@ust.hk.

‡Risk Management Institute, National University of Singapore. Email: zgui@nus.edu.sg.

§University of Mannheim and CEPR. Email: vthadden@uni-mannheim.de.

¶Department of Economics, Monash Business School, Monash University. Email: xjzhao81@gmail.com.

1 Introduction

Supply chains are a major element of international trade and make decentralized manufacturing possible, often across considerable distances. Yet, exactly because of the large distances that typically separate the firms in a supply chain, these relationships are subject to inherent frictions. Two important such frictions are the mismatch between supply and demand and the information asymmetry between upstream and downstream participants. Supply-demand mismatch arises naturally because the interaction between upstream manufacturers and downstream users of their products typically is subject to unavoidable long lead times and demand uncertainty. Accordingly, production upstream must occur before downstream demand is realized. When demand is large, the downstream firm can only use up to the quantity delivered, and the excess demand is lost. When downstream demand is small, unsold production may have second-best uses downstream or be salvaged upstream. As transactions with final buyers take place downstream, information about realized demand is typically available downstream, but not upstream due to separation. This gives rise to the second friction, which is the asymmetry of information between the member firms of the supply chain. In a decentralized supply chain, these frictions are central to the working of supply chains and they naturally interact. This calls for a systematic investigation and is the objective of the present paper.

We examine the simplest upstream-downstream relationship possible, a model in which a manufacturer sells its products through a retailer operating in a remote market.[1] Two key factors characterize this relationship. First, production precedes sales, thus the contracting parties have to fix the contractual terms before demand uncertainty is resolved. This assumption is widely used in the literature of vertical contracting (e.g., Deneckere et al., 1996, 1997; Montez, 2015) and is the origin of the supply-demand mismatch. Second, the manufacturer cannot directly observe the retailer's sales and revenue. As a result, a vertical contract should: (1) determine the volume of trade, (2) specify the amount and timing of payments, and (3) deal with the questions of salvaging and re-ordering, the including incentive-compatible pricing of return shipments, so as to induce the retailer to execute the contract as intended, and to realize the envisaged gains from trade.

In practice, a number of contracts are used in supply chains, including buyback, franchise, quantity flexibility, revenue sharing, service commitment, two-part tariff, wholesale price, and others. There is large literature investigating these contracts and their consequences on supply chain performance in various contexts (see Cachon (2003) and Shen et al. (2019) for extensive discussions). Existing studies typically focus on particular contract forms and

[1] In its simplest form, this is the classical *news vendor problem* in Operations Management.

examine their impact on mitigating supply-demand mismatch, supply chain inefficiencies, or information asymmetry. In fact, the survey paper Shen et al. (2019) mentions 455 papers and categorizes 131 papers based on the contract form investigated. While they provide various angles to understand the practical use of these contracts, the conceptual question *what is the optimal contract* under the typical frictions present in supply chains, however, has received less attention. As Cachon (2003) put it succinctly: "practice has been used as a motivation for theoretical work, but theoretical work has not found its way into practice".

In order to make progress on this front, we study a game with two dates. Contracting happens at date 0, which determines the quantity, an immediate cash transfer, and rules for the execution of the future interaction. Retail demand is realized at date 1, after which there can be further transfers and a return or salvages of unsold inventories. If date-1 cash transfers and return shipments are contingent on realized demand, they must be incentive-compatible. Hence, ex-post, the contracting parties face a tradeoff between cash and returns. These two channels differ in several respects. Cash payments are bounded by the retailer's initial wealth plus his date-1 revenue, which is increasing in realized demand, but private information of the retailer. This limited liability constraint has been considered in the literature on contracts in industrial organization (e.g., Brander and Lewis, 1986) and captures the fundamental feature of small and medium enterprises: they are typically resource-constrained and thus the only collateral that can be pledged is the business value they create. Return shipments, on the other hand, cannot exceed the total amount of leftover inventories, which is negatively related to ex-post demand. Finally, cash transfers are efficient, while returning unsold inventories typically is not.[2] This assumption, together with the retailer's limited liability, makes the return of unsold inventories an imperfect screening device. As a result, the manufacturer's objective is to minimize the use of returns by appropriate incentive-compatible contracts.

Without assuming any functional form of contracts, we find that the optimal contract takes a rather simple form. At date 0, in exchange for the shipment of goods, the retailer makes an upfront payment to the manufacturer. After the realization of demand at date 1, there is either no more obligation for transfers or shipments from the retailer (wholesale contracts), or, under a buyback contract, the retailer transfers state-dependent quantities of cash and inventory back to the manufacturer. When realized demand is high, the cash payment is a high, fixed target and there are no inventory returns, while the cash payment is low and returns are positive when the realized demand is too low for the retailer to pay the fixed cash target in full. The rationale for this result is the following. Facing the adverse

[2] In part of the literature, the salvage value of unsold inventories is assumed to be zero (e.g., Marvel and Peck, 1995; Arya and Mittendorf, 2004). This simplification is less realistic for non-perishable goods, such as clothes and electronic devices.

selection problem, the manufacturer wants to elicit the retailer's private information, so the return of unsold inventory is used as a disciplining device when the reported demand is low. This inefficiency is also used to discipline the size of the initial order. However, ex ante the manufacturer also wants to minimize inefficient inventory returns by the retailer. Therefore, return policies will be used only when the reported demand is sufficiently low. Notably, and differently from related contractual structures in corporate finance, which we discuss below, the associated buyback price may not be constant. Hence, optimal contracts can involve variable or constant ex-post pricing.

Proposition 4 summarizes our main results. It shows that the optimal contract shifts from wholesale to buyback as the retailer's ex-ante liquidity or his bargaining power decrease. The buyback contract then features a constant buyback price, which switches to non-linear pricing when the retailer's ex-ante position deteriorates further. Our measure of bargaining power describes the competitiveness of the retailer's supply market, and our theory therefore links the contractual form endogenously to the competitiveness of that market. Therefore, our paper can be viewed as a unified micro-foundation for both wholesale contracts and buyback contracts observed in practice. Since the second-best optimal quantity under a buyback contract is strictly lower than the first-best level, the parties prefer wholesale contracts, but these cannot be implemented if the retailer does not have enough bargaining power.

This result is surprisingly robust to various extensions of our benchmark model. When the retailer can salvage unsold inventory by firesales without delay and frictions, such firesales may become part of the optimal contract, but the buyback structure with inefficient returns remains optimal. When the retail price is endogenous and influences the distribution of demand, Proposition 7 shows that the contract optimally reduces the price below the monopoly level. In other words, information asymmetry restricts the manufacturer's market power. When the manufacturer contracts with multiple retailers, Proposition 8 characterizes a set of symmetric optimal contracts. The "sum" of these contracts is equivalent to the optimal contract in a single-retailer model in which the retailer has larger bargaining power. It is as if the retailers are merged into one big entity and contract with the manufacturer, after which they split the contract terms equally. Based on this observation Corollary 1 states that when the number of retailers is sufficiently large, optimal contracts will switch from buyback to wholesale contracts without returns, and price and quantity will revert to the producers' optimal level. Put differently, introducing extra retailers in a vertical relationship pushes the market supply towards the first-best level, but will be accompanied by an increase in price.

1.1 Related literature

Our paper studies supply chains under general contracting subject to information asymmetry and limited liability. There is a lot of interesting prior work on specific contracts in such environments, for example on buyback/return contract in the presence of information asymmetry. Noteworthy are, in particular, Yue and Raghunathan (2007), Hsieh et al. (2008), Taylor and Xiao (2009), and Babich et al. (2012), which all assume that the retailer privately knows the demand distribution. Yue and Raghunathan (2007) compare two specific contract forms: no return and full return, where the latter guarantees that the buyback price equals the wholesale price. Hsieh et al. (2008) do not consider screening contracts but instead compare across three scenarios: the retailer shares demand information truthfully under an all-unit quantity discount contract with buyback, the retailer withholds demand information under an all-unit quantity discount contract with buyback, and the centralized supply chain. They focus on whether structural properties of ordering decisions are preserved in these scenarios. In Taylor and Xiao (2009), the retailer can exert efforts to improve demand information accuracy, and only two exogenous contract forms are considered: rebates contract that compensates the retailer for each unit sold to end consumers, and returns contract that specifies a buyback price for each unsold unit. In Babich et al. (2012), the supplier designs a menu of contracts, each of which comprises a wholesale price, a buyback price, and a lump-sum transfer. Kumar and Srinivasan (2007) consider the case in which the retailer who decides the order quantity and retail price is risk-averse. Our paper is distinct from all these papers by studying the general contract space without assuming any specific contract forms. Our proposed nonlinear return contract appears to be novel to the supply chain contracting literature[3] and may offer new guidance to practitioners for the design of optimal contracts.

The key feature of our model is to consider the frictions caused by information asymmetry and production-in-advance jointly. There is a large body of literature studying firms' inventory choices and production capacity, originating from Kreps and Scheinkman (1983) and then followed by many others (e.g., Davidson and Deneckere, 1986; Deneckere and Peck, 1995; Deneckere et al., 1996; Maggi, 1996). In a recent paper, Montez and Schutz (2021) apply techniques from all-pay auctions to study price competition where firms have private information about their inventory levels. Our study departs from these papers by arguing that market demand is likely to be the retailer's private information, which gives rise to the

[3] In the survey paper by Shen et al. (2019, Table 1), a return contract is defined by the right to "return the unsold products at the end of selling season at a return price". This *constant* return price specification is used in all the references therein.

informational role of buyback contracts.[4] A number of papers introduce information asymmetry into vertical contracts (e.g., Rey and Tirole, 1986; Blair and Lewis, 1994; Arya and Mittendorf, 2004), but do not specifically consider the tension between the retailer's private information about demand and inventory decisions.

Technically, we model an ex-post screening problem with hidden characteristics. When the type set is a continuum, the standard methodology used in the literature is control theory, pioneered by Guesnerie and Laffont (1984) and further developed by Hellwig (2010). However, the control-theoretic approach cannot be applied in the present paper. In our model, each type of retailer's set of deviations is bounded by his limited liability and feasibility constraint and thus depends on the endogenous contract. Therefore, the retailer's incentive constraint cannot be simplified into a local differential equation. This feature is similar to the financial contracting literature by Townsend (1979) and Gale and Hellwig (1985), but in their settings, there is no feasibility constraint, which substantially complicates the problem in our retail contracting context.[5] We overcome this difficulty using an ironing approach in the spirit of Myerson (1981), and recently Loertscher and Muir (2022), but we differ from the Myersonian approach in two aspects. First, our incentive constraint does not imply the monotonicity of allocation rules, so our ironing process can be applied to problems with weaker incentive constraints than that of standard screening problems. Second, the "ironed" contract in our model need not be optimal and must be further optimized by global techniques, while in the existing literature the "ironed" mechanism is already optimal.

The rest of this paper is organized as follows. Section 2 introduces the model setup. Section 3 provides the benchmark with symmetric information. Section 4 analyzes the model of asymmetric information and determines optimal contracts including their comparative statics. Section 5 discusses several key assumptions of the model and extends our benchmark model to different environments. Section 6 concludes. Some technical proofs are given in Appendix A. We provide a more comprehensive review of our ironing approach and the proofs of two extensions in the internet Appendix B.

[4] In this context, Wang et al. (2020) examine the signaling role of buyback contracts, while they take the buyback contractual form as exogenous.

[5] Relatedly, Gui et al. (2019) provide a detailed discussion on how the presence of limited liability affects the analysis of incentive constraints in the financial contracting literature. In particular, that paper shows that ignoring limited liability off the equilibrium path may lead to an over-simplified analysis and sub-optimal contracts.

2 Model

A manufacturer (she) contracts with a retailer (he) on the delivery of a homogeneous product. Production has no fixed costs and constant marginal costs $c > 0$. Given any retail price p, retail demand ω is stochastic and characterized by the distribution function $F(\cdot; p)$ over $[0, +\infty)$.[6] $F(\cdot; p)$ admits a density function f that is positive and bounded almost everywhere. In our baseline model, we assume that p is exogenous and observable. Therefore, we drop the reference to p in this section and the next two.

Retail demand ω is realized after the quantity q has been produced and delivered to the retailer and can only be observed by the retailer. The manufacturer only knows $F(\cdot)$. This results in two distortions. First, production must take place prior to the realization of demand, thus there will be a supply-demand mismatch. Second, the realization of demand is the retailer's private information, so all contractual obligations after sales must be incentive-compatible for the retailer. By applying the Revelation Principle, we focus on direct mechanisms in which the retailer simply reports his demand (or type) $\hat{\omega}$ and the contract is executed correspondingly.

After observing ω, the retailer determines the volume of sales, s. When there is a supply shortage, i.e., $q < \omega$, the retailer can only sell up to the quantity q, and the excess demand is lost. When there is insufficient demand, i.e., $q > \omega$, the retailer can only sell up to ω. Thus realized sales satisfy

$$s = s(\omega) \in [0, \min(\omega, q)]. \tag{FS}$$

s, too, is unobservable to the manufacturer.

The retailer is able to salvage unsold inventories at a constant salvage value v_r per unit. If instead the manufacturer possesses unsold inventories, her per unit salvage value is v_m. However, because of local knowledge, transaction costs or other frictions, it is more efficient for the retailer to keep unsold inventories.[7], i.e., $v_m < v_r$. To make the analysis non-trivial, we assume

$$p > c > v_r. \tag{1}$$

Hence, producing to sell at salvage value is not profitable, but normal sales are profitable

[6] This formulation of stochastic outcomes, the "parameterized-distribution-function approach", was pioneered by Mirrlees (1974) and Holmström (1979). Early usages in the IO/price theory literature include Burns and Walsh (1981). See Section 5.2 for a brief structural discussion of this model.

[7] In the framework developed here, this feature can hold even if $v_r \leq v_m$ as long as transportation costs for return shipments are taken into account.

and are known to be profitable.

A contract $\Gamma = (q, T_0, s, T_1, R)$ specifies: (1) the quantity q delivered to the retailer; (2) the cash transfer from the retailer to the manufacturer prior to sales (T_0), (3) the level of sales by the retailer (s), (4) after sales payments by the retailer to the manufacturer (T_1); (5) the return shipment of unsold inventory R. The last three components depend on ω and are therefore to be understood as functions.[8] This description captures many different types of retail contracts in practice.

Example 1 (Wholesale price). In a wholesale price contract, the manufacturer charges the retailer a constant wholesale price p_w per unit purchased at date 0, with no state-contingent transfer at date 1. The corresponding transfers and returns are, respectively,

$$T_0 + T_1 = p_w q, \ \ R = 0.$$

Example 2 (Buyback). In a buyback contract, the manufacturer charges the wholesale price p_w and buys back unsold units at the price $b < p_w$ per unit. Therefore,

$$T_0 = p_w q, \ \ T_1 = -bR, \ \ R = q - s.$$

Example 3 (Revenue sharing). In a revenue-sharing contract, in addition to the wholesale price, the manufacturer also obtains a fraction α of the retailer's revenue. In this case,

$$T_0 = p_w q, \ \ T_1 = \alpha p s, \ \ R = 0.$$

If demand is commonly observable and the retailer faces no limited liability, all these contracts are enforceable. We assume that this is not possible, as the manufacturer does not have enough information about the retailer's activity. As discussed in the introduction, this is the case in many applications in practice.

The timing of the game is depicted in Figure 1. At date 0, the manufacturer offers the retailer a take-it-or-leave-it contract Γ. If the retailer accepts the contract, he makes an initial payment T_0 to the manufacturer in exchange for the delivery of q units of the product. At date 1, retail demand ω is realized. The retailer observes ω and sells the quantity s. He then makes a report $\hat{\omega}$ to the manufacturer, pays her T_1, and returns R units, based on $\hat{\omega}$.

For simplicity, we assume that both contracting parties are risk-neutral, and there is no discounting. Let the retailer's initial wealth be $W \geq 0$. Under contract Γ, the retailer's

[8] As usual, when there is no risk of confusion, we shall denote the quantity (a number) and the function (a mapping into these quantities) by the same symbol.

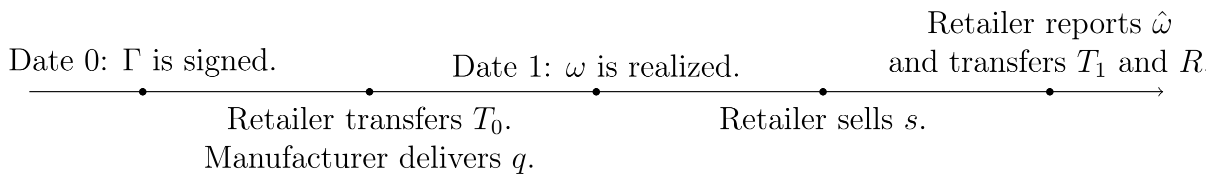


Figure 1: Timeline

ex-post profit from realized demand ω, reported demand $\hat{\omega}$ and sales s is

$$u_r(\omega, \hat{\omega}, s) = W - T_0 + ps - T_1(\hat{\omega}) + v_r[q - s - R(\hat{\omega})].$$

Here, $W - T_0$ is the retailer's cash position at date 0, ps is the gross revenue from sales, so $ps - T_1(\hat{\omega})$ is the retailer's cash flow at date 1, and $v_r[q - s - R(\hat{\omega})]$ is the salvage value of the retailer's inventory after returns.

Since the manufacturer has no fixed costs and constant marginal costs,[9] her ex-post payoff is

$$u_m(\hat{\omega}) = T_0 - cq + T_1(\hat{\omega}) + v_m R(\hat{\omega}).$$

Note that the problem is one of private values: the manufacturer is exposed to the demand shock ω only through the retailer's ex-post actions (T_1, R).

Ex-post, for each realization of ω, the retailer chooses not only his report strategically optimally, but also his sales level s. The contracting problem becomes interesting because of the feasibility, liquidity and information restrictions faced by the retailer.

The first such restriction is that the retailer cannot return more than the amount of unsold inventory he has and cannot re-order after strong demand. This implies the following feasibility constraint for returns:

$$0 \leq R \leq q - s. \tag{FR}$$

The second feasibility restriction is that the retailer cannot pay the manufacturer more than what he has at any time of the game. This implies the following liquidity constraints, at dates 0 and 1, respectively:

$$T_0 \leq W, \tag{FT$_0$}$$

$$T_1 \leq W - T_0 + ps. \tag{FT$_1$}$$

[9] Given our explicit consideration of non-linear pricing, it is not difficult to include fixed costs of production here.

In practice, liquidity constraints (feasibility with respect to payment) arise for various reasons, such as the retailer's inability to raise additional external finance, his option to quit the relationship ex-post, or legislation banning exploitative contracts. Note that we do not consider the salvage value of unsold inventory on the right-hand side of (FT_1), because in practice liquidating leftover inventory typically takes time. In Section 5.1, we discuss a variation of the model where the retailer can use cash generated by salvaging.

Third, sales by the retailer must satisfy the feasibility constraint (FS). Fourth, by the revelation principle, the retailer must have the incentive to report his type ω truthfully. And fifth, he must have the incentive to carry out sales $s(\omega)$ as planned.[10] Overall, this leads to the following incentive-compatibility constraint:

$$ps(\omega) - T_1(\omega) + v_r[q - s(\omega) - R(\omega)] \geq p\hat{s} - T_1(\hat{\omega}) + v_r[q - \hat{s} - R(\hat{\omega})] \tag{IC}$$

for all ω, $\hat{\omega}$, and $\hat{s}$ such that

$$0 \leq \hat{s} \leq \min(\omega, q) \tag{IC-FS}$$

$$0 \leq R(\hat{\omega}) \leq q - \hat{s} \tag{IC-FR}$$

$$T_1(\hat{\omega}) \leq W - T_0 + p\hat{s} \tag{IC-FT$_1$}$$

Note that as the type-ω retailer misreports to be type-$\hat{\omega}$, the transfer and the return shipment change accordingly. Hence, deviations of transfers and returns, $(\hat{T}_1, \hat{R})$, are restricted to lie in the range of the functions T_1, R. However, any deviation of the retailer's ex-post choice of sales, $\hat{s}$, is unobserved and therefore unrestricted, as long as it satisfies the incentive-feasibility constraint (IC-FS). We therefore face a problem of partially verifiable mechanism design, where the disclosure of private information (ω) through observable actions (T_1, R) is obfuscated by some other unobservable action (s).

The incentive constraint (IC)-(IC-FT_1) is special in that it restricts the choice of possible deviations to *feasible lies*. We only require each type of retailer to have no incentive to choose the contract designed for other types when his after-sales wealth, $W - T_0 + p\hat{s}$, and unsold inventory, $q - \hat{s}$, permit this. Hence, not only does the incentive condition (IC) depend on the retailer's type, but also the feasibility of her deviations in (IC-FS)-(IC-FT_1). The overall incentive constraint therefore is weaker than in standard problems where (IC) holds for all $\omega, \hat{\omega}$. Incorporating the qualifications (IC-FS)-(IC-FT_1) makes it difficult to simplify the global incentive constraint to a set of local first-order conditions, and apply the well-established control-theoretic approach to solve for the optimal contract, as in the literature

[10] Note that the last two points are distinct because sales $s(\omega)$ are unobservable upstream.

of mechanism design with hidden characteristics.[11] We discuss how we address this problem in Section 4.

Finally, the retailer has a monetary outside option, denoted by $\underline{u}$. Naturally, $\underline{u} \geq 0$. The manufacturer's outside option is normalized to zero. $\underline{u}$ therefore measures the retailer's relative bargaining power vis-à-vis the manufacturer in a take-it-or-leave-it offer in the contract proposal game. More generally, $\underline{u}$ is a measure of the competitiveness of the supplier market and will play an important role in our comparative statics analysis of contract design. Thus, the contracting parties have the participation (or individual-rationality) constraints

$$\mathrm{E}_\omega\, u_r(\omega, \omega, s(\omega)) \geq W + \underline{u} \tag{IR$_r$}$$

$$\mathrm{E}_\omega\, u_m(\omega) \geq 0. \tag{IR$_m$}$$

Hence, a full statement of the contracting problem between manufacturer and retailer is

$$\begin{aligned} \max_{\Gamma} \quad & \mathrm{E}_\omega\, u_m(\omega) \\ \text{subject to} \quad & \text{(FS), (FR), (FT}_0\text{), (FT}_1\text{), (IC)-(IC-FT}_1\text{), (IR}_r\text{), (IR}_m\text{).} \end{aligned}$$

To simplify the analysis, we restrict attention to schedules T_1 and R with finitely many discontinuities. This assumption allows us to simplify the characterization of contracts in Lemma 5 below when considering its Lagrangian dual problem. This or similar assumptions are standard, either explicitly or implicitly, in much of the literature on contracting with asymmetric information (Guesnerie and Laffont, 1984; Lacker and Weinberg, 1989).

We call a contract *feasible* if it satisfies the constraints (FS), (FR), (FT$_0$), and (FT$_1$), *admissible* if it satisfies all the constraints, and *optimal* if it is a solution to this problem. Moreover, if two contracts differ only on a set of states with zero measure, we say they are *equivalent*. If an admissible contract Γ generates less expected payoff to the manufacturer than an admissible contract $\hat{\Gamma}$, we say Γ is *dominated* by $\hat{\Gamma}$. And finally, we say that a quantity q can be *implemented* if there is an admissible contract $\Gamma = (q, T_0, s, T_1, R)$.

3 Symmetric Information Benchmark

The contracting problem described above features two main frictions: one physical in the sense that quantities must be determined before the realization of demand, the other informational in the sense that demand is private information of the retailer. We take the former friction as given and immutable, and in this section investigate the benchmark of symmetric

[11] See the large literature following Guesnerie and Laffont (1984).

full information.

In this case, it is not efficient to return merchandise to the manufacturer, as this is value-reducing compared to salvaging by the retailer and provides no other benefit. Furthermore, sales are equal to maximum feasible demand, $s = \min(\omega, q)$. Social surplus from producing quantity q therefore is

$$S(q) = \int_0^{+\infty} p \min(\omega, q) + v_r(q - \min(\omega, q))dF(\omega) - cq.$$

Denote by $Q(q)$ the expected feasible demand given q and price p,

$$Q(q) = \int_0^{+\infty} \min(\omega, q)dF(\omega) = q - \int_0^q F(\omega)d\omega, \tag{2}$$

where the last equality follows by partial integration. Then

$$S(q) = (p - v_r)Q(q) - (c - v_r)q. \tag{3}$$

This reformulation of total surplus has a natural interpretation. Since the product can always be salvaged with a per unit value v_r, $p - v_r$ and $c - v_r$ are the "real" price and marginal cost of the retailer, respectively. Therefore, $S(q)$ is similar to the standard profit function of a monopolist facing a demand function $Q(q)$. Moreover, $Q'(q) = 1 - F(q)$ and $Q''(q) = -f(q)$. Hence, $Q'(0) > 0$, and it is optimal to produce a positive quantity. By (1), the first-best quantity therefore is uniquely pinned down by the first-order condition.[12]

Proposition 1. *The first-best quantity q^{FB} is unique and satisfies*

$$F(q^{FB}) = \frac{p - c}{p - v_r}. \tag{4}$$

The first-best surplus is

$$S(q^{FB}) = (p - c)q^{FB} - (p - v_r) \int_0^{q^{FB}} F(\omega)d\omega > 0. \tag{5}$$

Proof. (4) follows directly from the first-order condition, (5) from (2) and (3). □

[12] The quantity is first-best from the production/distribution side, as these are the frictions we focus on. Since we work only with reduced-form demand, we cannot make statements about overall optimality including consumers surplus.

4 Optimal Contract under Asymmetric Information

For the second-best analysis it is useful to distinguish two parts of a contract. The first part consists of q and T_0, deliveries and transfers at date 0. The second part consists of T_1, R and s; they are functions of ω and $\hat{\omega}$ and are subject to the incentive-compatibility constraint. We will sometimes refer to (q, T_0) as the *date-0 component*, and the triple (s, T_1, R) as the *date-1 component*. It is important to realize that the choice of s and $\hat{\omega}$ at date 1 must be optimal for each ω given the schedules T_1 and R, since the retailer has private information and no commitment power.

4.1 Implementation by wholesale contracts

Under a wholesale contract, there are no state-contingent transfers. In our framework, for an optimal contract to be wholesale the date-1 component must therefore satisfy

$$T_1(\omega) = T_1 \text{ and } R(\omega) = 0 \text{ for all } \omega.$$

In this case, the contract provides no costly incentives, so social surplus is split between contracting parties without efficiency loss. More generally, the following characterization of wholesale contracts is useful.

Lemma 1. *The quantity q can be implemented by a wholesale contract with full surplus extraction if and only if*

$$S(q) + cq \leq W + \underline{u}. \tag{6}$$

Proof. Under a wholesale contract that implements q, the retailer gets

$$u_r = W - T_0 + S(q) + cq - T_1.$$

Under the constraint $T_1 \leq W - T_0$, the total payment $T_0 + T_1$ necessary to achieve a binding (IR_r) is feasible if and only if (6) holds. □

Lemma 1 characterizes the situations in which the retailer's initial liquidity constraint (FT_0) does not bind. This occurs if she either has sufficient funds or sufficiently high bargaining power.

Note that the left-hand side of (6) is strictly monotone in q. Let $\underline{q}$ be the greatest q for which (6) holds. Hence, if $q^{FB} \leq \underline{q}$, the first-best can be implemented by a wholesale contract. If $q^{FB} > \underline{q}$, the first-best cannot be implemented by a wholesale contract with full surplus extraction. By Lemma 1 the condition $q^{FB} \leq \underline{q}$ is therefore necessary and sufficient for the first-best to be second-best optimal. By (5) and (6), this condition is equivalent to

$$pq^{FB} - (p - v_r) \int_0^{q^{FB}} F(\omega) d\omega \leq W + \underline{u}. \tag{7}$$

If (7) does not hold, the manufacturer has two options. First, she can implement the quantity $\underline{q}$ by an optimal wholesale contract with full surplus extraction, as given in Lemma 1. And second, she can offer a contract that implements a quantity $q > \underline{q}$ by requiring some ex-post state-contingent repayment T_1 supported by costly incentives.[13] In the following three subsections, we assume that $q > \underline{q}$ and investigate this second option.

4.2 The sales decision

In this subsection, we simplify the contracting problem by eliminating the in- and off-equilibrium sales decisions $s(\omega)$ and $\hat{s}$, respectively. To begin with we note that it is redundant to consider the choice of T_0 and T_1 separately, since (FT_0) and all payoffs involve only $T_0 + T_1(\omega)$ and T_1 is unbounded below. Hence, we can add any constant $C \leq W - T_0$ to T_0 and subtract it from $T_1(\omega)$ for all ω without changing any of the analysis and results. Without loss of generality, we assume from now on that $T_0 = W$.[14]

Next, denote by V the total ex-post value transfer to the manufacturer as evaluated by the retailer:

$$V(\omega) = T_1(\omega) + v_r R(\omega). \tag{8}$$

Using $T_0 = W$ and (8), the retailer's ex-post utility is

$$u_r(\omega, \hat{\omega}, s) = ps - T_1(\hat{\omega}) + v_r[q - s - R(\hat{\omega})] \tag{9}$$

$$= (p - v_r)s - V(\hat{\omega}) + v_r q. \tag{10}$$

As (10) shows, $u_r(\omega, \hat{\omega}, s)$ is strictly increasing in s for each ω, and the total transfer V is not affected by increasing s. Hence, it is optimal for the retailer ex-post to set s as large as possible. But as (9) shows, for a given deviation $\hat{\omega}$ the maximum sales volume that the

[13] Clearly, the option of supplying q^{FB} and leaving some rents over and above $W + \underline{u}$ to the retailer is dominated by the first alternative.

[14] If W is large or q small, as in the preceding subsection, this implies negative T_1, i.e. payments from the manufacturer to the retailer, in some states at date 1.

retailer can choose is

$$\hat{s} = \min(q - R(\hat{\omega}), \min(\omega, q)) \tag{11}$$
$$= \min(q - R(\hat{\omega}), \omega). \tag{12}$$

By (11), if $R(\hat{\omega}) \leq q - \min(\omega, q)$, the maximum sales volume for deviation $\hat{\omega}$ is $\min(\omega, q)$, the maximum feasible volume given by (FS). If $R(\hat{\omega}) > q - \min(\omega, q)$, the maximum sales volume for deviation $\hat{\omega}$ is $q - R(\hat{\omega})$ because of the feasibility constraint for returns (FR).

The conceptual difficulty with the IC constraint (IC)-(IC-FT$_1$) is that the observable off-the-equilibrium choices $T_1(\hat{\omega})$ and $R(\hat{\omega})$ depend on the unobservable action $\hat{s}$. The preceding distinction allows to eliminate $\hat{s}$ from the IC constraint .

Lemma 2. *A feasible contract is incentive-compatible if and only if for any $\omega, \hat{\omega} \geq 0$ with $T_1(\hat{\omega}) \leq p\min(\omega, q)$,*

(IC1) $R(\hat{\omega}) \leq q - \min(\omega, q)$ *implies* $V(\omega) + (p - v_r)(\min(\omega, q) - s(\omega)) \leq V(\hat{\omega})$,

(IC2) $R(\hat{\omega}) > q - \min(\omega, q)$ *implies* $V(\omega) + (p - v_r)(q - R(\hat{\omega}) - s(\omega)) \leq V(\hat{\omega})$.

Proof. **"Only if":** Suppose that a feasible contract Γ satisfies (IC)-(IC-FT$_1$). Consider any $\omega, \hat{\omega} \geq 0$ with $T_1(\hat{\omega}) \leq p\min(\omega, q)$. If $R(\hat{\omega}) \leq q - \min(\omega, q)$, $\hat{s} = \min(\omega, q)$ satisfies (IC-FS)-(IC-FT$_1$) by definition. Inserting this $\hat{s}$ into (IC) gives (IC1). If $R(\hat{\omega}) > q - \min(\omega, q)$, let $\hat{s} = q - R(\hat{\omega})$. The assumption then reads $\hat{s} < \min(\omega, q)$, which implies (IC-FS). (IC-FR) holds trivially by construction. Finally, use feasibility (FR) and (FT$_1$), evaluated at $\hat{\omega}$, twice to obtain $T_1(\hat{\omega}) \leq ps(\hat{\omega}) \leq p(q - R(\hat{\omega})) = p\hat{s}$, which is (IC-FT$_1$). Inserting this $\hat{s}$ into (IC) then gives (IC2).

"If": Suppose that Γ satisfies (IC1) and (IC2) for any $\omega, \hat{\omega} \geq 0$ with $T_1(\hat{\omega}) \leq p\min(\omega, q)$. Consider ω, $\hat{\omega} \geq 0$, and $\hat{s}$ that satisfy the constraints (IC-FS)-(IC-FT$_1$). (IC-FS) and (IC-FT$_1$) imply $T_1(\hat{\omega}) \leq p\min(\omega, q)$. By the argument leading to (12), when $R(\hat{\omega}) \leq q - \min(\omega, q)$, the best possible ex-post deviation is $\hat{s} = \min(\omega, q)$. Since (IC) holds for this $\hat{s}$ by (IC1), it holds for all admissible $\hat{s}$. When $R(\hat{\omega}) > q - \min(\omega, q)$, the best possible ex-post deviation is $\hat{s} = q - R(\hat{\omega})$. Since (IC) holds for this $\hat{s}$ by (IC2), it holds for all admissible $\hat{s}$. Hence, Γ is incentive-compatible. □

Turning to the in-equilibrium sales decision $s(\omega)$, the argument leading up to (12) shows that, if there is a $\hat{\omega}$ such that $q - R(\hat{\omega}) \leq \omega \leq q$, with at least one inequality being strict, incentive-compatibility alone does not suffice to conclude that the retailer chooses $s = \min(\omega, q)$ ex post. However, as the following lemma shows, such a contract structure would not be optimal ex ante.

Lemma 3. *A contract is optimal only if* $s(\omega) = \min(\omega, q)$ *for all* ω.

Proof. See Appendix A.1. □

Intuitively, the only reason that the retailer may want to undersell in some state ω is that the contract specifies a large return shipment of unsold inventory in order to relax the incentive constraint for the report of some $\hat{\omega}$. However, this yields lower profits on the equilibrium path. Therefore, the manufacturer can be made better off by reducing the return shipment without violating the incentive constraint, which is possible because if the retailer in any state has the ability and incentive to misreport ω in $\hat{\Gamma}$, he would have done so already in the original contract Γ by selling less.

4.3 Implementing $q > \underline{q}$: Local buyback contracts

Lemmas 2 and 3 taken together now make it possible to identify several key features of optimal contracts that can be used to characterize them constructively. The following result follows directly from these two lemmas.

Lemma 4. *If* T_1 *and* R *are the ex-post components of an optimal contract, then:*

(O1) For any $\omega, \hat{\omega} \leq q$ *with* $T_1(\hat{\omega}) \leq p\omega$,

- *(a) if* $R(\hat{\omega}) \leq q - \omega$, *then* $V(\omega) \leq V(\hat{\omega})$,
- *(b) if* $R(\hat{\omega}) > q - \omega$, *then* $V(\omega) \leq V(\hat{\omega}) + (p - v_r)[R(\hat{\omega}) - (q - \omega)]$.

(O2) For all $\omega \geq q$, $T_1(\omega) = T_1(q)$ *and* $R(\omega) = 0$.

Lemma 4 implies that in an optimal contract, any demand realization higher than q is irrelevant to the retailer's incentive problem, because beyond q payments are flat at $T_1(q)$ and returns are zero. Moreover, for any demand realization lower than q, an optimal contract only needs to prevent two types of deviation: First, if $\hat{\omega}$ is a feasible deviation under $s(\omega) = \omega$, then the retailer's total transfer $V(\omega)$ must be lower than that of $\hat{\omega}$. This is the case of (O1a) in Lemma 4. Second, if $\hat{\omega}$ is infeasible given the optimal sales at the realized state ω, then the retailer who wants to deviate must deliberately sell less, i.e., $\hat{s} = q - R(\hat{\omega})$, to fulfill the return requirements $R(\hat{\omega})$. In this case, $V(\omega)$ can be higher than $V(\hat{\omega})$, but is still bounded by a (affine) linear function of ω. This is the case of (O1b) in Lemma 4. But as we shall see momentarily, *(O1)* does not imply monotonicity.

Using Lemma 3 and Lemma 4, for any given q, we can relax the problem of finding the optimal date-1 component of the contract as follows:

$$\max_{T_1,R} \mathrm{E}_\omega[T_1(\omega) + v_m R(\omega)],$$

subject to

- the pointwise constraints

$$T_1(\omega) = T_1(q) \text{ and } R(\omega) = 0, \tag{O2}$$

  for all $\omega \geq q$, and

$$0 \leq R(\omega) \leq q - \omega, \tag{FR}$$

$$T_1(\omega) \leq p\omega, \tag{FT$_1$}$$

  for all $\omega \in [0, q]$,

- the incentive constraints (O1) of Lemma 4,
- and the retailer's participation constraint

$$(p - v_r)\,\mathrm{E}_\omega[\min(\omega, q)] - \mathrm{E}_\omega[V(\omega)] + v_r q \geq W + \underline{u}. \tag{IR$_r$}$$

In the relaxed problem, using the incentive constraints (O1) is still difficult because of the inherent off-the-equilibrium feasibility constraints. In classical adverse selection models, incentive-compatibility usually reduces to an envelope formula, which implies that the agent's indirect utility function is absolutely continuous (Milgrom and Segal, 2002; Hellwig, 2010). However, this is not the case in our model due to the presence of (FR) and (FT$_1$). To wit, consider the following date-1 component of a contract for some given quantity $q > \underline{q}$:

$$T_1(\omega) = \begin{cases} \alpha\omega & \omega < q/2, \\ pq/2 & \omega \geq q/2; \end{cases} \quad R(\omega) = \begin{cases} q - \omega & \omega < q/2, \\ 0 & \omega \geq q/2. \end{cases}$$

The retailer's total transfer function V is

$$V(\omega) = \begin{cases} \alpha\omega + v_r(q - \omega) & \omega < q/2, \\ pq/2 & \omega \geq q/2; \end{cases}$$

and his indirect utility function is

$$U_r(\omega) = u_r(\omega, \omega, \min(\omega, q)) = \begin{cases} (p-\alpha)\omega & \omega < q/2, \\ p\omega + v_r(q-\omega) - pq/2 & q/2 \le \omega < q, \\ pq/2 & \omega \ge q. \end{cases}$$

If $\alpha \in [p - v_r, p]$, then this contract is admissible. Note that $R(\omega) > q/2$ for any $\omega \in [0, q/2)$, while $T_1(\omega) \ge pq/2$ for any $\omega \ge q/2$. The retailer with type $\omega \in [0, q/2)$ cannot exaggerate his type above $q/2$ due to the liquidity constraint for cash. The retailer with type $\omega \ge q/2$ does not want to understate his type below $q/2$ due to the incentive constraint. However, if $\alpha > p - v_r$, U_r has an upward jump at $q/2$. If also $\alpha > v_r$, the function V associated with T_1, R is non-monotonic and discontinuous. Figure 2 provides a graphical illustration of this example.

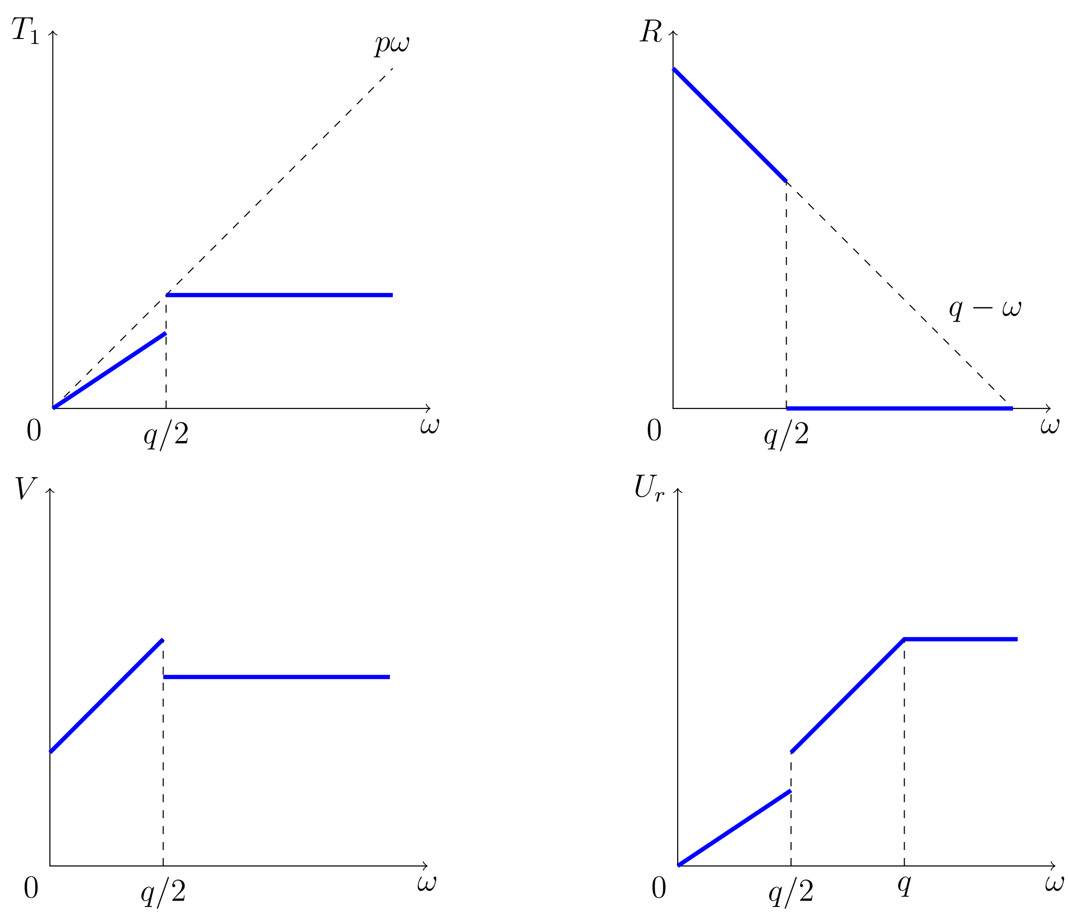


Figure 2: The non-monotonicity of V and discontinuity of the indirect utility function

Moreover, unlike in classical models such as Guesnerie and Laffont (1984), incentive-

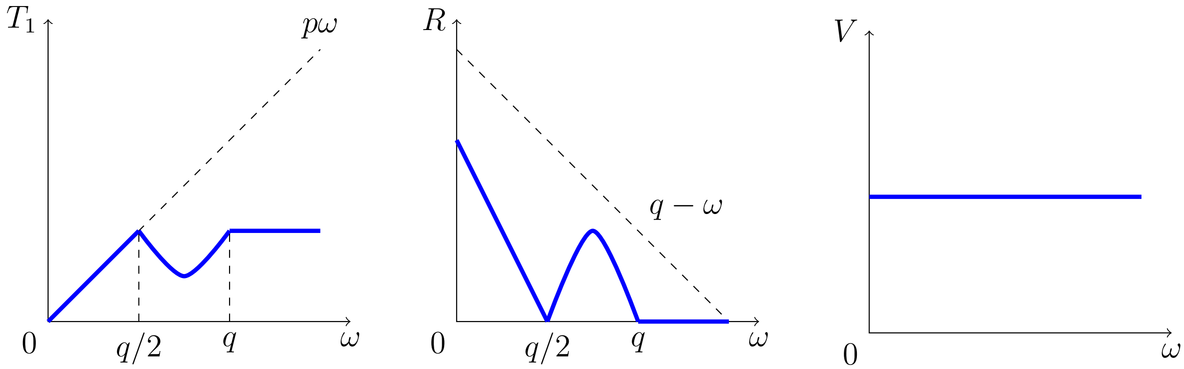


Figure 3: The non-monotonicity of T_1 and R.

compatibility does not imply the monotonicity of the decision function R. To see this, it suffices to consider a particularly simple case where V is constant on the whole state space, as depicted in Figure 3. Such a contract is clearly incentive-compatible because no type has an incentive to misreport other types (formally, both conditions of Lemma 2 are always satisfied). Given that $V(\omega) = T_1(\omega) + v_r R(\omega)$ is constant, the only constraint for T_1 is (FT_1) and for R is (FR). Hence, there is considerable leeway in choosing the shapes of T_1 and R.

We therefore propose a more complex constructive approach that combines local and global optimization techniques. The main idea is to show that any admissible contract that meets the necessary conditions in Lemma 3 and Lemma 4 is strictly dominated by an admissible contract of a specific form, if it does not already have this form. This implies that optimal contracts exist and are of this form. In order to construct such contracts, we apply the ironing technique used in traditional screening models (Baron and Myerson, 1982; Guesnerie and Laffont, 1984) and auction theory (Myerson, 1981), using some results from real analysis and the theory of convex functions (see, e.g., Rockafellar, 1970; Rudin, 1987).

Our ironing approach works as follows (we provide a full technical description in online Appendix B.1). Since return shipments are inefficient, it is better to replace them, over any small interval, with cash transfers. If the retailer's expected utility remains unchanged after this process, then we get an improved contract. However, this improvement must satisfy feasibility. Suppose that we reduce return shipments at (a small interval around) $\hat{\omega}$ from $R(\hat{\omega})$ to $\hat{R}(\hat{\omega})$, and let ω be a state satisfying $p\omega \geq T_1(\hat{\omega})$ and $\hat{R}(\hat{\omega}) < q - \omega < R(\hat{\omega})$. In the original contract, the relationship between $V(\omega)$ and $V(\hat{\omega})$ is governed by (O1b) of Lemma 4, while in the improved contract, $V(\omega)$ and $V(\hat{\omega})$ should satisfy (O1a) of Lemma 4, which is tighter than (O1b).

Therefore, we start by ironing the total transfer function V to make it nondecreasing on the state space. This is followed by a restructuring of the function R to make it nonincreasing. The ironed contract has an important property: $p\omega \geq T_1(\hat{\omega})$ implies $V(\omega) \leq V(\hat{\omega})$. In other words, V satisfies (O1a) of Lemma 4 irrespective of R. We therefore can use cash transfers to replace return shipments on those ironed intervals.

Figure 4 provides a graphical illustration of the ironing procedure when $\mathcal{N} = \{1, 2\}$.

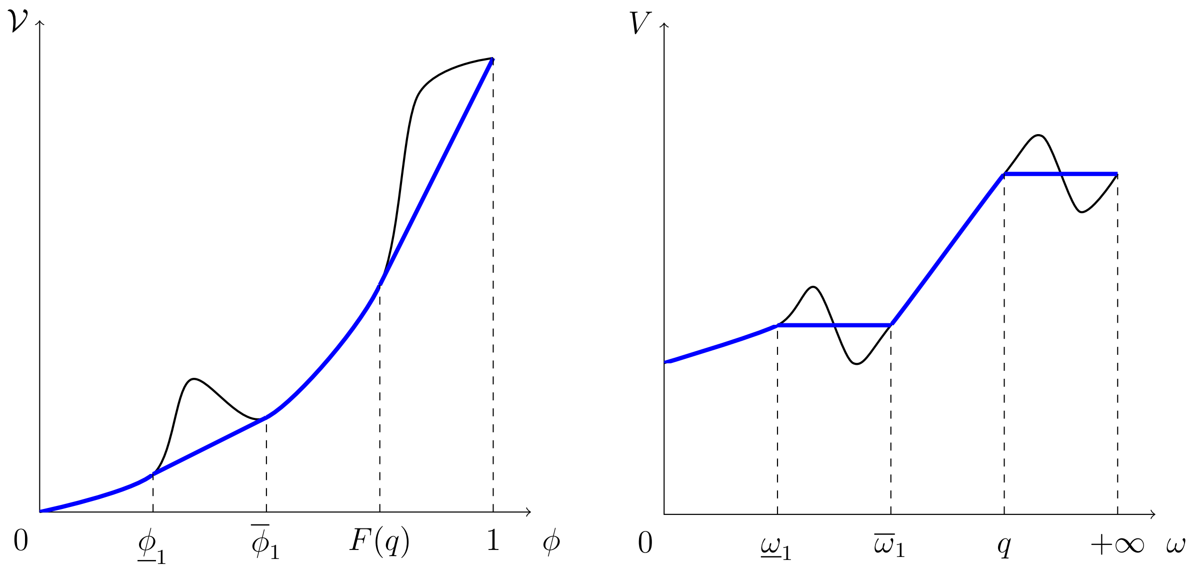


Figure 4: The ironing approach

The ironed contract has countably many intervals $[\underline{\omega}_n, \overline{\omega}_n)$, indexed by $n \in \mathcal{N}$, such that on each of them: (1) $V(\omega)$ equals a constant $t_n + v_r \max(q - \overline{\omega}_n, 0)$; and (2) there is a minimum level $\max(q - \overline{\omega}_n, 0)$ of return shipments. Intuitively, for realizations in $[\underline{\omega}_n, \overline{\omega}_n)$ the contract can be interpreted as setting a reference repayment level of t_n in cash and of $\max(q - \overline{\omega}_n, 0)$ of returns. If the retailer's cash is insufficient (which implies that there are more unsold units), the manufacturer buys back additional units at a marginal buyback price v_r. Hence, locally in $[\underline{\omega}_n, \overline{\omega}_n)$, a retailer who considers deviating marginally trades off cash payments and returns at the rate $1 : v_r$, as the unrestricted incentive constraint (IC) would require. In this sense, the new contract, which locally resembles a buyback contract on the interval $[\underline{\omega}_n, \overline{\omega}_n)$, provides optimal incentives. This gives rise to the following definition.

Definition 1. *A contract is a* ***local buyback contract*** *if there exist countably many disjoint intervals $[\underline{\omega}_n, \overline{\omega}_n)$ and an equal number of constants t_n, indexed by $n \in \mathcal{N}$, such that:*

(a) For any ω,

(a.1) if $\omega \in [\underline{\omega}_n, \overline{\omega}_n)$, then $T_1(\omega) = \min(p\omega, t_n)$, $R(\omega) = \max(q - \overline{\omega}_n, 0) + \max((t_n - p\omega)/v_r, 0)$,

(a.2) if $\omega \notin \bigcup_{n \in \mathcal{N}} [\underline{\omega}_n, \overline{\omega}_n)$, then $T_1(\omega) \leq p \min(\omega, q)$, $R(\omega) = \max(q - \omega, 0)$.

(b) For any $n \in \mathcal{N}$, $t_n \leq (p - v_r)\underline{\omega}_n + v_r \min(\overline{\omega}_n, q)$.

(c) $V(\omega)$ is nondecreasing and continuous.

One can immediately verify from Definition 1 that any local buyback contract is admissible. By (a) and (b), it satisfies (FT_1) and (FR). For incentive-compatibility, note that the retailer has no incentive to exaggerate his type because $V(\omega)$ is nondecreasing. If $\omega \in [\underline{\omega}_n, \overline{\omega}_n)$, the retailer cannot understate his type below $\underline{\omega}_n$, because $R(\hat{\omega}) > q - \underline{\omega}_n$ for all $\hat{\omega} < \underline{\omega}_n$, and he does not want to deviate to $\hat{\omega} \in [\underline{\omega}_n, \omega)$, because V is constant on $[\underline{\omega}_n, \overline{\omega}_n)$. If $\omega \notin \bigcup_{n\in\mathcal{N}}[\underline{\omega}_n, \overline{\omega}_n)$, he cannot understate his type because $R(\hat{\omega}) > \max(q - \omega, 0)$ for all $\hat{\omega} < \omega$.

Lemma 5 shows why these contracts are important.

Lemma 5. *If a contract implementing $q > \underline{q}$ is optimal, then it is a local buyback contract.*

Proof. Assume that Γ implementing $q > \underline{q}$ is optimal, and $\hat{\Gamma}$ is the corresponding local buyback contract constructed from the ironing procedure. By construction, the retailer's total transfer function in $\hat{\Gamma}$ is exactly $\hat{V}$. Since $\hat{\mathcal{V}}(1) = \mathcal{V}(1)$, the retailer's expected total transfer in $\hat{\Gamma}$ is identical to that of Γ.

Suppose that $T_1(\omega) > \hat{T}_1(\omega)$ for some $\omega \in [\underline{\omega}_n, \overline{\omega}_n)$. Then $T_1(\omega) > t_n$. There are three possible cases to be discussed.

Case 1: $V(\omega) \leq k_n$. Then $R(\omega) < q - \overline{\omega}_n$ for some $\overline{\omega}_n < q$. Recall that in the right neighborhood of $\overline{\omega}_n$, V is strictly increasing, so there exists $\hat{\omega} > \overline{\omega}_n$ such that $\hat{\omega} \leq q - R(\omega)$ and $V(\hat{\omega}) > k_n \geq V(\omega)$, a contradiction to (O1a) of Lemma 4.

Case 2: $V(\omega) > k_n$ for some $\overline{\omega}_n \leq q$. Then there exists $\hat{\omega} \in (\omega, \overline{\omega}_n]$ such that $V(\hat{\omega}) \leq k_n < V(\omega)$. By (O1a) of Lemma 4, $T_1(\hat{\omega}) > p\omega \geq T_1(\omega) > t_n$, which implies that $R(\hat{\omega}) < q - \overline{\omega}_n$. The state $\hat{\omega}$ fits the assumption of Case 1 and will again lead to the same contradiction.

Case 3: $V(\omega) > k_n$ for some $\overline{\omega}_n = +\infty$. Then $T_1(\omega) > t_n = k_n$. Since $\lim_{\omega\to+\infty} V(\omega) = k_n$, there exists $\hat{\omega} > \omega$ such that $V(\hat{\omega}) = T_1(\hat{\omega}) < T_1(\omega) \leq V(\omega)$, a contradiction to (O1a) of Lemma 4.

Therefore, $T_1(\omega) \leq \hat{T}_1(\omega)$ for any $\omega \in [\underline{\omega}_n, \overline{\omega}_n)$. Since $\mathrm{E}_\omega[\hat{V}(\omega)|\underline{\omega}_n \leq \omega < \overline{\omega}_n] = \mathrm{E}_\omega[V(\omega)|\underline{\omega}_n \leq \omega < \overline{\omega}_n]$, we have $\mathrm{E}_\omega[\hat{R}(\omega)|\underline{\omega}_n \leq \omega < \overline{\omega}_n] \leq \mathrm{E}_\omega[R(\omega)|\underline{\omega}_n \leq \omega < \overline{\omega}_n]$. Moreover, for any $\omega \notin \bigcup_{n\in\mathcal{N}}[\underline{\omega}_n, \overline{\omega}_n)$, Γ and $\hat{\Gamma}$ are identical. In expectation, $\hat{\Gamma}$ uses (weakly) more cash and (weakly) less return shipments than Γ. Thus, $\hat{\Gamma}$ is also optimal, and there must be $T_1(\omega) = \hat{T}_1(\omega) = \min(p\omega, t_n)$ for almost all $\omega \in [\underline{\omega}_n, \overline{\omega}_n)$.

If $T_1(\omega) = p\omega$, then $R(\omega) < \max(q - \omega, 0)$. Otherwise, by (O1b) of Lemma 4, both (FT_1) and (FR) bind for any states lower than ω, V is therefore strictly increasing from 0 to ω, and ω cannot be an interior point of an ironed interval. Now that (FT_1) is binding while

(FR) is slack at ω, V is weakly decreasing in the neighborhood of ω. If $T_1(\omega) = t_n$, then V is constant in the neighborhood of ω. Thus, on the whole interval $[\underline{\omega}_n, \overline{\omega}_n)$, V is weakly decreasing. Note that $V(\overline{\omega}_n+) \geq k_n$. If $V(\overline{\omega}_n-) < k_n$, then V has an upward jump at $\overline{\omega}_n$, a contradiction to (O1b) of Lemma 4. Hence, $V(\overline{\omega}_n-) \geq k_n$. Recall that $\hat{V}$ equals the constant k_n on $[\underline{\omega}_n, \overline{\omega}_n)$. The monotonicity of V must imply that V also equals the constant k_n on $[\underline{\omega}_n, \overline{\omega}_n)$. In summary, Γ is equivalent to $\hat{\Gamma}$. □

Figure 5 depicts how we construct the local buyback contract used in Lemma 5. In the top panel, the black line is an arbitrary admissible contract. The blue lines illustrate our ironing approach, which yields the local buyback contract in the bottom panel.

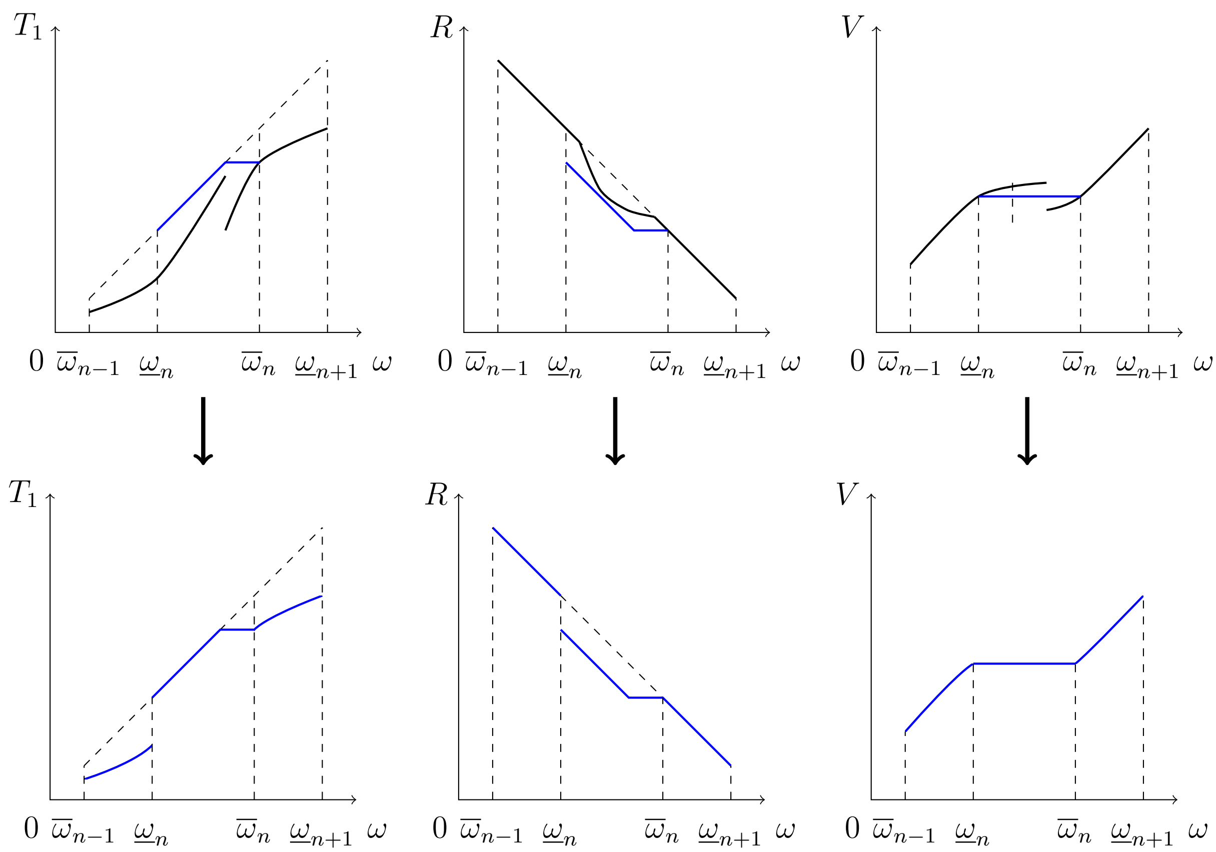


Figure 5: Constructing a local buyback contract

4.4 Implementing $q > \underline{q}$: Buyback contracts

Lemma 5 transforms the problem of finding optimal contracts into the problem of finding optimal local buyback contracts. In any local buyback contract, T_1 and R can have discontinuities only at $\underline{\omega}_n$. Therefore, the assumption that T_1 and R have only finitely many discontinuities restricts our attention to local buyback contracts with $|\mathcal{N}| = N < +\infty$. The analysis is further simplified if one assumes that the distribution of retail demand has a nondecreasing hazard rate, i.e., that $f(\omega)/[1 - F(\omega)]$ is nondecreasing. From now on we make this assumption, which, while not without loss of generality, is standard in the mechanism design literature and is satisfied by many commonly used distributions (see, e.g., Bagnoli and Bergstrom, 2005).

When $|\mathcal{N}|$ is finite, without loss of generality denote $\mathcal{N} = \{1, 2, \ldots, N\}$. For convenience, we rank the set of cutoffs $\{\underline{\omega}_n, \overline{\omega}_n | 1 \leq n \leq N\}$ by ascending order of n, and denote $\overline{\omega}_0 = 0$. Thus, $N = n(q)$, and

$$\overline{\omega}_{n-1} \leq \underline{\omega}_n \leq \overline{\omega}_n \leq \underline{\omega}_{n+1} \text{ for any } 1 \leq n \leq N - 1. \tag{13}$$

The following result completes our characterization of local buyback contracts, by "filling in the holes" between the buyback intervals.

Lemma 6. *If a local buyback contract is optimal, then for each $\omega \in [\overline{\omega}_n, \underline{\omega}_{n+1})$ and $0 \leq n \leq N - 1$,*

$$T_1(\omega) = t_n + p(\omega - \overline{\omega}_n). \tag{14}$$

Proof. (O1b) of Lemma 4 requires that $V'(\omega) \leq p - v_r$, and Definition 1 requires that $R(\omega) = q - \omega$, so (14) is equivalent to $V'(\omega) \leq p - v_r$ being binding on any $[\overline{\omega}_n, \underline{\omega}_{n+1})$. Suppose that (14) does not hold for some n. We can reduce $\underline{\omega}_{n+1}$ to $\underline{\omega}'_{n+1}$ and increase $T_1(\omega)$ to make (14) hold on $[\overline{\omega}_n, \underline{\omega}'_{n+1})$. Here, $\underline{\omega}'_{n+1}$ is chosen to keep t_{n+1} unchanged and $V(\omega)$ continuous.[15] The resulting contract is still a local buyback contract which uses more cash and less return shipments in expectation. The retailer is worse off, but he can be compensated by the manufacturer with lump-sum transfers. The manufacturer is strictly better off even after the compensation. Therefore, a binding (14) is necessary for optimality. □

Figure 6 graphically illustrates our discussion. In the top panel, the blue line is an arbitrary local buyback contract. The red lines depict how we increase $T_1(\omega)$ to make (14) bind, which yields the improved contract in the bottom panel.

Using the continuity of V, as well as (13) and (14), we can write $\{t_n\}_{2 \leq n \leq N}$ as functions

[15] If $n = N$, we can simply raise $T_1(\omega)$ on $[\overline{\omega}_N, +\infty)$ and the argument continues to hold.

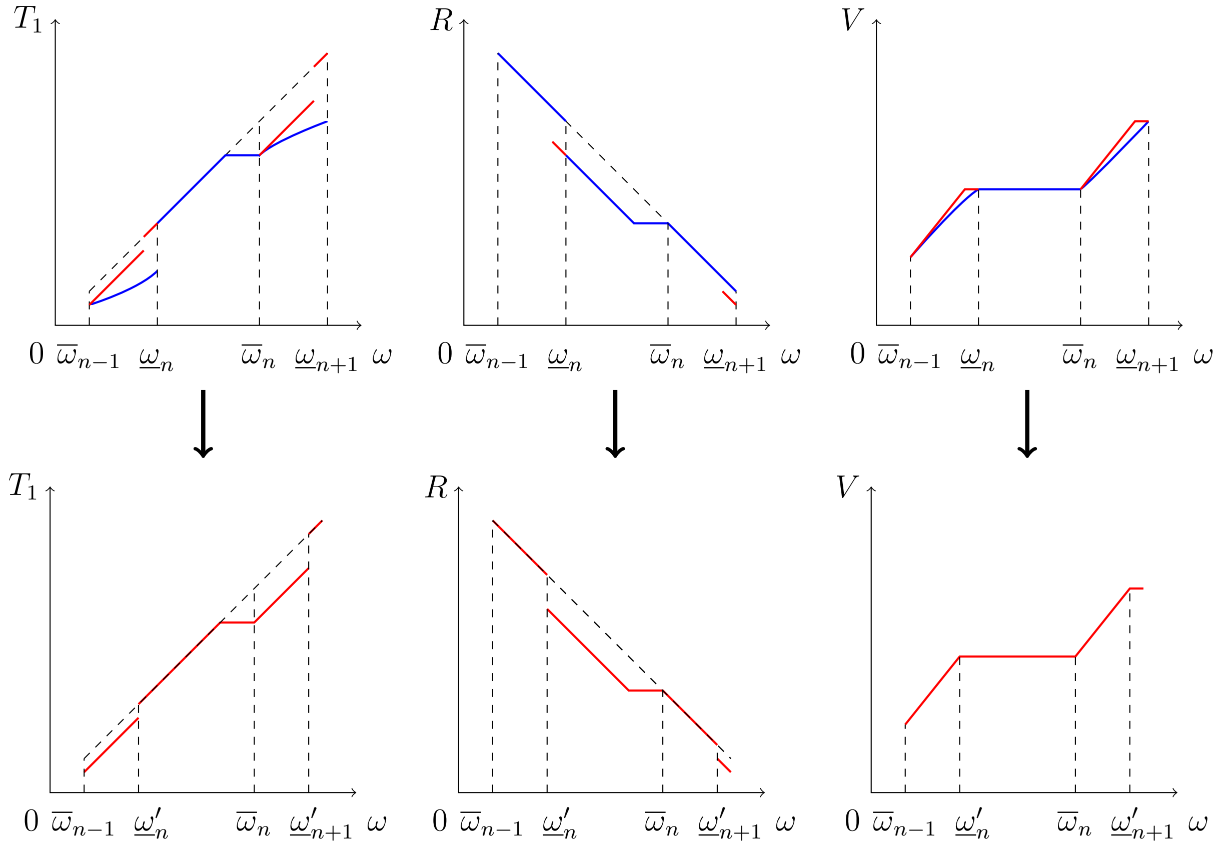


Figure 6: Improving a local buyback contract

of the cutoffs $\{\underline{\omega}_n, \overline{\omega}_n\}_{1\leq n\leq N}$ and t_1. To see this, one can compute that

$$V(\omega) = \begin{cases} t_n + v_r(q - \overline{\omega}_n) & \omega \in [\underline{\omega}_n, \overline{\omega}_n),\ 0 \leq n \leq N-1, \\ t_n + v_r(q - \overline{\omega}_n) + (p - v_r)(\omega - \overline{\omega}_n) & \omega \in [\overline{\omega}_n, \underline{\omega}_{n+1}),\ 0 \leq n \leq N-1, \\ t_N & \omega \in [\underline{\omega}_N, +\infty). \end{cases}$$

Continuity implies that

$$t_n + v_r(q - \overline{\omega}_n) + (p - v_r)(\underline{\omega}_{n+1} - \overline{\omega}_n) = t_{n+1} + v_r(q - \overline{\omega}_{n+1}),\ 1 \leq n \leq N-1,$$
$$t_{N-1} + v_r(q - \overline{\omega}_{N-1}) + (p - v_r)(\underline{\omega}_N - \overline{\omega}_{N-1}) = t_N,$$

which, by recursive substitution, can be simplified to

$$t_n = t_1 - p\overline{\omega}_1 + (p - v_r)\underline{\omega}_n + v_r \min(\overline{\omega}_n, q) - (p - v_r)\sum_{j=2}^{n-1}(\overline{\omega}_j - \underline{\omega}_j). \tag{15}$$

Thus, (b) and (c) of Definition 1 boil down to the recursive relationship (15) plus the following constraint for t_1:

$$t_1 \leq (p - v_r)\underline{\omega}_1 + v_r\overline{\omega}_1, \text{ with equality if } \underline{\omega} > 0. \tag{16}$$

As a result, both contracting parties' expected payoffs in any local buyback contract can be pinned down by the cutoffs $\{\underline{\omega}_n, \overline{\omega}_n\}_{1\leq n\leq N}$, t_1, and the quantity q. Thus, standard techniques for constrained optimization problems can be applied.

Let L be the Lagrangian of the manufacturer's optimization problem, and λ be the Lagrangian multiplier of the retailer's participation constraint (IR_r). If Γ maximizes $\text{E}_\omega[u_m(\omega)]$ subject to (IR_r), the sequence $\{\underline{\omega}_n, \overline{\omega}_n\}_{1\leq n\leq N}$, the parameters t_1, q, and λ must jointly be a stationary point of the Lagrangian

$$L = \text{E}_\omega[u_m(\omega)] + \lambda(\text{E}_\omega[u_r(\omega, \omega, \min(\omega, q))] - W - \underline{u}),$$

subject to (13)-(16), as well as the complementary slackness constraints for (IR_r):

$$\lambda \geq 0, \quad \lambda(\text{E}_\omega[u_r(\omega, \omega, \min(\omega, q))] - W - \underline{u}) = 0.$$

Omitting constant terms in L, we have:

$$L = \int_0^{+\infty} [(1 - \lambda)T_1(\omega) - (\lambda v_r - v_m)R(\omega)]\text{d}F(\omega) - cq. \tag{17}$$

It is now straightforward to complete the optimization problem. We start by discussing the range of λ. If $\lambda \leq v_m/v_r$, then L is strictly increasing in $\text{E}_\omega[R(\omega)]$ (unless $\lambda = v_m/v_r$) and $\text{E}_\omega[T_1(\omega)]$. If $\lambda \geq 1$, then L is strictly decreasing in $\text{E}_\omega[T_1(\omega)]$ (unless $\lambda = 1$) and $\text{E}_\omega[R(\omega)]$. In both cases, the Lagrangian has no interior stationary point. Therefore, $\lambda \in (v_m/v_r, 1)$, which immediately tells us that (IR_r) binds in any optimal contracts.

An examination of the first-order necessary conditions shows that the Lagrangian has an interior stationary point only when $N = 1$. The resulting contract thus has the following structure.

Definition 2. *A contract is a* ***buyback contract*** *if there exist a constant* $t \in [0, pq]$*, such* *that for any* ω*,*

$$T_1(\omega) = \min(p\omega, t);$$
$$R(\omega) = \begin{cases} q - \omega & \omega < \underline{\omega}, \\ \max((t - p\omega)/v_r, 0) & \omega \geq \underline{\omega}; \end{cases}$$

where $\underline{\omega} = \max\left(0, \frac{t - v_r q}{p - v_r}\right)$*.*

Put simply, Definition 2 comes from taking $N = 1$ and $\overline{\omega}_1 = q$ in Definition 1. We now have:

Proposition 2. *The optimal contract implementing* $q > \underline{q}$ *is a buyback contract with a binding* (IR_r)*.*

Proof. See Appendix A.2. □

To see the intuition why more than one interval in Definition 1 is not optimal, consider increasing both $\underline{\omega}_1$ and $\overline{\omega}_1$ by $\varepsilon > 0$ sufficiently small. According to the recursive relationship (15), this increases t_1 by εp while keeping all other t_ns unchanged. By Definition 1, these increments will reduce return shipments by ε and increase cash repayments by εp on the interval $[t_1/p, \overline{\omega}_1)$, and will increase return shipments by $\varepsilon(p - v_r)/v_r$ on the interval $[\underline{\omega}_1, t_1/p)$. The gross effect on the Lagrangian is, approximately,

$$\varepsilon[p(1-\lambda) + (\lambda v_r - v_m)][F(\overline{\omega}_1) - F(t_1/p)] - \frac{\varepsilon(p - v_r)(\lambda v_r - v_m)}{v_r}[F(t_1/p) - F(\underline{\omega}_1)].$$

Note that t_1/p is a convex combination of $\underline{\omega}_1$ and $\overline{\omega}_1$. Therefore, when F is not too concave, which is ensured by the nondecreasing hazard rate, the gross effect will be positive. This pushes $\overline{\omega}_1$ to its largest possible value and implies that the optimal local buyback contract must have $N = 1$ and $\overline{\omega}_1 = q$.

By Definition 2, in a buyback contract the retailer must repay t to the manufacturer in cash, regardless of the state of demand. If $p\omega \geq t$, the retailer can fulfill this obligation and no return is needed. If $p\omega < t$, return shipments must be added to make up the difference in value. This is as if the manufacturer buys back unsold inventories from the retailer.

Buyback price. Our analysis yields further insights into the pricing of optimal buyback contracts. In fact, there are two types of optimal buyback contracts. To see this, with a slight abuse of notation, let us write T_1 as a function of R, denoted by $T_1 = T_1(R)$. One can therefore define the *buyback price* $b(R)$ as

$$b(R) = \frac{t - T_1(R)}{R}.$$

Intuitively, $b(R)$ is the menu of per-unit prices offered by the manufacturer if the retailer returns R units of unsold inventory. The price is set to ensure that the retailer can repay exactly t in cash together with the returns R. Depending on the buyback price scheme, a buyback contract may now exhibit two different structures.

Constant buyback price. If $\underline{\omega} = 0$, then $T_1(\omega) = p\omega$ and $R(\omega) = (t - p\omega)/v_r$ for any $\omega < t/p$. The buyback price is a constant $b(R) = v_r$ for all R.

Variable buyback price. If $\underline{\omega} > 0$, then when $\omega \in [\underline{\omega}, t/p)$, the buyback price is still v_r, but when $\omega < \underline{\omega}$, $T_1(\omega) = p\omega$, $R(\omega) = q - \omega$, so $T_1(R) = p(q - R)$. The buyback price is therefore

$$b(R) = \frac{t - p(q - R)}{R} = p - \frac{pq - t}{R}.$$

which is increasing in R.

Proposition 2 shows that the pricing strategy of the optimal buyback contract depends on order quantity. In particular, let $\overline{q}$ be the solution for

$$S(q) + (c - v_r)q = W + \underline{u}. \tag{18}$$

Clearly, comparing (6) and (18), $\underline{q} < \overline{q}$. As the following proposition shows, the buyback price becomes variable if and only if the order quantity is larger than $\overline{q}$.

Proposition 3. *The optimal contract implementing $q > \underline{q}$*

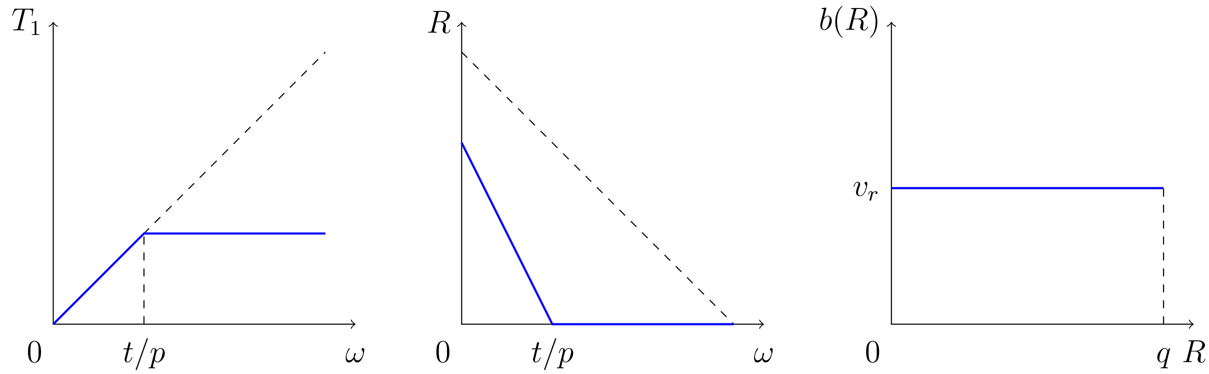
(a) has a constant buyback price when $\underline{q} < q \leq \overline{q}$;

(b) has a variable buyback price when $q > \overline{q}$.

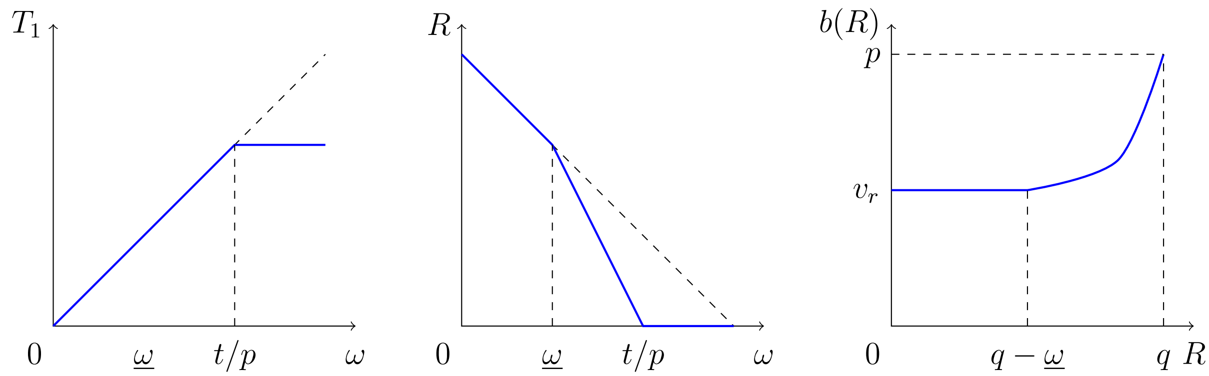
Proof. See Appendix A.3. □

Figure 7 provides a graphical illustration of Proposition 3.

The above analysis exhibits some interesting similarities and differences between our constructive approach and other ironing approaches in the literature. All these approaches start with taking the convex closure of certain components of a mechanism. The ironing approach in Myerson (1981), as well as other papers in screening, monopoly pricing, and auction, makes use of this convexity to obtain a monotone allocation rule, which is both necessary and sufficient for incentive-compatibility and thus serves as a natural base for



(a) Constant buyback price ($q \leq \overline{q}$)



(b) Variable buyback price ($q > \overline{q}$)

Figure 7: Two types of buyback contracts

overall optimization. However, in our model, incentive-compatibility itself does not imply monotonicity neither of T_1 and R, nor of the indirect utility U_r, nor their continuity. We use ironing to get a monotonic V, but additional structures of T_1 and R (the local buyback structure) are still needed for local optimality. In this sense, our approach can be applied to problems with incentive constraints weaker than that of the standard screening problems.

4.5 Optimal order quantity

Going back to Subsection 4.1, our final step is to investigate whether the manufacturer indeed benefits from implementing a quantity higher than $\underline{q}$.

Proposition 4. *(a) When $q^{FB} \leq \underline{q}$, the optimal contract is a wholesale contract implementing $q^* = q^{FB}$.*

(b) When $q^{FB} > \underline{q}$, the optimal contract is a buyback contract implementing a quantity $q^* \in (\underline{q}, q^{FB})$. Moreover, if $q^* > \overline{q}$, the buyback price is variable; if $q^* \leq \overline{q}$, the buyback price is constant.

Proof. See Appendix A.4. □

In a nutshell, when (6) holds, optimal contracts are wholesale contracts with the first-best quantity. When (6) fails, optimal contracts are buyback contracts, with inefficient returns of unsold inventories at date 1, which are necessary for ex-post truth-telling and incentivize the contracting parties to reduce the probability of oversupply ex ante. Consequently, the second-best quantity is smaller than the first-best. The date-0 cash repayment T_0 does not enter the manufacturer's objective function, thus it is irrelevant to the optimality of contracts.

The explicit characterization of the solution in the proof of Proposition 4 makes it possible to study the comparative statics regarding the retailer's initial cash holding W and reservation utility $\underline{u}$. Here, the sum $W + \underline{u}$ can be interpreted as a measure of the retailer's bargaining power.

Note that q^{FB} depends on neither W nor $\underline{u}$. By (6) and (18), the two cutoffs $\underline{q}$ and $\overline{q}$ are increasing in $W + \underline{u}$. Moreover, it is easy to show that the optimal order quantity q is weakly increasing in $W + \underline{u}$. This implies the following comparative statics with respect to $W + \underline{u}$, which can be easily proved by using the explicit expressions from the proof of Proposition 4.

Proposition 5. *There are two thresholds $W^{VB} > W^{CB} > 0$ such that*

(a) If $W + \underline{u} < W^{VB}$, the optimal contract implements $q^* > \overline{q}$ and is buyback with variable buyback pricing.

(b) If $W^{VB} < W + \underline{u} < W^{CB}$, the optimal contract implements $\underline{q} < q^* < \overline{q}$ and is buyback with constant buyback pricing.

(c) If $W + \underline{u} > W^{CB}$, the optimal contract implements $q^* = q^{FB}$ and is wholesale.

In words: When the retailer has a lot of liquidity and/or good alternative sources, i.e. when $W + \underline{u}$ is large, the optimal contract is a wholesale contract with a fixed date-1 cash transfer and zero return shipment. As the retailer's bargaining power decreases, the optimal contract shifts from wholesale to buyback, and the retailer must return inventory at a constant buyback price. When the retailer's bargaining power decreases further, the overall date-1 return obligation increases, and when it passes a certain threshold, the optimal contract becomes a buyback contract with variable pricing. The manufacturer then raises the buyback price in low-demand states to extract more revenue from the retailer in high-demand states.

This relationship between bargaining power and contract structures actually corresponds to observed practice. Large retailers such as Walmart or Target are less likely to delay payments to suppliers and can salvage their inventory directly as they face weaker financial constraints and better sourcing alternatives, while smaller groceries or bookstores often specify buyback terms in their contracts with producers. Our analysis provides further insights on the optimal pricing of these buyback contracts, which to our knowledge have not yet been investigated systematically. The existing supply chain contracting literature usually compares various contracts observed in practice (e.g., Cachon, 2003; Chen, 2003) and derives managerial implications. Our approach speaks directly to the question of contract optimality. Remarkably, even though salvaging unsold inventory at the retailer is more efficient, the manufacturer optimally can buy back some of it to alleviate the ex-post adverse selection problem.

5 Discussion and Extensions

The benchmark model of the previous sections is a bare-bones version of a standard contracting model in a vertical relationship. In this section we discuss several modifications and extensions and show how our technique and results can be adapted to these different scenarios.

5.1 Firesales

The limited liability constraint at date 1 results from the assumption that at the market price p the retailer has a limited market and may therefore be cash-constrained ex post. This is a fairly typical situation. Yet, as an alternative it may be instructive to consider a modified setting where the retailer has the option to salvage unsold inventories to a third party at a firesale price v_r *before* making payments to the manufacturer at date 1. This changes his liquidity constraint as follows.

Denoting by $s_f(\omega)$ the amount of inventory thus salvaged by the retailer, we have $s_f(\omega) \leq q - s(\omega)$. Without loss of generality we continue to assume that $T_0 = W$. Now the incentive-compatibility constraint becomes

$$ps(\omega) - T_1(\omega) + v_r[q - s(\omega) - R(\omega)] \geq p\hat{s} - T_1(\hat{\omega}) + v_r[q - \hat{s} - R(\hat{\omega})] \qquad (\text{IC}')$$

for all ω, $\hat{\omega}$, $\hat{s}$, and $\hat{s}_f$ such that

$$0 \leq \hat{s}_f \leq q - \hat{s} \tag{IC$'$-FS$_f$}$$

$$0 \leq \hat{s} \leq \min(\omega, q) \tag{IC$'$-FS}$$

$$0 \leq R(\hat{\omega}) \leq q - \hat{s} - \hat{s}_f \tag{IC$'$-FR}$$

$$T_1(\hat{\omega}) \leq p\hat{s} + v_r \hat{s}_f. \tag{IC$'$-FT$_1$}$$

Note that neither s_f nor $\hat{s}_f$ enters (IC$'$). They only affect (IC$'$-FS$_f$), (IC$'$-FR), and (IC$'$-FT$_1$).

Two observations are key for understanding the ex-post problem in this setting. First, suppose that (s, s_f, T_1, R) is feasible in the model with firesales. Define $\tilde{T}_1 = T_1 - v_r s_f$ and $\tilde{R} = R + s_f$. Then $(s, \tilde{T}_1, \tilde{R})$ is feasible in the model *without* firesales. To see this, note that the only function of s_f is to enlarge the set of possible deviations for the retailer. Allowing for firesales therefore actually tightens his incentive-compatibility constraint. On the other hand and second, firesales conducted by the retailer are more efficient than salvaging by the manufacturer ($v_r > v_m$). Therefore, to the extent that a contract in the baseline model uses salvaging by the manufacturer with positive probability, contracts in the extended model will try to avoid this.

This implies that the optimal contract of Propositions 4 and 5 will be optimal in the extended model if $W + \underline{u} > W^{CB}$, because it does not make use of inefficient liquidations. For $W + \underline{u} < W^{CB}$, however, the optimal contract of Section 4 does use inefficient liquidation. Hence, in the extended model there is a tradeoff between truth-telling incentives through returns and efficiency achieved by retail firesales. Note that the optimal contract of Section 4 is feasible and incentive-compatible in the extended model, because the optimal buyback price is always (weakly) greater than the firesale price v_r. Hence, under these contracts the retailer never has a (strict) incentive to use firesales as an instrument to generate ex-post cash. However, this contract is no longer optimal. By arguments similar to those in the previous section, one can show that the tradeoff is optimally resolved by a "revised" buyback contract that combines (inefficient) returns and (efficient) retail salvaging. Without proof, we provide this result in the following proposition.

Proposition 6. *In the model in which the retailer can use firesales to generate cash, if a contract with date-1 component* (s, s_f, T_1, R) *is optimal for implementing* $q > \underline{q}$ *then there* *exists a constant* $t \in [0, pq]$*, such that for any* ω*,*

$$T_1(\omega) = \begin{cases} p\omega & \text{if } \omega < \underline{\omega}, \\ t & \text{if } \omega \geq \underline{\omega}; \end{cases}$$

$$R(\omega) = \begin{cases} q - \omega & \text{if } \omega < \underline{\omega}, \\ 0 & \text{if } \omega \geq \underline{\omega}; \end{cases}$$

$$s_f(\omega) = \begin{cases} (t - p\omega)/v_r & \text{if } \omega \in [\underline{\omega}, t/p), \\ 0 & \text{otherwise.} \end{cases}$$

where $\underline{\omega} = \max\left(0, \frac{t - v_r q}{p - v_r}\right)$*, and* ($\text{IR}_r$) *binds.*

Comparing Proposition 6 with Proposition 2, one can see that the two optimal contracts are distinct only when $\omega \in [\underline{\omega}, t/p)$, which corresponds to a positive s_f. Since $t > p\underline{\omega}$, the manufacturer's utility from the optimal contract with firesales is always strictly higher than the optimal contract without firesales. But although returns are inefficient, the optimal contract continues to use them as a truth-telling incentive if $\underline{\omega} > 0$, i.e. if $W + \underline{u}$ is small. If $W + \underline{u}$ is larger, such that $\underline{\omega} = 0$, the optimal contract uses no returns and essentially mimics the efficiency properties of a wholesale contract by means of retail firesales.

5.2 Price-dependent demand

The model described in Section 2 has assumed that the retail price p is exogenous. This is probably a good assumption if the manufacturer is sufficiently remote and unacquainted with the retailer's local market, and if that market is sufficiently competitive. Alternatively, the retail price could be contractible and therefore endogenous to the contracting problem. In this section, we therefore relax our restriction and allow for a price-dependent stochastic demand function $F(\cdot; p)$. This parameterized-distribution-function approach provides greater flexibility than standard state-space models and encompasses different specific state-space formulations.[16] That pricing is determined before the demand realization is consistent with the long-standing literature on price-setting newsvendor problems (Petruzzi and Dada, 1999).

[16] As an example, consider the basic demand model $Q = Q(p, \theta)$ where $Q \geq 0$ is market demand and $\theta \in \Theta$ a random variable with probability measure μ. For any p and $0 \leq \omega_1 < \omega_2$, we have $0 \leq \mu(\{\theta | Q(p;\theta) \leq \omega_1\}) \leq \mu(\{\theta | Q(p;\theta) \leq \omega_2\}) \leq 1$. Hence, $F(\omega; p) = \mu(\{\theta | Q(p;\theta) \leq \omega\})$ is a well-defined family of distribution functions. Other examples can easily be constructed.

We assume that $F(\cdot;\cdot)$ is atomless and differentiable in both ω and p, and write

$$F_p(\omega;p) = \frac{\partial F(\omega;p)}{\partial p},\ f(\omega;p) = \frac{\partial F(\omega;p)}{\partial \omega} > 0.$$

By definition, $F_p(\cdot;p)$ is the marginal effect of price on the distribution of demand, $f(\cdot;p)$ is the density function of F given price p.[17] In line with traditional models, such as the state-space model sketched above, we assume that $F(\cdot;p)$ satisfies first-order stochastic dominance, i.e., for any p and $\omega > 0$, $F_p(\omega;p) > 0$. This assumption implies that retail demand is more likely to be higher when the price is lower. To avoid unbounded solutions, we also assume that $\lim_{p\to\infty} F(\omega;p) = 1$ for all $\omega > 0$. That is, when the price is sufficiently high, retail demand becomes arbitrarily small.

Expected feasible demand and social surplus are as in Section 3,

$$Q(q;p) = q - \int_0^q F(\omega;p)d\omega,$$
$$S(q;p) = (p - v_r)Q(q;p) - (c - v_r)q,$$

respectively. To avoid excessive technical details, we assume that social surplus is concave in p and q, and $S_{pq}(q;p) \geq 0$. Assuming that p is observable and contractible, the definition of a retail contract must be extended to $\Gamma = (p, q, T_0, s, T_1, R)$.

When information is symmetric, the first-order conditions for maximizing surplus are

$$F(q;p) = \frac{p-c}{p-v_r}, \tag{19}$$

$$Q(q;p) = -(p - v_r)Q_p(q;p). \tag{20}$$

The concavity of S ensures that these conditions are sufficient and the solution unique. Denote the solution by (p^{FB}, q^{FB}). Then the first-best contract generates surplus $S(q^{FB};p^{FB})$.

When information is asymmetric, we maintain the assumption in Section 4.4 that F has a nondecreasing hazard rate, i.e., given any p, $f(\omega;p)/[1 - F(\omega;p)]$ is nondecreasing in ω. Since p is observable and determined before the realization of demand, it continues to be true that optimal contracts must be either wholesale or buyback. Following an argument similar to Lemma 1, for any p, let $\underline{q}(p)$ be the largest solution of

$$S(\underline{q}(p);p) + c\underline{q}(p) \leq W + \underline{u}. \tag{21}$$

$\underline{q}(p)$ is the maximum quantity that can be implemented by a wholesale contract (with full

[17] We will use subscripts p and q to denote partial derivatives throughout this section.

surplus extraction) under price p. As a result, the first-best quantity q^{FB} can be implemented by a wholesale contract under the first-best price p^{FB} if and only if $q^{FB} \leq \underline{q}(p^{FB})$.

Now we consider any p and hold it fixed. To implement a $q > \underline{q}(p)$, the parties have to turn to a buyback contract. In this case, the optimal price and quantity are characterized in Proposition 7.

Proposition 7. *Suppose the retail price is endogenous and contractible. Then*

(a) when $q^{FB} \leq \underline{q}(p^{FB})$, the optimal Γ is a wholesale contract implementing q^{FB} at a price p^{FB};

(b) when $q^{FB} > \underline{q}(p^{FB})$, the optimal Γ is a buyback contract implementing $q^* < q^{FB}$ at a price $p^* < p^{FB}$.

Proof. See Appendix B.2. □

Intuitively, the efficiency loss in a buyback contract comes from return shipments, so the manufacturer is more reluctant to have "excess supply" rather than "excess demand". Consequently, she will deliver less products ex-ante and request a lower retail price to reduce the probability of oversupply. This logic also applies to Proposition 4.

5.3 Multiple retailers

It is common in practice that a manufacturer sells her products through different retailers. The manufacturer may want to maintain a relatively high retail price for her products, but retailers usually compete with each other and attract customers by cutting down retail prices. As a result, the manufacturer sometimes fixes the retail price through contracts. This mechanism is the so-called Resale Price Maintenance (RPM) that has been well studied in the literature (e.g., Marvel and McCafferty, 1984; Shaffer, 1991; Deneckere et al., 1996; Jullien and Rey, 2007; Asker and Bar-Isaac, 2014) and intensively discussed in legal practice.[18] However, there is still fierce debate about whether RPM is anti-competitive and should be prohibited by policymakers. In this section, we extend our benchmark model to allow for multiple retailers and see whether downstream competition changes the manufacturer's incentive to control retail price and quantity.

Consider an environment that is identical to the benchmark model in Section 2 with the only exception that now there are n symmetric retailers, indexed by superscript $j \in \{1, 2, \ldots, n\}$. At date 0, the manufacturer offers a contract to each retailer. The contract for

[18]See, e.g., Leegin Creative Leather Products, Inc. v. PSKS, Inc., dba Kay's Kloset...Kay's Shoes, 551 U.S. 877 (2007). https://www.supremecourt.gov/opinions/06pdf/06-480.pdf.

retailer j specifies the date-0 price p^j, quantity q^j, cash transfer T_0^j, the date-1 cash repayment T_1^j and the return shipment R^j. The last two components are contingent on retailer j's report $\hat{\omega}^j$.[19] Retailers then decide whether to accept their corresponding contracts simultaneously. At date 1, the retail demand ω is realized, and retailers make their reports. In the spirit of Kreps and Scheinkman (1983), we assume that demand is allocated according to efficient rationing, and when some retailers post the same price, their allocated demand should be equal. Moreover, the distribution of demand $F(\omega; p)$ is determined by the highest price in the market, i.e., $\max\{p^1, p^2, \ldots, p^n\}$. We say that a collection of contracts $\Gamma^1, \Gamma^2, \ldots, \Gamma^n$ are optimal if they maximize the manufacturer's profits subject to all the constraints listed in Section 2.

Optimal contracts are then characterized by Proposition 8.

Proposition 8. *If $\Gamma^1, \Gamma^2, \ldots, \Gamma^n$ are optimal, then they are identical. Moreover, let*

$$p^* = p^1,\ q^* = nq^1,\ T_0^* = nT_0^1,\ T_1^*(\omega) = nT_1^1(\omega),\ R^*(\omega) = nR^1(\omega).$$

Then $\Gamma^* = (p^*, q^*, T_0^*, T_1^*, R^*)$ is optimal when there is only one retailer with initial wealth nW and reservation utility $n\underline{u}$.

Proof. See Appendix B.3. □

According to Proposition 8, optimal contracts with multiple retailers are closely related to the optimal buyback contract in the single-retailer model. It is as if that retailers are merged together before contracting with the manufacturer. Therefore, by Proposition 4, the structure of optimal contracts as well as the equilibrium price and quantity depends on $n\underline{u}$. In particular, by part (a) of Proposition 4, the first-best quantity will be implemented when the single retailer's reservation utility is sufficiently high, which translates into sufficiently many retailers in the present model. We formally state this result in Corollary 1.

Corollary 1. *When n is sufficiently large, the manufacturer distributes the first-best quantity q^{FB} evenly to all retailers. In this case, her profits decreases with n.*

Corollary 1 describes the effect of competition under RPM. Since the manufacturer fully controls the retail price through contracts, she distributes her products equally among retailers. As the number of retailers increases, the manufacturer has to produce more to make sure that each retailer receives at least $\underline{u}$. The total supply thus increases to the first-best level q^{FB}. After this point, the total supply becomes constant, so the manufacturer's profit decreases as competition becomes more intensive.

[19]For simplicity, we assume that retailer j's contract cannot depend on the other retailer's report.

6 Concluding Remarks

In this paper, we have analyzed optimal contracts in the archetypal model of a supply chain between a manufacturer and retailer with demand-supply mismatch. If downstream demand is private information, the optimal contract takes the form of either a wholesale or a buyback contract, which provides a unified microeconomic foundation for retail contracts. We generalize existing analytical tools by combining the ironing approach of Myerson (1981) and others with global Lagrangian optimization tools and can thus study the comparative statics of the optimal contract explicitly.

The model of this paper is very simple and already presents a number of technical difficulties. We have discussed a few extensions, which can be accommodated in our modelling framework and solved with the techniques developed here. Generalizations to more complex supply chains with richer interactions will require further work. An example is the problem of supply chain coordination, to alleviate the problem of demand-supply mismatch. Our simple assumption has been that production precedes sales, which is certainly appropriate for many production-in-advance industries, in particular when geographic distance is important (see, e.g., Ganapati and Wong (2023)). But in other situations some coordination may be feasible, for example through reordering additional units or using wholesale intermediaries. This type of model would introduce a more intricate dynamic dimension into the problem and link it to the theory of dynamic screening models. Further developments of the theory presented here should address the relation between supply chain contracts and financial contracts, the determination of the optimal size of a chain, or the question of how vertical relationships, demand volatility, and inventory management affect the market structure of the retail sector. In this respect, (e.g., Hortaçsu and Syverson, 2015) leaves us a promising research agenda of combining theory and practice in the future.

References

Arya, A. and B. Mittendorf (2004): "Using Return Policies to Elicit Retailer Information," *RAND Journal of Economics*, 35, 617–630.

Asker, J. and H. Bar-Isaac (2014): "Raising Retailers' Profits: On Vertical Practices and the Exclusion of Rivals," *American Economic Review*, 104, 672–686.

Babich, V., H. Li, P. Ritchken, and Y. Wang (2012): "Contracting with Asymmetric Demand Information in Supply Chains," *European Journal of Operational Research*, 217, 333–341.

Bagnoli, M. and T. Bergstrom (2005): "Log-Concave Probability and its Applications," *Economic Theory*, 26, 445–469.

Baron, D. P. and R. B. Myerson (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50, 911–930.

Blair, B. F. and T. R. Lewis (1994): "Optimal Retail Contracts with Asymmetric Information and Moral Hazard," *RAND Journal of Economics*, 25, 284–296.

Brander, J. A. and T. R. Lewis (1986): "Oligopoly and Financial Structure: The Limited Liability Effect," *American Economic Review*, 76, 956–970.

Burns, M. E. and C. Walsh (1981): "Market Provision of Price-excludable Public Goods: A General Analysis," *Journal of Political Economy*, 89, 166–91.

Cachon, G. P. (2003): "Supply Chain Coordination with Contracts," in *Handbooks in Operations Research and Management Science*, ed. by A. G. de Kok and S. C. Graves, Elsevier, vol. 11, 227–339.

Chen, F. (2003): "Information Sharing and Supply Chain Coordination," in *Handbooks in Operations Research and Management Science*, ed. by A. G. de Kok and S. C. Graves, Elsevier, vol. 11, 341–421.

Davidson, C. and R. Deneckere (1986): "Long-Run Competition in Capacity, Short-Run Competition in Price, and the Cournot Model," *RAND Journal of Economics*, 17, 404–415.

Deneckere, R., H. P. Marvel, and J. Peck (1996): "Demand Uncertainty, Inventories, and Resale Price Maintenance," *Quarterly Journal of Economics*, 111, 885–913.

——— (1997): "Demand Uncertainty and Price Maintenance: Markdowns as Destructive Competition," *American Economic Review*, 87, 619–641.

Deneckere, R. and J. Peck (1995): "Competition Over Price and Service Rate When Demand is Stochastic: A Strategic Analysis," *RAND Journal of Economics*, 26, 148–162.

Gale, D. and M. F. Hellwig (1985): "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, 52, 647–663.

Ganapati, S. and W. F. Wong (2023): "How Far Goods Travel: Global Transport and Supply Chains from 1965 – 2020," *Journal of Economic Perspectives*, 37, 3–30.

Guesnerie, R. and J.-J. Laffont (1984): "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329–369.

Gui, Z., E.-L. von Thadden, and X. Zhao (2019): "Incentive-Compatibility, Limited Liability and Costly Liquidation in Financial Contracting," *Games and Economic Behavior*, 118, 412–433.

Hellwig, M. F. (2010): "Incentive Problems with Unidimensional Hidden Characteristics: A Unified Approach," *Econometrica*, 78, 1201–1237.

Holmström, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74–91.

Hortaçsu, A. and C. Syverson (2015): "The Ongoing Evolution of US Retail: A Format Tug-of-War," *Journal of Economic Perspectives*, 29, 89–112.

Hsieh, C.-C., C.-H. Wu, and Y.-J. Huang (2008): "Ordering and Pricing Decisions in a Two-Echelon Supply Chain with Asymmetric Demand Information," *European Journal of Operational Research*, 190, 509–525.

Jullien, B. and P. Rey (2007): "Resale Price Maintenance and Collusion," *RAND Journal of Economics*, 38, 983–1001.

Kreps, D. M. and J. A. Scheinkman (1983): "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *Bell Journal of Economics*, 14, 326–337.

Kumar, S. and G. Srinivasan (2007): "Manufacturer's Pricing Strategies in a Single-Period Framework under Price-Dependent Stochastic Demand with Asymmetric Risk-Preference Information," *Journal of the Operational Research Society*, 58, 1449–1458.

Lacker, J. M. and J. A. Weinberg (1989): "Optimal Contracts under Costly State Falsification," *Journal of Political Economy*, 97, 1345–1363.

Loertscher, S. and E. V. Muir (2022): "Monopoly Pricing, Optimal Randomization, and Resale," *Journal of Political Economy*, 130, 566–635.

Maggi, G. (1996): "Strategic Trade Policies with Endogenous Mode of Competition," *American Economic Review*, 86, 237–58.

Marvel, H. P. and S. McCafferty (1984): "Resale Price Maintenance and Quality Certification," *RAND Journal of Economics*, 15, 346–359.

Marvel, H. P. and J. Peck (1995): "Demand Uncertainty and Returns Policies," *International Economic Review*, 36, 691–714.

Milgrom, P. and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

Mirrlees, J. A. (1974): "Optimum Accumulation Under Uncertainty: The Case of Stationary Returns to Investment," in *Allocation Under Uncertainty: Equilibrium and Optimality*, ed. by J. H. Drèze, Springer, 36–50.

Montez, J. (2015): "Controlling Opportunism in Vertical Contracting when Production Precedes Sales," *RAND Journal of Economics*, 46, 650–670.

Montez, J. and N. Schutz (2021): "All-Pay Oligopolies: Price Competition with Unobservable Inventory Choices," *Review of Economic Studies*, 88, 2407–2438.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Petruzzi, N. C. and M. Dada (1999): "Pricing and the Newsvendor Problem: A Review with Extensions," *Operations Research*, 47, 183–194.

Rey, P. and J. Tirole (1986): "The Logic of Vertical Restraints," *American Economic Review*, 76, 921–939.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton University Press.

Rudin, W. (1987): *Real and Complex Analysis*, McGraw-Hill.

Shaffer, G. (1991): "Slotting Allowances and Resale Price Maintenance: A Comparison of Facilitating Practices," *RAND Journal of Economics*, 22, 120–135.

Shen, B., T.-M. Choi, and S. Minner (2019): “A Review on Supply Chain Contracting with Information Considerations: Information Updating and Information Asymmetry,” *International Journal of Production Research*, 57, 4898–4936.

Taylor, T. A. and W. Xiao (2009): “Incentives for Retailer Forecasting: Rebates vs. Returns,” *Management Science*, 55, 1654–1669.

Townsend, R. M. (1979): “Optimal Contracts and Competitive Markets with Costly State Verification,” *Journal of Economic Theory*, 21, 265–293.

Wang, S., H. Gurnani, and U. Subramanian (2020): “The Informational Role of Buy-back Contracts,” *Management Science*, 67, 279–296.

Yue, X. and S. Raghunathan (2007): “The Impacts of the Full Returns Policy on a Supply Chain with Information Asymmetry,” *European Journal of Operational Research*, 180, 630–647.

Appendix

A.1 Proof of Lemma 3

For any admissible Γ, denote by $A = \{\omega : s(\omega) < \min(\omega, q)\}$ the set of states ω in which sales are not maximal. Suppose that A has a positive measure.

Ignoring some measure-theoretic fine points, choose a measurable function $\varepsilon(\omega) > 0$ on A and an alternative contract $\tilde{\Gamma}$ such that for all $\omega \in A$, $\tilde{s}(\omega) = s(\omega) + \varepsilon(\omega) < \min(\omega, q)$, $\tilde{R}(\omega) = R(\omega) - \varepsilon(\omega) > 0$,[20] and $\tilde{T}_1(\omega)$ increases $T_1(\omega)$ in a way that keeps the retailer's utility unchanged:

$$\tilde{T}_1 = (p - v_r)(\tilde{s} - s) + v_r(R - \tilde{R}) + T_1 = p\varepsilon + T_1.$$

Thus, for any $\omega \in A$, $\tilde{V} = \tilde{T}_1 + v_r\tilde{R} = V + (p - v_r)\varepsilon$. By construction, the new contract satisfies the feasibility constraints (FS), (FR), and (FT_1). We now use Lemma 2 to verify that it is incentive-compatible.

For $\omega \in A$ and $\hat{\omega} \notin A$, note that $\tilde{V}(\omega) - V(\omega) = (p - v_r)(\tilde{s}(\omega) - s(\omega))$, so (IC1) and (IC2) of Lemma 2 follow from the original contract. For $\omega \notin A$ and $\hat{\omega} \in A$, note that $\tilde{V}(\hat{\omega}) - V(\hat{\omega}) = (p - v_r)(\tilde{R}(\hat{\omega}) - R(\hat{\omega}))$, so (IC1) and (IC2) hold when $R(\hat{\omega}) - (q - \min(\omega, q))$ is of the same sign as $\tilde{R}(\hat{\omega}) - (q - \min(\omega, q))$. When $R(\hat{\omega}) > q - \min(\omega, q) > \tilde{R}(\hat{\omega})$, we have

$$\begin{aligned} V(\omega) + (p - v_r)(\min(\omega, q) - s(\omega)) &\leq V(\omega) + (p - v_r)(q - \tilde{R}(\hat{\omega}) - s(\omega)) \\ &\leq V(\hat{\omega}) + (p - v_r)\varepsilon(\omega) \\ &= \tilde{V}(\hat{\omega}). \end{aligned}$$

For $\omega, \hat{\omega} \in A$, (IC1) and (IC2) follow from combining the previous two cases. The case when $\omega, \hat{\omega} \notin A$ is trivial. Thus

The retailer's utility is unchanged in $\tilde{\Gamma}$, so $\tilde{\Gamma}$ satisfies (IR_r). Since the extra revenue from $\tilde{\Gamma}$ goes entirely to the manufacturer, she is strictly better off:

$$\tilde{T}_1 + v_m\tilde{R} = T_1 + p\varepsilon + v_m\tilde{R} > T_1$$

for all $\omega \in A$. Therefore, Γ is strictly dominated by $\tilde{\Gamma}$, a contradiction.

Hence, any optimal contract has $s(\omega) = \min(\omega, q)$ for almost all ω. The piecewise continuity of T_1 and R implies that this is true pointwise.

[20] Note that by construction of A, $R(\omega) \geq q - s(\omega) > \max(0, q - \omega) \geq 0$.

A.2 Proof of Proposition 2

Suppose that $N \geq 2$. We first compute partial derivatives for the Lagrangian by ignoring the recursive relationship (15). For $1 \leq n \leq N-1$,

$$\frac{\partial L}{\partial \underline{\omega}_n} = -\left[(1-\lambda)(p\overline{\omega}_{n-1} - t_{n-1}) + (\lambda v_r - v_m)\left(\overline{\omega}_n - \underline{\omega}_n - \frac{t_n - p\underline{\omega}_n}{v_r}\right)\right] f(\underline{\omega}_n),$$
$$\frac{\partial L}{\partial \overline{\omega}_n} = -(1-\lambda)p[F(\underline{\omega}_{n+1}) - F(\overline{\omega}_n)] + (\lambda v_r - v_m)[F(\overline{\omega}_n) - F(\underline{\omega}_n)],$$
$$\frac{\partial L}{\partial t_n} = (1-\lambda)[F(\underline{\omega}_{n+1}) - F(t_n/p)] - \frac{\lambda v_r - v_m}{v_r}[F(t_n/p) - F(\underline{\omega}_n)],$$

and for $n = N$,

$$\frac{\partial L}{\partial \underline{\omega}_N} = -\left[(1-\lambda)(p\overline{\omega}_{N-1} - t_{N-1}) + (\lambda v_r - v_m)\left(q - \underline{\omega}_N - \frac{t_N - p\underline{\omega}_N}{v_r}\right)\right] f(\underline{\omega}_N),$$
$$\frac{\partial L}{\partial t_N} = (1-\lambda)[1 - F(t_N/p)] - \frac{\lambda v_r - v_m}{v_r}[F(t_N/p) - F(\underline{\omega}_N)].$$

After accounting for the recursive relationship (15), two cases are to be discussed.

Case 1. Suppose that $\underline{\omega}_1 > 0$. Then $t_1 = (p - v_r)\underline{\omega}_1 + v_r\overline{\omega}_1$. Inserting this to the expression of $\partial L/\partial \underline{\omega}_1$ gives us $\partial L/\partial \underline{\omega}_1 = 0$. Moreover, $\underline{\omega}_1$ has an interior solution, so

$$\frac{\mathrm{d}L}{\mathrm{d}\underline{\omega}_1} = \frac{\partial L}{\partial \underline{\omega}_1} + (p - v_r)\sum_{n=1}^{N}\frac{\partial L}{\partial t_n} = 0 \implies \sum_{n=1}^{N}\frac{\partial L}{\partial t_n} = 0.$$

Therefore,

$$\frac{\mathrm{d}L}{\mathrm{d}\overline{\omega}_1} = \frac{\partial L}{\partial \overline{\omega}_1} + p\frac{\partial L}{\partial t_1} - (p - v_r)\sum_{n=1}^{N}\frac{\partial L}{\partial t_n} = \frac{\partial L}{\partial \overline{\omega}_1} + p\frac{\partial L}{\partial t_1}.$$

Case 2. Suppose that $\underline{\omega}_1 = 0$. Then t_1 becomes an independent variable, so

$$\frac{\mathrm{d}L}{\mathrm{d}t_1} = \sum_{n=2}^{N}\frac{\partial L}{\partial t_n}.$$

Therefore,

$$\frac{\mathrm{d}L}{\mathrm{d}\overline{\omega}_1} = \frac{\partial L}{\partial \overline{\omega}_1} - p\sum_{n=2}^{N}\frac{\partial L}{\partial t_n} = \frac{\partial L}{\partial \overline{\omega}_1} + p\frac{\partial L}{\partial t_1}.$$

In both cases,

$$\begin{aligned}\frac{\mathrm{d}L}{\mathrm{d}\overline{\omega}_1} &= \frac{\partial L}{\partial \overline{\omega}_1} + p\frac{\partial L}{\partial t_1} \\ &= [(1-\lambda)p + (\lambda v_r - v_m)][F(\overline{\omega}_n) - F(t_n/p)] - \frac{(p-v_r)(\lambda v_r - v_m)}{v_r}[F(t_n/p) - F(\underline{\omega}_n)].\end{aligned} \tag{A.1}$$

Observe that

$$\frac{\partial L}{\partial q} = -(\lambda v_r - v_m)F(q) + (1-\lambda)p[1-F(q)] - c.$$

Since $\partial L/\partial q = 0$ is necessary for optimality, we have

$$1 - F(q) = \frac{c + \lambda v_r - v_m}{(1-\lambda)p + \lambda v_r - v_m}. \tag{A.2}$$

By nondecreasing hazard rate, for any $\omega \in [t_1/p, \overline{\omega}_1]$,

$$f(\omega) \geq \frac{[1-F(\omega)]f(t_1/p)}{1-F(t_1/p)} \geq \frac{[1-F(q)]f(t_1/p)}{1-F(t_1/p)},$$

which implies that

$$F(\overline{\omega}_1) - F(t_1/p) = \int_{t_1/p}^{\overline{\omega}_1} f(\omega)\mathrm{d}\omega \geq \frac{(\overline{\omega}_1 - t_1/p)[1-F(q)]f(t_1/p)}{1-F(t_1/p)}. \tag{A.3}$$

Similarly, for any $\omega \in [\underline{\omega}_1, t_1/p]$,

$$f(\omega) \leq \frac{f(t_1/p)[1-F(\omega)]}{1-F(t_1/p)} \leq \frac{f(t_1/p)}{1-F(t_1/p)},$$

which implies that

$$F(t_1/p) - F(\underline{\omega}_1) = \int_{\underline{\omega}_1}^{t_1/p} f(\omega)\mathrm{d}\omega \leq \frac{(t_1/p - \underline{\omega}_1)f(t_1/p)}{1-F(t_1/p)}. \tag{A.4}$$

Plugging (A.2), (A.3), and (A.4) into (A.1) yields

$$\frac{\mathrm{d}L}{\mathrm{d}\overline{\omega}_1} \geq \frac{c(p-v_r)(\overline{\omega}_1 - \underline{\omega}_1)f(t_1/p)}{p[1-F(t_1/p)]} > 0,$$

which contradicts $N \geq 2$. Hence, an optimal contract must have $N = 1$ and a binding (IR_r).

A.3 Proof of Proposition 3

If the buyback price is constant, the manufacturer's expected utility is

$$\begin{aligned}
\mathrm{E}_{\omega}\, u_m(\omega) &= W - cq + \int_0^{t/p} \left[p\omega + v_m \left(\frac{t - p\omega}{v_r} \right) \right] dF(\omega) + \int_{t/p}^{+\infty} t dF(\omega) \\
&= W - cq + t - \left(1 - \frac{v_m}{v_r} \right) p(t/p - Q\,(t/p)),
\end{aligned}$$

where t is determined by a binding (IR_r),

$$\begin{aligned}
\mathrm{E}_{\omega}\, u_r(\omega, \omega, \min(\omega, q)) &= S(q) + cq - t \\
&= W + \underline{u}.
\end{aligned} \tag{A.5}$$

Hence,

$$\mathrm{E}_{\omega}\, u_m(\omega) = S(q) - \underline{u} - \left(1 - \frac{v_m}{v_r} \right) p(t/p - Q\,(t/p)). \tag{A.6}$$

If the buyback price is variable, the manufacturer's expected utility is

$$\begin{aligned}
\mathrm{E}_{\omega}\, u_m(\omega) = W - cq &+ \int_0^{\underline{\omega}} [p\omega + v_m(q - \omega)] dF(\omega) \\
&+ \int_{\underline{\omega}}^{t/p} \left[p\omega + v_m \left(\frac{t - p\omega}{v_r} \right) \right] dF(\omega) + \int_{t/p}^{+\infty} t dF(\omega),
\end{aligned}$$

where t is determined by a binding (IR_r),

$$\begin{aligned}
\mathrm{E}_{\omega}\, u_r(\omega, \omega, \min(\omega, q)) &= S(q) + cq - \int_0^{\underline{\omega}} [p\omega + v_r(q - \omega)] \mathrm{d}F(\omega) - \int_{\underline{\omega}}^{+\infty} t \mathrm{d}F(\omega) \\
&= W + \underline{u}.
\end{aligned} \tag{A.7}$$

Hence,

$$\mathrm{E}_{\omega}\, u_m(\omega) = S(q) - \underline{u} - \left(1 - \frac{v_m}{v_r} \right) \left[p(t/p - Q\,(t/p)) - (p - v_r)(\underline{\omega} - Q(\underline{\omega})) \right]. \tag{A.8}$$

The cutoff between the case of constant price and that of variable price can be derived from taking $\underline{\omega} \to 0$ (and thus $t \to v_r q$) on the right-hand side of (A.5) (or equivalently the right-hand side of (A.7)). Both approaches will give us (18). Then, $\underline{\omega} = 0$ if and only if $q \leq \overline{q}$, and $\underline{\omega} > 0$ if and only if $q > \overline{q}$. It is immediate from (6) and (18) that $\underline{q} < \overline{q}$.

A.4 Proof of Proposition 4

Note that if the manufacturer offers a wholesale contract implementing $\underline{q}$, her payoff is simply $W - c\underline{q}$. If she offers a buyback contract that implements $q \in (\underline{q}, \overline{q}]$, her payoff is determined by (A.6). The first-order derivative of $\mathrm{E}_\omega\, u_m(\omega)$ is

$$\begin{aligned}\frac{\mathrm{d}\,\mathrm{E}_\omega\, u_m(\omega)}{\mathrm{d}q} &= S'(q) - \left(1 - \frac{v_m}{v_r}\right)[1 - Q'(t/p)]\, t'(q) \\ &= S'(q) - \left(1 - \frac{v_m}{v_r}\right) F(t/p)[p - (p - v_r)F(q)]. \end{aligned} \tag{A.9}$$

In the second equality, we use $t'(q) = S'(q) + c = v_r + (p - v_r)(1 - F(q))$, which comes from (A.5). When $q \to \underline{q}$ from the right, we have $t \to 0$, and, more importantly,

$$\lim_{q \to \underline{q}+} \mathrm{E}_\omega\, u_m(\omega) = S(\underline{q}) - \underline{u} = W - c\underline{q},$$
$$\lim_{q \to \underline{q}+} \frac{\mathrm{d}\,\mathrm{E}_\omega\, u_m(\omega)}{\mathrm{d}q} = p - (p - v_r)F(\underline{q}) - c > 0.$$

The last inequality follows from (4) and $q^{FB} > \underline{q}$. In other words, there must be some $q > \underline{q}$ that gives the manufacturer a strictly higher payoff than the wholesale contract.

If the manufacturer offers a buyback contract that implements $q > \overline{q}$, her payoff is determined by (A.6). The first-order derivative of $\mathrm{E}_\omega\, u_m(\omega)$ is

$$\begin{aligned}\frac{\mathrm{d}\,\mathrm{E}_\omega\, u_m(\omega)}{\mathrm{d}q} &= S'(q) - \left(1 - \frac{v_m}{v_r}\right)[v_r + (p - v_r)Q'(\underline{\omega})\underline{\omega}'(q) - Q'(t/p)t'(q)] \\ &= S'(q) - \left(1 - \frac{v_m}{v_r}\right)\left[v_r F(t/p) + (p - v_r)\frac{(F(t/p) - F(\underline{\omega}))(1 - F(q))}{1 - F(\underline{\omega})}\right]. \end{aligned} \tag{A.10}$$

In the second equality, we use:

$$t'(q) = \frac{S'(q) + c - v_r F(\underline{\omega})}{1 - F(\underline{\omega})} = v_r + (p - v_r)\frac{1 - F(q)}{1 - F(\underline{\omega})},$$
$$\underline{\omega}'(q) = \frac{t'(q) - v_r}{p - v_r} = \frac{1 - F(q)}{1 - F(\underline{\omega})},$$

both of which come from (A.7).

Since q^* is the optimal quantity that the manufacturer wants to implement using a buyback contract, it should be a stationary point of $\mathrm{E}_\omega\, u_m(\omega)$, which is given by either (A.9) or (A.10). In both cases, $S'(q^*) > 0$, which implies $q^* < q^{FB}$ from (4).

Online Appendix

B.1 The Ironing Approach

In this section, we provide formal statements and proofs of the ironing approach and the derivation of local buyback contracts in Section 4.3.

Assume that Γ implements $q > \underline{q}$ and satisfies the necessary conditions in Lemma 3 and Lemma 4. Let V be the associated total transfer as defined in (8). We first construct a new function $\hat{V}$ by mapping V into the quantile space. For any $\phi \in [0,1]$, let $\mathcal{V}(\phi)$ be the accumulated total transfer for all types below $F^{-1}(\phi)$.[21] That is,

$$\mathcal{V}(\phi) = \int_0^{F^{-1}(\phi)} V(\omega)\mathrm{d}F(\omega) = \int_0^{\phi} V(F^{-1}(\hat{\phi}))\mathrm{d}\hat{\phi}.$$

By construction, $\mathcal{V}(\phi)$ is increasing, absolutely continuous, and admits a (Radon–Nikodym) derivative $\mathcal{V}'(\phi) = V(F^{-1}(\phi))$. Using $\phi = F(\omega)$, we have $\mathcal{V}'(F(\omega)) = V(\omega)$.

Denote by $\hat{\mathcal{V}}$ the lower convex envelope of $\mathcal{V}$, which is the largest convex function below $\mathcal{V}$, formally defined by

$$\hat{\mathcal{V}} = \sup\{\mathcal{U} | \mathcal{U} \text{ is convex and } \mathcal{U}(\phi) \le \mathcal{V}(\phi) \text{ for all } \phi \in [0,1]\}.$$

Since $\hat{\mathcal{V}}$ is convex, it is absolutely continuous and admits a nondecreasing (Radon–Nikodym) derivative $\hat{\mathcal{V}}'$. The "ironed total transfer function" is defined as

$$\hat{V}(\omega) = \lim_{\hat{\omega} \to \omega+} \hat{\mathcal{V}}'(F(\hat{\omega})). \tag{B.1}$$

Clearly, $\hat{V}$ is right-continuous and nondecreasing. Furthermore, $\hat{\mathcal{V}}(1) = \mathcal{V}(1)$, so $\hat{V}$ has the same expectation as V with respect to the probability measure given by F. In what follows, we establish several properties of $\hat{V}$ that will allow us to construct an alternative contract that dominates Γ.

Note that by construction, there exists countably many disjoint intervals $[\underline{\phi}_n, \overline{\phi}_n)$, indexed by $n \in \mathcal{N}$, such that $\hat{\mathcal{V}}(\phi)$ is linear on every $[\underline{\phi}_n, \overline{\phi}_n)$ and is strictly convex otherwise. Since by (O2) of Lemma 4, $\mathcal{V}$ is linear on $[F(q), 1]$, there exists one $n \in \mathcal{N}$, say $n(q)$, such that: (1) $[F(q), 1) \subseteq [\underline{\phi}_{n(q)}, \overline{\phi}_{n(q)})$; and (2) for all $n \neq n(q)$, $\overline{\phi}_n \le F(q)$. By applying the inverse mapping F^{-1}, which maps $[0,1]$ to the extended real line, we define $\underline{\omega}_n = F^{-1}(\underline{\phi}_n)$ and $\overline{\omega}_n = F^{-1}(\overline{\phi}_n)$ for all n. Thus, $\underline{\omega}_{n(q)} \le q$, $\overline{\omega}_{n(q)} = +\infty$, and for $n \neq n(q)$, $\underline{\omega}_n \le \overline{\omega}_n \le q$.

[21] Recall that F is an atomless distribution, so $F^{-1} : [0,1] \mapsto [0, +\infty)$ is well-defined.

Several observations follow from this construction. First, on each $[\underline{\phi}_n, \overline{\phi}_n)$, $\hat{\mathcal{V}}(\phi)$ is linear, so $\hat{\mathcal{V}}'(\phi)$ is constant, and we denote this constant value by k_n. By (B.1), for any $\omega \in [\underline{\omega}_n, \overline{\omega}_n)$, $\hat{V}(\omega) = k_n$. Since $\mathcal{V}$ is enveloped by $\hat{\mathcal{V}}$ from below,

$$k_n \leq V(\underline{\omega}_n) \leq p\underline{\omega}_n + v_r(q - \underline{\omega}_n). \tag{B.2}$$

Second, for any $\phi \notin \bigcup_{n\in\mathcal{N}}[\underline{\phi}_n, \overline{\phi}_n)$, $\hat{\mathcal{V}}(\phi) = \mathcal{V}(\phi)$. By convexity, for almost all $\omega \notin \bigcup_{n\in\mathcal{N}}[\underline{\omega}_n, \overline{\omega}_n)$, $\hat{V}(\omega) = V(\omega)$.

Third, for any $\omega \notin \bigcup_{n\in\mathcal{N}}[\underline{\omega}_n, \overline{\omega}_n)$, $R(\omega) = \max(q - \omega, 0)$. Suppose contrary to the assertion that $R(\omega) < q - \omega$ for some $\omega < q$. Let $\hat{\omega}$ be a type satisfying $\omega \leq \hat{\omega} \leq q - R(\omega)$. Then $T(\omega) \leq p\hat{\omega}$ and $R(\omega) \leq q - \hat{\omega}$, which, by (O1a) of Lemma 4, implies that $V(\hat{\omega}) \leq V(\omega)$. Since the analysis applies to any $\hat{\omega} \in [\omega, q - R(\omega)]$, $\mathcal{V}$ is concave on $[F(\omega), F(q - R(\omega))]$. Passing to the convex envelope of $\mathcal{V}$ shows that $\phi = F(\omega)$ must belong to some $[\underline{\phi}_n, \overline{\phi}_n)$ on which $\hat{\mathcal{V}}$ is linear, a contradiction.

Finally, $\hat{V}$ is continuous. Suppose contrary to the assertion that $\hat{V}(\omega-) < \hat{V}(\omega) = \hat{V}(\omega+)$ at some ω, where $\hat{V}(\omega-)$ and $\hat{V}(\omega+)$ represent the left and right limit of $\hat{V}$ at ω, respectively. Then $\hat{\mathcal{V}}$ has subdifferential $[\hat{V}(\omega-), \hat{V}(\omega+)]$ at $\phi = F(\omega)$, implying that $\hat{\mathcal{V}}$ has a kink at ϕ. Hence, $\mathcal{V}(\phi) = \hat{\mathcal{V}}(\phi)$. Since $\mathcal{V}$ is enveloped by $\hat{\mathcal{V}}$ from below, $V(\omega-) \leq \hat{V}(\omega-) < \hat{V}(\omega+) \leq V(\omega+)$. However, by (O1b) of Lemma 4, for any $\hat{\omega} < \omega$,

$$V(\omega) \leq V(\hat{\omega}) + (p - v_r)[R(\hat{\omega}) - (q - \omega)] \leq V(\hat{\omega}) + (p - v_r)(\omega - \hat{\omega}).$$

When $\hat{\omega}$ converges to ω from the left, $V(\omega) \leq V(\omega-)$. Similarly, for any $\hat{\omega} > \omega$,

$$V(\hat{\omega}) \leq V(\omega) + (p - v_r)(\hat{\omega} - \omega).$$

When $\hat{\omega}$ converges to ω from the right, $V(\omega+) \leq V(\omega) \leq V(\omega-)$, a contradiction.

Using the function $\hat{V}$ and the set of cutoffs $\{\underline{\omega}_n, \overline{\omega}_n\}_{n\in\mathcal{N}}$, we are ready to construct an alternative contract $\hat{\Gamma}$, defined as[22]

$$\hat{T}_1(\omega) = \begin{cases} \min(p\omega, t_n) & \omega \in [\underline{\omega}_n, \overline{\omega}_n), \\ V(\omega) - v_r \max(q - \omega, 0) & \omega \notin \bigcup_{n\in\mathcal{N}}[\underline{\omega}_n, \overline{\omega}_n), \end{cases} \tag{B.3}$$

$$\hat{R}(\omega) = \begin{cases} \max(q - \overline{\omega}_n, 0) + \max((t_n - p\omega)/v_r, 0) & \omega \in [\underline{\omega}_n, \overline{\omega}_n), \\ \max(q - \omega, 0) & \omega \notin \bigcup_{n\in\mathcal{N}}[\underline{\omega}_n, \overline{\omega}_n), \end{cases} \tag{B.4}$$

[22]To save notation, we allow the operators min and max to take arguments on the extended reals. That is, for any $z \in \mathbb{R}$, $\min(z, +\infty) = z$ and $\max(z, -\infty) = z$.

where $t_n = k_n - v_r \max(q - \overline{\omega}_n, 0)$. This implies

$$t_n \leq (p - v_r)\underline{\omega}_n + v_r \min(\overline{\omega}_n, q) \text{ for any } n \in \mathcal{N}. \tag{B.5}$$

Clearly, $\hat{\Gamma}$ as constructed from (B.3)-(B.5) is a local buyback contract.

B.2 Proof of Proposition 7

First, note that $S(q;p)$ becomes negative for sufficiently large p and q, so it is without loss to solve the manufacturer's optimization problem under the assumption that p and q are both bounded. In this case, the optimal (p,q) must be an interior stationary point of $\mathrm{E}_\omega\, u_m(\omega)$. Hence we can use first-order necessary conditions to quantify (p,q).

When $0 < t \leq v_r q$, (A.6) becomes

$$\mathrm{E}_\omega\, u_m(\omega) = S(q;p) - \underline{u} - \left(1 - \frac{v_m}{v_r}\right) p\left[t/p - Q\left(t/p;p\right)\right]. \tag{B.6}$$

By first-order conditions,

$$\begin{aligned}
S_q(q;p) &= \left(1 - \frac{v_m}{v_r}\right) F(t/p;p) t_q, \\
S_p(q;p) &= \left(1 - \frac{v_m}{v_r}\right) [F(t/p;p)t_p + (t/p)Q_q(t/p;p) - pQ_p(t/p;p) - Q(t/p;p)].
\end{aligned}$$

By (A.5), $t_q = S_q(q;p) + c = p - (p - v_r)F(q;p) > 0$, so $S_q(q;p) > 0$. Also, $t_p = S_p(q;p) = Q(q;p) + (p - v_r)Q_p(q;p)$, so

$$\begin{aligned}
S_p(q;p) &= \left(1 - \frac{v_m}{v_r}\right) [F(t/p;p)S_p(q;p) + (t/p)Q_q(t/p;p) - S_p(t/p;p) - v_r Q_p(t/p;p)] \\
&\geq \left(1 - \frac{v_m}{v_r}\right) [F(t/p;p)S_p(q;p) - S_p(q;p) + (t/p)Q_q(t/p;p) - v_r Q_p(t/p;p)].
\end{aligned}$$

where the inequality comes from $S_{pq} \geq 0$. Thus, $S_p(q;p) \geq 0$.

When $t > v_r q$, (A.8) becomes

$$\mathrm{E}_\omega\, u_m(\omega) = S(q;p) - \underline{u} - \left(1 - \frac{v_m}{v_r}\right) \left\{p\left[t/p - Q\left(t/p;p\right)\right] - (p - v_r)[\underline{\omega} - Q(\underline{\omega};p)]\right\}. \tag{B.7}$$

By first-order conditions,

$$
\begin{aligned}
S_q(q;p) &= \left(1 - \frac{v_m}{v_r}\right)[F(t/p;p) - F(\underline{\omega};p)]t_q, \\
S_p(q;p) &= \left(1 - \frac{v_m}{v_r}\right)[(t/p - t_p)Q_q(t/p;p) - pQ_p(t/p;p) - Q(t/p;p) \\
&\qquad + (t_p - \underline{\omega})Q_q(\underline{\omega};p) + (p - v_r)Q_p(\underline{\omega};p) + Q(\underline{\omega};p)].
\end{aligned}
$$

By (A.7),

$$
\begin{aligned}
t_q &= \frac{S_q(q;p) + c - v_r F(\underline{\omega};p)}{1 - F(\underline{\omega};p)} = v_r + (p - v_r)\frac{1 - F(q;p)}{1 - F(\underline{\omega};p)} > 0, \\
t_p &= \frac{S_p(q;p) - \int_0^{\underline{\omega}} \omega \mathrm{d}F(\omega;p)}{1 - F(\underline{\omega};p)} = \underline{\omega} + \frac{S_p(q;p) - Q(\underline{\omega};p)}{Q_q(\underline{\omega};p)}.
\end{aligned}
$$

Thus, $S_q(q;p) > 0$, and $S_p(q;p)$ can be further simplified as:

$$
\begin{aligned}
S_p(q;p) = \left(1 - \frac{v_m}{v_r}\right)&[(t/p - t_p)Q_q(t/p;p) - S_p(t/p;p) - v_r Q_p(t/p;p) \\
&+ (t_p - \underline{\omega})Q_q(\underline{\omega};p) + S_p(\underline{\omega};p)].
\end{aligned}
$$

Furthermore,

$$
S_{pq}(q;p) = 1 - F(q;p) - (p - v_r)F_p(q;p) < 1 - F(q;p),
$$

which implies

$$
S_p(t/p;p) - S_p(\underline{\omega};p) = \int_{\underline{\omega}}^{t/p} S_{pq}(q;p)dq < \int_{\underline{\omega}}^{t/p} [1 - F(\underline{\omega};p)]dq = [1 - F(\underline{\omega};p)](t/p - \underline{\omega}).
$$

Therefore,

$$
S_p(q;p) > \left(1 - \frac{v_m}{v_r}\right)[(t/p - t_p)Q_q(t/p;p) - v_r Q_p(t/p;p)].
$$

Assume that $S_p(q;p) \leq 0$, then $t_p \leq \underline{\omega} \leq t/p$, which again implies $S_p(q;p) > 0$, a contradiction. Hence, $S_p(q;p) > 0$.

Finally, recall that $S_{pq}(q;p) > 0$, so $S_q(q;p) > 0$ and $S_p(q;p) > 0$ jointly imply that $q^* < q^{FB}$ and $p^* < p^{FB}$.

B.3 Proof of Proposition 8

Since retailers are symmetric, it suffices to prove the proposition when $n = 2$. First, we show $p^1 = p^2$ by contradiction. Suppose that $p^1 < p^2$. Then increasing p_1 will not change the distribution of ω as demand is determined by the higher price p_2. If the manufacturer increases p_1 and the date-1 cash repayment T_1^1 uniformly so that the retailer is indifferent, she can extract more surplus from Γ^1 without affecting her payoff from Γ^2. Therefore, the manufacturer optimally offers $p^1 = p^2$. It is then straightforward to see that Γ^1 and Γ^2 are identical. Moreover, they are both buyback or wholesale contracts, because by Proposition 4, the optimality of both contracts is robust to any distribution of demand (as long as the monotone hazard rate condition is satisfied).

Suppose that Γ^1 and Γ^2 are both buyback contracts with constant buyback prices. Then the manufacturer's expected payoff is

$$\begin{aligned}\mathrm{E}_\omega\, u_m(\omega) &= 2\left\{W - cq^1 + \int_0^{\overline{\omega}^1} \left[\frac{1}{2}p^1\omega + \frac{v_m}{v_r}\left(t^1 - \frac{1}{2}p^1\omega\right)\right] \mathrm{d}F(\omega;p^1) + \int_{\overline{\omega}^1}^{+\infty} t^1 \mathrm{d}F(\omega;p^1)\right\} \\ &= 2(W - cq^1) + \int_0^{\overline{\omega}^1} \left[p^1\omega + \frac{v_m}{v_r}(2t^1 - p^1\omega)\right] \mathrm{d}F(\omega;p^1) + \int_{\overline{\omega}^1}^{+\infty} 2t^1 \mathrm{d}F(\omega;p^1),\end{aligned} \tag{B.8}$$

where $\overline{\omega}^1 = 2t^1/p^1$, and t^1 is determined by a binding (IR_r),

$$\begin{aligned}t^1 &= \int_0^{2q^1} \left[\frac{1}{2}p^1\omega + v_r\left(q^1 - \frac{1}{2}\omega\right)\right] \mathrm{d}F(\omega;p^1) + \int_{2q^1}^{+\infty} p^1q^1\mathrm{d}F(\omega;p^1) - W - \underline{u} \\ &= \frac{1}{2}(p^1 - v_r)Q(2q^1;p^1) + v_rq^1 - W - \underline{u}.\end{aligned} \tag{B.9}$$

By comparing (B.8) and (B.9) with (B.6), we can conclude that the manufacturer's expected utility is equivalent to that from our benchmark model where the only retailer has reservation utility $2(W + \underline{u})$. A similar argument applies to the case where the buyback prices for both contracts are variable. Hence, the proposition is proved.