Vertical Relational Contracts and Trade Credit

Marta Troya-Martinez*
University of Oxford

June 21, 2012

Abstract

This paper analyzes the impact of trade credit on a relational contract between two vertically related firms. The downstream firm operates in an environment with unobservable shocks, like a developing country or a black market, which creates moral hazard in the repayment decision.

Trade credit limits the supplier’s possibilities to punish the downstream firm and termination occurs on the equilibrium path. A larger quantity increases the incentives of the constrained firm to steal the credit and needs to be accompanied with a more severe termination policy. Since termination is also costly for the upstream firm, the quantity sold in the market is distorted downwards (i.e., there is double marginalization despite the use of a quantity forcing contract). The quantity is bounded away from efficiency even as the firms become arbitrarily patient.

The optimal contract resembles a debt contract: when the revenues are relatively large, the upstream firm asks for a fixed payment and never terminates. Otherwise, the downstream firm repays all his revenues and the smaller this repayment, the larger the termination period imposed.

*JEL Classification: C73, D82, L14.*

**Keywords:** Relational contracts, trade credit, imperfect monitoring.

1 Introduction

In developing countries, firms have to deal with a number of constraints that are not present in developed countries and that necessarily interfere in the way firms compete and interact with each other along the vertical chain. Among other adversities, they face weak institutions, underdeveloped legal system and capital markets, corruption and

---

*University of Oxford, e-mail: marta.troyamartinez@economics.ox.ac.uk. I am indebted to my supervisors Paul Klemperer and especially David Myatt for their guidance. I have also benefited from helpful comments by Mikhail Drugov, Guido Friebel, Lakshmi Iyer, James Malcomson, Meg Meyer, Patrick Rey and the participants of the Gorman workshop in Oxford, the CSAE Conference 2011 in Oxford, HOC 2011 in Boston, JEL 2011 in Valencia and the RES 2012 meetings in Cambridge. Any remaining errors are my own.
non-prosecuted criminal offences such as robberies. This makes the conclusions of the standard literature on industrial organization less useful to understand the functioning of markets in these countries.

As can be seen from Figure 1, the number of competition authorities in developing countries has rocketed from under ten in 1990 to almost eighty in 2008. Therefore, understanding the industrial organization of developing countries has become a timely issue as these young antitrust authorities need to base their policies on theories that fit their environments rather than the ones created for developed countries.

The goal of this paper is to incorporate some stylized facts of developing countries into the analysis of vertical relations. We want to explore how these stylized facts affect firms’ behavior and market outcomes and to generate testable predictions.

The first stylized fact we want to consider is the underdevelopment of legal systems. An important consequence of the firms not being able to resort to the rule of law to enforce their contracts is the need to rely on relational contracts, where the agreement is self-enforced thanks to the gains from future repeated trade. McMillan and Woodruff (1999a and 1999b) find evidence of relational contracting using surveys from 259 non-state firms in Vietnam. They show how prior experience with a trading partner and prior information gathering on that partner are associated with increased provision of credit.

Figure 1: Time-line of adopting competition laws in the developing world. *Source:* Waked (2010)

Indeed, supplier’s credit (the temporal dissociation between delivery and payment) is particularly common due to the underdevelopment of financial markets. Among others, McMillan and Woodruff (1999a and 1999b) and Fafchamps (1997 and 2000) point out the important role played by supplier credit\(^1\) as compared to bank credit in developing countries. This is the second feature we want to include in our model.

The third and last ingredient that we want to incorporate is the existence of adverse shocks that make the firms unable to honour their agreements, such as a warehouse being robbed or unexpected low revenues\(^2\). These adverse shocks are not always observed by the other contracting party which creates a moral hazard (with hidden knowledge) problem, as the downstream firms may not return the trade credit and pretend that was a shock.\(^3\)

With our analysis, we want to address the following questions. How is the contract offered by the manufacturer affected by the trade credit and the asymmetric information problem? What are the resulting market distortions?

Although we have framed the introduction in a developing country setting, the results are also useful to understand inter-firm relations in developed countries, where trade credit is also very common, with firms that operate in the shadow economy, that do not declare all their transactions or whose trade volume is small and does not justify the cost of going to the courts. In all these instances, firms are left to rely on relational contracts.

Another application of this framework are black markets like drug trade. In these markets, obviously, contracts cannot be legally enforced, traders are usually credit constrained and are also subject to many unobservable shocks. Since little is known of these markets due to their undercover aspect, these results are valuable for shedding some light in the understanding of these markets.

In the model that we consider there is an upstream firm that supplies a good and offers trade credit to a cashless downstream firm. To fix ideas, imagine that the upstream firm ("she") is a manufacturer and the downstream firm ("he") is a retailer. For instance, it can be the case that the manufacturer’s machinery can be used as collateral and this makes her less credit constrained than the retailer. The lack of access to the financial system for the retailer may also be explained by the retailer operating in the shadow economy. The manufacturer proposes a quantity forcing contract of the form of a quantity and a repayment\(^4\) and the retailer pays back to the manufacturer only after having sold the

---

\(^1\)Trade credit is normally defined as the dissociation between invoicing and payment and it is also widely used in developed countries.

\(^2\)Fafchamps et al. (forthcoming) document the volatility of sales and profits of microenterprises in Ghana and show how this volatility is not created by measurement errors.

\(^3\)Note that the shocks may be observable and the trading party who offers the trade credit may still not be able to determine whether the other party is unable to fulfill the agreement because of the shock - an expensive or unavailable legal process is needed to have such information disclosed.

\(^4\)Since the quantity is delivered before any private information becomes known to the retailer, it is not used for sorting purposes. As a result, using a two-part tariff or a more complex nonlinear scheme is equivalent to choosing a quantity and a total repayment.
good. However, a shock may occur before the payment, making him unable to repay either part or the whole amount. The manufacturer cannot observe if the failure of payment is due to the shock or to the retailer keeping the money\(^5\) and she punishes him by terminating the relational contract for a number of periods.

We first solve a simple example (Section 3), where the shock makes the retailer lose all the revenues. Therefore, the retailer can either pay back all the credit or make no payment at all (if he received the shock). Then we introduce a second demand into the example. In this extended example (Section 4), the retailer places the order in the market and receives high or (if the shock occurs) low revenues. In this setup, the retailer can then either return the whole credit, or if unlucky, give back only part of it (we also allow for the possibility that the retailer just walks away with all the revenues). Finally, we analyze a general model where there is a continuum of states of demand (Section 5).

In a perfect world with symmetric information (i.e. observable demand shocks), both firms maximize their joint profits and no punishment is used. The quantity sold in the market is smaller than in the absence of shocks because the firms take into account the "effective" marginal cost which not only accounts for the production cost but also for the likelihood of the shocks.

When demand shocks are not observed by the manufacturer, trade credit, by postponing the payment until the shock is realized, creates a source of asymmetric information. If contracts can be legally enforced and the retailer has access to unlimited funds, it is possible to replicate the first best outcome by asking the retailer to repay a fixed payment (equal to the expected revenues) regardless of the state. The first best is no longer achievable, however, when either the retailer has access to unlimited funds but contracts cannot be legally enforced or contracts can be enforced but the retailer is cashless. This paper explores the consequences of the first departure, although we have also (unnecessarily) assumed that the retailer is cashless to make the setup closer to a developing country environment (See Fafchamps (1997 and 2000)). Section 4.1 illustrate the persistence of the distortions even when the cash constraint does not bind.

From the analysis of Example 1, it emerges that in the optimal contract, the manufacturer distorts the quantity downwards. The reason why she does not choose the profit maximizing quantity and then shares the extra profits with the retailer through a lower repayment is because the efficient quantity generates large revenues which increases the retailer’s incentives to steal them. In order to curb these incentives, the manufacturer needs to increase the termination period which is also costly for her. Thus, despite having enough available instruments to set the quantity double marginalization occurs. The downward distortion in the quantity is entirely driven by the retailer’s impatience. Moreover, it is also present in Example 2 and the general model and hence it is one of the general characteristics of the relational contract in our setting.

In Example 2, we find that if the high and the low demands are very similar (i.e. the shock is mild), then the manufacturer prefers to charge the same amount in both situations and to terminate the contract forever only if there is no repayment. As the

\(^5\)In order to find out, the manufacturer would need to use (if existent) a legal system that is too costly or too corrupt.
low demand shrinks (i.e. the shock increases), then the limited liability constraint starts binding and the manufacturer asks for a larger repayment in the high demand state than in the low demand state. In order to implement this contract, however, she needs to interrupt the relationship with the retailer for a positive but finite number of periods following a low repayment. To minimize the punishment imposed, she asks for the maximum possible repayment in the low state. When the limited liability constraint binds, termination is used and the outcome is bounded away from efficiency and the downward distortion persist in the quantity even as the firms become arbitrarily patient.

In the general model of Section 5, we find that the optimal contract offered by the manufacturer resembles a debt contract. In particular, if the states of demand are good enough (above some threshold optimally chosen by the manufacturer), the manufacturer ask for the same repayment and never terminates the contract. The larger this fixed repayment, the smaller the quantity distortion imposed by the manufacturer. Otherwise, the manufacturer asks for the highest possible repayment (which in the absence of retail costs is equal to the revenues) and punishes for a number of periods. Asking for the maximum possible repayment is optimal as it allows the manufacturer to soften the termination policy while keeping the retailer’s incentives unaffected. Furthermore, the smaller the repayment, the larger the termination period. Indeed, the manufacturer makes the retailer value the relationship by distributing rents in the good times. The retailer always keeps a minimum amount of rents (equal to the current period revenues) due to the fact that he can always choose not to return the credit. If the value of the future (i.e. the discount factor) is large enough, then the manufacturer actually prefers to leave even more rents to the retailer. This is because trying to extract these extra rents would imply larger termination periods which are very costly for the manufacturer as she also loses future trade. In this last case, the contract does not depend on the particular discount factor. As in the Examples, since the limited liability constraint binds for the bad states, punishments are used in equilibrium and the quantity remains distorted even as the discount factor becomes arbitrarily close to 1.

1.1 Related literature

This paper belongs to the literature on inter-firm relational contracts with asymmetric information, where incentives are usually given by discretionary payments conditional on good performance (see Malcomson 2010). In this setup, however, the use of trade credit makes the transfer of continuation payoffs unfeasible and instead incentives are provided by the threat of termination, and hence by burning value.

Levin (2003) is the first paper to introduce moral hazard or adverse selection in a principal-agent relational contract in order to identify in which way contract self-enforceability reduces the provision of high-powered incentives. In our model, however, there is a problem of moral hazard with hidden knowledge, where the retailer’s actions are observable (whether he repaid or not) but not the information on which their are based (whether he received a shock and of which size). This asymmetric information is

---

6This problem is also known as post-contractual adverse selection, where the type of the agent (which
reminiscent of the model of Green and Porter (1984) where an oligopoly is colluding in a market with noisy prices. When a low price is observed, firms do not know with certainty if this bad outcome is due to a market shock or a firm deviating to a larger quantity. As in our model, incentives to collude are created by value burning (i.e. the threat of triggering a price war if the observed price is too low).

In Levin’s model, the agent undertakes a costly action and the principal rewards him with a fixed fee (to which she can commit) plus a discretionary bonus (that can be positive or negative). The fixed fee enables the parties to share the total surplus in an arbitrary way, and hence their objective is to maximize it (subject to the self-enforceability constraint) and whether is the principal or the agent who offers the contract does not matter. Another consequence of this assumption is that there is no loss of generality in restricting attention to stationary contracts where incentives are provided by punishing and rewarding the agent with transfers that depend on the particular outcome, but not of the time.\(^7\) If deviation occurs, there is no need to terminate the contract as the principal can ask for a transfer that leaves the agent indifferent between staying in the relation and the outside option.

Trade credit bundled with limited liability imposes important restrictions on the contract that the principal can use. Loosely speaking, trade credit is a large up-front payment to the agent equal to some uncertain revenues. However, contrary to Levin’s fixed fee, this payment is not independent of the repayment amount and hence it affects the total surplus (indeed, the size of the trade credit will depend on the quantity sold). As a result, who offers the contract makes a difference and exploring how the relational contract changes when it is offered by the retailer is left for future work. The quantity of the good is delivered to the retailer before the private information is learnt and hence it cannot be used as an instrument for sorting. Moreover, the repayment is always positive\(^8\) and the principal has no other means to reward the agent. Due to the limited liability, the manufacturer cannot punish the agent in monetary terms. As a result termination is necessary to give the agent incentives to repay and stationary contracts may not be optimal. Indeed, Fong and Li (2010) show in their simplified moral hazard version of Levin (2003) how the introduction of limited liability (i.e. a minimum wage needs to be paid) makes stationary contracts non-optimal. As in Green and Porter (1984), we have imposed the use of a stationary two-state (trading and termination) T-stage Markov contract. In the Appendix, we show that in Example 1, the manufacturer does not want to offer a second

\(^7\)This is because, following a low outcome, the principal can punish the agent by: (1) moving to a future state that is worse for the agent than the present one; or (2) to pay less to the agent today. Since in this setup both instruments are perfect substitutes, there is no loss of generality in restricting attention to the second option only.

\(^8\)In terms of Levin (2003), the bonus is always negative and it is paid by the agent when the outcome is high (i.e. there is no shock) rather than when it is low.
chance to the retailer by selling a potentially different quantity in the period following no repayment for a large parameter space\textsuperscript{9}. Keeping contracting with the retailer after no repayment, although it reduces the inefficiency following a shock, it also increases the incentives not to repay in the first period which will push the manufacturer to distort downwards this quantity even more. Indeed, the optimal punishment in the Green and Porter model involves choosing larger quantities than in the stage-game Nash equilibrium in order to allow equilibria that are "more collusive"\textsuperscript{10}. Moreover, the strongly symmetric equilibria of Abreu, Pearce and Stacchetti (1986), which only use value burning to create incentives, show how one can restrict attention to equilibria with the bang-bang property which, under some conditions, implement the maximum pure perfect public equilibrium payoffs using continuation values drawn exclusively from the maximum and minimum pure perfect public equilibrium payoffs.

This paper is closely related to the literature where an entrepreneur is wealth constrained and needs to be financed by an investor, who cannot observe the resulting cashflows of the investment. Hence, the entrepreneur can potentially divert or steal them and incentives to repay are given by liquidating the entrepreneur’s assets or by threatening to withhold the investment in the second (and last) period. The early papers in this literature are by Bolton and Scharfstein (1990) and Hart and Moore (1998). Faure-Grimaud (2000) and Povel and Raith (2004a, 2004b) extend the basic two-state model to the case where the cashflows are distributed continuously on a bounded interval. Debt contracts are optimal in this environment because they minimize the probability of inefficient liquidation while inducing the entrepreneur to repay.

Debt contracts\textsuperscript{11} are also known to be optimal in the literature of costly state verification (see Townsend (1979) and Gale and Hellwig (1985)), where the cashflows of an investment are also unobserved to the investor but the parties can agree to carry on a costly audit to make the realized cashflow verifiable. If the entrepreneur repays the agreed fixed payment, no inspection takes place; otherwise, an inspection is carried out and the entrepreneur needs to repay all the realized cashflows. Asking for the maximum repayment is optimal as this relaxes the participation constraint of the investor, who can then reduce the repayment and hence the audit set.

Indeed, the probability of liquidation (or not refinancing) and the probability of inspection play the same role as the length of termination imposed by the manufacturer. First, it relaxes the incentive compatibility of the retailer. Second, like the auditing or liquidating, imposing the termination policy is costly for the manufacturer, because the relation is always profitable (and hence socially valuable).

The more closely related paper to ours is Povel and Raith (2004a)\textsuperscript{12}. The authors

\textsuperscript{9}For the parameter specification for which the manufacturer gives a second opportunity to the retailer, we still do not know whether the manufacturer earns more profits than in Example 1.

\textsuperscript{10}See Mailath and Samuelson (2006) page 349.

\textsuperscript{11}Innes (1990) finds that debt contracts are optimal in an environment with moral hazard with limited liability. This is because a debt contract gives the best incentives as it makes the agent residual claimant in the good times and penalises him in the bad times by extracting all the surplus.

\textsuperscript{12}Faure-Grimaud (2000) and Povel and Raith (2004b) focus on the effect that financial constraints have on the choice of output when the firm is competing a la Cournot with another (financially unconstrained)
allow for the size of the investment to be chosen by the entrepreneur as well as how much

to repay and the probability of liquidating. Their focus is on showing that the optimal

contract is still a debt contract even when the investment choice (as well as the resulting
cashflows) is not observed by the investor. They also find that the entrepreneur under-
invests as compared to the first best as this decreases the fixed repayment and hence the
likelihood of defaulting which may result in an inefficient liquidation. In the same way,
the manufacturer of our model (the "investor") chooses to sell less output than the first
best to the retailer to soften the (inefficient) termination policy. Our contribution is to
endogenize the future value that accrues to the retailer (i.e. the "entrepreneur") if he does
not default, as it corresponds to the potential profits generated within the relationship.
This allows us to explore the impact of the future on the contract. For instance, when the
parties value the future enough, the optimal quantity and the repayment do not depend
on the discount factor or the observation that the impatience of the retailer is the only
driver of the downward distortion in the quantity.

The goal of this paper is not to explain why and how much trade credit is offered by the
manufacturer\textsuperscript{13}. We take this decision as given, and rather we explore how trade credit
affects the different contract characteristics, such as late payment penalty, quantity and
price of the good sold. The main prediction of this analysis is that the use of trade credit
does have an important impact on the market outcome. In particular, the quantity sold
in the market is expected to be lower than the efficient one in order to give incentives
to the retailer to repay without having to use too severe termination policies. In the
setup of Green and Porter (1984), when quantities are chosen from a sufficiently fine
grid of points, a similar result emerges. In particular, firms "produce quantities larger
than the monopoly output to reduce the incentives to deviate from equilibrium play,
which in turn allows equilibrium punishments to be less severe. Because punishments
actually occur in equilibrium, this reduced severity is valuable". (Mailath and Samuelson
(2006), p. 353). Another implication of this result is that as the firms become arbitrarily
patient, the outcome is still bounded away from efficiency. Intuitively, the imperfection
in the monitoring implies that on the equilibrium path, the manufacturer must face low
continuation payoffs with positive probability. Because the retailer is credit constrained,
it is impossible to use inter-temporal transfers of payoffs to maintain efficiency while still
providing incentives.

Finally, Buehler and Gartner (forthcoming) also explore the use of relational contracts
within two vertically related firms, however their question is very different. They show,
how in a vertical relationship where the manufacturer has private information about the
manufacturing cost, recommending a retail price may be necessary in order to maximize
total profits. The reason is that in order to induce the manufacturer to report the true

\textsuperscript{13}Answers to these questions can, for instance, be found in Cunat (2007) who shows how suppliers of
services and differentiated goods are more willing to sell on credit than suppliers of standardized goods
because they may be harder to replace and hence the downstream firm is more reluctant to default. Also,
Smith (1987) finds that trade credit may be a consequence of an agency problem. Indeed, if there is
quality variation in the good supplied, the downstream firm may be more reluctant to pay before having
had the time to inspect the good.
cost, he needs to be the residual claimant of the costs. This can result in the wholesale price being non-monotonic in the cost and hence the need for recommended retail prices.

The paper proceeds as follows. Section 2 introduces the model. Section 3 explores a simple example where an unlucky retailer may lose all his revenues before paying to the manufacturer. Section 4 extends this example to the case where the retailer faces to possible demands: one is high and the other is low. Section 5 generalizes the previous examples to the case where the retailer faces a continuum of demand states. Section 6 discusses the empirical implications of the analysis. Finally, Section 7 concludes. Tedious computations are relegated to the Appendix.

2 The Setup

A manufacturer \((M)\) and a retailer \((R)\) have the opportunity to trade at dates \(t = 0, 1, 2, \ldots\). In each period, \(M\) produces a good at a marginal cost \(c > 0\) and needs a retailer to market the product to the final consumer. The retailer can sell the good at no cost but he is completely credit constrained and needs to be fully financed by \(M\) in order to sell. As a result, \(M\) offers trade credit to \(R\), who will then pay \(M\) back after selling the good but within the same period (so no interest rate is charged).

In order to remove distortions coming from the manufacturer not having enough instruments to determine the final quantity, we let the manufacturer offer a quantity forcing contract. When contracts are not legally enforced and the retailer does not repay, there are not many instruments available to the manufacturer to punish for misbehavior. For instance, audits or liquidation of the retailer’s assets cannot be implemented. Therefore, the manufacturer is left to use the threat of a \(T\) period termination as a mean to provide incentives. McMillan and Woodruff (1999a and 1999b) find that such retaliation occurs in Vietnam, although it is not as forceful as one would expect in a standard repeated game framework. An alternative interpretation to the periods of termination is to trade in less profitable terms (for instance by diminishing the quality of the good).\(^{14}\)

We denote by \(0 < \delta < 1\) the discount factor and we assume that in the periods of no trade, both firms get a constant outside option which is normalized to 0.

The timing is summarized in Figure 2. In each period, the manufacturer offers a contract to the retailer. The retailer rejects or accepts and if he accepts, he places an order in the market. Then an \(iid\) shock is realized (and this information is only observed by the retailer) which determines the size of the revenues. Finally, \(R\) decides how much to repay and the contract is terminated for a number of periods if it is specified in the contract.

The effect of the shock is to add randomness to the revenues of the retailer. We can interpret this shock as an uncertainty with the respect to the willingness to pay of final consumers. For instance, the retailer could be placing the quantity offered by the manufacturer in the market and some periods he is paid a high price for it and other

\(^{14}\)See Baker, Gibbons and Murphy (1994) for an example in the employer-employee relationship framework.
periods a low price. Another interpretation is the one of an adverse shock whereby demand is certain but either the goods or the revenues are stolen now and then (for instance, by an organized crime group). Finally, the setup could also represent a situation where the uncertainty refers to how many units of a non-perishable (where it is not possible to give back the unsold units) or perishable good are demanded every period in the market.

3 Example 1

As an example, we consider the case where the retailer faces a demand \( p(q) \) and with probability \( s \), there is a shock that makes him lose all the revenues, so he can either repay the whole credit or nothing. The manufacturer offers a quantity forcing contract \( \{q, D, T\} \) where \( q \) is the quantity, \( D \) the repayment and \( T \) periods of termination following the lack of payment.

In a benchmark situation with no asymmetric information and law enforcement, the parties would maximize their joint one-period profits:

\[
\max_q (1 - s) p(q)q - cq
\]

The first best quantity is determined by:

\[
q^{FB} : p'(q)q + p(q) = \frac{c}{1 - s} = \tilde{c}
\]

where \( \tilde{c} \) is the effective marginal cost, which accounts for the likelihood of the shock. This is the relevant marginal cost against which we will make comparisons. The parties want to sell less than in the absence of the shock because with some probability they will not receive the revenues but incur the production costs anyway.

With the relational contract, since the manufacturer offers the contract, she maximizes her profits ensuring that the retailer has incentives to repay. Let \( \pi_R \) denote the retailer’s present discounted value of selling the good and repaying to the manufacturer from date \( t \) on.

\[
\pi_R = (1 - s) \left[ p(q)q - D + \delta \pi_R \right] + s\delta^{T+1} \pi_R
\]

\( ^{15} \text{If the good was non-perishable and the retailer could give back the unsold units, the problem would become trivial as the source of asymmetric information would disappear.} \)
The previous equation says that with probability \( 1 - s \) there is no shock, and hence the retailer receives the revenues and pays back the credit to the manufacturer; in which case she renews the relational contract and hence, the game remains in this cooperative phase in the next period. However, with probability \( s \), there is a shock that "destroys" the revenues, the retailer cannot pay back the credit and the game moves to the termination phase in the next period. In this case, the retailer will earn again \( \pi_R \) only after the end of the punishment phase of \( T \) periods.

In a similar way, let \( \pi_M \) be the present discounted value for the manufacturer:

\[
\pi_M = (1 - s) [-cq + D + \delta \pi_M] + s [-cq + \delta^{T+1} \pi_M]
\]

With probability \( 1 - s \), \( R \) repays so the relationship move on to the next period and with the complementary probability, \( R \) cannot repay and \( M \) terminates the contract for \( T \) periods.

The problem of \( M \) is to choose \( \{q, D, T\} \) that maximizes \( \pi_M \) subject to the incentive compatibility constraint (1) and \( R \)'s choice of \( q \):

\[
\pi_M = \frac{(1 - s) D - cq}{1 - \delta (1 - s) - s \delta^{T+1}}
\]

s.t.

\[
p(q)q - D + \delta \pi_R \geq p(q)q + \delta^{T+1} \pi_R
\]  

(1)

Condition (1) reflects the following retailer’s trade-off: if he does not pay back, he keeps \( D \) but this will automatically trigger the termination phase, which yields valuation \( \pi_R \) only after \( T \) periods. For the retailer to pay back, the manufacturer needs to ensure that tomorrow’s gains from not being terminated are larger than the payment today: \( \delta (\pi_R - \delta^T \pi_R) \geq D \).

Note that an increase in \( D \) decreases the expected payoff of repaying directly and indirectly through a decrease in \( \pi_R \) (LHS of (1)). It also decreases the expected payoff of not repaying (RHS of (1)) but only indirectly and to less extent (i.e. discounted by \( \delta^T \)). As a result, to minimize the rents given to \( R \), \( M \) can choose \( D \) to make (1) bind, which allows us to rewrite it as follows:

\[
D = \frac{1 - \delta^T}{1 - \delta^{T+1}} \delta (1 - s) p(q)q
\]

The tightness of this restriction depends on the discount factor \( \delta \), the probability of the shock \( s \), as well as on the profitability of the market, \( p(q) \). The incentive compatibility constraint imposes an upper bound on how much \( M \) can ask from \( R \). In particular, the payment today should be smaller than what \( R \) can steal tomorrow if there is no shock, adjusted by the coefficient \( f(T) = \frac{1 - \delta^T}{1 - \delta^{T+1}} \). Note that \( f(T) \) is increasing in \( T \) and that at \( T = 0 \) it takes value 0. Therefore, (1) is not satisfied if the manufacturer never punishes. Similarly, the tougher the punishment, the easier it is for \( M \) to satisfy the constraint.

\(^{16}\)Note that the participation constraint is always satisfied as in expectation \( R \) can ensure a minimum amount of rents, \( (1 - s)p(q)q \), by walking away with the generated revenues.
The manufacturer chooses $q$ so that:

$$p'(q)q + p(q) = \frac{c}{\frac{1-\delta^r}{1-\delta^r} \delta (1 - s)^2}$$

(2)

Therefore, the marginal costs (RHS) in (2) are larger than $\bar{c}$, and hence the resulting quantity is smaller than in the benchmark.

**Proposition 1** Despite the use of a quantity forcing contract, there exists double marginalization. Furthermore, the smaller the length of the punishment, the larger the double marginalization.

The intuition is as follows: imagine that $M$ were to offer the surplus maximizing $q$, and share the extra generated profits by charging a lower $D$. The reason why this is not optimal is because the resulting larger revenues increase the retailer’s incentives to steal them. In order to curb these incentives, $M$ needs to make the termination policy tougher but this is also costly for her in terms of lost future trade. Hence, the manufacturer prefers to distort the wholesale price at the cost of not maximizing the total profits in order to relax the incentive compatibility constraint. Thus, there is a trade-off between increasing the length of the punishment and distorting the wholesale price above the effective marginal cost.

When the manufacturer is deciding for how long to stop contracting with the retailer following a no repayment, she is facing the following trade-off: on one hand, by increasing the punishment length, she decreases the incentives of the retailer to steal, relaxing in this way the incentive compatibility constraint. On the other hand, she increases the inefficiency following a shock as she terminates profitable trading with the unlucky retailer.

**Lemma 1** The manufacturer terminates with the retailer forever if $s < \frac{(1-\delta)^2}{\delta^2}$. Otherwise, depending on the parameters, she may also terminate forever (for instance when $s \to 1$), or set a positive and finite $T$ (for instance when $\delta \to 1$).

**Proof.** See the Appendix. ■

Finally, it is worth highlighting that it is the impatience of the retailer the one that is creating the downward distortion in the quantity. Indeed, if the discount factor for $R$ were different from the one for $M$, then it would be retailer’s discount factor appearing in condition (2). In contrast, both discounts factors would affect the choice of the termination policy.

**4 Example 2**

This example extends the previous one by decreasing the severity of the shock; namely, the retailer faces a low demand $p_L(q)$ with probability $s$ and a high demand $p_H(q)$ with probability $(1 - s)$, where $p_L(q) \leq p_H(q) \forall q$. The manufacturer chooses the quantity $q$,
together with a repayment \( D_H \) if demand is high and a repayment \( D_L \) if it is low. In order to give incentives to repay the true amount, \( M \) terminates the contract for \( T \) periods following \( D_L \) and forever following no payment\(^{17}\). Finally, let the expected revenues be 
\[
R(q) = [(1 - s) p_H(q) + s p_L(q)] q.
\]

In the first best, the parties maximize their joint one-period profits, \( R(q) - cq \), so the resulting optimal quantity is determined by:
\[
q^{FB} : R'(q) = c
\]
and since there is no asymmetric information, no termination policy is imposed.

As an example, assume that \( p_H(q) = p(q) \) and \( p_L(q) = lp(q) \), where \( 0 < l < 1 \).\(^{18}\) The optimal quantity is determined by:
\[
p'(q)q + p(q) = \frac{c}{1 - s + ls} = \tilde{c}
\]
and hence, as in Example 1, the new effective marginal cost is larger. As the size of the shock, \( 1 - l \), decreases; the downward quantity distortion decreases.

With the relational contract, and as we did in Example 1, we can define the present discounted value of the manufacturer:
\[
\pi_M = \frac{(1 - s) D_H + s D_L - cq}{1 - \delta (1 - s) - s \delta^{T+1}}
\]
and the one of the retailer:
\[
\pi_R = \frac{R(q) - [(1 - s) D_H + s D_L]}{1 - \delta (1 - s) - s \delta^{T+1}}
\]

With two state-demand uncertainty, \( M \) needs to ensure that the three incentive compatibility constraints are satisfied. The first one says that the retailer who faces a high demand does not want to pretend that he is facing a low one:
\[
p_H(q)q - D_H + \delta \pi_R \geq p_H(q)q - D_L + \delta^{T+1} \pi_R \quad (IC_H)
\]
For this constraint to be satisfied, the profits in the high demand state plus tomorrow’s value of staying in the relationship should be larger than the profits when making a \( D_L \) payment but being punished for \( T \) periods.

The second constraint says that a low demand retailer does not want to walk away with all the revenues:
\[
p_L(q)q - D_L + \delta^{T+1} \pi_R \geq p_L(q)q \quad (IC_L)
\]
\(^{17}\)This is because \( M \) knows that, provided \( p_L(q) > 0 \) for the chosen \( q \), \( R \) can repay some amount, so if he does not that means that he is stealing and hence inflicts the maximum punishment. It is optimal to impose the maximum punishment because it helps to relax the incentive compatibility constraint and it is not imposed in equilibrium.

\(^{18}\)We use this multiplicative functional form of the demand function very often in the paper because of its simplicity. This functional form is, for instance, used in Green and Porter (1984).
or, equivalently, the future expected value of being within the non-cooperative phase (rather than being terminated forever) should be larger than the low demand payment.

Finally, the last constraint says that a high demand retailer does not want to keep all the revenues:

$$p_H(q) q - D_H + \delta \pi_R \geq p_H(q) q$$

In other words, the future expected value of being within the cooperative phase (rather than being terminated forever) should be larger than the high demand payment.

Note that if $IC_H$ and $IC_L$ are satisfied, then $IC_H^{Global}$ is also satisfied. Since an $H$-type retailer never wants to walk away with all his revenues, we can ignore $IC_H^{Global}$.

In addition to the incentive compatibility constraints, the retailer’s limited liability requires the payment done by $R$ to be smaller than his revenues. That is, $p_L(q) q \geq D_L$ for the low state and $p_H(q) q \geq D_H$ for the high state, respectively. We call them $LLs$.

Using the above information, and since the manufacturer directly chooses $q$, $M$’s problem becomes:

$$\max_{q,D_H,D_L,T} \pi_M = \frac{(1 - s) D_H + s D_L - cq}{1 - \delta (1 - s) - s \delta^{T+1}}$$

s.t.

$$\delta (\pi_R - \delta^T \pi_R) \geq D_H - D_L \quad (IC_H)$$
$$\delta^{T+1} \pi_R \geq D_L \quad (IC_L)$$
$$p_H(q) q \geq D_H \quad (LL_H)$$
$$p_L(q) q \geq D_L \quad (LL_L)$$

Inspecting $\pi_M$, we see that $M$ always wants to increase $D_H$ and $D_L$ as much as possible, therefore at least two constraints need to bind to bound $D_H$ and $D_L$ from above. The limited liability constraint in the high state, $LL_H$, is never binding because $M$ has to give away rents in "the good times" to make the relationship valuable for $R$, so he has incentives to repay and stay within the relationship. Finally, the case where only $IC_L$ and $LL_L$ bind is not possible either. This is because an increase in the punishment period not only decreases $\pi_M$ but also makes $IC_L$ more binding. Therefore, $M$ chooses no punishment ($T = 0$) and as a result $IC_H$ is violated.

Therefore, we are left with three possible combinations of binding constraints: $IC_L - IC_H$, $IC_H - LL_L$ and $IC_L - IC_H - LL_L$. We deal with each of them in turn relegating tedious computations to the Appendix. After characterizing the three regimes, we construct a numerical example using the linear demand to determine under which parameter specification each of the regime occurs.

### 4.1 $IC_H$ and $IC_L$ bind

If the revenues obtained in the low demand state are nonetheless relatively large, $LL_L$ will not be binding. Then $IC_H$ and $IC_L$ determine the payments in each state as a function of $q$ and $T$: $D_H = \delta R(q)$ and $D_L = \delta^{T+1} R(q)$. The manufacturer asks to repay in the
high state what \( R \) can steal in expected terms in the next period if they keep trading (i.e. tomorrow’s discounted expected revenues) and in the low state what he can steal after the termination period ends.

Plugging \( D_H \) and \( D_L \) into \( \pi_M \), \( M \) is left to maximize:

\[
\max_{q,T} \pi_M = \frac{((1-s)\delta + s\delta^{T+1})R(q) - cq}{1 - \delta (1-s) - s\delta^{T+1}}.
\]

Note that \( T \) always decreases the objective function, hence \( M \) never punishes following \( D_L \): \( T = 0 \), and therefore she can only ask for the same payment in both states: \( D_H = D_L = \delta R(q) \). The \( LL_L \) not binding implies \( \frac{\delta(1-s)}{1-\delta} p_H(q) < p_L(q) \) and a lower \( \delta \) or a larger \( s \) makes it easier for this condition to be satisfied.\(^{19}\) Finally, the optimal quantity is defined by: \( R'(q) = \frac{c}{s} \). As in Example 1, there is a downward quantity distortion in the quantity with respect to the benchmark. In order to compare this distortion with the one in Example 1, let us assume that \( p_H(q) = p(q) \) and \( p_L(q) = lp(q) \), where \( 0 < l < 1 \). Then the previous condition becomes:

\[
q : p'(q)q + p(q) = \frac{c}{\delta (1-s+ls)} > \bar{c}
\]

The quantity distortion is lower that the one in Example 1 (condition (2)) because, since the repayment is the same for both states, the manufacturer only needs to prevent the retailer from stealing the revenues.\(^{20}\) Finally, a larger \( \delta \) or a smaller \( s \) increases the quantity sold and hence the total repayment.

4.2 \( IC_H, IC_L \) and \( LL_L \) Bind

When the difference between the high and low demand becomes too large, the limited liability constraint, \( LL_L \), starts binding, which determines the payment in the low state as a function of the quantity: \( D_L = p_L(q)q \). The incentive constraints determine the high demand payment and the punishment period also as a function of \( q \): \( D_H = \delta R(q) \) and \( T = \frac{\ln(\frac{\delta p_H(q)}{R(q)})}{\ln \delta} - 1 \). Therefore, a difference in repayment, \( D_L < D_H \), requires \( M \) to terminate with \( R \) following a low payment for a positive number of periods. In order to be in this case, the shock size should be large enough: \( p_L(q) < \frac{\delta(1-s)}{1-\delta} p_H(q) \).

Plugging the above information into \( M \)'s objective function, \( M \) is left to maximize:

\[
\max_{q} \pi_M = \frac{\delta (1-s) R(q) + sp_L(q)q - cq}{1 - \delta (1-s) - \frac{sp_L(q)q}{R(q)}}
\]

The first order condition can be found in the Appendix. If we use the multiplicative functional form again, then \( l \) should be small enough so the differences in demands are

\(^{19}\)Note that if \( LL_L \) is satisfied, \( LL_H \) is also satisfied because the last price is higher for any \( q \).

\(^{20}\)That is, \( \frac{c}{\delta(1-s+ls)} < \frac{c}{1-\delta \delta(1-s)^2} T \).
large: \( l < \frac{\delta (1-s)}{1-\delta s} \) and the quantity is determined by:

\[
q : p'(q)q + p(q) = \frac{c}{\delta (1-s)(1-s+sl) + sl} > \tilde{c}
\]

(4)

Since \( 0 < l < 1 \) the quantity downward distortion is larger than in the case where only the incentive compatibility constraints bind. This is because the manufacturer, in addition to preventing an \( L \)-type retailer from stealing the revenues, she also needs to prevent an \( H \)-type retailer from pretending to be an \( L \)-type. For this particular example, the termination policy is:

\[
T = \ln \left( \frac{1}{\ln \delta} \right) - 1 \text{ where } 0 < T < +\infty^{21}.
\]

Finally, we can see from (4) that \( q \) increases following an increase in \( \delta \). Since the quantity increases, so does the repayment amounts \( D_H \) and \( D_L \). To enforce these larger repayments, \( M \) increases the length of the termination period. When \( s \) increases, \( q \) decreases as well as both repayment amounts and \( T \) (see Appendix).\(^{22}\)

### 4.3 \( IC_H \) and \( LL_L \) Bind

When the demands are still different enough so that the limited liability constraint is still binding, but when the probability of the shock is low or the value of the future is large, then the incentive compatibility constraint for the \( L \)-type stops binding, that is, the retailer does not want to walk away with the small revenues. To be precise, we are in this case when:

\[
p_L(q) < \frac{\delta^{T+1}(1-s)}{1-s^{T+1}}p_H(q),
\]

which requires \( T < +\infty \) to be satisfied.

As in the previous regime, the limited liability determines the low demand payment:

\[
D_L = p_L(q)q.
\]

The payment in the high demand state is determined by \( IC_H \): \( D_H = \frac{\delta (1-\delta T)R(q) + (1-\delta)\delta p_L(q)q}{1-\delta^{T+1}} \). Using this information, \( M \) maximizes:

\[
\max_{q,T} \pi_M = \frac{\delta (1-s)(1-\delta T)R(q) + (1-\delta)(1-s)-s\delta^{T+1}p_L(q)q - cq}{1-\delta (1-s) - s\delta^{T+1}}
\]

and again the first order conditions are relegated to the Appendix.\(^{23}\) Using the multiplicative example, the high demand payment becomes:

\[
D_H = \frac{(1-\delta T)\delta (1-s+sl) + (1-\delta)l}{1-\delta^{T+1}} p(q)q
\]

\(^{21}\)\( T \) is positive because \( \frac{l}{1-s+sl} < 1 \) and \( \frac{l}{1-s+sl} \) is constrained by \( l < \frac{\delta (1-s)}{1-\delta s} \). Moreover, it is never infinity because \( \delta < 1 \).

\(^{22}\)Note that \( LL_H \) is always satisfied as it requires:

\[
\frac{\delta s}{1-\delta (1-s)}p_L(q) < p_H(q)
\]

\(^{23}\)Note that \( TC_H \) is always satisfied as it requires:

\[
p_L(q) < p_H(q)
\]
and that the quantity is determined by:

\[ p'(q)q + p(q) = \frac{c}{\delta(1-s)\delta^T + (1-\delta)(1-s)(1-s\delta^T) - s\delta^{T+1}l} > \bar{c} \]  

(5)

where \(q\) and \(T\) are the optimal choices. Finally, we are in this case when \(l < \frac{\delta^{T+1}(1-s)}{1-s\delta^{T+1}}\).

Setting \(l = 0\), it is easy to check that this solution corresponds to the one in Example 1. Since in Example 1 the low demand is equal to zero, \(M\) cannot ask for a repayment in this state and hence cannot distinguish between \(R\) keeping the money or being in the low state. This case is equivalent to Example 1 because the \(IC_L\) is not binding, that is, an \(L\) - type retailer has no incentives to walk away with his revenues (so this possibility is irrelevant). Note that because of the existence of a positive low demand, the manufacturer can sell a larger quantity \(q\) and ask for a larger \(D_H\).

4.4 Comparison of regimes

From the previous analysis, it emerges the same conclusion that we found in Example 1: the manufacturer, even though she has enough instruments to set the quantity, chooses to sell less than the efficient amount in each of the three previous regimes. However, this example is useful to shed light on the reasons for such a downward distortion of \(q\).

**Proposition 2** As \(\delta\) becomes arbitrary large, the outcome is bounded away from efficiency only if the limited liability constraint binds.

**Proof.** See Appendix.  ■

Indeed, as the firms become arbitrarily patient, the downward distortion of \(q\) remains only in the regimes where \(LL_L\) binds. Intuitively, when \(LL_L\) does not bind, \(M\) is better off asking for the same quantity regardless of the demand realization to avoid using a termination policy to separate the types, as this practice destroys surplus. When the retailer values more the future, the manufacturer can offer a larger quantity and ask for a larger repayment because the retailer has more incentives to repay and stick with the relationship in the future. However, as the size of the shock increases, it becomes unprofitable to ask for such a low repayment regardless of the revenues. Hence, \(M\) is pushed to ask for a larger payment in "good times". Since the retailer is credit constrained and inter-temporal transfers of payoffs are impossible to use, termination occurs in equilibrium, which bounds \(q\) away from efficiency for no matter which \(\delta\).

Note that for any \(T\), \(\frac{\delta^{T+1}(1-s)}{1-s\delta^{T+1}} \leq \frac{\delta(1-s)}{1-\delta s}\), therefore from the previous sections we know that if \(0 < l < \frac{\delta^{T+1}(1-s)}{1-s\delta^{T+1}}\) we are in case \(IC_H + LL_L\), if \(\frac{\delta^{T+1}(1-s)}{1-s\delta^{T+1}} < l < \frac{\delta(1-s)}{1-\delta s}\) we are in case \(IC_H + IC_L + LL_L\) and if \(\frac{\delta(1-s)}{1-\delta s} < l < 1\) we are in case \(IC_H + IC_L\). Let us assume that the demand is linear and equal to: \(p_H(q) = a - bq\) and \(p_L(q) = l(a - bq)\), where \(0 < l < 1\). Figures 3-5 depict which regime yields the largest profits for \(M\) in the space \(\delta\) and \(l\). The difference between figures is the probability of the shock \(s\).
Figure 3: Optimal regime when $s = 0.1, a = 10, b = 1$ and $c = 2$

Figure 3 depicts the case where the likelihood of the shock is low and equal to $s = 0.1$. The line that separates the green and the red area is defined by $l = \frac{\delta(1-s)}{1-s^a}$ and the line that separates the red and the blue area is $l = \frac{\delta^{T+1}(1-s)}{1-s^a}$ where $T$ is the optimal termination policy. When the value of the future is small (i.e. $\delta$ is small) no contract can be implemented (and hence no trade occurs). If the demands in both states are similar (i.e. $l$ large and above $\frac{\delta(1-s)}{1-s^a}$), then $M$ asks for the same payment and hence does not need to use any termination policy. As the low demand becomes smaller ($l$ decreases below $\frac{\delta(1-s)}{1-s^a}$), then $LL_L$ starts binding and $M$ asks for a lower payment in the low state than in the high state ($D_L < D_H$) and punishes $R$ following the low payment for a finite number of periods. Finally, as $\delta$ increases, the value of the relationship increases and $IC_L$ stops binding (when $l < \frac{\delta^{T+1}(1-s)}{1-s^a}$), that is, the $L$-type retailer does not have incentives to walk away with the revenues.

Figure 4 and Figure 5 depict a shock of likelihood $s = 0.4$ and $s = 0.9$, respectively. Note that as $s$ increases, less contracts can be self-enforced (for low $l$ and $\delta$). Since the shock is more likely and the manufacturer does not want to punish an unlucky retailer, the regime $IC_H + IC_L$, where the payment is the same for both states and there is no termination, becomes more common. Finally, when $\delta$ is high, $M$ can ask for a larger payment from $R$ (i.e. we are in the case $IC_H + LL_L$) thanks to the increased value of the future.

---

24 Although this is out of the scope of this model, the no trade situation could also correspond to a manufacturer that vertically integrates downwards and serves the consumers.
Figure 4: Optimal regime when $s = 0.4$, $a = 10$, $b = 1$ and $c = 2$

Figure 5: Optimal regime when $s = 0.9$, $a = 10$, $b = 1$ and $c = 2$
5 The Model

In this Section, we generalize the examples presented above by exploring the scenario where the retailer faces a continuum of demand states. In particular, the demand takes the following form:

\[ p(q; s) \]

where \( s \) is a random variable distributed on the interval \([0, \bar{s}]\). We denote by \( h(s) \) and \( H(s) \) the density and cumulative distribution functions, respectively. In contrast with the examples above, a larger \( s \) means a better state of demand for a given \( q \) (that is, \( \frac{\partial p(q; s)}{\partial s} > 0 \)). As in the examples, \( M \) offers a quantity \( q \), a repayment \( D(\tilde{s}, q) \) and a termination policy \( T(\tilde{s}, q) \), for each particular shock reported \( \tilde{s} \). Denote the revenues for a given state \( s \) as \( R(q; s) = p(q; s)q \) and the expected revenues as \( R(q) = \int_0^{\bar{s}} R(q; s)h(s)ds \).

5.1 Benchmark

The firms maximize their one-period joint profits:

\[
\max_q R(q) - cq
\]

The first best quantity is determined by:

\[
q^{FB} = \int_0^{\bar{s}} \left[ \frac{\partial p(q; s)}{\partial q} q + p(q; s) \right] h(s)ds = c
\]

If the demand has the multiplicative functional form: \( p(q; s) = sp(q) \), the optimal quantity is then given by

\[
p'(q)q + p(q) = \frac{c}{\int_0^{\bar{s}} sh(s)ds}
\]

and hence we observe a reduction in quantity as in Examples 1 and 2. Now, the effective marginal cost is \( \bar{c} = \frac{c}{E(s)} \).

5.2 Relational contract

In this Section we consider the case of relational contracts. We first find the structure of an incentive compatible contract in this context and then we proceed to solve the problem of the manufacturer.

5.2.1 Incentive compatible contract

Since the quantity is chosen before the state of the demand is realized, it is not used for sorting purposes. Thus, the quantity is fixed at the time that the retailer is reporting how much he can repay. Then, for a given state of the demand \( s \) and a given quantity \( q \), \( R \) chooses a report \( \tilde{s} \) to maximize his profits:

\[
\pi_R(\tilde{s}; s) = R(q; s) - D(\tilde{s}, q) + \delta^{T(\tilde{s}, q) + 1} \pi_R
\]

Today's payoff of reporting \( \tilde{s} \)  Expected payoff of reporting \( \tilde{s} \)
where \( \pi_R \) is the expected discounted profits from staying in the relationship. Note that if the retailer were not credit constrained, the choice about which \( e \) to report would not depend on the true \( s \) as it is equally costly for any type of the retailer to report any \( \tilde{s} \). In other words, there is no sorting condition in this model, \( \frac{\partial^2 \pi_R(\tilde{s}; s)}{\partial e \partial s} = 0 \). This feature makes this setup different from Levin (2003). Indeed, repaying is like the "burning money" story of advertising\(^{25}\), where expending resources in advertising is equally costly for firms with different levels of efficiency. Because \( R \) is credit constrained, however, he cannot repay a larger \( D(\tilde{s}, q) \) than his actual revenues \( R(q; s) \):

\[
D(\tilde{s}, q) \leq R(q; s) \quad \forall s, \tilde{s}
\]

The limited liability condition (6) links the choice of the report with the true state.

For a given \( q \), the manufacturer has two instruments to deter \( R \) from under-reporting the state of the demand: following a low \( \tilde{s} \), she can either increase the payment today \( D(\tilde{s}, q) \) or increase the length of the termination period \( T(\tilde{s}, q) \), which decreases \( R \)'s value of the future trade. Given that increasing \( T(\tilde{s}, q) \) is also costly for \( M \) (because she loses future trade), whenever it is possible, \( M \) will set the payment today as high as possible: \( D(\tilde{s}, q) = R(\tilde{s}, q) \).

In equilibrium, however, \( M \) cannot extract all the revenues from \( R \) in all the states, as this would make the relationship worthless to the retailer (i.e., \( \pi_R = 0 \)). There must be a report \( s^* \) for which \( M \) does not ask for all the revenues and, consequently, does not terminate, \( T(s^*, q) = 0 \):

\[
\pi_R(s^*; s) = R(q; s) - R(q; s^*) + \delta \pi_R
\]

Hence for \( \tilde{s} \geq s^* \), \( M \) can only offer the contract: \( T(\tilde{s}, q) = 0 \) and \( D(\tilde{s}, q) = R(q; s^*) \), because she can no longer decrease the punishment period to compensate \( R \) for a larger repayment. The following Lemma summarizes the structure of the contract:

**Lemma 2** The manufacturer offers the following repayment quantities:

\[
D(\tilde{s}, q) = \begin{cases} 
R(q; \tilde{s}) & \text{if } \tilde{s} < s^* \\
R(q; s^*) & \text{if } \tilde{s} \geq s^*
\end{cases}
\]

and length of contract termination:

\[
T(\tilde{s}, q) = \begin{cases} 
T(\tilde{s}, q) & \text{if } \tilde{s} < s^* \\
0 & \text{if } \tilde{s} \geq s^*
\end{cases}
\]

where \( T(\tilde{s}, q) \) is to be defined.

\(^{25}\)See, for instance, Nelson (1974).

\(^{26}\)If \( M \) were to terminate for some periods following the report \( s^* \), she would strictly prefer to increase the repayment to decrease the termination length (and hence to reduce the inefficiency generated by the no-trade situation).
Let us proceed to lay down the conditions under which this contract induces truth-telling. Since \( \frac{\partial^2 \pi_R(\tilde{s}, \bar{s})}{\partial s \partial \bar{s}} = 0 \), let us denote by \( u(\tilde{s}) \) the part of \( R' \)'s payoff that does not depend on his type \( s \):

\[
u(\tilde{s}) = -D(\tilde{s}, q) + \delta^{T(\tilde{s}, q)+1} \pi_R
\]  

(8)

The independence between the incentives to report a demand state and the actual demand state makes the task of inducing truth-telling quite simple. Intuitively, if it were feasible, the retailer would always report the demand state \( \tilde{s} \) with the highest \( u(\tilde{s}) \), regardless of the true \( s \). Therefore, to retailer will report the truth if the rents are not dependent on \( s \).

**Proposition 3** To induce truth-telling from the retailer, the manufacturer gives him a non-negative rent that does not depend on the state of the demand (i.e., his type): \( u(\tilde{s}) \geq 0 \) and \( u'(\tilde{s}) |_{\tilde{s}=s} = 0 \forall s \).

**Proof.** See Appendix. ■

Since the rent does not depend on the report, let us redefine \( u(\tilde{s}) = u \). Using Proposition 3, equation (7) becomes \( u = -R(q; s^*) + \delta \pi_R \) when \( u > 0 \). We can combine it with (8) to find:

\[
\delta^{T(s, q)} = \frac{\delta \pi_R - R(q; s^*)}{\delta \pi_R} + \frac{R(q; s)}{\delta \pi_R}
\]  

(9)

Equation (9) defines \( T(\tilde{s}, q) \) from Lemma 2. Note that when \( R \) reports \( s = 0 \), if \( u > 0 \), then the \( M \) terminates for a finite number of periods, while if \( u = 0 \), she terminates the contract forever. In the last case, equation (9) becomes \( \delta^{T(s, q)} = \frac{R(q; s)}{\delta \pi_R} \). In both cases, a larger report is coupled with a shorter punishment period.

Note that, as we mentioned in Section 1.1, the contract offered by the manufacturer resembles to what it is known in the corporate finance literature as a debt contract. Debt contracts leave no money to the borrower in the bad states while making the borrower the residual claimant in the good states. The debt contract is optimal in this model because it minimizes the inefficiency associated with the termination (by trading-off larger repayments for lower termination periods in the default states) while inducing the retailer to repay.

Before solving the problem of the manufacturer, let us find \( \pi_R \) using Lemma 2:

\[
\pi_R = \int_0^{s^*} \delta^{T(s, q)+1} \pi_R h(s) ds + \int_{s^*}^{\bar{s}} [R(q; s) - R(q; s^*) + \delta \pi_R] h(s) ds
\]

Therefore, the retailer gives back all his revenues if the shock is smaller than \( s^* \) and has the contract terminated for \( T(s, q) \) periods. Otherwise, he repays the constant amount \( R(q; s^*) \) and keeps trading with the manufacturer in the next period. Using condition
(9), \( \pi_R \) can be simplified to:

\[
\pi_R = \int_0^{s^*} R(q; s) h(s) ds - R(q; s^*) + \delta \pi_R
\]

Consider a first best world where the retailer is not credit constrained. Then an optimal contract could be a quantity (chosen so that the expected revenues minus the production costs are maximized) and a fixed up-front payment from the retailer. If the manufacturer has all the bargaining power, this payment would be equal to the expected revenues (and hence the manufacturer would extract all the rents, \( \pi_R = 0 \)) and the retailer would never have his contract terminated.\(^28\) We can interpret the above equation in these terms, where the retailer keeps the expected profits minus a fixed payment, \( R(q; s^*) \), and he never has his contract terminated. The difference from the first best world would come in terms of a different quantity and a smaller repayment amount in order to leave some rents to the retailer (\( \pi_R > 0 \)) so it is worth for him to stay in relationship. Solving for \( \pi_R \) yields

\[
\frac{\pi_R}{R(q) - R(q; s^*)} = \frac{1}{1 - \delta}
\]

Hence, for the expected discounted profits to be positive, it needs to be the case that the revenues at \( s^* \) are smaller than the expected revenues.

Finally, using (10), the incentive compatibility condition of Proposition 3 becomes:

\[
u = \frac{\delta R(q) - R(q; s^*)}{1 - \delta} \geq 0
\]

In order to ensure the repayment, the manufacturer needs to guarantee that the retailer obtains at least the difference between what he can steal tomorrow if he stays in the relationship and the maximum repayment that he can potentially do today.

5.2.2 The problem of the Manufacturer

Using Lemma 2, the profits of the manufacturer are:

\[
\pi_M = \int_0^{s^*} \left[ R(q; s) + \delta^{T(s,q)+1} \pi_M \right] h(s) ds + (1 - H(s^*)) \left[ R(q; s^*) + \delta \pi_M \right] - cq
\]

that is, the manufacturer receives all the revenues in the bad states (\( s < s^* \)) and terminates the contract with the retailer for \( T(s, q) \) periods. Otherwise, the manufacturer charges a fixed repayment, \( R(q; s^*) \), and never terminates the contract. Finally, the manufacturer incurs the production costs.

\(^{27}\)If \( u = 0 \), then

\[
\pi_R = \int_0^{s^*} R(q; s) h(s) ds
\]

\(^{28}\)Note that an ex-ante equivalent possibility is for the manufacturer to ask for the obtained revenues in all the states.
When choosing \( s^* \), \( M \) faces the following trade-off: a larger \( s^* \) allows \( M \) to extract a larger expected repayment from the retailer today (and hence \( \pi_R \) decreases); however, by (9), and in order to keep the incentives unchanged for all other types, this comes at a cost of a longer termination period.\(^{29}\) On the other hand, an increase in the quantity \( q \) leads to an increase in the total payment as well as an increase in the production costs. Since it also leads to an increase of \( \pi_R \), the effect on the termination policy is less clear and depends on the particular form of the demand function.

Let us plug (9) in the objective function and single out \( \pi_M \):

\[
\pi_M = \pi_R \frac{\int_0^{s^*} R(q; s)h(s)ds + (1 - H(s^*)) R(q; s^*) - cq}{R(q) - \left( \int_0^{s^*} R(q; s)h(s)ds + (1 - H(s^*)) R(q; s^*) \right)}
\]

where \( \pi_R \) is defined in (10). The expected discounted profits of the manufacturer can be rewritten as the product of the retailer’s earnings and the manufacturer’s expected earnings today, divided by the loss in the surplus due to the relational contract.

The problem of the manufacturer is then to choose \( s^* \) and \( q \) to maximize \( \pi_M \) subject to (11). After replacing \( \pi_R \), the problem of the manufacturer is\(^{30}\)

\[
\max_{q, s^*} \pi_M = \frac{\int_0^{s^*} R(q; s) h(s) ds + (1 - H(s^*)) R(q; s^*) - cq}{(1 - \delta) R(q) - \left( \int_0^{s^*} R(q; s) h(s) ds + (1 - H(s^*)) R(q; s^*) \right)} \tag{12}
\]

\[
\text{s.t. } \delta R(q) - R(q; s^*) - \frac{\delta R(q) - R(q; s^*)}{1 - \delta} \geq 0
\]

Note that as \( \delta \to 1 \), the constraint (11) does not bind. When this is the case, the optimal \( s^* \) and \( q \) do not depend on \( \delta \). Conversely, if (11) binds, both choices depend on \( \delta \). In particular, and in line with Example 1 and 2, the fixed repayment that the manufacturer asks for the types above \( s^* \) will be equal to what the retailer expects to obtain tomorrow if he does not repay: \( R(q; s^*) = \delta R(q) \). Therefore, a larger discount factor allows for a larger \( s^* \) that the manufacturer can impose on the retailer. Also, when (11) binds, an increase in \( s^* \) does not unambiguously increase the expected repayment because the fixed repayment for larger types \( R(q; s^*) \) is constrained by \( \delta R(q) \) and hence

\(^{29}\) Indeed, after substituting for \( \pi_R(s^*, q) \), the sign of the derivative is negative:

\[
\frac{\partial \pi_T(s, q)}{\partial s^*} = -\frac{(1 - \delta) \partial R(s^*, q)}{\partial s^*} \frac{R(q) - R(s, q)}{(R(q) - R(s^*, q))^2} < 0 \forall s \leq s^*
\]

where the inequality follows from the fact that in order to have \( \pi_R(s^*, q) > 0 \), it needs to be the case that \( R(q) > R(s^*, q) \).

\(^{30}\) Note that the retailer’s participation constraint is never binding as he always has the option of walking away with the current revenues \( R(q; s) \) by reporting \( \tilde{s} = 0 \) and hence not repaying anything. Thus, \( R \) receives at least the current revenues:

\[
\pi_R(\tilde{s}; s) = R(q; s) + \delta^{T(0, q) + 1} \pi_R \geq 0 \forall s
\]
remains unchanged. Since the expected repayment remains unchanged, the termination policy does not change either. Thus, the only instrument available for the manufacturer is the choice of \( q \). Note as well that the manufacturer will terminate with the retailer forever if there is no repayment at all.

The following Proposition summarizes this discussion and characterizes the optimal choices of the manufacturer.

**Proposition 4** The optimal \( s^* \) and \( q \) are determined by

\[
(1 - H(s^*)) \pi_R - H(s^*) \pi_M E(s) - E(s | s \leq s^*) = \hat{\lambda} \tag{13a}
\]

\[
\pi_R \frac{\partial R(q; s^*)}{\partial q} - \delta \frac{\partial R(q)}{\partial q} + \frac{\int_0^{s^*} \left( \frac{\partial R(q; s)}{\partial q} - \frac{\partial R(q; s^*)}{\partial q} \right) h(s) ds}{R(q) - R(q; s^*)} \left( \frac{\partial R(q)}{\partial q} - \delta \frac{\partial R(q)}{\partial q} \right) = \hat{\lambda} \tag{13b}
\]

where \( \hat{\lambda} \) is the shadow cost of the incentive compatibility constraint adjusted by the second best loss of revenues, \( \hat{\lambda} = \lambda \left( R(q) - \int_0^{s^*} R(q; s) h(s) ds - (1 - H(s^*)) R(q; s^*) \right) \). If \( \delta \) is large enough \( \left( \frac{R(q; s^*)}{R(q)} < \delta \right) \), then \( \hat{\lambda} = 0 \) and the optimal \( s^* \) and \( q \) are independent of \( \delta \).

**Proof.** See Appendix. \( \blacksquare \)

Since the solution depends on the particular demand function, in what follows, we illustrate these results with an example. For simplicity and comparability with the previous examples we use again the multiplicative demand function.

### 5.2.3 Example

Let us consider the case where \( p(q; s) = sp(q) \). Then the retailer’s profits are: \( \pi_R = \frac{(E(s) - s^*) p(q) q}{1 - \delta} \), and thus, in order for \( \pi_R > 0 \), the threshold \( s^* \) needs to be smaller than the expected shock. The problem of the manufacturer (12) becomes:

\[
\max_{q, s^*} \pi_M = \frac{\hat{E}(s, s^*) p(q) q - cq}{1 - \delta} \frac{E(s) - E(s, s^*)}{E(s) - s^*}
\]

\[
\text{ s.t. } \frac{\delta E(s) - s^*}{1 - \delta} \geq 0 \tag{14}
\]

where \( \hat{E}(s, s^*) = H(s^*) E(s | s \leq s^*) + (1 - H(s^*)) s^* \) is the expected shock that matters to \( M \) in terms of the repayment. Indeed, for \( s \leq s^* \) the manufacturer gets all the revenues repaid while for \( s > s^* \) she gets the revenues of the state \( s^* \). Clearly, for any \( s^* < \bar{s} \), \( \hat{E}(s, s^*) < E(s) \), which leads to the downward distortion of \( q \) compared to the benchmark.
Figure 6: \( p(s, q) = sp(q) \), \( s \sim U \, [0,1] \)

The incentive compatibility constraint (11) becomes (14). It does not bind if \( s^* \) is smaller than the discounted value of the expected shock tomorrow, which happens for a large \( \delta \).

Figure 6 depicts the current and future profits of the manufacturer as a function of the realization of the shock for the case where the shock is distributed like a Uniform on [0, 1].

The payment today is the current revenues, \( R(q; s) \), up to \( s^* \), and a fixed payment, \( R(q; s^*) \), from \( s^* \) onwards. The future discounted benefits from maintaining the relationship are constant and equal to \( \delta \pi_M \) for \( s \) larger than \( s^* \), and \( \delta^{T(q,s)+1} \pi_M \) for \( s \) smaller than \( s^* \).

Before we state the first order conditions, let us understand the trade-offs that the manufacturer is facing when choosing \( s^* \) and \( q \). If the incentive compatibility constraint (14) binds, then \( s^* \) is determined by it so \( M \) is left to choose only \( q \). In particular, \( M \) sells the quantity that maximizes her profits taking into account the repayment that she is expected to obtain and that it is determined by \( s^* \).

When the constraint (14) does not bind, \( M \) can also choose \( s^* \). A larger \( s^* \), on one hand, increases the expected repayment\(^31\) and hence \( \pi_M \). On the other, because it also decreases \( \pi_R \) and \( R \)'s incentives to repay, it increases the inefficiency due to tougher punishments\(^32\).

**Corollary 3** If \( p(q; s) = sp(q) \), then the optimal \( s^* \) and \( q \) are determined by the following

\(^{31}\) Indeed, \( \frac{\partial \hat{E}(s,s^*)}{\partial s^*} > 0 \).

\(^{32}\) This is reflected in the denominator of \( \pi_M \), which is increasing in \( s^* \) because \( \frac{\partial \hat{E}(s,s^*)}{\partial s^*} < 1 \).
first order conditions when $\delta E(s) > s^*$:

$$ q : p'(q)q + p(q) = \frac{c}{E(s, s^*)} \quad (15) $$

$$ s^* : \frac{p(q)}{c} = \frac{H(s^*)}{1 - H(s^*) \frac{E(s) - E(s \mid s \leq s^*)}{E(s) - E(s^*)}} \quad (16) $$

Otherwise, $s^* = \delta E(s)$ and $q$ is determined by (15). $s^*$ and $q$ are strategic complements.

Note that (15) and (16) do not contain $\delta$. Thus if the value of the future is large, the constraint (14) does not bind, and the choice of $q$ and $s^*$ is not affected by $\delta$. In this case, the manufacturer would set the first best quantity only if $s^* = \bar{s}$, which is not possible as this makes the retailer’s profits, $\pi_R$, zero. Therefore, Proposition 2 applies and the outcome is bounded away from efficiency even for $\delta$ arbitrarily close to 1. This is because in order to make the relationship profitable for the manufacturer, the limited liability constraint (6) binds for some states (i.e. $0 < s^*$). This in turn implies that the termination is used in equilibrium and thus the surplus is bounded away from efficiency.

When the value of the future is low enough, $\delta < \frac{s^*}{E(s)}$, then the constraint (14) binds, in which case $s^*$ and $q$ do depend on $\delta$. Then an efficient quantity will require $\delta > 1$, which is not possible either. Finally, note that, since $\frac{\partial E(s, s^*)}{\partial s^*} > 0$, the downward distortion in the quantity decreases with $s^*$, regardless of whether the constraint binds. This is because, as in the previous examples, a larger quantity needs to be associated with a tougher punishment and hence a larger $s^*$. Therefore, $q$ and $s^*$ are strategic complements.

Using Corollary and condition (9), it is possible to derive the termination policy implemented by the manufacturer. In particular, if (14) binds:

$$ T(s, q) = \begin{cases} \ln(\frac{\pi_{tr}}{\ln s}) & \text{if } s < \delta E(s) \\ \ln s & \text{if } \delta E(s) \leq s \end{cases} $$

If (14) does not bind:

$$ T(s, q) = \begin{cases} \frac{\ln(\delta E(s) - s^* + (1 - \delta) E(s) - s^*)}{\ln s} & \text{if } s < s^* \\ 0 & \text{if } s^* \leq s \end{cases} $$

where $s^*$ is defined by (16).

Finally, in order to compare the general model with the numerical example in Section 4, we assume the linear demand function $p(q; s) = s(a - bq)$ and that $s$ is distributed

---

33 Because only then: $E(s) = \bar{E}(s, s^*)$.

34 Otherwise, $M$ would produce no quantity because $\bar{E}(s, s^*) = 0$.

35 Indeed, $\delta = \frac{s^*}{E(s)} > 1$. 

27
uniformly on the interval \([0, 1]^{36}\). For \(a = 10\), \(b = 1\) and \(c = 2\), the optimal \(s^* = 0.41\) and \(q = 1.92\) if \(\delta\) is at least \(0.82\). Otherwise, \(s^*\) and \(q\) are increasing in the discount factor. If \(\delta < 0.45\), then there is no trade. Figures 7 and 8 depict \(s^*\) and \(q\) as a function of \(\delta\) for this particular example.

6 Empirical implications

This analysis suggests that we might observe fewer contracts renewed, lower quantity sold and greater wholesale markups when demand uncertainty is higher. Demand uncertainty could potentially be proxied by macroeconomic volatility (across places). Another approximation to demand uncertainty could be the differences in the ease of verification. For instance, demand shocks could be easier to verify with non-perishable goods as compared to perishable ones. The distance between wholesale and retail markets could also be used to proxy the difference in the ease of verification as the upstream firm may have more information about the state of the demand the closer she is to the retail market. Finally, another possibility could be the presence or absence of ethnic or other links be-

\[ q = \frac{p'(q)q + p(q)}{c} \]

\[ \frac{p(q)}{c} = \frac{s^*}{s^* (1 - \frac{s^*}{2}) (1 - s^*) - \left( \frac{1}{2} - s^* \right)} \]

\[ \frac{q}{0.8} \]

\[ \frac{q}{0.6} \]

\[ \frac{q}{0.4} \]

\[ \frac{q}{0.2} \]

\[ \frac{q}{0} \]

Figure 7: \(s^*\) when \(a = 10\), \(b = 1\) and \(c = 2\)

\[ \text{Benchmark} \]

\[ \text{Relational contract} \]
between wholesalers and retailers which would make the upstream or downstream firm more informed.

If we interpret the "no trade" situations of the model as the upstream firm directly selling to the consumer, then we expect to see more vertical integration when demand uncertainty is higher.

Finally, it would be interesting to test whether higher discount rates (proxied for instance, by larger interest rates) or greater revenues are associated with fixed repayments.

7 Conclusions

The goal of this paper has been to study the impact of trade credit and adverse selection on the vertical relational contract between an upstream and a downstream firm. Indeed, trade credit limits the available instruments that can be used within a relational contract and this affects the form of the emerging contract.

Contrary to Levin (2003), the upstream firm does not want to maximize the joint surplus, even though more surplus could be used to credibly punish and reward the downstream firm. This is because maximizing the joint surplus would give the downstream firm more incentives to steal which would require larger termination periods. Since the termination policy is also costly for the upstream firm (because of the resulting loss of future trade), the upstream firm prefers to distort the quantity downwards, away from the efficient quantity, in order to decrease the revenues held by the downstream firm. Because of this, we observe double marginalization despite the use of two-part tariffs. It is left for future work to determine in which way the contract would change if the
upstream firm could not perfectly choose the final quantity sold in the market.

Not renewing the contract for a number of periods is costly for the upstream firm and this also has two clear implications in terms of the contract offered. First, if the revenues are large enough (above the state of the demand $s^*$), the upstream firm will never terminate the contract and ask for a fixed repayment. Otherwise, the upstream firm will ask for the highest possible repayment (equal to the generated revenues) and punish for a number of periods. Furthermore, the smaller the repayment, the larger the termination period. The choice of the threshold state $s^*$ is linked to the choice of the quantity sold to the retailer. In particular, the larger this threshold, the smaller will be the downward distortion of the quantity because the larger is the share accrued by the upstream firm.

Finally, since the use of trade credit does not allow the firms to share the joint profits in an arbitrary way (i.e. using fixed transfers), the contract is expected to change depending on whether it is the upstream or downstream firm making the offer. It is left for future work to explore in which particular way the contract would change if it were offered by the downstream firm.

In our analysis, we have assumed that the quantity offered by the downstream firm does not change depending on the past repayment history. This framework would be suitable for industries where it is very costly to adjust the production quantity from one period to the other. In the future, it would be interesting to explore the form of non-stationary contracts and determine in which particular way the upstream firm will increase or decrease the quantity in each period as a function of the previous period repayment.

8 Appendix

Remark about non-stationarity. To get a sense about how restrictive it is the assumption of not letting the manufacturer offer a second opportunity to the retailer following an nonpayment, let us explore the following setup: imagine that the manufacturer starts selling the quantity $q$ to the retailer and asks him to repay $D_q$. If the retailer does not pay back, then he gets a second chance where he is offered a quantity $x$ and a repayment $D_x$. If after this second opportunity, the retailer repays, the manufacturer offers again $\{q, D_q\}$; otherwise, he is punished for $T$ periods, after which the trading starts again with the contract $\{q, D_q\}$. Denote by $\pi_{Rq}$ and $\pi_{Rx}$ the present discounted value of the terms of trade $\{q, D_q\}$ and $\{x, D_x\}$, respectively. The retailer repays in the first period if:

$$(1 - s) \left[p(q)q - D_q + \delta \pi_{Rq} \right] + s \delta \pi_{Rx} \geq (1 - s) p(q)q + \delta \pi_{Rx} \quad ((IC_1))$$

and in the second period if:

$$(1 - s) \left[p(x)x - D_x + \delta \pi_{Rq} \right] + s \delta^{T+1} \pi_{Rq} \geq (1 - s) p(x)x + \delta^{T+1} \pi_{Rq} \quad ((IC_2))$$
Note that these constraints limit the repayment that M is expected to receive in each state respectively: \( D_q \leq \delta (\pi_{Rq} - \pi_{Rx}) \) and \( D_x \leq \delta (1 - \delta^T) \pi_{Rq} \), where \( \pi_{Rq} = (1 - s) \frac{p(q)\alpha - D_x + s\delta(p(x)x - D_x)}{1 - \delta(1-s) - s\delta(\delta(1-s) + \delta^T)} \) and \( \pi_{Rx} = (1 - s) \frac{(1 - \delta^T) p(q)x - D_q + \delta(1-s) + \delta^T}{1 - \delta(1-s) - s\delta(\delta(1-s) + \delta^T)} \).

The respective present discounted values for the manufacturer are defined by:

\[
\pi_{Mq} = (1 - s) (D_q + \delta \pi_{Mq}) + s \delta \pi_{Mx} - cq
\]

and:

\[
\pi_{Mx} = (1 - s) (D_x + \delta \pi_{Mq}) + s \delta^T \pi_{Mq} - cx
\]

The problem of the manufacturer is then:

\[
M_{q,x,T} \pi_{Mq} = \frac{(1 - s) D_q - cq + s \delta [(1 - s) D_x - cx]}{1 - \delta (1 - s) - s \delta [\delta (1 - s) + \delta^T]} \quad \text{s.t.}
\]

\[
D_q = \delta (\pi_{Rq} - \pi_{Rx}) \quad (IC_1)
\]

\[
D_x = \delta (1 - \delta^T) \pi_{Rq} \quad (IC_2)
\]

therefore introducing the second chance, decreases the incentives to repay in the first period (see \( IC_1 \)), but reduces the inefficiency following the shock (see \( \pi_{Mq} \)). Plugging in \( IC_1 \) and \( IC_2 \) into the objective function, we obtain the following first order condition for \( x \):

\[
\frac{\partial \pi_{Mq}}{\partial x} = -\frac{1 - \delta - s \delta^2 (1 - \delta^T)}{A + BC} (1 - s)^2 [p'(x)x + p(x)] - sc = 0
\]

where \( A = \delta^2 (1 - s)^2 (1 - \delta) (1 - \delta^T) \), \( B = 1 - \delta (1 - s) - s \delta^T + 2 \) and \( C = 1 - \delta^2 (1 - s) - s \delta^T + 2 \). Note that \( \frac{\partial \pi_{Mx}}{\partial x} < 0 \) if \( 1 - \delta - s \delta^2 (1 - \delta^T) > 0 \), which is obviously true for \( T = 0 \). Since the expression is decreasing in \( T \), the lower bound is provided when \( T = +\infty \). It easy to check that \( \frac{\partial \pi_{Mq}}{\partial x} < 0 \) when \( \delta < 0.62 \) and if \( \delta \geq 0.62 \), then when \( \frac{1}{\delta} (\frac{1}{\delta} - 1) > s \). In the shaded area of Figure 9 the manufacturer offers \( x = 0 \). Note that this is for \( T = 0 \) and that the shaded area will increase if one consider the optimal \( T \). ■

**Proof of Lemma 1.** The first order condition for the length of the punishment is:

\[
\frac{\partial \pi_M}{\partial T} = \frac{-\delta^T \ln \delta}{(1 - \delta - s \delta^2 (1 - \delta^T))} \left[ \frac{(1 - \delta)^2 - s \delta^2 (1 - \delta^T)^2}{(1 - \delta^T)^2} (1 - s)^2 p(q) + sc \right] q
\]

This derivative is always positive at \( T = 0 \). It tends to \( 0^+ \) as \( T \to +\infty \) if \( s < \frac{(1 - \delta)^2}{\delta^2} \), therefore, in the white region of Figure 10 \( M \) terminates forever. If \( s > \frac{(1 - \delta)^2}{\delta^2} \), \( M \) still terminates forever if:

\[
\frac{c}{\delta (1 - s)^2} > \left[ \frac{\delta^2 s - (1 - \delta)^2}{\delta s} \right] p(q)
\]

Since by (2), \( \frac{c}{\delta (1 - s)^2} < p(q) \), and because \( \frac{\delta^2 s - (1 - \delta)^2}{\delta s} \in [0,1] \), for some parameter values, this inequality may be true (for instance when \( s \to 1 \)). When the inequality does not hold
Figure 9: The shaded area is the minimum region in which the optimal $x = 0$

(for instance when $\delta \to 1$), the optimal termination period is positive and finite. Since the second derivative (evaluated at $\frac{\partial \pi_M}{\partial T} = 0$) is negative, the solution is a maximum.

$$\frac{\partial^2 \pi_M}{\partial T^2} = - \frac{\delta^2 (T+1) (\ln \delta)^2 (1-s)^2 \left( (1-\delta)^2 + \delta s (1-\delta^T) (1-\delta + 1 - \delta^{T+1}) \right)}{(1-\delta + s\delta (1-\delta^T))^2 (1-\delta^{T+1})^4} p(q)q$$

**Operations for Example 2.** When the three constraints bind, the first order condition for $q$ is:

$$\frac{\partial \pi_M}{\partial q} = \frac{\delta (1-s) R'(q) + s (p'_L(q)q + p_L(q)) - c (R(q) (1-\delta (1-s)) - sp_L(q)q)}{R(q) \left( 1-\delta (1-s) - sp_L(q)q \right)^2}$$

$$\frac{\partial \pi_M}{\partial q} = \frac{s (\delta (1-s) R(q) + sp_L(q)q - cq) \left( p'_L(q)q + p_L(q) - p_L(q)q \frac{R'(q)}{R(q)} \right)}{R(q) \left( 1-\delta (1-s) - sp_L(q)q \frac{R'(q)}{R(q)} \right)^2} = 0$$

substituting $p_L(q)$ by $lp(q)$ and $p_H(q)$ by $p(q)$ give us condition (4). The comparative statics of $s$ are:

$$\frac{\partial^2 \pi_M}{\partial q \partial s} \propto - \frac{c \left( \delta [1-s+sl+(1-s)(1-l)] - l \right)}{(\delta (1-s)(1-s+sl)+sl)^2}$$

therefore, $q$ decreases following an increase in $s$ if:

$$\delta > \frac{l}{1-s+(1-s)(1-l)+sl}$$

which is always true because the RHS is increasing in $l$, and the inequality holds at the upper bound of $l$, $\frac{\delta (1-s)}{1-\delta s}$. By inspection, it is easy to see that $\frac{\partial D_H}{\partial s} < 0$, $\frac{\partial D_L}{\partial s} < 0$ and $\frac{\partial T}{\partial s} < 0$. 32
For the case where $IC_H$ and $LL_L$ bind, the first order condition for $q$ is:

\[
\frac{\partial \pi_M}{\partial q} = \frac{\delta (1 - s) \left(1 - \delta^T\right) R'(q) + \left(1 - \delta (1 - s) - s\delta^{T+1}\right) \left(p'_L(q)q + p_L(q)\right)}{1 - \delta^{T+1}} - c
\]

and for $T$ is:

\[
\frac{\partial \pi_M}{\partial T} = \frac{-\delta^{T+1} \ln \left[ \frac{\left(1-\delta\right)^2 - s\delta^2 \left(1-\delta^T\right)^2}{\delta^{T+1}} \right] \left[ \left(1-s\right)R(q) - \left(1-\delta(1-s) - s\delta^{T+1}\right) \right] p_L(q)q + scq}{\left(1 - \delta (1 - s) - s\delta^{T+1}\right)^2}
\]

Note that $\frac{\partial \pi_M}{\partial T}$ evaluated at $T = 0$ is:

\[
\frac{\partial \pi_M}{\partial T} = \frac{-\delta \ln \delta \left[ \left(1-s\right)R(q) - p_L(q)q + scq\right]}{\left(1 - \delta\right)^2}
\]

which is positive if the expression in brackets is positive. Plugging in the upper bound of $p_L(q)$ given by $IC_L$, we obtain: $\frac{1-s-\delta}{1-s\delta} (1 - s)p_H(q) + sc > 0$. Thus a sufficient condition for $T > 0$ is $1 - s > \delta$. ■

**Proof of Proposition 2.** By inspection, the RHS of (3) and (4) are larger than $\tilde{c}$. Using l’Hopital’s rule, the RHS of (5) tends to $\frac{(1-s)^2 T + (1-s + Ts(1-s) + (T+1)s)}{T+1}$ as $\delta \to 1$. Noting that it is increasing in $T$ and that it is larger than $\tilde{c}$ at $T = 0$ completes the proof.

■

**Proof of Proposition 3.** Because of condition (6), a retailer can lie only downwards. To avoid the retailer underreporting his type, $M$ needs to ensure that $u'(\hat{s}) \geq 0$, that is, the payoff $u(\hat{s})$ increases with reported $\hat{s}$. However, $M$ cannot design a contract that satisfies $u'(\hat{s}) > 0$ for $s \in [\hat{s}, \overline{s}]$ because she has no means to increase $u(\hat{s})$ (i.e., she is already not terminating with the retailer and asking for a smaller repayment $R(q; s') < R(q; s^*)$).
would make types $s \in [s', s^*]$ overreport their types). Furthermore, for $s \in [0, s^*)$ the contract cannot satisfy $u'(\tilde{s}) > 0$ either, because $M$ is already extracting the maximum payment and increasing the payoff with $\tilde{s}$ would imply a costly (for $M$) increase in the termination period for smaller types. Indeed, $M$ wants to set $T(s, q)$ as low as possible as long as there is truth-telling. Finally, to prevent $R$ walking away with the earnings, $M$ need to ensure that $u(\tilde{s})$ is non-negative, which completes the proof.

### Proof of Proposition 4.

The Lagrangian function is:

$$
\mathcal{L} = \int_{s'}^{s^*} R(q; s) h(s) ds + (1 - H(s^*)) R(q; s^*) - cq
\frac{1 - H(s^*)}{(1 - \delta) \left( 1 + \frac{H(s^*) R(q; s^*) - \int_{s'}^{s^*} R(q, s) h(s) ds}{R(q) - R(q; s^*)} \right)} - \lambda (R(q; s^*) - \delta R(q))
$$

where $\lambda$ is the shadow cost associated with the constraint. The first order conditions are:

$$
\frac{\partial \mathcal{L}}{\partial s^*} = \pi_R \frac{\partial R(q; s^*)}{\partial s^*} \left( 1 - H(s^*) \right) R(q) - \int_{s'}^{s^*} R(q; s) h(s) ds - (1 - H(s^*)) R(q; s^*) - \pi_M \frac{\partial R(q; s^*)}{\partial s^*} \frac{H(s^*) R(q) - \int_{s'}^{s^*} R(q, s) h(s) ds}{R(q) - R(q; s^*)} - \lambda \left( R(q; s^*) - \delta R(q) \right) = 0
$$

$$
\frac{\partial \mathcal{L}}{\partial q} = \pi_R \int_{s'}^{s^*} \frac{\partial R(q, s)}{\partial q} h(s) ds + (1 - H(s^*)) \frac{\partial R(q, s^*)}{\partial q} - c
\frac{1 - H(s^*)}{R(q) - \int_{s'}^{s^*} R(q, s) h(s) ds - (1 - H(s^*)) R(q; s^*)} - \pi_M \frac{H(s^*) \frac{\partial R(s^*)}{\partial q} - \int_{s'}^{s^*} \frac{\partial R(q, s)}{\partial q} h(s) ds - \left( H(s^*) R(q, s^*) - \int_{s'}^{s^*} R(q, s) h(s) ds \right) \left( \frac{\partial R(q)}{\partial q} - \frac{\partial R(q, s^*)}{\partial q} \right)}{R(q) - R(q; s^*)} - \lambda \left( \frac{\partial R(q; s^*)}{\partial q} - \delta \frac{\partial R(q)}{\partial q} \right)
$$

and

$$
\lambda (R(q; s^*) - \delta R(q)) = 0
$$

Setting the first two conditions equal to zero, give us equation (13a) and (13b), respectively.

We assume that the second order conditions hold. Note that the optimal $s^*$ cannot be in a corner as this would give zero profits to either the manufacturer (if $s^* = 0$) or to the retailer (if $s^* = 1$). We assume that the demand and the cost function are such that production takes place, i.e., $0 < q < +\infty$.

### Proof of Corollary 3.

Suppose that the constraint does not bind, the first order conditions are:

$$
\frac{\partial \pi_M}{\partial q} = \frac{\hat{E}(s, s^*) (p'(q) q + p(q)) - c}{(1 - \delta) E(s) - \hat{E}(s, s^*)} E(s) - s^* = 0
$$
\[
\frac{\partial \pi_M}{\partial s^*} = (1 - H(s^*)) \frac{E(s) - s^*}{E(s) - \hat{E}(s, s^*)} p(q)q \\
+ H(s^*) \frac{E(s \mid s \leq s^*) - E(s)}{(E(s) - \hat{E}(s, s^*))^2} \left( \hat{E}(s, s^*) p(q)q - cq \right) = 0
\]

The cross derivative is:
\[
\frac{\partial^2 \pi_M}{\partial s^* \partial q} = (1 - H(s^*)) \frac{E(s) - s^*}{E(s) - \hat{E}(s, s^*)} (p'(q)q + p(q)) \\
+ H(s^*) \frac{E(s \mid s \leq s^*) - E(s)}{(E(s) - \hat{E}(s, s^*))^2} \left( \hat{E}(s, s^*) (p'(q)q + p(q)) - c \right) = 0
\]

Using \( \frac{\partial \pi_M}{\partial q} = 0 \), it is easy to see that \( \frac{\partial^2 \pi_M}{\partial s^* \partial q} > 0. \)

**References**


