Vertical Relational Contracts and Trade Credit

Marta Troya-Martinez*
University of Oxford

Job Market Paper

November 12, 2013

Abstract

Trade credit plays a very important role in inter-firm transactions. Because formal contracts are often unavailable, it is granted within an ongoing relationship. We characterize the optimal relational contract, when the ability to repay is unknown to the supplier and the threat of trade suspension is used to discipline the buyer. The optimal contract resembles a debt contract: if the fixed repayment is met, the contract is renewed. Otherwise, the supplier demands the highest feasible repayment and suspends trade for some time. The length of the trade suspension is contingent on the repayment. We provide a novel explanation for why the quantity is rationed, even when a repayment is met. We also show how the need for contract self-enforceability can be good for welfare because it limits how inefficiently tough the supplier can be.

JEL Classification: C73, D82, L14.
Keywords: Relational contracts, trade credit, imperfect monitoring.

*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 Vicente Cuñat, Mikhail Drugov, Marcel Fafchamps, Matthias Fahn, Guido Friebel, Rachel Griffith, Lakshmi Iyer, James Malcolmson, Meg Meyer, Ines Moreno de Barreda, Massimo Motta, Michael Powell, Patrick Rey, Andrew Rhodes, John Thanassoulis and the participants of the Gorman workshop in Oxford, the CSAE Conference 2011 in Oxford, JEl 2011 in Valencia, the RES 2012 meeting in Cambridge, the Econometric Society NASM 2012 in Evanston, RES mentoring meeting 2012 in York, EARIE 2012 in Rome, Barcelona GSE Winter Workshop 2012, IIOC 2013 in Boston, 2013 Workshop on Institutions in Alghero and the FIRMS seminar at CEMFI. I gratefully acknowledge financial support from Fundación Ramón Areces. Any remaining errors are my own.
1 Introduction

Vertically-related firms find it extremely costly to conduct all their business-to-business transactions on a purely cash-and-carry basis, and the capacity to enter into contracts with delayed obligations is essential for a good business environment. The delay in the payment of goods already delivered, what is known as trade credit, accounts for about 11.5 to 17 percent of the assets of G-7 traded non-financial firms (Rajan and Zingales (1995)). In developing countries, limited access to capital markets makes trade credit even more important (Fafchamps 2000).

Trade credit is rarely secured on collateral. Moreover, enforcing repayment through the courts can be problematic.\textsuperscript{1} As a result, trade credit is usually granted within an ongoing relationship.\textsuperscript{2} A large body of work has found evidence of the link between relational contracts and the provision of trade credit; especially in developing countries, or in international transactions, where relational contracts substitute for missing laws of contract or differences in legal systems.\textsuperscript{3} This practice is also expected to be pervasive when transactions are not entirely legal. For example, when firms operate in the shadow economy or in black markets, such as the drug trade.

This is the first paper that takes into account the relational contracting nature of trade credit. It does so in an environment where adverse shocks may make the cashless downstream firm unable to honor the credit agreement. Since these shocks are not observable to the supplier, trade credit, by postponing the payment until the shock is realized, makes asymmetric information matter. Once we take into account the relational

\textsuperscript{1}Legal costs may be too high relative to the size of the transaction, and outstanding credit is usually placed at the end of the debt priority queue in case of bankruptcy. Furthermore, the buyer may have been affected by a shock, leaving nothing to the supplier to foreclose on.

\textsuperscript{2}As Cuñat and García-Appendini (2012) put it: "The frequent occurrence of late payment highlights that it is hard to understand trade credit as a fully contractual, independent, one-off transaction. (...) In most cases trade credit has to be understood as a multi-period, highly non-contractual type of credit that interacts with an ongoing commercial relationship." p. 543

nature of trade credit, a whole new set of important questions arises. For instance, how is the contract affected by the lack of legal enforcement? How does the contract self-enforceability interact with the asymmetric information problem? How does the contract respond to an increase in the level of trust between the firms? When addressing these questions, we take the provision of trade credit as given and look at the impact on the different contract characteristics, such as non-payment penalty, quantity of the good sold and repayment. Instead, the main focus of the literature has been on explaining why, when and how much trade credit should be granted.\(^4\)

The model uses an agency setting where an upstream firm repeatedly supplies a good and offers trade credit to a cashless downstream firm.\(^5\) For instance, the upstream firm ("she") can be a manufacturer and the downstream firm ("he") a retailer. The manufacturer’s machinery is used as collateral, making her less credit constrained than the retailer.\(^6\) The manufacturer has all the bargaining power in dictating the terms of trade. She proposes a quantity forcing contract which establishes the quantity supplied to the final consumers and a repayment that is postponed until the sale is made. The retailer places the good in the market and obtains some stochastic revenues. Depending on the revenues, the retailer may be unable to repay either part of or the whole amount.

The manufacturer faces two problems. First, she cannot distinguish a genuine from a strategic default where the retailer privately diverts revenues.\(^7\) To induce repayment,

\(^4\)For instance, Burkart and Ellingsen (2004) find that trade and bank credit are complements because goods are less divertible to private benefits than money. In the same spirit, Giannetti, Burkart and Ellingsen (2011) document 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. Smith (1987) suggests that trade credit may be a consequence of the ex-ante unobservable quality variation in the good supplied as it allows the buyer to inspect the good first. Another example is Daripa and Nilsen (2011) who point out that trade credit mitigates the negative externality on the manufacturer from the retailer’s trade-off between loss of sales and inventory costs. Cuñat (2007) shows how trade credit allows suppliers to act as insurers against liquidity shocks.

\(^5\)In Section 5 we explore the role of exclusivity, by assuming that there is a competitive fringe of downstream firms.

\(^6\)High credit quality suppliers have a comparative advantage in securing outside finance that they can pass on small, credit-constrained buyers (Boissay and Gropp (2007)).

\(^7\)The asymmetric information problem considered is 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.
the supplier ensures that it is worthwhile for the downstream firm to repay the credit rather than face retaliation. It is widely accepted that the "extra enforceability power of suppliers (as compared to banks) comes from the fact that they can threaten to stop supplying intermediate goods to their customers" (Cuñat and Garcia-Appendini (2012), page 545). McMillan and Woodruff (1999a and 1999b) find that such retaliation is used by firms in Vietnam. Fafchamps (1997) highlights the "importance of regular purchases as a determinant of supplier credit and the use of stopped deliveries when payment is not forthcoming" in Zimbabwean manufacturing (page 812). More generally, Fafchamps (2004) finds that 52% of a sample of Sub-Saharan African manufacturers suspend trade following (late or non) payment disputes. Informed by this evidence, we restrict attention to the refusal to transact for some time as the form of retaliation.\footnote{The model can be reinterpreted as the manufacturer choosing the probability of a permanent termination rather than the length of a temporary trade suspension.} \footnote{The T-period trade suspension contract that we consider is similar to how credit reference agencies operate: they "simplify the information about each agent i with a credit report showing when the agent last ‘cheated’ (e.g., paid late or not at all). This information is kept on the agent’s record for a set number of years T, after which time it is erased." Fafchamps (2010), page 57.} \footnote{A retailer with market power charges a mark-up (on top of the manufacturer’s) because he does not internalise the negative effect that a smaller quantity has on the manufacturer. The resulting quantity is then smaller than the monopoly quantity. This is ruled out in our setting by letting the manufacturer set the final quantity.} \footnote{The quantity is delivered before the uncertainty is resolved.}

The second problem that the manufacturer faces is the need to make the contract self-enforceable. Since it is difficult to enforce repayments legally, the retailer can potentially leave the relationship at any time, that is, "take the money and run". The manufacturer needs to ensure that the relationship is valuable enough to the retailer so that she does not walk away from it.

The optimal contract has important market outcome implications. First, the manufacturer, inefficiently, always sells too little. This result is not related to well-known sources of quantity distortion occurring between vertically-related firms such as double marginalization\footnote{A retailer with market power charges a mark-up (on top of the manufacturer’s) because he does not internalise the negative effect that a smaller quantity has on the manufacturer. The resulting quantity is then smaller than the monopoly quantity. This is ruled out in our setting by letting the manufacturer set the final quantity.} or screening\footnote{The quantity is delivered before the uncertainty is resolved.}. We identify a novel explanation for quantity rationing: to ensure the retailer’s good behavior, the manufacturer needs to leave him a surplus and this surplus increases with the quantity supplied. It is as if the manufacturer were
facing an additional marginal cost for making the contract enforceable and truth-telling. Why does the surplus left to the retailer need to increase with the quantity? What prevents the manufacturer from choosing the efficient quantity and using the extra generated profits to credibly punish and reward the retailer? Because it is costly to produce the good, the manufacturer would like to accompany a larger quantity with larger repayments. However, increasing the repayments decreases the surplus left to the retailer, who will then run away with the money. In other words, the manufacturer would like to sell more, but cannot do it without either getting a smaller share of the surplus (i.e., keeping the repayment constant) or having the retailer run away with the money (i.e., increasing the repayment). As a result, the manufacturer faces a trade-off between maximizing and sharing the surplus and the resulting quantity is too small. This result has implications for the structural empirical industrial organization literature. It shows that having enough instruments to remove vertical externalities from market power may not be enough to assume that efficient quantities are sold in the market when trade credit and limited liability play an important role.

The reasons for sharing the surplus are twofold. First, suppose the shocks are observable. The manufacturer needs to leave the retailer a surplus for the repayment to be honored. Second, suppose the shocks are not observable but, because the future is very valuable, the retailer does not want to walk away from the contract. To tackle the asymmetric information problem, the manufacturer needs to impose a long-enough trade suspension in the low revenue states and leave enough surplus in the high revenue states so that the true state is reported. A tougher trade suspension policy is accompanied by larger repayments and a smaller quantity distortion because it can be used, instead of giving away the surplus, to provide incentives to report the truth.

The model makes it possible to identify subtle effects between the need to make the contract enforceable and truth-telling. When the supplier addresses both problems at the same time, she faces a trade-off. Increasing the repayment in low revenue states or toughening the trade suspension following small repayments decreases the retailer’s incentives to underreport revenues, but at the same time it decreases the value of the re-
relationship and makes it less worthwhile to continue. The self-enforceability problem limits how tough the manufacturer can be by removing the possibility to implement a larger repayment through a longer trade suspension. In particular, the manufacturer would like to increase the repayment (and hence the quantity supplied) through a longer trade suspension, but cannot do so without having the retailer walk away from the relationship.

The repeated game approach formalizes an economic concept of trust (or quality of contract enforcement) and the discount factor is a good proxy for the trust level in the relationship. When the level of trust increases, the retailer more values staying in the relationship. The manufacturer can then afford to increase the repayment and the quantity via a longer trade suspension. As a result, both the quantity and the toughness of the termination policy increase with the level of trust. However, welfare increases with the quantity supplied and decreases with the length of the termination. We find that the welfare is non-monotonic in the level of trust. The need to make the contract self-enforced can be good for welfare because it prevents the manufacturer from inefficiently increasing her share of the surplus at the expense of decreasing the total surplus (by imposing an excessively tough trade suspension from the welfare point of view). This result suggests that when the level of trust reaches an intermediate level, policy efforts are best spent in relaxing firms’ credit and liquidity constraints as this will soften the asymmetric information problem.

Finally, the optimal contract resembles a debt contract. The manufacturer asks for a fixed repayment that guarantees the continuation of the contract. Otherwise, the manufacturer asks for the highest feasible repayment, forgives the rest of the debt and suspends trade for a number of periods. To provide incentives to repay as much as possible, larger repayments are associated with shorter trade suspension. By extracting all the retailer’s surplus in the default states, the manufacturer can soften the trade suspension policy and reduce inefficiency.

McMillan and Woodruff (1999b) find evidence of debt contracts being used for their sample of Vietnamese firms: “One (manager) (case #12) sent employees to visit a cus-

\footnote{See, for instance, Kvaløy and Olsen (2009).}
tomer every day to ask for a late payment. ‘After a few weeks of negotiation, the firm got back part of the debt and stopped selling to this customer.’ Another (case #10) after some negotiation accepted 70% of the amount owed, and another (case #8), owed money by a firm in Taiwan, after protracted negotiations was paid a year late. What is important, another said (case #4), ‘is to forget about the debt and keep social relations with the customer’.” (page 642). The authors find it surprising that “firms often try to keep the relationship going despite defection” (page 642) whereby the supplier tries to get restitution which sometimes involves forgiving part of the debt. As a result, they conclude that retaliation is not “as forceful as in the standard repeated-game story” (page 637). We improve on the existing literature by showing that debt contracts are able to rationalize such a puzzling behavior.

The paper proceeds as follows. Section 2 introduces the model. Section 3 explores a simple example with two states of demand. Section 4 solves the model with a continuum of demand states. The role of downstream competition is studied in Section 5. For comparability purposes, the literature review is deferred until Section 6. Finally, Section 7 concludes. The proofs are relegated to the Appendix.

2 The Setup

A manufacturer and a retailer have the opportunity to trade at dates $t = 0, 1, 2, \ldots$. In each period, the manufacturer produces $q$ units of a good at a marginal cost $c > 0$ and needs a retailer to market the product to the final consumer. The retailer sells the good at no cost and obtains some stochastic and continuous revenues $R(q; s)$ where $s$ denotes the size of the shock. We assume that $s$ is an iid random variable distributed on the interval $[s, \overline{s}]$ with $h(s)$ and $H(s)$ as the density and cumulative continuum distribution functions, respectively. Furthermore, we assume that the revenues increase in $q$ and $s$, and denote the expected revenues by $R_E(q) = \int_{s}^{\overline{s}} R(q; s)h(s)ds$.

\footnote{Boissay and Gropp (2007) find indirect evidence that some relationships continue despite defaults. In particular, they find a repeated default on the same supplier occurs about 15% of the time when the default is due to lack of liquidity.}
In terms of the information structure, the quantity of the good is publicly observable. The revenues, however, are the private information of the retailer. For instance, there may be uncertainty with respect to the willingness to pay of final consumers, or the demand may be certain but either the goods or the revenues are stolen now and then (for instance, by an organized crime group). Similarly, the uncertainty could refer to how many units of a perishable or non-perishable but non-returnable good are demanded every period in the market.\textsuperscript{14} If selling the good would be costly for the retailer, the shock could be on the retailer’s costs as well. Note that the quantity is delivered before the revenues are known and hence, it is not used to screen retailers with different revenue levels.

The retailer is completely credit and liquidity constrained. The manufacturer sells the good on trade credit, that is, the retailer pays the manufacturer back after obtaining the revenues and within the same period (no interest rate is charged). The manufacturer offers a quantity forcing contract which consists in a quantity, \( q \), and a repayment, \( D(\tilde{s}) \), for each shock reported \( \tilde{s} \) by the retailer. We consider this contract to remove quantity distortions coming from the double marginalization problem.\textsuperscript{15} Since the retailer is cashless, there are not many instruments available to the manufacturer to punish a default. Informed by the trade credit literature, we restrict attention to stationary \( T \)-period trade suspension contracts. These contracts establish a length of trade suspension, \( T(\tilde{s}) \), for each particular shock reported. We look for the contract that maximizes the manufacturer’s profits within this class.\textsuperscript{16} Choosing the length of trade suspension is equivalent to choosing a probability of a permanent termination.\textsuperscript{17} An alternative interpretation to suspending trade is to keep trading but in less profitable terms (for instance by diminishing the quality of the good).\textsuperscript{18} No repayment can be legally enforced, including \( D(\tilde{s}) \).

\textsuperscript{14}A manufacturer cannot repossess a good that has been transformed.
\textsuperscript{15}Since the quantity is delivered before the uncertainty is resolved, using a two-part tariff or a more complex nonlinear scheme is equivalent to offering a quantity forcing contract.
\textsuperscript{16}We discuss in footnote 25 the implications of giving the bargaining power to the retailer instead.
\textsuperscript{17}Denote by \( t(\tilde{s}) \) the probability of terminating forever with the retailer if \( \tilde{s} \) is reported. Then \( \delta^T(\tilde{s}) = 1 - t(\tilde{s}) \).
\textsuperscript{18}See Baker, Gibbons and Murphy (1994) for an example in the employer-employee relationship framework.
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.\(^{19}\) To keep the analysis simple, we assume that the retailer is not able to save (i.e. any profits are consumed within the same period) and therefore we can restrict attention to Perfect Public Equilibria.

The timing is summarized in Figure 1. In each period, the manufacturer offers a contract to the retailer. The retailer rejects or accepts and if he accepts, he places the order in the market. Then an iid shock is realized (and observed only by the retailer) which determines the size of the revenues. Finally, the retailer decides how much to repay and the contract is terminated for a number of periods if it is specified in the contract.\(^{20}\)

### 3 Example

We first consider an example where the revenues take only two values. The goal of this example is to illustrate the generality of our first result: the quantity supplied by the manufacturer is always rationed. In Section 4, we solve the model with a continuum of values.

Consider the case where the retailer receives revenues $R(q; 1) = R(q)$ from selling the good. However with probability $p$, there is a shock and the revenues are $R(q; s) = sR(q)$

\(^{19}\)We relax this assumption in Section 5.

\(^{20}\)Termination involves inefficiencies just as in the subjective performance evaluation of Levin (2003) or in Fuchs (2007). Committing to the trade suspension policy improves the manufacturer’s profits. One simple way to make the contract renegotiation-proof is to give the bargaining power to the retailer. In this case, the manufacturer receives no surplus and is indifferent between terminating or continuing to trade. See footnote 25 for more details.
instead, where $0 \leq s < 1$ and $E(s) = 1 - p + ps$. The contract consist in the 4-tuple: \{q, D(1), D(s), T(s)\}. To simplify the notation, denote $\overline{D} = D(1)$, $D(s) = D$ and $T(s) = T$. In order to give incentives to repay $\overline{D}$, trade is suspended for $T$ periods following $D$ and forever following a smaller payment.

The quantity that maximizes the joint surplus, $E(s)R(q) - cq$, when the revenues are observable and contracts can be enforced is defined by:

$$q^{FB} : R'(q) = \frac{c}{E(s)} = \tilde{c}$$

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 make comparisons. The parties produce less than in the absence of the shock because with some probability they do not receive the entire revenues but incur the production costs anyway.

To explore the implications of asymmetric information and contract enforcement, let $\pi_R$ denote the retailer’s present discounted value from date $t$ onwards when respecting the contract:

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

The previous equation says that with probability $1 - p$ there is no shock so the retailer can pay back $\overline{D}$ and the contract is renewed in the next period. However, with probability $p$, a shock reduces the revenues. The retailer can only pay back $D$ and the game moves to the trade suspension 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.

---

21 We use this multiplicative functional form for the revenues because of its simplicity, but the results are robust enough to use a general function $R(q; s)$. This functional form is used, for instance, in Green and Porter (1984).

22 When $s > 0$, repaying something smaller than $D$ or $\overline{D}$ implies stealing, in which case it is optimal to impose the maximum punishment because it relaxes the dynamic enforcement constraint and it is not imposed in equilibrium. When $s = 0$, then $D = 0$ and the manufacturer punishes for $T$ periods following no repayment.
In a similar way, let \( \pi_M \) be the present discounted value for the manufacturer:

\[
\pi_M = (1 - p) \left[ \overline{D} - cq + \delta \pi_M \right] + p \left[ \underline{D} - cq + \delta^{T+1} \pi_M \right]
\]

With probability \( 1 - p \), the retailer repays \( \overline{D} \) so the relationship moves on to the next period and with the complementary probability, the retailer only repays \( \underline{D} \) and the manufacturer suspends the contract for \( T \) periods.

Since the shocks are not observed by the manufacturer, she needs to ensure that the retailer has incentives to repay \( \overline{D} \) when possible. Because the retailer is cashless, only a high revenue retailer can pretend to have obtained low revenues:

\[
R(q) - \overline{D} + \delta \pi_R \geq R(q) - \underline{D} + \delta^{T+1} \pi_R
\]  \hfill (IC)

The incentive compatibility condition, \( IC \), reflects the following trade-off: if the retailer does not pay back \( \overline{D} \), he keeps \( \overline{D} - \underline{D} \) but this will automatically trigger the trade suspension phase, which yields valuation \( \pi_R \) only after \( T \) periods. For the retailer to pay back \( \overline{D} \), the manufacturer needs to ensure that tomorrow’s gains from not being terminated are larger than the difference in payments today: \( \delta (\pi_R - \delta^T \pi_R) \geq \overline{D} - \underline{D} \). The longer the trade suspension, the easier it is for the manufacturer to satisfy this constraint.

Since contracts cannot be legally enforced, the dynamic enforceability constraints ensure that a low (respectively, high) revenue retailer does not want to walk away from the contract with the revenues:

\[
sR(q) - \underline{D} + \delta^{T+1} \pi_R \geq sR(q)
\]  \hfill (DE)

and

\[
R(q) - \overline{D} + \delta \pi_R \geq R(q)
\]  \hfill (DE)

respectively. The future expected value of being within the trade suspension (trade renewal) phase, rather than being terminated forever, needs to be larger than the relevant
Note that if \( DE \) is satisfied, a low revenue retailer does not want to walk away from the relationship. If \( IC \) holds, a high revenue retailer does not want to pretend to be a low demand one, therefore, he does not want to walk away either so we can ignore \( DE \).\(^{23}\) Note as well that \( DE \) remains relevant even if there was no asymmetric information (for instance, if \( s \) were to be 1) as this constraint addresses the lack of ability to enforce repayment rather than the ability to observe outcomes and actions.

In addition to the \( IC \) and \( DE \) constraints, a cashless retailer cannot be forced to make a repayment exceeding the revenues he reports in the current period: \( \overline{D} \leq R(q) \) and \( D \leq sR(q) \). We call them limited liability constraints: \( \overline{LL} \) and \( LL \) respectively.

To sum up, the manufacturer’s problem is:

\[
\max_{q,D,D,T} \pi_M = \frac{(1-p) \overline{D} + pD - cq}{1-\delta(1-p)-p\delta^{T+1}}
\]

s.t. \((IC), (DE), (\overline{LL}) \) and \((LL)\)

where \( \pi_R = \frac{E(s)R(q) - (1-p)\overline{D} + pD}{1-\delta(1-p)-p\delta^{T+1}} \).

By inspection of \( \pi_M \), the manufacturer always wants to increase \( \overline{D} \) and \( D \) as much as possible, therefore at least two constraints need to bind to bound \( \overline{D} \) and \( D \) from above. We show in the Appendix that the \( IC \) is always binding and that the \( \overline{LL} \) is never binding.

The asymmetric information problem is always present and, since the manufacturer needs to give rents in at least one state to make the relationship valuable, she prefers to do so in the high state in order to relax the informational problem. Therefore, we are left with three possible combinations of binding constraints: \( IC - DE \), \( IC - DE - LL \) and \( IC - LL \).

The optimal contract is summarized in the next Proposition.

**Proposition 1** The optimal contract is as follows:

1. If the revenues are similar \((s \) large), the probability of the shock is large \((p \) large)
or the future is not very valuable (\( \delta \) small), the manufacturer offers a flat contract with a unique repayment, \( D^F = D^F = \delta E(s)R(q^F) \), and a permanent termination for a non-payment.

(2) If the size of the shock is large (\( s \) small), the probability of the shock is small (\( p \) small) or the future is very valuable (\( \delta \) large), the retailer repays all the revenues in the event of a shock. The manufacturer offers either:

a) a separating contract: \( D^S = \delta E(s)R(q^S) > D^S \) and \( \delta^{TS+1} = \frac{s}{E(s)} \), where \( T^S \in (0, +\infty) \).

b) a separating or flat contract: \( D^{S/F} = \alpha(T^{S/F})\delta E(s)R(q^{S/F}) \geq D^{S/F} \), where \( \alpha(T^{S/F}) \in [s/E(s), 1) \) and \( T^{S/F} \in [0, T^S) \).

In any case, the quantity is always distorted downwards despite the use of a quantity forcing contract.\(^{24}\)

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

One of the messages of the paper is that the manufacturer always sells too little. By rewriting \( \pi_M \) in terms of \( \pi_R: \pi_M = \frac{E(s)R(q) - cq}{1 - \delta(1-p) - \rho^{T+1}} - \pi_R \), we see that the manufacturer maximizes the total surplus minus how much she shares with the retailer. She cannot appropriate all the surplus and the ultimate amount of surplus left to the retailer depends on which of the constraints are binding. Because it is costly to produce, larger quantities will be associated with larger repayments. However, larger repayments imply less surplus left to the retailer, which triggers the retailer misbehavior. Therefore, if the manufacturer want to sell a larger quantity and ensure that the repayment is met, she cannot increase the repayments, which means that she is sharing more surplus with the retailer. It is as if the manufacturer had an extra marginal cost due to the asymmetric information and/or the enforcement problem and, hence, quantity distortions emerge.\(^{25}\) It is worth highlighting that it is the retailer’s impatience which is creating the downward distortion

---

\(^{24}\)The full version of this Proposition in the Appendix defines the optimal quantities.

\(^{25}\)When the retailer has the bargaining power, he maximizes: \( \pi_R = \frac{E(s)R(q) - cq}{1 - \delta(1-p) - \rho^{T+1}} - \pi_M \) subject to \( IC, DE, LL, LL \) and the participation constraint of the manufacturer, \( \pi_M = 0 \). We conjecture the quantity distortion to be significantly smaller. The reason is twofold: first, the retailer does not need to share the surplus with the manufacturer (i.e. the second part of the objective function is zero), and second, since the retailer keeps all the profits, we expect \( DE \) to be binding for a smaller set of parameters.
in the quantity.\textsuperscript{26} In contrast, both discounts factors would affect the choice of the termination policy.

The flat or pooling contract occurs when \( \delta E(s) \leq s \)\textsuperscript{27} and involves the IC and DE binding. When the low demand is nonetheless large or the shock is likely, inducing separation is too costly in terms of wasteful trade suspension and the manufacturer asks for the same repayment to avoid it. This is also the case when \( \delta \) is small and the dynamic enforceability problem is very tight as a result. This problem is then relaxed by both decreasing \( D_F \) (below the limit established by LL) and getting rid of the retaliation policy. The repayment is bounded by the amount the retailer can "take and run with" the next time he has the chance to do it (i.e. in the next period). The implication of asking for the same repayment is that the asymmetric information problem disappears. This case is equivalent to the one where the shocks are observable and LL does not bind.\textsuperscript{28}

The quantity distortion is exclusively the result of the rents that the manufacturer needs to give to the retailer to address the dynamic enforcement problem.

When \( s < \delta E(s) \), the limited liability constraint, LL, always binds. This means that the manufacturer recovers all the revenues in the event of a shock. Then we can have two possible cases depending on whether DE binds (case a) or not (case b).

In case a, the manufacturer asks for two repayments. As with the flat contract, the high repayment is bounded by the amount obtained from walking away in the next period. The difference in repayments, \( D^S < \overline{D}^S \), requires the manufacturer to terminate the contract for a positive number of periods after the low repayment. The termination length is chosen so that the retailer is indifferent between stealing today (keeping \( qS \)) or the next time he has the chance to do it, that is, after \( T^S \) periods (\( \delta^{T^S+1} E(s)R(qS) \)).

The trade suspension is temporary unless the low revenues are zero (\( s = 0 \)), where the trade suspension is permanent.

\textsuperscript{26}If the discount factor for the retailer were different from the one for the manufacturer, then it would be retailer’s discount factor appearing in conditions (6), (8) and (7) in the Appendix.

\textsuperscript{27}The condition is implied by a slack LL once the optimal \( \overline{D} \) is introduced.

\textsuperscript{28}To see this, imagine that the shocks are observable. The manufacturer asks for \( \overline{D} \) if demand is high, \( \overline{D} \) if demand is low and terminates forever if the appropriate payment is not made. Therefore, the manufacturer needs to replace (IC) by (DE) and transform (DE) in the following way: \( \delta R \geq \overline{D} \). These two constraints imply: \( \overline{D} = \overline{D} = \delta E(s)R(q) \), and hence we obtain the same solution.
The quantity distortion is the result of addressing the asymmetric information problem and giving incentives to the retailer to keep trading. This example illustrates the tension between both problems. From the IC, the manufacturer would like to increase $D^S$ and $T^S$ as much as possible\(^{29}\) to give incentives to the high demand retailer to tell the truth. At the same time, from the DE, a larger $D^S$ and $T^S$ decreases the value of the relationship for a low demand retailer and hence increases his incentives to walk away. When the event of a shock brings no revenues to the retailer, this tension is relaxed. The retailer can repay nothing and steal nothing. As a result, the manufacturer is free to implement the toughest trade suspension policy, that is, a permanent termination.

When the value of the future, $\delta$, is very large, the dynamic enforceability problem becomes irrelevant, and this brings us to case b. The quantity distortion that emerges in this setup is only due to the asymmetric information problem. Indeed, the manufacturer has to leave the surplus in the high revenue state so that this state is reported. There is a trade-off between a tougher punishment and a smaller quantity. With a tougher punishment, the manufacturer can ask for a larger $D^{S/F}$ and offer a larger $q^{S/F}$.\(^{30}\) However, increasing the punishment comes at a cost: it increases the inefficiency following a shock as she terminates profitable trading with the unlucky retailer. This cost is very important when the future is very valuable and the trade suspension is softer than in case a, that is, $T^{S/F} < T^S$.

When $T^{S/F} = 0$, the manufacturer offers a flat contract with a unique repayment equal to $sR(q^{S/F})$. This occurs when the demand elasticity at the optimal quantity, $\varepsilon$;\(^{31}\)

\(^{29}\)An increase in $D$ decreases the RHS of IC. It also decreases the LHS indirectly through a decrease in $\pi_R$ but to a less extent so, overall, the first effect dominates. Note that $\frac{1-\delta^T}{1-\delta(1-p) - p\delta^{T+1}}$ is increasing in $T$.

\(^{30}\)The weight $\alpha(T^{S/F})$, defined in the Appendix, is increasing in the trade suspension length:

$$\frac{\partial \alpha(T^{S/F})}{\partial T^{S/F}} = \frac{-\delta^{T^{S/F}} (\ln \delta)(1-p)(1-s)(1-\delta)}{E(s) \left( 1 - \delta^{T^{S/F+1}} \right)^2} > 0$$

\(^{31}\)We break down the revenue function as follows: $R(q) = P(q)q$ where $P(q)$ is the inverse demand function. The demand elasticity, $\varepsilon$, is defined as: $\frac{1}{\varepsilon} = -\frac{P'(q)q}{P(q)}$.
is inelastic enough:

\[ \varepsilon \leq \frac{p}{1 - p} \frac{s}{E(s) - s} \]  

(1)

where \( \varepsilon \) is the elasticity of demand at the optimal quantity. Intuitively, when the demand is more inelastic, it is relatively less costly to distort the quantity than when the demand is more elastic. As a result, the cost associated with termination is larger than the benefit of increasing the repayment in the high demand state, and hence the quantity. Otherwise, two different repayments are offered as in case a.\(^{32}\)

**Corollary 1** When the manufacturer only addresses the asymmetric information problem, a more inelastic demand is associated with a flat contract being offered more often and a smaller quantity being supplied.

Finally, to illustrate how the interaction between the different parameters determines the regimes, we use the linear demand. Figure 2 depicts which regime yields the largest profits for the manufacturer in the space \( \delta \) and \( p \) when the shock cuts by half the revenues \( (s = 0.5) \).\(^{33}\) When the value of the future is small, no contract can be implemented and hence no trade occurs.\(^{34}\) As the value of the future increases, the manufacturer offers a flat contract when the likelihood of the shock is high (to avoid termination with an unlucky retailer) and a separating contract when the likelihood of the shock is low. When the value of the future is even larger, the self-enforceability problem becomes irrelevant and the manufacturer can either offer a separating or a flat contract.

\(^{32}\)Note that \( \varepsilon < 1 \), because of the downward distortion in the quantity. Using the elasticity where a monopolist prices, \( \varepsilon = 1 \), we provide a sufficient (but not necessary) condition for a flat contract to be offered: \( \frac{(1-p)^2}{p + (1-p)} \leq s. \)

\(^{33}\)The line that separates the "Flat contract" from the "Separating contract" area is \( s = \delta \frac{(1-p)}{1 - \delta p} \) and the line that separates the "Separating contract" from the "Flat/Separating contract" area is \( s = \delta^T \frac{(1-p)}{1 - p \delta^T + \delta} \) where \( T \) is the optimal termination policy at the border between the regimes, \( T = \log_\delta \left( \frac{1}{2} \left( \frac{1}{2 \delta} \right) \right) - 1. \)

\(^{34}\)Although this is out of the scope of this model, the no-trade situation could also correspond to a manufacturer that vertically integrates downwards.
In this Section we solve the model, presented in Section 2, where the retailer faces a continuum of demand states. We assume that, in the worse state of the world $s = \underline{s}$, the retailer receives epsilon revenues or no revenues at all. We do this to rule out the case where the limited liability constraint does not matter (case 1 in Section 3) as our goal is to explore the effect that liquidity constraints have on the optimal contract.\footnote{We conjecture that it is possible to find a distribution $h(s)$ such that if $R(q; \underline{s})$ is large enough, the limited liability constraint does not bind and the manufacturer asks for a unique repayment equal to $D = \delta R_E(q)$ where $q$ is the optimal quantity.}

When $R(q; \underline{s}) > 0$, the absence of repayment triggers the maximum punishment, that is permanent termination, as it relaxes the dynamic enforceability constraint and it is not imposed in equilibrium. The present discounted value for the retailer, $\pi_R$, and the manufacturer, $\pi_M$, when the retailer respects the contract are:

$$\pi_R = R_E(q) - \int_\underline{s}^\bar{s} D(s) h(s) ds + \int_\underline{s}^\bar{s} \delta^{T(s)+1} \pi_R h(s) ds$$

$$\pi_M = \int_\underline{s}^\bar{s} D(s) h(s) ds + cq + \int_\underline{s}^\bar{s} \delta^{T(s)+1} \pi_M h(s) ds$$

We first find the conditions under which the contract elicits the true demand and is...
self-enforceable and then we proceed to characterize the optimal contract. We conclude
the Section with an example.

4.1 Incentive compatibility and dynamic enforceability

We explore the retailer’s incentives to report the revenues associated with \( \tilde{s} \), when the
current level of revenues is determined by \( s \). By the time the retailer learns about \( s \), the
quantity has already been supplied and hence does not depend on \( s \) nor its report \( \tilde{s} \). For
a given \( s \) and \( q \), the retailer chooses a report \( \tilde{s} \) to maximize his profits:

\[
\pi_R(\tilde{s}; s) = \underbrace{R(q; s) - D(\tilde{s})}_{\text{Today’s payoff of reporting } \tilde{s}} + \underbrace{\delta^{T(\tilde{s})+1} \pi_R}_{\text{Expected payoff of reporting } \tilde{s}}
\]

where \( \pi_R \) is the expected present discounted profits from staying in the relationship.\(^{36}\) If
the retailer were not credit and liquidity constrained, the choice of \( \tilde{s} \) would not depend on
the true \( s \) as it is equally costly for any type of retailer to report any \( \tilde{s} \). Formally, there
is no sorting condition in this model, \( \frac{\partial^2 \pi_R(\tilde{s}; s)}{\partial \tilde{s} \partial s} = 0 \), as in the standard adverse selection
model in Levin (2003). Because the retailer is credit and liquidity constrained, however,
he cannot repay more than his current revenues \( R(q; s) \):

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

The limited liability condition (\( LL_s \)) indirectly links the choice of the report with the
true state of demand.

Let us denote by \( u(\tilde{s}) \) the part of the retailer’s payoff that does not directly depend
on \( s \):

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

The independence between the incentives to report a demand state and the actual demand
state makes the task of inducing truth-telling simple. Intuitively, if it were feasible, the

\(^{36}\)Formally, it is \( \pi_R = \int_a^\infty \pi_R(\tilde{s}; s) h(s) ds \).
retailer would always report the $\tilde{s}$ with the highest $u(\tilde{s})$ regardless of $s$. Hence, to induce truth-telling, the retailer needs to be indifferent to reporting any $s$. Truth-telling is achieved when $u(\tilde{s})$ is constant across reports: $u'(\tilde{s}) \mid_{\tilde{s}=s} = 0 \ \forall \tilde{s} \leq s$.

Furthermore, the contract needs to be self-enforceable. The retailer can always guarantee himself the current revenues $R(q; s)$ by stopping the contract after the sale. In order to prevent the retailer from walking away with these revenues, the repayment needs to be weakly smaller than the continuation value: $u(\tilde{s}) \mid_{\tilde{s}=s} \geq 0 \ \forall s$.

**Lemma 2** The dynamic enforcement is guaranteed if:

$$
\left(D(\tilde{s}) \leq \delta^{T(\tilde{s}) + 1} \pi_R \right) \mid_{\tilde{s}=s} \forall \ s \tag{DE_s}
$$

Truth-telling is achieved when $\pi_R(\tilde{s}; s)$ is constant in the report $\tilde{s}$:

$$
\frac{\partial \pi_R(\tilde{s}; s)}{\partial \tilde{s}} \mid_{\tilde{s}=s} = 0 \ \forall \tilde{s} \leq s \tag{IC_s}
$$

$IC_s$ binds for all $s$.

**Proof.** See Appendix. ■

For a given $q$, the manufacturer has two instruments to penalize low reported revenues: she can either increase the payment today $D(\tilde{s})$ or increase the length of trade suspension $T(\tilde{s})$, which decreases the retailer’s continuation value. Given that increasing $T(\tilde{s})$ is also costly for the manufacturer (because she loses future trade), whenever it is possible, the manufacturer asks for the largest repayment: $D(\tilde{s}) = R(\tilde{s}, q)$.

In equilibrium, however, the manufacturer cannot implement a repayment equal to the revenues in all the states. The manufacturer shares the surplus with the retailer through smaller repayments and such a repayment strategy would render the relationship worthless to the retailer (i.e., $\pi_R = 0$) which violates $DE_s$. There must be a report $s^*$ above which the manufacturer asks for a fixed repayment, $R(q; s^*)$, and, consequently, does not terminate, $T(s^*) = 0$. 

19
The resulting contract resembles what 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 by asking a fixed payment.

**Proposition 2** The optimal contract is a debt contract.

**Proof.** See Appendix. ■

In particular, for a given \( q \) and \( s^* \), the contract is defined as follows:

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

and

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

where \( T(\tilde{s}) \) is to be defined\(^{37} \) and \( \tilde{s} = s \) when all \( IC_s \) hold. Note that \( LL_s \) bind for \( s \leq s^* \). A debt contract is optimal in this model because it minimizes the inefficiency associated with trade suspension (by trading-off larger repayments for shorter termination periods in the default states), while inducing the retailer to report the truth.

Let us use Proposition 2 to simplify \( \pi_R \). The retailer repays all the revenues if the shock is smaller than \( s^* \) and has the contract terminated for \( T(s) \) periods. Otherwise, he repays the constant amount \( R(q; s^*) \) and keeps trading in the next period.

\[
\pi_R = \int_{2}^{s^*} \left[ R(q; s) - R(q; s) + \delta^{T(s)+1} \pi_R \right] h(s)ds + \int_{s^*}^{\pi} \left[ R(q; s) - R(q; s^*) + \delta \pi_R \right] h(s)ds
\]

By Lemma 2, telling the truth when \( s < s^* \) results in \( u = -R(q; s) + \delta^{T(s)+1} \pi_R \) and when \( s^* \leq s \) in \( u = -R(q; s^*) + \delta \pi_R \). Realizing that \( u \) is constant give us:

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

Consider a first best world where the retailer is not credit and liquidity constrained. Then an optimal contract could be a quantity (chosen so that the total surplus is maximized) and a fixed up-front payment from the retailer (i.e. no trade credit is granted).

\(^{37} \)In equation \((IC)\) below.
If the manufacturer has all the bargaining power, then 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. We can interpret the above equation in these terms, where the retailer keeps the expected profits minus a fixed payment, \( R(q; s^*) \), and never has his contract terminated. The difference to the first best world comes 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 worthwhile to stay in the relationship. In particular, if the dynamic enforceability constraint binds, the retailer obtains a payoff equal to one-period expected revenues in one period. Otherwise, \( u > 0 \), and the retailer can obtain more than \( R_E(q) \). Solving for \( \pi_R \) yields:

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

(2)

Hence, for \( \pi_R \) to be positive, it needs to be the case that the revenues at \( s^* \) are smaller than the expected revenues. Notice as well that \( \pi_R \) increases in the quantity supplied.

Similarly, \( \pi_M \) can be simplified to:

\[
\pi_M = \int_{s^*}^{s_2} \left[ R(q; s) + \delta^{T(s)} \pi_M \right] h(s)ds + (1 - H(s^*)) [R(q; s^*) + \delta \pi_M] - cq
\]

that is, the manufacturer receives all the revenues in the bad states \((s \in [s_2, s^*])\) and terminates the contract with the retailer for \( T(s) \) periods. Otherwise, the manufacturer charges a fixed repayment, \( R(q; s^*) \), that guarantees the continuation of the contract. Finally, the manufacturer incurs the production costs. Rewriting \( \pi_M \) in terms of \( \pi_R \), illustrates the fact that the manufacturer cannot appropriate all the surplus:

\[
\pi_M = \frac{R_E(q) - cq}{1 - \delta (1 - H(s^*)) - \int_0^{s^*} \delta^{T(s)+1} h(s)ds} - \pi_R
\]

As in the Example of Section 3, since \( \pi_R \) increases with \( q \) by (2), it is as if the manufacturer had an extra marginal cost associated with the asymmetric information and enforceability problems. As a result, quantity rationing occurs.
Equipped with Proposition 2 and equation (2), we proceed to simplify Lemma 2. Since $IC_s$ holds for all $s$, we just need to ensure that $DE_s$ is satisfied for one state. For convenience, we choose $s = s^*$ because we already know that $T(s^*) = 0$. Using (2), $DE_{s^*}$ becomes:

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

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

Finally, we use the fact that $u$ is constant together with (2), to find the relationship between $T(s)$ and the remaining parts of the contract: $s^*$ and $q$.

$$\delta^{T(s)+1} = \frac{R(q; s) + u}{(R_E(q) - R(q; s^*)) / (1 - \delta)}$$

The $IC$ ensures that the contract elicits truth-telling for a given $s^*$ and $q$. Two observations are in order. First, note that $T(s)$ decreases in the current revenues as more revenues imply a larger repayment. Second, the manufacturer terminates the contract forever if the demand is such that no repayment is feasible in the worse demand state (i.e., $R(q; \bar{s}) = 0$) and $DE$ binds (i.e., $u = 0$). Otherwise, trade suspension is temporary.\(^{38}\) If $DE$ binds, $IC$ becomes $\delta^{T(s)+1} R_E(q) = R(q; s)$. The manufacturer would like to suspend trade for longer periods but is constrained by the contract self-enforceability problem: the retailer needs to be left indifferent between stealing the current revenues today ($R(q; \bar{s})$) or next time he has the chance to do it if he is to repay ($R(q; s)$), that is, after $T(s) + 1$ periods. When $R(q; \bar{s}) = 0$, this tension is soften because there are no current revenues to run away with. The manufacturer can then be as tough as she wishes and termination is permanent. As a result, the all-or-nothing shock structure considered in Li and Matouschek (forthcoming)’s paper (with limited liability) is not very convenient for our framework. When the value of the future increases, the dynamic enforceability problem

\(^{38}\) When $R(q; \bar{s}) > 0$, the permanent termination is only reserved for the absence of repayment, since a repayment equal to $R(q; \bar{s})$ is always possible.
becomes irrelevant. However, because the future is very valuable for the manufacturer as well, permanent termination becomes too costly and a temporary trade suspension is implemented if the worse state is reported.

**Corollary 3** The manufacturer adopts a policy of contingent contract renewal whereby a larger repayment is associated with a shorter trade suspension. The absence of repayment triggers permanent trade suspension if: (1) a minimum repayment is always feasible, or (2) a minimum repayment is not feasible and the contract self-enforceability problem is present.

### 4.2 Optimal contract

The optimal contract consists in the bundle $q$ and $s^*$, such that $\pi_M$ is maximized subject to $DE$ and $IC$.\(^{39}\) When choosing $s^*$, the manufacturer faces the following trade-off: a larger $s^*$ involves a larger expected repayment from the retailer today; however, this comes at the cost of longer termination\(^{40}\) to keep the truth-telling incentives unchanged and a tighter self-enforceability problem. An increase in the quantity leads to an increase in the total payment as well as an increase in the production costs. Since it also leads to an increase in $\pi_R$, the effect on the termination policy is less clear and depends on the particular form of the demand function.\(^{41}\)

Since the asymmetric information problem is always present, it is useful to plug $IC$ in $\pi_M$ to obtain:

$$
\pi_M = \frac{\int_{s}^{s^*} R(q; s)h(s)ds + (1 - H(s^*)) R(q; s^*) - cq}{(1 - \delta) \int_{s}^{s^*} \frac{(R(q; s) - R(q; s^*))h(s)ds}{R_E(q) - R(q; s^*)}}
$$

\(^{39}\)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.

\(^{40}\)The length of trade suspension increases in $s^*$:

$$
\frac{\partial T(s)}{\partial s^*} = -\frac{1 - \delta}{\delta} \frac{\partial R(q; s^*)}{\partial s^*} \frac{R_E(q) - R(q; s)}{(R_E(q) - R(q; s^*))^2} < 0 \forall s \leq s^*
$$

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

\(^{41}\)In the multiplicative example that we present below, $T(s)$ does not depend on $q$.  


As the future becomes more and more valuable ($\delta \to 1$), $DE$ no longer binds. Note that when this happens, the optimal $s^*$ and $q$ do not depend on $\delta$. The optimal termination length, $T(s)$, does depend on $\delta$ by $IC$. The limited liability is preventing the manufacturer from increasing all the repayments following an increase in the value of the future. Provided that the retailer does not want to walk away, a more valuable future does not help the manufacturer address the asymmetric information problem.

In contrast, when $DE$ binds, $s^*$ and $q$ do depend on $\delta$. The fixed repayment established by the debt contract is bounded by what the retailer expects to keep tomorrow if he does not repay: $R(q; s^*) = \delta R_E(q)$. A larger discount factor allows the manufacturer to implement a larger fixed payment. As we will see in the next Section, the presence of the dynamic enforceability problem prevents the manufacturer from choosing an excessively large $s^*$ (from the welfare point of view).

The following Proposition summarizes this discussion.

**Proposition 3** The optimal $s^*$ and $q$ are independent of the discount factor, $\delta$, if and only if $\delta$ is large enough: $\delta > \frac{R(q; s^*)}{R_E(q)}$, where $R(q; s^*)$ is the fixed repayment and $R_E(q)$ the expected revenues.\(^\text{42}\)

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

Since the choice of $q$ and $s^*$ depends on the particular revenue function, in what follows, we illustrate these results with the multiplicative example.

### 4.3 Example

For simplicity and comparability with the Example in Section 3, we use the multiplicative revenue function, $R(q; s) = sR(q)$, with $s = 0$. As in the Example, the first best quantity that maximizes the one-period joint profits is determined by: $R'(q) = \frac{c}{E(s)} = \tilde{c}$ where $E(s) = \int_0^\pi sh(s)ds$. If on average, the shocks are detrimental, there is a reduction in the quantity.

\(^\text{42}\)The full version of this Proposition in the Appendix contains the first-order conditions from which $s^*$ and $q$ are found both for this and lower $\delta$. 

24
With this revenue function, the retailer’s profits in (2) become: \( \pi_R = \frac{(E(s) - s^*)R(q)}{1-\delta} \), which means that for \( \pi_R > 0 \), \( s^* \) needs to be smaller than the expected shock. The problem of the manufacturer becomes:

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

s.t. \( \frac{\delta E(s) - s^*}{1-\delta} \geq 0 \) \quad (DE)

where \( \hat{E}(s, s^*) = H(s^*) E(s \mid s \leq s^*) + (1 - H(s^*)) s^* \) is the expected shock that matters to the manufacturer in terms of the repayment established by the debt contract. Since for any \( s^* < \pi, \hat{E}(s, s^*) < E(s) \), the manufacturer cannot appropriate all the surplus and there is an under-supply of \( q \).

The \( DE \) does not bind if \( s^* \) is smaller than the discounted value of the expected shock, which happens for a large \( \delta \). When the dynamic enforcement constraint binds, \( s^* \) is bounded by it so the manufacturer only chooses \( q \). The manufacturer would like to punish more but cannot because that would decrease the value of the relationship making the retailer walk away. In particular, the manufacturer sells the quantity that maximizes her profits taking into account the repayment that she is expected to obtain given \( s^* \). Conversely, when \( DE \) does not bind, the manufacturer also chooses \( s^* \). A larger \( s^* \), on one hand, increases the expected repayment\(^{43} \) and hence \( \pi_M \). On the other, it also decreases \( \pi_R \) and the retailer’s incentives to truthfully report the demand state. Hence, a tougher termination policy is needed\(^ {44} \), which increases the inefficiency from the trade suspension.

**Corollary 4** If \( R(q; s) = sR(q) \), the optimal \( s^* \) and \( q \) are determined by these first order conditions when \( DE \) does not bind:

\[
q : R'(q) = \frac{c}{E(s, s^*)} > \bar{c}
\]

\(^{43}\) Note that \( \frac{\partial \hat{E}(s, s^*)}{\partial s^*} = 1 - H(s^*) > 0 \).

\(^{44}\) This is reflected in the denominator of \( \pi_M \), which is increasing in \( s^* \) because \( \frac{\partial \hat{E}(s, s^*)}{\partial s^*} < 1 \).
Figure 3: Optimal \( q \) and \( s^* \) when \( R(q) = (10 - q) q \), \( c = 2 \) and \( s \sim U(0, 1) \)

\[
s^* : \frac{R(q)}{cq} = \frac{H(s^*)}{1 - H(s^*)} \frac{E(s) - E(s \mid s \leq s^*)}{E(s \mid s \geq s^*) - s^* - E(s) (E(s) - s^*)}
\]

(4)

Otherwise, \( s^* = \delta E(s) \) and \( q \) is determined by (3). In both cases, the quantity is distorted downwards. The optimal \( s^* \) and \( q \) are strategic complements. The length of the trade suspension is defined by \( \delta^{T(s)+1} = \frac{(1-\delta)E(s) - s^*}{E(s) - s^*} \).

Despite using a quantity forcing contract, the quantity sold by the manufacturer is always rationed.\(^{45}\) The largest quantity is offered when the dynamic enforceability constraint does not bind because \( s^* \) is then not bounded by it. Since \( \hat{E}(s, s^*) \) increases in \( s^* \), the downward distortion in the quantity decreases with \( s^* \), regardless of whether the constraint binds or not. Therefore, \( q \) and \( s^* \) are strategic complements, that is, a tougher trade suspension\(^{46}\) policy is accompanied by a smaller quantity distortion. This is because a tougher punishment can be used, instead of giving away surplus, to provide incentives to report the truth.

\(^{45}\) The manufacturer sells the first best quantity only if \( s^* = \pi \), which violates \( DE \). With a binding constraint, an efficient quantity will require \( \delta > 1 \), which is not possible either.

\(^{46}\) Trade suspension occurs for a larger set of states and more periods:

\[
\frac{\partial \delta^{T(s)+1}}{\partial s^*} = -\frac{(1 - \delta) (E(s) - s)}{(E(s) - s^*)^2} < 0
\]
Figure 3 depicts the optimal $s^*$ and $q$ as a function of $\delta$ for $R(q) = (10 - q)q$, $c = 2$ and $s \sim U(0, 1)$. Trade occurs if the value of the future or level of trust is large enough ($0.45 \leq \delta$). Provided that trade occurs, if the level of trust is intermediate, $\delta < \frac{s^*}{E(s)} = 0.82$, then the $DE$ binds. Both $s^*$ and $q$ increases with $\delta$ because more trust relaxes the self-enforceability problem. This enables the manufacturer to ask for a larger repayment (larger $s^*$) and offer a larger quantity in exchange. For a larger level of trust, the $DE$ does not bind. Since (3) and (4) do not contain $\delta$, the choice of $q$ and $s^*$ are not affected by $\delta$ as in Proposition 3. Therefore, the quantity is distorted downwards even for $\delta$ arbitrarily close to 1.

The implications of this result are important for policy makers. It suggests that when the level of trust is intermediate, policy efforts at increasing the level of trust or making the contracts more enforceable (that is, increasing the discount factor) will be ineffective at increasing the volume of trade. This is because the limited liability bundled with the asymmetric information problem makes it very costly for the manufacturer to implement a larger quantity. Without the limited liability constraint, the manufacturer could increase all the repayments in the same proportion, appropriate more surplus and offer more quantity, leaving the asymmetric information problem unchanged. However, the presence of limited liability prevents the manufacturer from increasing lower repayments (for states below $s^*$) as the debt contract is asking for the maximum payment already. Increasing the repayment for large states only (above $s^*$) needs to be accompanied with a tougher termination so that the retailer still reports the truth. But a tougher punishment becomes too costly for the manufacturer when the future is valuable and she decides not to increase $s^*$ and $q$.

**Corollary 5** When the level of trust is intermediate, policy efforts should be targeted at relaxing liquidity and credit constraints if the quantity in the market is to be increased.

Finally, we look at the surplus generated in the market and how it is shared between both firms. Figure 4 depicts the joint per-period average profits as well as the per-period
average profits of the retailer and the manufacturer.\footnote{The join per-period average profits are: $(1 - \delta)(\pi_R + \pi_M) = \frac{1}{2} p(q)q - cq \frac{1}{1-s^2}$. The retailer’s and manufacturer’s per-period average profits are: $(1 - \delta)\pi_R = \left(\frac{1}{2} - s^*\right) p(q)q$ and $(1 - \delta)\pi_M = s^*\frac{(1 - \frac{1}{2} s^*) p(q)q - cq}{1 - s^2}$, respectively.}

For the manufacturer, an increase in the (inefficiently small) quantity unambiguously increases her profits. Instead, an increase in $s^*$ increases the repayment but also the inefficiency for terminating with an unlucky retailer who is not able to repay $s^* R(q)q$. Since the manufacturer has all the bargaining power, as $\delta$ increases, she keeps increasing $s^*$ and $q$ as long as her profits increase (which happens up to $\delta = 0.82$, which is when $DE$ stops binding). After that, she keeps her choices of $s^*$ and $q$ fixed.

For the retailer, an increase in $s^*$ means keeping a smaller share of the surplus, so his profits decrease as a result. However, a larger $s^*$ also means a larger $q$ and hence larger expected revenues and profits. Since both $s^*$ and $q$ increase with $\delta$, the retailer’s per period profits are non-monotonic in the level of trust and are maximized when $\delta = 0.65$. Before $\delta = 0.65$, the positive effect of a larger $q$ dominates the negative one of a larger $s^*$.

Overall, the total surplus is non-monotonic in the level of trust. An increase in $s^*$ decreases the total surplus because of the inefficiency created by the trade suspension, while
an increase in $q$ increases the total surplus because the quantity is under-supplied. As a result, the total surplus is inversely U-shaped in the discount factor and it is maximized when $\delta = 0.69$.

**Corollary 6** The expected welfare decreases with $s^*$ and increases with $q$. Since they both increase with the level of trust, the expected welfare is non-monotonic in it. When the level of trust is intermediate, the self-enforceability problem is good for welfare because it limits how inefficiently tough the manufacturer can be with the retailer.

For intermediate level of trust below $\delta = 0.69$, further increases in $\delta$ make the manufacturer increase $q$ and terminate for longer periods. In this region, the increase in efficiency from a larger volume of trade compensates the inefficiency from a longer trade suspension. After $\delta = 0.69$, the opposite is true. The manufacturer, nonetheless, keeps increasing both $q$ and $s^*$ because this allows her to keep a larger share of the albeit smaller surplus.

This result reinforces our previous message. When there is a minimum level of contract enforceability or trust, improving it, not only is unsuccessful at increasing the quantity but is welfare detrimental. The lack of legal enforceability is protecting the retailer (and the welfare) from the inefficient policy of the retailer.

## 5 Downstream competition

This Section explores the role of exclusivity using the multiplicative example from Section 4.3. The goal is to determine if the contract is substantially affected by making another retail chain available to the manufacturer. In particular, we let the manufacturer sell through an inefficient competitive fringe of sellers while she is suspending trade with the retailer. We model it by allowing the manufacturer to obtain a non-negative outside option, $\pi_O$, every period that she is in the trade suspension phase. This possibility affects the manufacturer’s profits and introduces a participation constraint, $\pi_M \geq \frac{\pi_O}{1-\delta}$, that the
contract needs to satisfy and that can be simplified, after some algebra, to:

\[
\hat{E}(s, s^*) R(q) - cq \geq \pi_O \tag{PC_M}
\]

We show that, provided that the one-period expected profits of selling to the retailer are larger than \(\pi_O\), the contract is qualitatively unaffected.

**Proposition 4** If \(PC_M\) binds, the manufacturer does not sell to the retailer and takes her outside option instead.

If \(PC_M\) and \(DE\) do not bind, the optimal \(q\) is determined by (3), \(s^*\) by:

\[
s^* : \frac{R(q)}{cq + \pi_O} = \frac{H(s^*)}{1 - H(s^*)} \hat{E}(s, s^*) \frac{E(s) - E(s \mid s \leq s^*)}{(E(s \mid s \geq s^*) - s^*) - E(s)(E(s) - s^*)}
\]

and a larger \(\pi_O\) is associated with a larger \(s^*\) and hence a larger \(q\). When \(DE\) binds, \(s^* = \delta E(s)\), \(q\) is determined by (3) and these choices are unaffected by \(\pi_O\).

**Proof.** See Appendix. ■

Introducing an outside option makes the manufacturer more reluctant to sell to the retailer when the value of the future is low. However, provided that the relationship with the retailer is valuable enough, the contract offered is almost unaffected. We illustrate these findings in Figures 5 and 6 where we reproduce the Figures of Section 4.3 in grey and plot the new solution for \(\pi_O = 0.8\) in black.

Suspending trade with the retailer is less onerous for the manufacturer. As a result, she chooses a larger fixed repayment (larger \(s^*\)) and suspends trade for more states of the demand. By appropriating more surplus from the retailer, the manufacturer is able to offer a larger (but albeit distorted) quantity. This applies only when the future is sufficiently valuable in that the retailer does not want to walk away from the relationship \((\delta > 0.82)\). Instead, when the contract self-enforceability is an issue, the contract offered is identical to the one offered in the absence of an outside option. The manufacturer would like to choose a larger \(s^*\), but cannot because that would decrease the value of the relationship making the retailer walk away. Note as well that the range of parameters
where the contract self-enforceability matters is larger. In Figure 6 we plot the per period profits from the relationship (that is, excluding the outside option that the manufacturer obtains when dealing with a third party). When the contract self-enforceability problem matters (0.72 < δ < 0.82), the per period profits are the same as in the absence of an outside option. Otherwise, the per period profits are smaller. Despite the larger quantity supplied, the manufacturer suspends trade more often, so less surplus is created within the relationship.

In the future, it would be interesting to explore in more detail the role of retail competition by looking at how the optimal contract and trade suspension policy change when the manufacturer can sell through two competing retailers simultaneously.

### 6 Related literature

This paper is related to the literature on inter-firm relational contracts with asymmetric information starting by Levin (2003). Trade credit bundled with limited liability imposes important restrictions on the contracts that this literature considers\footnote{See Malcomson (2012) for a recent survey.}. Li and Matouschek (forthcoming) is the closest related paper. They characterize the optimal
Figure 6: Per period profits when $R(q) = (10 - q)q$, $c = 2$, $s \sim U(0, 1)$ and $\pi_O$

dynamic contract in a setup where a manager rewards a worker for exerting an observable (but not contractible) effort with an ex-ante fixed wage and an ex-post positive discretionary bonus. With some probability, it is inefficient to reward the worker, but whether this has occurred is unobservable to the worker. When the manager is able to borrow enough, the optimal contract does not involve termination, as the manager can be punished for failing to pay the bonus with larger future payments and lower effort today. If the manager faces borrowing constraints, the characterization of the contract becomes very complex and the authors assume that in the event of a shock, the manager cannot pay anything to the worker. The optimal contract may involve permanent termination as a means to provide incentives when the borrowing constraints are very tight and the expected profits of the manager are small.

The first important difference concerns the timing of the shock. The retailer’s revenues are stochastic, while in their setup, the output is deterministic and the shock occurs at the payment stage.\footnote{The Public Perfect Equilibrium Pareto frontier becomes non-differentiable.} Another important difference is the manager’s ability to commit\footnote{This is similar to setting the size of the shock, $s$, to zero in our Example in Section 3.} \footnote{To illustrate the importance of this difference, take their optimal contract when borrowing is not feasible. The contract involves a zero wage, no bonus in the event of a shock and a bonus equal to the output otherwise. With this contract, their manager keeps all the output in the event of a shock; however, the retailer in our setup obtains zero revenues. Such a contract yields no profits to the retailer.}
to a fixed wage, which in our setting is equivalent to a minimum repayment. The fixed wage plays two important roles. First, it allows the parties to choose a contract that maximizes the joint surplus as the wage can be used to share it in an arbitrary way. Instead, we give the bargaining power to the manufacturer and show how she inefficiently increases her share of the surplus at the expense of decreasing the total surplus. Second, the wage is used as a punishment following the non-payment of the bonus when the effort is already low because it has been consecutively used as a punishment in the past. Finally, they study the dynamics of the relational contract. Instead, and informed by the trade credit literature, we restrict attention to stationary T-period trade suspension contracts. By doing this, we are able to enrich the model in other dimensions. Notably, in our framework, making a similar all-or-nothing assumption concerning the distribution of the shocks waters down the tension between the asymmetric information and the self-enforceability problems. A retailer hit by a shock does not want to walk away from the contract (with empty pockets). We consider instead a continuum of shocks and a general distribution. Also, we are able to provide a more explicit solution and relate it to important characteristics of the environment such as the demand elasticity (Corollary 1) and the level of trust (Corollaries 5 and 6).

This paper is also related to the literature where an investor finances an entrepreneur who can divert the investment cashflows. Incentives to repay are given by liquidating the entrepreneur’s assets (Hart and Moore (1998)), threatening to withhold the second (and last) period investment (Bolton and Scharfstein (1990)) or carrying on an audit as in the costly state verification models (Townsend (1979) and Gale and Hellwig (1985)). As in this literature, we find that the optimal contract is a debt contract. Povel and Raith (2004) allow for the size of the investment to be secretly 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. As in our model, they also

and hence would trigger strategic default.

Innes (1990) finds that debt contracts are optimal in an environment with moral hazard with limited liability and observable output. 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.
find that the entrepreneur under-invests to decrease the fixed repayment and hence the likelihood of defaulting\textsuperscript{53}. However, in our model, because it is the uninformed party who has the bargaining power, the distortion is not only motivated to soften the (inefficient) termination policy but also to increase her share of the profits. As a result, the quantity distortion is even larger.\textsuperscript{54} Another fundamental difference is the use of a relational contract in the analysis, which has been shown to play a crucial role in trade credit provision. Formally, we endogenize the future value that accrues to the retailer if he does not default, as it corresponds to the potential profits generated within the relationship and highlight how it constrains the optimal quantity and trade suspension choices when the trust level is low.

Finally, DeMarzo and Fishman (2007a, 2007b) and the literature thereafter, considers the use of a contract termination threat in a multi-period model with a fixed one-time investment and diversion costs. They find the optimal long-term contract can be implemented by a combination of equity, long-term debt and a line of credit. Important differences with this literature are the absence of long-term contracts and the possibility of choosing the quantity and the termination length by the uninformed party.

7 Conclusions

The goal of this paper has not been to explain why, where and how much trade credit is offered by a supplier. Taking this decision as given, we explore how trade credit affects the different contract characteristics. The main prediction of this analysis is that 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.\textsuperscript{55} The supplier shares

\textsuperscript{53}In 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).

\textsuperscript{54}See footnote 25.

\textsuperscript{55}Note that this result is not in contradiction with the view of trade credit as a way to foster sales as without trade credit no sale could have been undertaken.
the surplus with the downstream firm to give him incentives to reveal the true demand and/or stay within the contract.

Another important result is that the quantity distortion remains even as the value of the future becomes arbitrarily large. Furthermore, for large values of the future, the welfare decreases as the manufacturer can impose inefficiently tough retaliation policies without worrying about the contract’s self-enforceability. This suggests that policy interventions that aim at making the contracts more enforceable when the level of trust is intermediate will not be successful in restoring efficiency in the market if downstream firms are credit constrained and their ability to repay is unobservable to the supplier. Instead, policy efforts are better expended in relaxing liquidity and credit constraints.

Finally, we find that the optimal contract resembles a debt contract. Debt contracts are successful in keeping the retaliation policy to the minimum while still providing the downstream firm incentives to repay the appropriate amount.

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 another. It would be interesting to explore the form of non-stationary contracts and determine in which way the upstream firm would increase or decrease the quantity in each period as a function of the previous period repayment as well as how this choice would interact with the trade suspension policy. We nonetheless expect quantity distortions to remain as the manufacturer still needs to share the total surplus with the retailer.

8 Appendix

**Proposition 5 (1)** The optimal contract is as follows:

1. If \( \delta E(s) \leq s \), the optimal contract is: \( D^F = D^F = \delta E(s)R(q^F) \), \( T^F = 0 \) and \( q^F \) is defined by:

\[
R'(q^F) = \frac{c}{\delta E(s)} > \tilde{c} \tag{6}
\]
(2) If $s < \delta E(s)$ and:

a) $DE$ binds, the optimal contract is: $D^S = sR(q^S)$, $\bar{D}^S = \delta E(s)R(q^S)$, $\delta^{T^S+1} = \frac{s}{E(s)}$ and $q^S$ is defined by:

$$R'(q) = \frac{c}{(1-p)\delta E(s) + ps} > \bar{c}$$

(7)

b) $DE$ does not bind, the optimal contract is: $D^{S/F} = sR(q^{S/F})$, $\bar{D}^{S/F} = \alpha(T^{S/F})\delta E(s)R(q^{S/F})$ where $\alpha(T^{S/F}) = \frac{1-\delta s}{1-\delta^{s+1}}$, $q^{S/F}$ is defined by:

$$R'(q^{S/F}) = \frac{c}{(1-p)\alpha(T^{S/F})\delta E(s) + ps} > \bar{c}$$

(8)

and $T^{S/F}$ is defined in the Proof.

Proof of Proposition 1. We first show that no other regime exists. Suppose that $LL$ binds, then $D = R(q)$ and either $DE$ or $LL$ need to bind to bound $D$. $LL$ binding implies $\pi_R = 0$, which triggers strategic default by the retailer. $DE$ binding makes the low repayment equal $\bar{D} = \frac{p(T+1)sR(q)}{1-\delta(1-p)}$. With these two repayments, $IC$ is satisfied only if $s > \frac{1-\delta(1-p)}{p^3}$, which leads to a contradiction as $\frac{1-\delta(1-p)}{p^3} > 1$. Suppose that only $DE$ and $LL$ bind. An increase in the punishment period not only decreases $M$ but also makes $DE$ more binding. Therefore, the manufacturer chooses $T = 0$ and as a result $IC$ is violated. Hence, $IC$ is always binding. Let us prove each of the remaining cases where we drop the superscripts.

- **Case 1 (Flat contract):** The binding constraints, $IC$ and $DE$, bound the payments by $\bar{D} = \delta E(s)R(q)$ and $\bar{D} = \delta^{T+1}E(s)R(q)$. Using these payments the objective function becomes: $\pi_M = \frac{(1-p)\delta^{T+1}E(s)R(q)^{-c}}{1-\delta(1-p)-ps\delta^{T+1}}$. Noticing that $T$ decreases the objective function completes the proof.

- **Case 2.a (Separating contract):** When the three constraints bind, $LL$ pins down $\bar{D} = sR(q)$ and the $IC$ and $DE$ jointly determine $\bar{D}$ and $T$: $\bar{D} = \delta E(s)R(q)$ and $\delta^{T+1}E(s) = s$. $T$ is positive because the parameter specification is such that $\delta < \frac{s}{E(s)}$. As a result, the objective function is $\pi_M = \frac{(1-p)\delta E(s)+ps)R(q)^{-c}}{1-\delta(1-p)-\frac{ps}{\delta^{T+1}}}$. 

- **Case 2.b (Flat/Separating contract):** When $IC$ and $LL$ bind, $LL$ pins down:
\[ D = sR(q) \text{ and IC: } \overline{D} = \alpha(T)\delta E(s)R(q), \text{ where } \alpha(T) = \frac{1 - \delta^T + (1-\delta)s}{1 - \delta^{T+1}}. \] The function \( \alpha(T) \) is increasing:

\[
\frac{\partial \alpha(T)}{\partial T} = -\delta^T (\ln \delta) (1-p) (1-s) (1-\delta) E(s) \left(1 - \delta^{T+1}\right)^2 > 0
\]

and is defined for \( T \in \left[0, \log_\delta \left(\frac{s}{E(s)}\right) - 1\right] \). We show that the optimal punishment is bounded by the optimal punishment in case 2.a. We plug \( D = sR(q) \) and \( \overline{D} = \alpha(T)\delta E(s)R(q) \) into \( DE \) to obtain:

\[
s < \delta^{T+1} \left[ \frac{E(s) - (1-p) \frac{(1-\delta^T)\delta E(s) + (1-\delta)s}{1 - \delta^{T+1}}}{1 - \delta (1-p)} \right]
\]

If we plug this inequality in the optimal punishment in case 2.a, \( \delta^{T+1}E(s) = s \), it is easy to show that the equation is violated.

The objective function in this case is \( \pi_M = \frac{(1-p)\alpha(T)\delta E(s) + ps)R(q) - cq}{1 - \delta(1-p) - p\delta^{T+1}} \). Simple algebra shows that the RHS of condition (8) is larger than \( \tilde{c} \) when \( 1 - \delta (1-p) - p\delta^{T+1} > 0 \), which is always true, and that it is decreasing in \( T \):

\[
\frac{\partial \text{RHS of (8)}}{\partial T} = \frac{c \ln \delta (1-p)^2 (1-\delta) (1-s) \delta^{T+1}}{(\delta (1-p)^2 + (1-\delta (1-p)^2) s - ((1-p) E(s) + ps) \delta^{T+1})^2} < 0
\]

The first order condition for \( T \) is:

\[
\frac{\partial \pi_M}{\partial T} = \frac{-\delta^{T+1} \ln \delta}{1 - \delta (1-p) - p\delta^{T+1}} \left[ (1-p) \frac{1 - \delta \alpha(T)}{1 - \delta^{T+1}} E(s)R(q) - p\pi_M \right] = 0
\]

It is easy to check that this condition, evaluated at \( T = 0 \), is negative if:

\[
\frac{cq}{R(q)} \leq \frac{s - (1-p) E(s)}{p}
\]

which gives us condition (1) when replacing \( c \) using (8).

Finally, we need to show that, even when the discount factor tends to 1, the quantity remains distorted downwards. When \( \delta \to 1 \), the constraint \( s < \delta E(s) \) is more likely to
be satisfied, so either case 2.a or 2.b occur. However, as \( \delta \to 1 \), no punishment in case 2.a could satisfy \( \delta^{T+1} = \frac{s}{E(s)} \) so only case 2.b emerges (indeed, when \( \delta \to 1 \), \( DE \) is slack).

Using Hopital’s rule, equation (8) becomes:

\[
\lim_{\delta \to 1} R'(q) = \frac{c}{s+[1-(1-p)E(s)+p_e]T} \tag{9}
\]

Note that at \( T = 0 \), the RHS of (9) is larger than \( \tilde{c} \) and that it is decreasing in \( T \). Noting that as \( \delta \to +1 \), the RHS of (9) tends to \( \frac{c}{(1-p)E(s)+p_e} \) which completes the proof. ■

**Proof of Lemma 2.** Because of condition \( (LL_s) \), a retailer can lie only downwards. The following needs to be true: \( \pi(s,s) \geq \pi(\tilde{s},s) \forall s \geq \tilde{s} \). To show that this condition holds with equality, suppose it does not. Then, substracting \( R(q;s) \) in both sides, we obtain: \(-D(s) + \delta^{T(s)+1}\pi_R > -D(\tilde{s}) + \delta^{T(\tilde{s})+1}\pi_R \), which means that state \( \tilde{s} \) will not be reported when \( s = \tilde{s} \), which leads to a contradiction. ■

**Proof of Proposition 2.** Suppose that Lemma 2 holds, the retailer’s and manufacturer’s payoffs are: \( \pi_R = \left( R_E(q) - \int_{\frac{\pi}{2}}^\pi D(s)h(s)ds \right) \left/ \left( 1 - \int_{\frac{\pi}{2}}^\pi \delta^{T(s)+1}h(s)ds \right) \right. \) and \( \pi_M = \left( \int_{\frac{\pi}{2}}^\pi D(s)h(s)ds - cq \right) \left/ \left( 1 - \int_{\frac{\pi}{2}}^\pi \delta^{T(s)+1}h(s)ds \right) \right. \), respectively. For a given \( s \), \( \frac{\partial \pi_R}{\partial D(s)} < 0 \) and \( \frac{\partial \pi_M}{\partial D(s)} < 0 \). It is then possible to find a \( \rho > 0 \) and \( \gamma > 0 \) so that Lemma 2 still holds for \( \hat{D}(s) = D(s) + \rho \) and \( \hat{T}(s) = T(s) - \gamma \). Since \( \frac{\partial \pi_M}{\partial D(s)} > 0 \) and \( \frac{\partial \pi_R}{\partial D(s)} < 0 \), the new contract \{ \( \hat{D}(s), \hat{T}(s) \) \} is strictly prefered by the manufacturer. We have established that the manufacturer wants increase \( D(s) \) as much as it is feasible, that is, \( D(s) = R(q;s) \).

However, if \( D(s) = R(q;s) \forall s \), then \( \pi_R = 0 \) and the dynamic enforcement in Lemma 2 is violated. Therefore, there is a state \( s^* < \pi \) such that \( D(s) = R(q;s^*) \forall s \geq s^* \) and \( T(s) = 0 \). If the manufacturer were to terminate for some periods following the report \( s^* \), she would strictly prefer to increase the repayment to reduce the trade suspension length and its consequent inefficiency, leaving the retailer’s incentives unaffected. ■
Proposition 3. The optimal $s^*$ and $q$ are determined by

\[
(1 - H(s^*)) \pi_R - H(s^*) \pi_M \frac{E(s) - E(s \mid s \leq s^*)}{E(s) - s^*} = \hat{\lambda} \tag{10a}
\]

\[
\pi_R \frac{\partial R(q;s^*)}{\partial q} - c + \int_q^{s^*} \left( \frac{\partial R(q;s)}{\partial q} - \frac{\partial R(q;s^*)}{\partial q} \right) h(s) ds = (\pi_R + \pi_M) \tag{10b}
\]

where $\lambda$ is the shadow cost of the dynamic enforcement constraint and $\hat{\lambda}$ is adjusted by the second-best loss of revenues, $\hat{\lambda} = \lambda \left( R_E(q) - \int_q^{s^*} R(q; s) h(s) ds - (1 - H(s^*)) R(q; s^*) \right)$.

If and only if $\delta$ is large enough, $\delta > \frac{R(q; s^*)}{R_E(q)}$, then $\hat{\lambda} = 0$ and the optimal $s^*$ and $q$ are independent of $\delta$. ■

Proof of Proposition 3. The Lagrangian function is:

\[
L = \left( \frac{\int_q^{s^*} R(q; s) h(s) ds + (1 - H(s^*)) R(q; s^*) - c q}{1 - \delta} \frac{1}{1 + H(s^*) \frac{R_E(q) - \int_q^{s^*} R(q; s) h(s) ds}{R_E(q) - R(q; s^*)}} \right) - \lambda \left( R(q; s^*) - \delta R_E(q) \right)
\]

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

\[
\frac{\partial L}{\partial s^*} = \pi_R \frac{\partial R(q; s^*)}{\partial s^*} - \pi_M \frac{\partial R(q; s^*)}{\partial s^*} \left( \frac{1 - H(s^*)}{R_E(q) - \int_q^{s^*} R(q; s) h(s) ds - (1 - H(s^*)) R(q; s^*)} \right) \]

\[
\frac{\partial L}{\partial q} = \pi_R \frac{\int_q^{s^*} \frac{\partial R(q; s)}{\partial q} h(s) ds + (1 - H(s^*)) \frac{\partial R(q; s^*)}{\partial q} - c}{R_E(q) - \int_q^{s^*} R(q; s) h(s) ds - (1 - H(s^*)) R(q; s^*)} \]

\[
-\lambda \left( \frac{\partial R(q; s^*)}{\partial q} - \frac{\partial R(q; s^*)}{\partial q} \right) = 0
\]

\[
-\lambda \left( \frac{\partial R(q; s^*)}{\partial q} - \frac{\partial R(q; s^*)}{\partial q} \right) = 0
\]

39
\[
\lambda(R(q; s^*) - \delta R_E(q)) = 0
\]

Setting the first two conditions equal to zero, gives us equation (10a) and (10b), 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 \( s^* = \hat{s} \)) or to the retailer (if \( s^* = \bar{s} \)). We assume that the demand and the cost function are such that production takes place, i.e., \( 0 < q < +\infty \). \(\Box\)

**Proof of Corollary 4.** Suppose \( DE \) does not bind, the first order conditions are:

\[
\frac{\partial \pi_M}{\partial q} = \frac{\hat{E}(s, s^*) R'(q) - c}{(1 - \delta) \frac{E(s) - E(s, s^*)}{E(s) - s^*}} = 0
\]

\[
\frac{\partial \pi_M}{\partial s^*} = (1 - H(s^*)) \frac{E(s) - s^*}{E(s) - \hat{E}(s, s^*)} R(q) - H(s^*) \frac{E(s) - E(s \leq s^*)}{(E(s) - \hat{E}(s, s^*))^2} \left( \hat{E}(s, s^*) R(q) - cq \right) = 0
\]

When \( DE \) binds, \( \frac{\partial \pi_M}{\partial q} \) remains unaffected and \( s^* \) is uniquely determined by \( DE \). The cross derivative is:

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

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

**Proof of Proposition 4.** The manufacturer maximizes:

\[
\pi_M = \frac{\hat{E}(s, s^*) R(q) + H(s^*) \frac{E(s^*) - E(s \leq s^*)}{E(s) - s^*} - cq}{(1 - \delta) \frac{E(s) - \hat{E}(s, s^*)}{E(s) - s^*}}
\]
subject to $DE$ and $PC_M$. Suppose none of the constraints bind, the first order conditions are (11) and:

$$
\frac{\partial \pi_M}{\partial s^*} = (1 - H(s^*)) \frac{E(s) - s^*}{E(s) - E(s, s^*)} R(q) - H(s^*) \frac{E(s) - E(s | s \leq s^*)}{(E(s) - \hat{E}(s, s^*))^2} \left[ \hat{E}(s, s^*)R(q) - cq - \pi_O \right] = 0
$$

and a larger $\pi_O$ has associated a larger $s^*$:

$$
\frac{\partial^2 \pi_M}{\partial s^* \partial \pi_O} = H(s^*) \frac{E(s) - E(s | s \leq s^*)}{(E(s) - \hat{E}(s, s^*))^2} > 0
$$

When $DE$ binds, (11) remains unaffected and $s^*$ is uniquely determined by $DE$. Because $\pi_M$ is larger when $DE$ does not bind, for a given $\pi_O$, $PC_M$ is more likely to bind when $DE$ also binds. Because there is downward quantity distortion, if the optimal quantity without considering $PC_M$ does not satisfy $PC_M$, no other quantity will and the manufacturer offers $q = 0$. ■

**References**


43


