Vertical Relations under Credit Constraints

Volker Nocke
University of Oxford, CEPR, and CESifo

John Thanassoulis
University of Oxford

March 2009

Abstract

We analyze the impact credit constraints have on how firms structure their dealings with their partners in the supply chain; and hence how imperfect capital markets can alter short run prices and long run investment levels. Credit constrained firms are shown to become endogenously risk averse and so seek to push some risk onto suppliers. Our predictions reflect risk sharing contracts in general, slotting fees (prevalent in groceries), and current evidence from trade credit. We demonstrate that the optimal contract has the effect of raising marginal costs and so credit constraints raise retail prices in the short run. We show that this is exacerbated by tight money; poor corporate governance; negative technology shocks and lack of diversification. Finally we show that credit constraints and market risk cause outsourcing to dominate in house provision.

1 Introduction

The current upheavals in financial markets have made credit scarce for a high proportion of firms. However credit constraints are not just a feature of today's economy. They have been found to be widespread for decades in the US, Europe and elsewhere (Hubbard 1998). We offer some new insight into the impact credit constraints have on how firms structure their dealings with their partners in the supply chain. Our hypothesis is that credit constraints affect vertical contract structure creating implications for retail prices in the short run, and for investment levels in the long run.

Consider a firm exposed to demand side risk which has some investment opportunities, and yet is credit constrained. To invest in the future, assets must be accrued which will serve as collateral. These assets are amassed by trading in the short run. Hence, as evidence corroborates, credit constrained firms' investment level is closely related to their cash flow (Gertler and Gilchrist 1994). If investment is subject to diminishing marginal returns then we note that this causes the firm to be endogenously risk averse during trading in the short run.[1] That is low demand realisations limit the collateral which the firm can use for investment and so result in very low investment levels.

We show that the firm, exposed to demand side risk, has an incentive to alter its contracts with its suppliers. It will seek some insurance from its suppliers. We show that this is achieved by requiring the supplier to make a payment independent of demand to the firm. In return the firm will make demand dependent payments back up to the supplier.

However, for the supplier to recoup her payment requires the per unit input price of the input to be in excess of cost. That is double marginalization is introduced. Hence the credit constraints cause contracts to be altered so that the retail price of the firm rises. The cost of the insurance made necessary by the credit constraints is in this sense partly paid for by the ultimate consumers.

The insurance service we model is, we argue, reflected in at least three common business practices. Firstly slotting fees which are common practice in the grocery market. These fees are payments many supermarkets require of their suppliers. Theoretical explanations for this practice have portrayed the slotting fee as a signalling device (Klein and Wright 2007). Empirical evidence suggests that an important part of the story is, however, the sharing of risk (Sudhir and Rao (forthcoming)) which accords with our hypothesis. Secondly risk sharing contracts are an apparently direct manifestation of our model in which the firm may receive explicit support for costs incurred which are repaid depending upon realised demand. Finally trade credit is a route by which insurance can be secured from a supplier – amongst other uses (Petersen and Rajan 1997). Petersen and Rajan offer evidence that the amount of trade credit used by a firm declines as the demand realisation improves - as our model predicts.

If a credit constrained firm exposed to market risk derives some insurance value from a supplier we show that it follows that such a firm has an incentive to outsource supply to a non credit constrained third party. This link between market risk and outsourcing is a further important theoretical contribution. The result is supported by empirical evidence (Harrigan (1985) and Sutcliffe and Zaheer (1988)); however the main theoretical arguments have worked in the opposite direction and cite incomplete contract problems

[1] A related point is made by Froot et al. (1993), though they do not explore the implications this insight has for vertical contracting.

(Mahoney 1992).

We show that there exist complementarities between the provision of insurance and lending. A supplier with the same monitoring technology as the banking sector could actually improve welfare if it conducted all lending. We also demonstrate that our results are in no way specific to demand side risk: the same intuitions apply with supply side risk. Further bargaining extensions we explore do not alter the insights our analysis generates.

The structure of the optimal contract can be approximated by a two part tariff as it consists of a payment from the supplier to the buyer which is only recouped if large volumes are subsequently sold. With this approach we can then explore the effect of changes in market structure on the level of input and hence retail prices in the short run; and on investment levels in the long run.

We investigate the implications of increasing interest rates for the investing firms; worsening corporate governance; shocks to the investment technology; alterations in the demand side risk; and implications of diversification. Tighter money and worsening corporate governance are both shown to lower long run expected investment levels; and to raise short run prices. Tighter money lowers the amount that can be borrowed and worsening corporate governance exacerbates the credit constraints; so both naturally lead to lower expected long run investment levels. But as marginal returns to investment are decreasing, the effect is to increase the cost of demand volatility to the firm in the short run. Hence the firm seeks to lock in more short run profit (secure more insurance) and this causes input and so retail prices to rise. In relation to tighter money this result is a new insight into the *price puzzle*: the macroeconomic link that has been noted between increases in the interest rate and increases in retail prices (Christiano et al. 1999).

The Model is offered in Section 2 and the fully optimal contract characterised and explored in Section 3. Section 4 then displays the complementarities between insurance provision and the supply of credit. Section 5 restricts to two party tariffs and conducts a comparative statics analysis as different features of the market environment are altered. Section 6 returns to fully optimal contracting and demonstrates that the results are unchanged if one considers supply side, rather than demand side, risks. Section 7 explores the implications of bargaining and Section 8 concludes. All omitted proofs are in the appendix.

2 The Model

We consider a model of a vertically related industry with two firms, a downstream firm D and an upstream firm U. D's marginal cost of production is normalized to zero, while the marginal cost of production of U is $c \geq 0$. There are two periods.

Period 0. In period 0, U supplies an input to D which D transforms into a final good on a one-to-one basis, and then sells on. When choosing output Q and facing market size z, D faces inverse demand $p(Q/z)$.[2] We assume that D is exposed to market risk in that market size z is a random variable with finite support $\{z_1, ..., z_n\}$.[3] A larger value of z implies that the volume supplied is a smaller proportion of the total market, and so a higher unit price results. We label states in increasing order so that $0 < z_1 < z_2 < \cdots < z_n$.The probability of state z_i is g_i, and $\zeta = \Sigma_{i=1}^{n} g_i z_i$ is the expected value of z.

Assumption 1 *We make the following standard assumptions on downstream demand:*

(i) *Marginal revenue $dQp\,(Q/z)\,/dQ$ is declining in quantity Q.*

(ii) *The reservation price at $Q = 0$ exceeds marginal cost: $p(0) > c$.*

Assumption 1 implies that, in any demand state z, industry profit $Q\,[p\,(Q/z) - c]$ is strictly concave in quantity Q for any Q such that $p(Q/z) > c$. Moreover, it implies that, in demand state z, industry profit is maximized at quantity $Q = zq\,(c)$, where $q(c)$ is the unique solution in q of $p\,(q) + qp'\,(q) = c$. That is, at the industry profit-maximizing quantity $zq\,(c)$, marginal revenue is equal to the industry's marginal cost of c. The industry profit-maximizing downstream price is $p\,(q\,(c))$ in every demand state z.

Before the demand state is realized, D and U agree to the contract $\{Q\,(z_i)\,, W\,(z_i)\}$, where $Q(z_i)$ is the input (and output) volume in state z_i, and $W(z_i)$ the associated transfer payment from D to U. We assume for now that D has all of the bargaining power. Then, D learns the realization of the demand state z and reports state $\hat{z}$ to U. D then receives $\hat{Q} = Q\,(\hat{z})$ units of input from U, transforms the input into a final good, and fetches a retail price of $p\left(\hat{Q}/z\right)$ per unit. Finally, D pays $W\,(\hat{z})$ to U. We assume for simplicity that D has no initial assets. D's asset level by the end of period 0, a, is therefore given by D's net profit in that period: $a = \hat{Q}p\left(\hat{Q}/z\right) - W(\hat{z})$.

Period 1. In period 1, D has to decide how much to invest in a project. Based on the moral hazard formulation offered by Holmstrom and Tirole (1997), we assume that

[2] D can equivalently be thought of as setting price p and facing demand $zQ(p)$.

[3] The finiteness assumption is for expositional reasons only.

D is endogenously credit constrained. Specifically, after choosing the investment level I, D's owner-manager can choose whether or not to shirk at the investment stage. If he does not shirk, D makes a gross profit of $\pi(I)$. If he does shirk, instead, the investment project fails (zero return) but the owner-manager receives a private benefit proportional to the size of the investment, $B \cdot I$, where $B \leq 1$.

If D wishes to invest more than its pledgeable assets, $I > a$, it can attempt to secure a loan of $I - a$ from an external banking sector. For now we set the market interest rate to zero so that D has to pay back only the amount of the loan, $I - a$. Any loan has to satisfy the no-shirking condition

$$BI \leq \pi(I) - (I - a) \tag{1}$$

otherwise, D's owner-manager would decide to shirk and D would be unable to pay back its loan.

Assumption 2 *We make the following assumptions on the gross profit function $\pi(\cdot)$:*

(i) *The marginal gross return of investment is positive but diminishing: $\pi(I)$ is strictly increasing and strictly concave in I.*

(ii) *We assume that $\pi(0) = 0$, $\pi'(0) > 1$, and $\pi'(I) < 1$ for I sufficiently large, so that the first-best level of investment, $\widehat{I} \equiv \arg\max_I \pi(I) - I$, is strictly positive.*

(iii) *In equilibrium, any realized value of a is smaller than the level necessary to finance the first-best investment level, $a < (B+1)\widehat{I} - \pi\left(\widehat{I}\right)$, so that the no-shirking constraint (1) is always binding in equilibrium.*

3 Equilibrium Analysis

We solve the model by backward induction. Suppose D's asset level at the beginning of period 1 is given by a. By Assumption 2(iii), D chooses an investment level $I(a)$ and an associated loan $I(a) - a$ so that the no-shirking constraint is just binding: while D would like to invest more, the banking sector would be unwilling to provide a larger loan. That is, $I(a)$ is the unique solution in I to

$$BI = \pi(I) - (I - a) \tag{2}$$

Note that at $I(a)$ the marginal gross return satisfies

$$1 < \pi'(I(a)) < 1 + B. \tag{3}$$

Since the no-shirking constraint is binding, D's net payoff at the end of period 2 is $\pi(I(a)) - (I(a) - a) = BI(a)$. The following lemma holds

Lemma 1 *D's payoff, $I(a)$ is increasing at a rate greater than unity and concave in assets.*

Proof. Implicitly differentiating $I(a)$ yields

$$\frac{dI}{da} = \frac{1}{1 + B - \pi'(I)} > 1 \quad \text{and} \quad \frac{d^2I}{da^2}\left[1 + B - \pi'(I)\right] = \pi''(I)\left[\frac{dI}{da}\right]^2$$

implying $\frac{d^2I}{da^2} < 0$ as required. ■

This is a very important preliminary result. It shows that risk neutral firm D is made endogenously risk averse in its objective function by virtue of the diminishing returns to investment. Credit constraints imply that the level of investment achievable is an increasing function of the pledgeable asset level a. Diminishing marginal returns to investment then imply that reductions in pledgeable assets relative to the expected level have a much bigger effect on the net present value of D than increases in pledgeable assets above the expected level. This is equivalent to stating that D is (endogenously) risk averse with respect to a.[4]

This endogenous risk aversion will affect the agreement D requires from its supplier. This will in turn affect the retail prices in period 0 (the short run) and the expected level of investment. Thus credit constraints via the supply chain relationships, will affect consumer surplus both in the short and long run. We now determine how.

3.1 The Optimal Contract under Symmetric Information

Before analyzing period-0 contracting under our assumption that D has private information about the realized demand state when choosing quantity (or price), it is instructive to consider first the case of symmetric information. Assuming the realized demand state is verifiable, the contract $\{Q(z_i), W(z_i)\}$ is a function of the realized demand state rather than the demand state reported by D. In this case there is no moral hazard for D at

[4] That credit constraints can imply risk aversion has been noted in Froot et al (1993). However this analysis is unique in exploring the effects of this on firm interactions and consumer surplus.

the quantity setting stage. There remains, however, a moral hazard problem for D at the investment stage. Hence D's problem becomes, from (2)

$$\max_{\{Q_i,W_i\}} \sum_{i=1}^{n} g_i B \cdot I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right)$$

subject to the individual rationality constraint for U,

$$\sum_{i=1}^{n} g_i \{W_i - Q_i c\} = 0. \tag{4}$$

Proposition 1 *When the demand state is verifiable, the equilibrium contract* $\{Q(z_i), W(z_i)\}$ *is such that industry profit is maximized in every demand state* z_i, $Q(z_i) = z_i q(c)$. *And* D *is fully insured always having pledgeable assets at the end of period 0 equal to* $\zeta q(c)[p(q(c)) - c]$.

Hence the optimal contract has U bearing all the risk, D delivers a quantity which always yields the industry profit maximizing price. Further, whatever the realization of risk, D completes period 0 with assets equal to the ex ante expected industry profit – U making good any shortfall and confiscating any excess.

So in the full information case consumers are unaffected by the market risk. The risk aversion created for D by the credit constraints is passed up to U and no inefficiency need be created.

Proof. The Lagrangian, which is to be maximized over $\{Q_i, W_i\}$, is given by

$$\mathcal{L} = \sum_{i=1}^{n} g_i \left\{B \cdot I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right) + \lambda [W_i - Q_i c]\right\}$$

This yields a set of first order conditions. With respect to W_i we have

$$\begin{aligned} -BI'\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right) + \lambda &= 0 \\ \Rightarrow I'\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right) &= \frac{\lambda}{B} \text{ (a constant)} \end{aligned} \tag{5}$$

As the investment function is strictly concave (Lemma 1), D has pledgeable income at the end of period 0 which is independent of the realized demand state – thus U is providing full insurance to D. $W_i = Q_i p(Q_i/z_i) + k$, where k is a constant (to be determined).

The first order condition with respect to Q_i yields

$$BI'\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right) \cdot \left[p\left(\frac{Q_i}{z_i}\right) + \frac{Q_i}{z_i} p'\left(\frac{Q_i}{z_i}\right)\right] - \lambda c = 0$$

Substituting in (5) we have

$$p\left(\frac{Q_i}{z_i}\right) + \frac{Q_i}{z_i} p'\left(\frac{Q_i}{z_i}\right) = c \quad \Rightarrow \quad Q_i = z_i q\left(c\right)$$

That is, the quantity is chosen to maximize industry profits in every state. Hence, the retail price is constant at the profit maximising level in every state. The transfer payment in state z_i is therefore $W_i = z_i q\left(c\right) p\left(q\left(c\right)\right) + k$. Inserting this term into U's individual rationality constraint, equation (4), we obtain that the constant k is given by $k = -\zeta q\left(c\right)\left[p\left(q\left(c\right)\right) - c\right]$ which yields the desired end of period 0 pledgeable asset level. ■

However, this outcome cannot hold when the demand state is not verifiable. This is because it would be in $D's$ interests to always claim the state of demand was weak and so extract extra payments from U - leaving the upstream firm with a payoff below her outside option. Thus full risk transfer is not in general possible.

3.2 The Optimal Contract Under Asymmetric Information

We now analyze period-0 contracting under our assumption that only D observes the (unverifiable) realized demand state. This creates moral hazard for D when she is setting her quantity as she could seek to deviate from reporting the true state of demand. That is D, once the market risk is revealed, will select a quantity which maximises her payoff given the agreed input tariff schedule.

If the state is z_i and D truthfully reports it then she would have a payoff of $I\left(Q_i p\left(Q_i/z_i\right) - W_i\right)$. Suppose instead D were to lie, and claim that the state is z_j and so request volumes Q_j in exchange for payment W_j. This would mean that the retail price received by D would be $p\left(Q_j/z_i\right)$. This yields D pledgeable income of $Q_j p\left(Q_j/z_i\right) - W_j$ at the end of period 0. The optimal program therefore requires the pledgeable income to be maximized when the truth is being told:

Program Bank The optimization program when D uses an independent banking sector is given by

$$\max_{\{Q_i, W_i\}} \sum_{i=1}^{n} g_i B \cdot I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right)$$

subject to the individual rationality constraint for U,

$$\sum_{i=1}^{n} g_i \{W_i - Q_i c\} = 0, \tag{6}$$

and the incentive constraint at the quantity setting stage for D,

$$Q_i p\left(\frac{Q_i}{z_i}\right) - W_i \geq Q_j p\left(\frac{Q_j}{z_i}\right) - W_j \text{ for all } j \neq i \tag{7}$$

This problem is isomorphic to one explored by Hart (1983) in the context of optimal labor contracts. In our model, U assumes the role of workers (the marginal cost c corresponding to workers' reservation wage in Hart's analysis) and D that of the firm (demanding a general input rather than labor specifically).

The following proposition follows from Hart (1983):

Proposition 2 *(Hart 1983, Proposition 2) The solution to Program Bank satisfies*

1. *No distortion at the top:*

$$\frac{\partial}{\partial Q}\left[Q_n p\left(\frac{Q_n}{z_n}\right)\right] = c$$

2. *Inefficiently low quantity demanded elsewhere*

$$\frac{\partial}{\partial Q}\left[Q_i p\left(\frac{Q_i}{z_i}\right)\right] > c \textit{ for all } i < n \tag{8}$$

3. *D's pledgeable income increases in the state*

$$Q_i p\left(\frac{Q_i}{z_i}\right) - W_i \geq Q_{i-1} p\left(\frac{Q_{i-1}}{z_{i-1}}\right) - W_{i-1} \textit{ for all } i > 1$$

4. *U's payoff increases in the state*

$$W_i - Q_i c \geq W_{i-1} - Q_{i-1} c \textit{ for all } i > 1 \tag{9}$$

Proof. Hart (1983) yields all 4 conditions.[5] We have a strict inequality in the second condition as U is risk neutral here. ■

[5] For D, explicitly, in Hart's notation, we have the revenue function

$$f(z, Q) = Qp\left(\frac{Q}{z}\right)$$

By exploring a general input into a downstream firm D a number of corollaries of the above proposition become novel and important results:

Proposition 3 *The optimal contract with a supplier U when D is subject to credit constraints and market risk results in:*

1. *Retail prices are too high relative to the level that would maximize joint period-0 profit (double marginalization) in all except the best demand state.*
2. *The optimal contract has the supplier making payments to D which are not recouped in low demand states. Hence, if marginal cost c is sufficiently small, $W(z_i)$ is negative for small realised demand states z_i and positive for large z_i.*

Proof. For part 1, note that equation (8) guarantees that the marginal revenue is above marginal cost at all demand states except for the highest. Hence as marginal revenue is declining we must have quantities being below the industry profit maximising levels. Hence retail prices are forced above the myopic industry profit maximising levels.

For part 2, note that U's individual rationality constraint yields $\sum_{i=1}^{n} g_i \{W_i - Q_i c\} = 0$ while $\{W_i - Q_i c\}$ is increasing in i (9). Hence we must have some state j such that

$$\begin{cases} W_i - Q_i c \leq 0 & \text{for } i \leq j \\ W_i - Q_i c \geq 0 & i > j \end{cases}.$$

Since U optimally shares in some of the risk, $W_1 - Q_1 c < 0$ and $W_n - Q_n c > 0$. ■

Proposition 3 yields new insights into the impact of credit constraints and market risk on vertical contracting. The downstream firm D is endogenously risk averse due to the credit constraints she faces. Had there been no credit constraints, the efficient supply

And this satisfies Hart's Assumptions 2 (as marginal revenue is positive and declining), and 6 (as profit grows in high demand states). For his Assumption 5 we require the marginal revenue to grow in high demand states. This is true as

$$\frac{\partial^2 f}{\partial Q \partial z} = \underbrace{\frac{\partial \left(\frac{\partial f}{\partial Q}\right)}{\partial \left(\frac{Q}{z}\right)}}_{\substack{\text{-ve as declining} \\ \text{marginal revenue}}} \cdot \frac{\partial \left(\frac{Q}{z}\right)}{\partial z} =_{\text{sign}} - \left[-\frac{Q}{z^2}\right] > 0$$

The other assumptions follow as U is assumed risk neutral and $I(\cdot)$ has been shown to be concave. D here is endogenously risk averse as the diminishing returns to investment while in Hart D is assumed risk averse.

contract would be for U to agree to supply the input at marginal cost of c, then D would require the myopic industry profit maximising quantity and set the myopic industry profit maximising price. However this contract leaves D with a risky income stream and exposes D to all the market risk. D would therefore like to convert at least some of her ex post risky profit stream into an ex ante certain one. That is D demands some insurance from U.

Formally D achieves this insurance by requiring U to make a payment independent of demand to D, and in return D makes a demand dependent payment back to U. The upstream firm agrees to this transfer of risk as she will recoup her payment to D, in expectation, through per unit input prices in excess of marginal cost. This is the intuition for result 2 of Proposition 3.

This general result appears to us to be relected in at least three common business practices. Firstly the common practice in the grocery market of slotting fees. Slotting fees are payments many supermarkets require of their suppliers - and a theoretical consensus has arguably emerged that slotting fees are to be thought of as part of a supplier signalling the quality of her product to the retailer (Klein and Wright 2007). However recent survey evidence suggests that risk sharing is a part of the rationale for slotting fees (Sudhir and Rao (forthcoming) and Bloom et al. (2000)). Second are risk sharing contracts. The payment made by U could be a financial transfer directly to D; or a sharing in some costs with repayments dependent upon realised demand. These are widespread in many industries - and indeed many outsourcing contracts have taken on risk sharing features.[6] Our model provides the first, to our knowledge, model of risk sharing (or slotting fee) contracts arising naturally outside of an exogenously imposed two-part tariff context. Finally our work is we would argue, reflected in the ubiquitous practice of trade credit. The payment from U to D can be seen as a short term loan with repayments dependent upon demand. Further the realised input cost per unit is higher than cost if the trade credit is used for more than (typically) 10 days when interest payments commonly start to apply (Petersen and Rajan 1997). In support of this final hypothesis that trade credit is used to provide insurance against fluctuartions, Petersen and Rajan demonstrate that the firm's ability to generate cash internally decreases its demand for trade credit. In particular each additional dollar of monthly profits lowers the firm's demand for trade credit by 23 cents and the estimate is significant at the 1% level.

Given that D optimally chooses a contract which sets the per unit input price in excess of marginal cost, she is driven to set a double marginalized price when the level of

[6]See "Outsourcing" in the Economist – September 29, 2008.

demand is revealed. Thus the price of the profit insurance is paid, in part, by consumers who receive prices in excess of the myopic profit maximising level. This is the intuition for result 1 of Proposition 3.[7,8]

A further corollary of Proposition 2 is the following:

Proposition 4 *A credit-constrained downstream firm exposed to market risk would strictly prefer to outsource input production to a non-credit constrained supplier U rather than produce in-house at the same cost.*

Proof. Suppose D were to produce the input in-house at marginal cost c. In this case, in effect the supply contract would satisfy $W_i = cQ_i$ for all states i. Hence for any demand state realization the integrated firm would maximize profits by solving

$$\max_{Q_i} \sum_{i=1}^{n} g_i B \cdot I \left(Q_i p \left(\frac{Q_i}{z_i} \right) - cQ_i \right).$$

This is solved where $\frac{\partial}{\partial Q} \left[Q_i p \left(\frac{Q_i}{z_i} \right) \right] = c$ for all z_i.That is the integrated firm would implement the efficient non-double-marginalized retail price.

However by Proposition 2 part 2,though implementable, this is not the optimal tariff when D is outsourcing input production to U. Hence, D strictly prefers outsourcing to U. ■

When D is facing market risk and is credit constrained then she becomes endogenously risk averse. A separate supplier can be used to provide some insurance against the worst market outcomes by defacto entering a profit sharing contract. The supplier is willing to offer the insurance as she is guaranteed some profit in good demand states; and so expects to at least break even overall.[9] Thus we exhibit a new rationale for credit constrained firms exposed to market risk to outsource supply: the suppliers can provide revenue insurance.

There are many reasons why outsourcing might be a good idea. But the relationship between market risk and outsourcing is still a topic of debate. Empirically there exists evidence supporting our theoretical results. For example both Harrigan (1985) analysing

[7] Some of this inefficient double marginalization is always optimal as, initially, it only has a second order effect on the downtsream period 0 retail profirts, while yielding a first order gain via the investment stage.

[8] These intuitions are new and different to the intuition offered for a parallel result in the context of labor contracting. The analogous result in labor economics is that too little labor is supplied by the union in low demand states to keep the firm honest when demand states are high.

[9] To the extent that input supply can be under-reported to a third party, this insurance cannot be provided by a third party. To see this suppose that the insurance were provided by a third party: D and U would then have an incentive to collude in under-reporting input supply.

executive interviews and Sutcliffe and Zaheer (1988) experimentally find evidence that firms do move more production outside the firm when exposed to demand uncertainty. However the dominant theoretical view is, arguably, that contractual incompleteness combined with demand risk would act to increase vertical integration (see Mahoney 1992 for a survey and discussion).[10] This model offers an important force pushing against integration which is responsive to market risk.

4 Complementarities Between Supplier Insurance and Banking

In the model as presented so far the supplier, U, offers her downstream buyer some pledgeable income insurance. The downstream firm D then goes to the banking sector to borrow to fund investment. If U also had access to the capital markets, and the same monitoring technology which banks have, then U could take the place of the bank providing the loan for investment as well as any pledgeable income insurance.

In fact this section shows that borrowing from U and committing to not use a separate banking sector strictly dominates using a banking sector. The reason is that by having to return to U for a loan D can commit to charge a lower price and therefore one which is less double marginalized. This is because if she misreports the state and so makes extra profits U can commit to ignoring the extra profits and so not allowing them to be leveraged. This allows D to credibly discipline herself.

It also follows from this analysis that U's dominance over a bank requires U to be able to commit to a contract which she (and D) would like to renegotiate should play arrive at a point off the equilibrium path. Without such commitment U cannot improve on a separate banking sector.

To derive these results suppose that D committed not to use a banking sector and only deal with U. D would now be proposing the contract $\{Q_i, T_0^i, T_1^i\}$ where Q_i is delivered in period 0 if the state is z_i in return for payment of T_0^i (which is net of any 'loan'). D then invests her available assets and after the investment returns are realized she pays U an amount T_1^i which is again conditional on the period 0 demand state.

As U is offering loans she must ensure that the amount she makes available satisfies D's investment moral hazard. U can therefore ask to see a given level of assets before providing the loan via T_0. This limits the states that D can misreport. Suppose that the

[10]Carlton (1979) offers the same conclusion but in a model of unadjustable input volumes.

state is z_i but D reports z_j. U will expect D to be able to show gross profits of $Q_j p\left(\frac{Q_j}{z_j}\right)$. However D will only be able to do this if her actual gross profits, $Q_j p\left(\frac{Q_j}{z_i}\right)$ exceed this level. This is only possible if $z_j < z_i$. Thus D can only report that the state is worse than it is - otherwise she will be found out at the end of period 0. The program to solve with no bank is therefore:

Program No Bank The optimal program when U provides the loan

$$\max_{\left\{Q_i, T_0^i, T_1^i\right\}} \sum_{i=1}^{n} g_i \left\{\pi\left(Q_i p\left(\frac{Q_i}{z_i}\right) - T_0^i\right) - T_1^i\right\}$$

Subject to

$$\sum_{i=1}^{n} g_i \left\{T_0^i + T_1^i - Q_i c\right\} = 0 \text{ (IR for } U\text{)} \quad (10)$$

$$\left[Q_i p\left(\frac{Q_i}{z_i}\right) - T_0^i\right] \cdot B \leq \pi\left(Q_i p\left(\frac{Q_i}{z_i}\right) - T_0^i\right) - T_1^i \text{ (no-shirk for } D\text{)} \quad (11)$$

$$\pi\left(Q_i p\left(\frac{Q_i}{z_i}\right) - T_0^i\right) - T_1^i \geq \pi\left(Q_j p\left(\frac{Q_j}{z_i}\right) - T_0^j\right) - T_1^j \quad \begin{array}{c}\text{for all } j < i \\ \text{(IC for } D\text{)}\end{array} \quad (12)$$

Note that if D should lie about the state and claim it is j when in fact it is $i > j$ then her assets will in truth be higher than she would have had under state j. However the size of her loan $\left(T_0^j\right)$ is not altered. These extra assets cannot therefore be leveraged.

Proposition 5 *Using U as a bank strictly dominates using a separate banking sector*

Proof. Consider the optimal tariff solving Program Bank: $\{Q_i, W_i\}$. This is the program when an independent banking sector is used. In state z_i, under this program D has pledgeable income of $Q_i p\left(\frac{Q_i}{z_i}\right) - W_i$. She invests an amount $I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right)$ and so borrows the difference between these two.

We first show that U can replicate the optimal contract D would set if using a banking sector. To achieve this set

$$\begin{aligned} T_1^i &= I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right) - \left[Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right] \\ T_0^i &= W_i - T_1^i \end{aligned}$$

So that T_1^i is the size of the loan provided. Keep Q_i as in the contract with the separate banks. Then (10), the individual rationality constraint of U, is satisfied by (6). By construction of T_1^i the credit constraint is binding in every state - therefore (11) is always tight. Finally (7) and the definition of the loan implies

$$\begin{aligned}\pi\left(Q_i p\left(\frac{Q_i}{z_i}\right) - T_0^i\right) - T_1^i &= B \cdot I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i\right) \\ &\geq B \cdot I\left(Q_j p\left(\frac{Q_j}{z_i}\right) - W_j\right) \text{ for all } j \neq i \text{ by (7)}\end{aligned}$$

The final term is the return available to D if her pledgeable assets are $Q_j p\left(\frac{Q_j}{z_i}\right) - W_j$ and she borrows to the point at which the credit constraint binds. We wish to show that this level of borrowing is greater than T_1^j for $j < i$. This is true if and only if having assets of $Q_j p\left(\frac{Q_j}{z_i}\right) - W_j$ and borrowing T_1^j (resulting in investment equal to the level in the right hand side of (12)) leaves the investment moral hazard slack. This is shown by noting that by definition:

$$\pi\left(Q_j p\left(\frac{Q_j}{z_j}\right) - T_0^j\right) - T_1^j = B \cdot \left[Q_j p\left(\frac{Q_j}{z_j}\right) - T_0^j\right]$$

now consider increasing z_j to z_i. As $\pi' > 1 \geq B$ we must have

$$\pi\left(Q_j p\left(\frac{Q_j}{z_i}\right) - T_0^j\right) - T_1^j > B \cdot \left[Q_j p\left(\frac{Q_j}{z_i}\right) - T_0^j\right]$$

hence borrowing T_1^j with pledgeable assets of

$$Q_j p\left(\frac{Q_j}{z_i}\right) - T_0^j - T_1^j = Q_j p\left(\frac{Q_j}{z_i}\right) - W_j$$

leaves the credit constraint slack. Therefore

$$B \cdot I\left(Q_j p\left(\frac{Q_j}{z_i}\right) - W_j\right) > \pi\left(Q_j p\left(\frac{Q_j}{z_i}\right) - T_0^j\right) - T_1^j \text{ for } j < i$$

as required. Hence (12) is actually slack (satisfied strictly).

But as the moral hazard condition on quantities is slack there is room for the transfer of some more risk upstream. Suppose that the quantities are altered to $Q_i + \varepsilon$ for all $i < n$ and the tariff W_i is increased by εc. The payment T_1^i and T_0^i retain the form given above. This new tariff satisfies (12) for small ε. U remains indifferent also satisfying (10). By definition of T_1 (11) is satisfied with equality. It therefore remains to note that the objective function has grown. This follows as by result 2 of Proposition 2, the marginal revenue at states below n exceeds c. ■

Proposition 5 shows that a supplier can, in principal, limit double marginalization by refusing to reward the gains from excessive double marginalization with a loan. However, renegotiation proofness would require that U always allows all assets to be fully leveraged - even if achieved off the equilibrium path. In such a case the optimal program would match that for an external banking sector and so the supplier could not beat a separate banking sector.

Of course, other than renegotiation proofness, a supplier may be unable to replicate a bank as its monitoring skills may not be as good and so it cannot borrow to lend to D as a bank could. We therefore restrict attention to D using a separate banking sector as in the core model.

5 Market Structure Comparative Statics

In this section we study how prices, investment levels and the supply contract are affected by changes in the market structure. For example, the impact of tight money; poor corporate governance; and alterations in the market risk profile. This analysis is not tractable in the fully optimal contracting framework characterised in Propositions 2 and 3. Motivated by these characetrisations we now turn to a subclass of contracts: two part tariffs.

Proposition 3 suggests that such a restriction is not overly onerous. The optimal contract requires U to pay a fixed cost which results in losses in poor demand states. The per unit prices are then above marginal cost so that at higher demand realisations U claws back her fixed payment and indeed makes a profit in the highest demand states. The two part tariff restricts the per unit markups required of U to be linear. The following subsection establishes some basic results. Subsequent subsections then conduct the comparative statics analyses.

5.1 Optimal Two Part Tariff Supply Contracts

We suppose that D offers U a contract of the form

$$W(Q) = f + wQ \tag{13}$$

where f is a fixed fee and w the per unit input price. Concurrently we extend our model to consider any distribution of demand states $G(z)$ supported on $[\underline{z}, \bar{z}]$, $\underline{z} \geq 0$ and we normalise so that $E[z] = 1$. This class of two-part tariffs (13) is particularly tractable as:

Lemma 2 *Under two part tariffs of the form (13) the retail price is independent of the demand state and the quantity delivered is proportional to the resultant demand state.*

Proof. Given tariff $W(Q) = f + wQ$, if realized demand is z then D will select Q to maximize $B \cdot I\left(Qp\left(\frac{Q}{z}\right) - wQ - f\right)$. This has first order condition given by

$$I'\left(Qp\left(\frac{Q}{z}\right) - wQ - f\right) \cdot \left\{p\left(\frac{Q}{z}\right) + \frac{Q}{z}p'\left(\frac{Q}{z}\right) - w\right\} = 0$$

as $I' > 1$ we require $p\left(\frac{Q}{z}\right) + \frac{Q}{z}p'\left(\frac{Q}{z}\right) = w$. This has solution $\frac{Q}{z} = q(w)$ (defined after assumption 1). And the retail price will be given by $p(q(w))$. Thus price is independent of the state and volume grows in proportion to the demand state as required. ■

Therefore once D selects the tariff, if the demand state is revealed as z D will require volumes $zq(w)$ resulting in a retail price of $p(q(w))$ whatever the state. We extend assumption 1 to give the industry profits a little more (yet standard) structure:

Assumption 3 *Industry profits $zq(w)[p(q(w)) - c]$ are strictly concave in w and maximized when there is no double marginalization ($w = c$).*
Further D's profits gross of the fixed fee are strictly positive, decline as the wholesale price increases,

$$\frac{d}{dw}[q(w)[p(q(w)) - w]] < 0$$

and strictly convex in w.

We can therefore rewrite D's problem (Program Bank) within this affine class of tariffs as

$$\max_w E_z [B \cdot I(zq(w)[p(q(w)) - w] - f)]$$

subject to

$$\begin{aligned} E_z [f + zq(w)(w - c)] &= 0 \\ \Leftrightarrow \quad f &= -\underbrace{E(z)}_{=1} q(w)(w - c) \end{aligned} \tag{14}$$

This formulation yields the slotting fee explicitly in terms of the per unit input price. Hence D's objective collapses to

$$\max_w B \cdot E_z [I(zq(w)[p(q(w)) - w] + q(w)(w - c))] \tag{15}$$

where I is defined implicitly by (2). Therefore under a two part tariff the expected payoff to D depends solely upon the wholesale price w. The pledgeable assets at the end of period 0 which are used to secure the loan are given by $a(z,w)$:

$$a(z,w) := zq(w)[p(q(w)) - w] + \underbrace{E[z]}_{=1} \cdot q(w)(w-c) \tag{16}$$

Alterations in the market structure (such as through the introduction of interest rates or alterations in the market risk) affect the contract agreed between U and D by altering the level of the per unit input price. To facilitate analysis it is to confirm that D's objective function is concave in w.

Lemma 3 *The downstream firm's objective function:* $E_z[I(a(z,w))]$ *is strictly concave in the wholesale price* w*.*

This result is proved in the appendix - as are all the technical results. An immediate application of this lemma is that it can be used to confirm our main results contained in Proposition 3. That is that retail prices are raised via double marginalised required as a response to the credit constraints D is facing; and U is required to pay a slotting fee.

Proposition 6 *The equilibrium contract in period 0,* (f^*, w^*)*, involves double marginalization,* $w^* > c$*, and payment of a slotting fee from the upstream firm to the downstream firm,*

$$f^* = -q(w^*)(w^* - c) < 0$$

The downstream firm D, which is exposed to market risk, is endogenously risk averse as there are diminishing returns to investment. These diminishing returns are captured by the fact that $\pi(I)$ is modeled as a concave function. We assume that the investment technology is regular in the following sense:

Assumption 4 We suppose that the curvature of the technology function is declining in magnitude at higher investment levels:

$$\frac{\partial \pi}{\partial I} > 0 \ , \ \frac{\partial^2 \pi}{\partial I^2} < 0 \ , \ \frac{\partial^3 \pi}{\partial I^3} \geq 0$$

Therefore the curvature π'' is negative and gradually increases towards 0 at higher investment levels.

5.1.1 Comparative Static Preliminaries

Suppose that there is a permutation of the market structure which, it can be demonstrated, results in a permutation θ of the investment function. That is D's period 1 profits are given by $B \cdot I(a;\theta)$. Restricting ourselves to cases where D remains credit constrained, Lemma 1 guarantees that $\frac{dI}{da} > 1$ and $\frac{d^2I}{da^2} \leq 0$. That is investment is worthwhile and there are diminishing marginal returns to investment.

At the optimal per unit input price, w, $E_z[I(a(z,w);\theta)]$ is maximized. Therefore the first order condition is given by

$$E_z\left[\frac{dI}{da}(a;\theta)\cdot\frac{da}{dw}\right] = 0$$

Note that $E\left[\frac{da}{dw}\right] = \frac{da}{dw}\Big|_{z=1} < 0$ as we have some double marginalization. The term $\partial a(z,w)/\partial w$ is strictly decreasing in z as

$$\frac{\partial a(z,w)}{\partial w \partial z} = \frac{d}{dw}\left[q(w)\left[p(q(w)) - w\right]\right] < 0$$

as D's profits decline as double marginalization is increased. Hence $\frac{\partial a}{\partial w}$ is potentially positive at small values of z and then becomes negative at larger values of z.

Suppose we move from parameter θ_1 to $\theta_2 > \theta_1$. We can define the function $x(a,\theta_2)$ by the identity:

$$\frac{dI}{da}(a,\theta_2) = x(a,\theta_2)\cdot\frac{dI}{da}(a,\theta_1)$$

which is well defined as $\frac{dI}{da} > 1$. We must therefore have $x(a,\theta_2) > 0$.

Lemma 4 *The following holds*

1. *If* $\frac{dx}{da} < 0$ *then* $w^*(\theta_2) > w^*(\theta_1)$
2. *If* $\frac{dx}{da} > 0$ *then* $w^*(\theta_2) < w^*(\theta_1)$

This lemma notes that, in case 1, moving to the permutation θ_2, lowers the marginal returns to pledgeable income at high asset levels. This is saying that at high asset levels there is little loss induced from a small reduction in pledgeable income due, for example, to double marginalization losses. Therefore the negative impact of double marginalization is reduced and so more of it becomes optimal.

We now suppose that the parameter θ_2 is close to θ_1. Then we have

$$\begin{aligned}\frac{dx}{da} &= \frac{d}{da}\left[x-1\right] = \frac{d}{da}\left[\frac{\frac{dI}{da}(a,\theta_2)-\frac{dI}{da}(a,\theta_1)}{\frac{dI}{da}(a,\theta_1)}\right] \approx (\theta_2-\theta_1)\cdot\frac{d}{da}\left[\frac{\frac{d^2I}{dad\theta}(a,\theta_1)}{\frac{dI}{da}(a,\theta_1)}\right] \\ &=_{\text{sign}} \frac{dI}{da}\frac{d^3I}{da^2d\theta}-\frac{d^2I}{dad\theta}\frac{d^2I}{da^2}\end{aligned}$$

We have therefore shown that

Lemma 5 *Consider a small increase in the parameter θ. If*

$$\frac{dI}{da}\frac{d^3I}{da^2d\theta}-\frac{d^2I}{dad\theta}\frac{d^2I}{da^2}>0 \text{ for all possible } a \tag{17}$$

then $w^*(\theta)$ declines in θ.
If the inequality is reversed $w^*(\theta)$ rises in θ.
If there is equality, w^* is unchanged.

Lemma 5 allows us to consider a number of changes to the market structure. By deducing the impact of these changes on the investment function I we can determine the effect on wholesale input prices – and hence on retail prices and expected investment levels.

5.2 The Impact Of Changes in Financial Markets - Tight Money

Suppose, in this subsection, that money borrowed from the external banking sector needs to be repaid at an interest rate r. As the downstream firm is assumed credit constrained, she will borrow as much as her end of period 0 assets (a) allow. Hence (2) is altered to

$$IB = \pi(I) - (I-a)(1+r) \tag{18}$$

which defines I implicitly. We require that $B+1+r-\frac{\partial\pi}{\partial I}>0$ (paralleling (3)) which is a corollary of the requirement that $\frac{dI}{da}>1$.

The raising of interest rates in the external capital markets has two opposite effects. The first is that higher interest rates make credit constraints tighter and so lower investment levels for a given level of assets. This pushes the marginal return to investment up and so makes the firm more risk averse as it seeks to move back up the technology function. However the raising of interest rates also increases the payment required for the loan and so lowers the marginal return to investment. If the curvature of the technology function is sufficiently great then the first effect dominates.

Lemma 6 *If the curvature of the technology function is sufficiently curved:*

$$\left[\underbrace{-\frac{\partial^2 \pi}{\partial I^2}}_{curvature} \right] \geq \frac{(B+r)(1+r)}{\pi(0)} \tag{19}$$

then $\frac{d}{dr}\left[\frac{\partial \pi}{\partial I} - r\right] \geq 0$

We are now in a position to derive the main result of this section:

Proposition 7 *As money becomes tighter (r rises), if the curvature of the technology function is great enough (19 holds), then:*

1. *[*cf. the price puzzle*] In the short run (period 0) wholesale input prices, and so also retail prices, rise.*

2. *In the long run the expected level of investment declines.*

Therefore if the interest rate against which businesses make their investment decisions is increased, consumer surplus declines in both the short and long run in this partial equilibrium model. The relevant interest rate will be the downstream firm's cost of debt which will be related to the long run real rate of interest in the economy. In the long run, as might be expected higher interest rates lead to lower firm investment on average. However in the short run the tighter money also pushes retail prices up as firms, responding to their endogenous risk aversion, introduce inefficiency into the supply chain to try and lock in what revenues and hence what investment levels they can.

Result 1 of Proposition 7 is closely related to the price puzzle. The price puzzle refers to a long standing observation in macroeconomics that retail prices appear to rise in the short run when interest rates are raised by the central bank. Standard macroeconomics would suggest that higher policy interest rates should raise the real interest rate for business investment decisions and so lead to a lowering of the price level (and so inflation) by contracting investment in the economy and so shrinking output below the economy's natural equilibrium rate. However before this macroeconomic effect occurs, prices (aggregated into an economy wide price level) seem to first rise for a number of months to a year by a statistically significant amount [Christiano et al. 1999]. The exact size of the price puzzle is in dispute as it varies depending on the extent to which the empirical estimation seeks to control for the link between interest rates and inflation expectations [Balke and Emery 1994]. But the existence of the price puzzle has become a broadly accepted stylized fact. Proposition 7 provides a novel explanation of the price puzzle grounded in optimizing firm behaviour.

5.3 The Impact Of Changes in Corporate Governance

Firms are credit constrained in this model as their managers suffer from moral hazard. They have the opportunity to avoid placing the highest effort into making a success of any investment and instead can garner some private benefits to themselves. If they do this, for expositional simplicity, we have assumed the project fails for sure. However the manager(s) amass a total private benefit of $B \cdot I$ where I is the size of the proposed investment.

Hence the larger is B the weaker is the corporate governance regime within which the managers of the firm work. Larger B's allow the managers to focus increasingly on their private benefit and less on the maximisation of the firm's profit. As a result poorly governed firms will find it more difficult to access capital markets - they are more tightly credit constrained. This section shows that weak corporate governance causes consumers to lose out in the present and the future: retail prices rise in the present and expected investment levels decline also.

Lemma 7 *If the curvature of the technology function is sufficient then* $\frac{\partial}{\partial B}\left[\frac{\partial \pi}{\partial I} - B\right] > 0$. *This is guaranteed if*

$$\left[\underbrace{-\frac{\partial^2 \pi}{\partial I^2}}_{curvature}\right] > \frac{B^2}{\pi(0)} \tag{20}$$

Lemma 7 parallels Lemma 6. As corporate governance gets worse (B rises) then less investment can be conducted for any given level of pledgeable assets due to the tighter credit constraints. This raises the marginal return to increased investment. However increased fraction B of extra unit of investment could be expropriated by the management. The lemma shows that if investment has sufficiently decreasing marginal returns the former effect dominates.

In this setting poor corporate governance is associated with poor results for consumers in the form of raised prices and lower investment:

Proposition 8 *As Corporate Governance deteriorates (B rises), if the curvature of the technology function is great enough (20 holds), then:*

1. *In the short run (period 0) wholesale input prices, and so also retail prices, rise.*
2. *In the long run the expected level of investment declines.*

Active corporate governance which limits management's discretion to misappropriate funds is therefore an unambiguously good thing in this model. Unsurprisingly it leads to higher levels of investment as the credit constraints facing the investing firm are loosened. Further, in the run up to investment, good corporate governance serves to lower the retail price and so raise consumer welfare. This follows as a manager working in a good corporate governance regime has less need to lock profits in due to her endogenous risk aversion. This reduces the need to pay to pass risk to the suppliers and so allows retail prices to fall.

5.4 The Impact Of Changes To Technology $[\pi(I)]$

In this subsection we explore how technology shocks, which alter the returns available from investment, affect relationships with suppliers and hence retail prices. To study this we consider rotations of the technology function $\pi(I)$.

Consider therefore a clockwise rotation of $\pi(I)$ which has the effect of pushing the function $\pi(I)$ upwards at every feasible level of investment I. Such an alteration in the potential of technology raises the return to investment available to the firm: D. One might therefore expect firms to become more risk averse and so for technological changes such as this to be associated with higher pre investment price levels. This reasoning is false.

To understand why consider a regular rotation of the sort described, parameterised by θ, and regular in the sense that

$$\frac{\partial \pi}{\partial \theta} > 0 \ , \ \frac{\partial^2 \pi}{\partial I \partial \theta} \leq 0 \ , \ \frac{\partial^3 \pi}{\partial^2 I \partial \theta} \geq 0$$

As we are describing a rotation of θ the increase in the return to investment is largest at low levels of investment and declines at higher levels of investment. Such an alteration to the investment technology is shown in Figure 1.

With such a shift in the technology we have the following result:

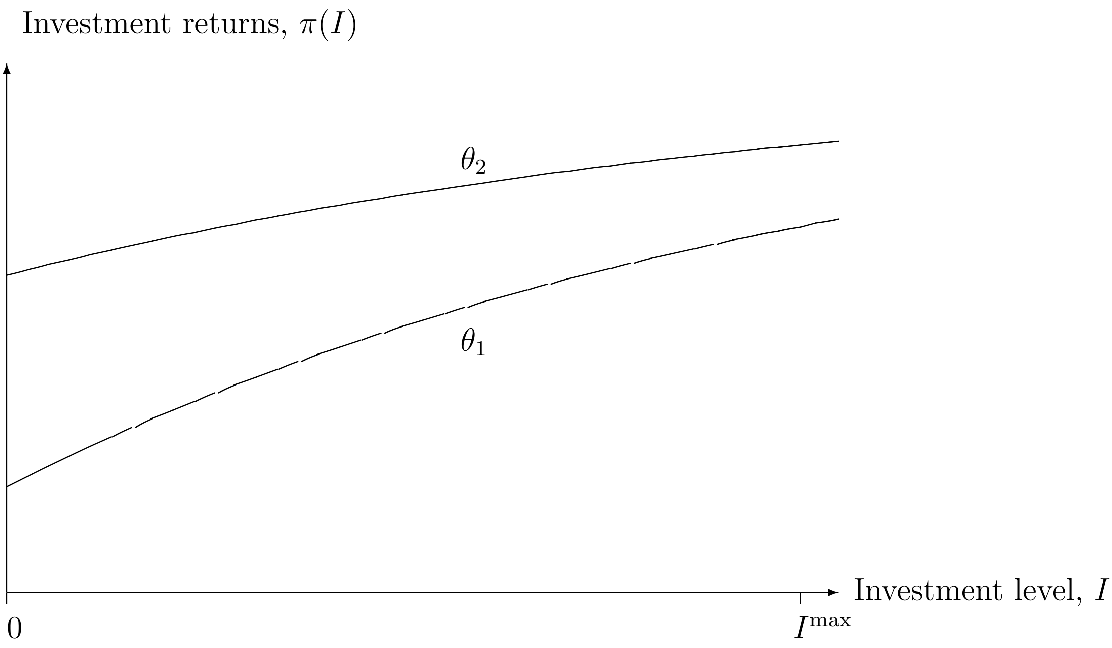

Figure 1: A positive shock to the investment technology.

Proposition 9 *If technology improves according to the rotation parametrised by θ then the optimal pre investment per unit price, w^*, falls with the rotation. Further the expected level of investment increases.*

As the improvement in the technology is largest at small investment levels it has the effect of dampening the curvature at these low levels. This reduces the need to be risk averse in the event of a bad market draw - and so the need to share risk and double marginalize declines. Further, though the return to investment has risen, the effect on curvature and hence on risk is the dominant force. Hence such a rotation in technology lowers consumer prices and so also raises the expected level of investment.

The case of an additive shock to technology is an immediate corollary:

Corollary 1 *Suppose that the ex post technology experiences a positive shock becoming $\tilde{\pi}(0, I; \theta) = \pi(0, I) + \theta$ then the optimal period 1 input price (and hence retail prices) decline. And the level of investment rises.*

Proof. We have

$$\frac{\partial \tilde{\pi}}{\partial \theta} = 1 > 0 \text{ and } \frac{\partial^2 \tilde{\pi}}{\partial I \partial \theta} = 0 = \frac{\partial^3 \tilde{\pi}}{\partial^2 I \partial \theta}$$

Hence the result follows from Proposition 9. ■

Hence credit constraints acting on firms cause them to be endogenously risk averse and so raise pre investment price levels as the firm seeks to lock certain levels of pledgeable assets in. However improvements in the technology function which increase the returns to low levels of investment at least as much as they increase the returns at high investment levels have the effect of lowering the extent of the risk aversion. This then results in lower pre-investment retail prices and higher expected investment levels overall.

5.5 Deep Pockets and Observed Buyer Power

In this section we explore the implication of a firm, while remaining credit constrained, having (exogenously) more assets at its disposal. In other words, suppose some downstream firms have deeper pockets than others. If one observed the per unit prices that these firms were securing from their suppliers it would seem that such well endowed firms possessed buyer power:

Proposition 10 *Suppose that D has exogenously extra assets $\theta > 0$ additional to those from period 0 trade. Then D will require a lower wholesale price from its supplier, implement a lower retail price and invest more on average in the long run.*

Proof. The incentive compatibility condition which defines the investment level, function I, is given by

$$\pi(I) - (I - [a + \theta]) = BI$$

This can be rewritten as

$$\pi(I) + \theta - (I - a) = BI$$

Thus the extra assets, as they are all used in the subsequent investment, act as a positive shock to the technology function. Corollary 1 then yields that increased assets θ translate into lower per unit input prices in the first period and hence lower retail prices also. The corollary also delivers the result that expected investment levels increase. ■

In this model therefore exogenously asset rich firms are unambiguously positive for consumer welfare. Their extra assets weaken their credit constraints and allow greater investment on average. However their advantage is greater than this. Such firms can expect to invest more on average and so the marginal returns to extra investment are lower. Therefore the need to insure pledgeable assets is less. This therefore translates into a requirement for lower (less double marginalized) input prices from the supplier. This lowers retail prices, which itself is good for consumers.

The lower per unit input prices required by asset rich downstream firms would look to the external observer like buyer power – but formally it is not. Buyer power is [usually] defined by competition authorities as an advantage granted to a buyer by its size allowing it to secure lower input prices. In this model D has all the bargaining power - whatever her exogenous asset level θ. And as a consequence U is always held to 0 utility. Rather size in our model causes the buyer to use her bargaining power in a different way which is unambiguously good for consumers.

5.6 Consumer Surplus Implications of Market Growth

In this subsection we analyse the case in which firm D enjoys a positive shock to the demand for her products. Thus the firm is still exposed to market risk - but increasing weight is placed on higher demand states being realised. We model this by supposing that the market size distribution, $G(z)$, undergoes a first order stochastically dominant shift to distribution $H(z)$ say with $E[z_H] > 1 = E[z_G]$.

We begin our analysis by noting that, as one might expect, the average level of investment in the second period (the long run) is higher if D experiences a positive shock to demand. This is partly because higher market size realizations occur more often due to the shock to demand. But also because the larger expected demand raises the size of the fixed fee the supplier can be made to pay for any given level of per-unit input price.

Of more interest is the analysis of the prices in the short run. Unfortunately a first order stochastically dominant change in the market risk does not have an unambiguous effect on the retail prices. Firstly, as high market sizes are more likely, the double marginalized losses are greater. This pushes towards a lower optimal input price as one might expect. However, as the expected size of the demand has risen U is willing to provide a greater slotting fee and further this slotting fee rises in the per unit input price at the optimal input price level w_G^*. To see this note that the first order condition at the optimal wholesale price requires $E\left(\frac{dI}{da}\frac{da}{dw}\right) = 0$. But then $\frac{da}{dw}$ must take positive and negative values. For $\frac{da}{dw}$ to take positive values is only possible if U gains from small increases in the per unit price (as D always loses from such gains). This therefore pushes towards a higher per unit price.

However, more can be said if the increase in expected market size comes exclusively from a reduction in the likelihood of small demand states. Formally define, under the benchmark distribution G, the cutoff between small and large demand states $\tilde{z}$ to be the market size at which the pledgeable assets for D are maximized when U is required to supply at a wholesale input price of w_G^*. Therefore

$$\left[\frac{\partial a}{\partial w}\right]_{w_G^*} = \underbrace{\frac{d}{dw}\left\{q(w)\left(w-c\right)\right\}_{w_G^*}}_{>0} + \tilde{z}\underbrace{\frac{d}{dw}\left\{q\left(w\right)\left[p\left(q\left(w\right)\right)-w\right]\right\}_{w_G^*}}_{<0} = 0$$

Such a point is unique as $\frac{\partial^2 a}{\partial w \partial z} < 0$ as D's profits decline in the wholesale price.

We therefore consider any first order stochastically dominant change in market risk, to distribution $H(z)$ say. However we restrict H to not lower the probability of any high demand states. Therefore in addition to the FOSD condition $H(z) \leq G(z)$ for all z, we require

$$H'(z) \geq G'(z) \text{ if } z \in (\tilde{z}, \bar{z}] \tag{21}$$

Under this stronger version of stochastic dominance we indeed have wholesale and retail prices being lower when the probability of higher market sizes is increased:

Proposition 11 *Suppose that the downstream firm D experiences a positive shock to market demand so that expected demand increases in a first order stochastically dominant way. Then:*

1. *The expected level of investment in the long run rises.*
2. *If the shock to demand is such that the probability of high demand states is not reduced – that is (21) is satisfied – then wholesale and so retail prices in the short run fall.*

5.7 Consumer Surplus Implications of Decreasing Risk

It is of value to analyse the effect on firm behaviour of a shock to the extent of market risk D is exposed to. To model this we consider a decrease in market risk according to a mean preserving contraction. Thus the expected size of the market is unaltered - but excessively low or high demand realizations become less likely.

Small reductions in market risk do not have an unambiguous effect on the retail and wholesale prices in period 0. When market risk falls it causes small demand realizations to be less likely. Thus insurance is less valuable and this creates a force towards lower (less double marginalized) wholesale prices. However large demand realizations are less likely too. This restricts the likely demand to levels where the investment function has larger marginal returns and so makes insurance more valuable acting to push up wholesale prices.

Once again therefore the direction in which wholesale and hence retail prices actually move is ambiguous without adding further structure.

We can however say a little more. It is clear that if there is no market risk at all then there is no benefit to double marginalization. The optimal two part tariff would have a wholesale price set at the cost level of c. However, given the competing effects outlined it is not clear that as we approach this limit point of no market size risk the inefficiency inserted into the supplier-buyer relationship can be made arbitrarily small. We show this result below.

A stronger result is available regarding investment and hence consumer surplus in the long run. *Any* decrease in market risk translates into a larger expected level of investment in the second period.

To formally capture these results we index the market risk by $g_K(z)$. As K grows we suppose that z undergoes a mean preserving contraction. In the limit of K becoming infinite, z will be almost surely 1.

Proposition 12 *Suppose that $K' > K$ so that $G_{K'}$ is a mean preserving contraction of distribution G_K :*

1. *The expected investment is greater under K' than under K.*
2. *There exists a threshold $\hat{K} > K$ such that any downstream firm facing demand risk $g_{K'}$ with $K' \geq \hat{K}$ sets a lower equilibrium wholesale price.*

Therefore any reduction in risk raises the expected levels of investment. This follows as the level of investment is a concave function of the realised demand – so realised investment levels are very sensitive to low demand realizations which become less likely if risk is reduced. This forces the expected investment level up. If the reduction in risk is substantial then the wholesale price and so lower the retail price are also lowered.

5.8 Welfare Impacts Of Downstream Acquisitions and Diversification

The analyses conducted have implications for any action of the downstream firm which alters the risk or expectation of demand. Downstream acquisitions and diversification certainly impact the risk faced by the firm; as well as the investment opportunities available. This section considers the impact of a strategy of diversification/acquisitions on the risk sharing requirements; retail pricing and investment levels of the firm D.

Suppose therefore that firm D is active in K markets in which she uses one unit of input from U to create one unit of product. The realised market size in market k is denoted z_k and is a random variable. This formulation is, thus far, without loss of generality. We also assume that in each market in which D operates she has the same investment opportunities yielding return $\pi(I)$ if an investment of I is conducted. Thus the returns to investment expenditure are the same across markets. This assumption allows us to focus on the short term risk and return implications of a strategy of diversification.

We suppose that D runs an internal capital market. This allows D to pool the returns from period 0 trade and then allocate the capital across her markets to maximise the investment return. The evidence that firms run some sort of internal capital markets is strong (Shin and Stulz 1998); though whether it is perfectly effective is not clear.

Adapting our model to D having multiple markets can therefore be accomplished as follows. Before period 0 D agrees a two part tariff with the upstream U of (F, w) where F is the total fixed fee. In period 0 D sees the demand state z_k in market k. She delivers volumes $z_k q(w)$ to market k. Individual rationality requires U to set the total fixed fee (payment to D which allows D to share risk) at the level $F = -\left(\sum_k E[z_k]\right) \cdot q(w)(w-c)$. As D is assumed to have the same long run investment opportunities with diminishing marginal returns in each market she will divide her end of period 0 pledgeable assets equally between the markets. She will then leverage these assets to conduct the same size investment project in each market. Hence D's objective collapses to

$$\max_w KB \cdot E_{\{z\}}\left[I\left(\frac{\sum_k z_k}{K} q(w)\left[p(q(w)) - w\right] + \left(\frac{\sum_k E[z_k]}{K}\right) \cdot q(w)(w-c)\right)\right]$$

Hence D's problem is isomorphic to the one market problem analysed above. However the internal capital market causes the risk from the K markets to be pooled.

We can therefore apply the previous results to this setting to develop the following insights:

The Conglomerate Benefit Suppose D increases the number of markets it is active in from K to K'. Suppose further that each market is equally risky and of the same expected size then:

1. D will invest more in each market in expectation.
2. There exists a threshold $\hat{K} > K$ such that if $K' > \hat{K}$ then D will lower the short run retail prices in all the markets she is active in.

This result is a consequence of the fact that diversification, combined with an internal capital market, lowers the variance of the average market size. That is $var\left(\sum z/K\right) < var(z)$. Hence the conglomerate benefit arises as an immediate corollary of Proposition 12.

It is not entirely controversial that conglomerate mergers can add value. Hubbard and Palia (1999) find that the largest returns made from conglomerate mergers in the 1960s occurred when capital unconstrained firms bought up capital constrained ones. Our analysis suggests that such mergers reduce the need for risk sharing with suppliers and so can be part of the explanation for higher investment, more efficient prices and hence greater market returns.

Expansion Into Risky Markets Suppose that D serves a safe market and decides to expand into a risky market of the same expected size. Both markets are assumed to have the same investment technology and potential. Then:

1. The expected investment levels in all markets would fall.
2. If the new market is sufficiently risky then the short run retail prices in all markets would rise.

This is again a consequence of Proposition 12. D's internal capital market pools the risk from the two markets. As the new market is risky the expansion raises the riskiness of the per market level of pledgeable income. As investment returns are concave this increased risk lowers the expected level of investment and can also cause D to seek more risk sharing in extremis.

6 Supply Side As Opposed To Demand Side Risk

The analysis so far has modeled market or demand risk relating to the profitability of the retail market for the produced goods. In this setting we have shown that market risk and credit constraints interact to produce an endogenously risk averse firm which seeks to share some of the risk, at a cost, with its vertical partners. Thus we have discovered how retail prices are pushed up; and how market risk creates an incentive to outsource supply processes off to a not financially constrained third party.

One is drawn to consider whether these results are robust to the risk being on the supply side rather than on the demand side of the production process. For example, if the risk was as to the costs of the input due to volatility in the price of some raw material, such as oil, would we still find that retail prices are pushed higher by credit constraints and that an incentive to outsource would remain?

Carlton (1977, 1979) would suggest not. He analyses input price risk and finds a theoretical rationale for increased vertical *integration* as market risk grows. Carlton's work is however underpinned by a key assumption which differs from our analysis. Namely Carlton assumes that the supplier must commit to production levels before the market parameters are revealed. Thus if demand is weak the input supplier must throw some production away. This risk of loss causes the input supplier to charge above marginal cost prices - an inefficiency which can be removed by vertical integration Carlton argues.

Our analysis offers the opposite conclusion and maintains the results derived earlier in this paper. We will confirm that whether risk is on the supply or demand side our retail price and outsourcing results stand. Thus the risk will cause the input prices (and hence retail prices) to rise. However this price rise allows risk to be shared through the supply chain and so outsourcing (of the marketing channel here) becomes desirable due to the insurance possibilities it creates.

Formally we consider the permutation of our core model to place risks on the upstream costs and not on the downstream demand:

U	$\rightarrow$	D	$\rightarrow$	Inverse demand $p(Q)$
$\Uparrow$		$\Uparrow$		
Marginal cost κ_i. Uncertain. Occurs with probability h_i		0 marginal cost		

We alter the location of bargaining power and assume U is in a position to offer take it or leave it contracts to D. In a robustness check we will demonstrate that this is not a crucial assumption as the intuitions already generated map straightforwardly to a full bargaining model (Section 7). For U to have any incentive to invest she must have some bargaining power so that she can secure positive utility after investment occurs. As previously, if U has pledgeable assets a then she gets return $BI(a)$ and the function $I(\cdot)$ has been shown to be concave increasing. The game proceeds analogously:

Period 0 U makes a take it or leave it offer to D of contract $\{Q(\kappa_i), W(\kappa_i)\}$. The cost state, κ_i, is then revealed. U announces the cost state which applies and the parties trade leaving U with some pledgeable assets. The cost states are numbered so that $\kappa_1 < \kappa_2 < ... < \kappa_n$.

Period 1 U levers the pledgeable assets from an external capital market, invests and generates payoff: $B \cdot I$.

The analysis of this model proceeds in a similar, but not identical manner, to that of our benchmark model. The difference arises as D (the party accepting the contract now) is risk neutral in the transfer payment - but not in the quantity delivered. Hence the proof used by Hart (1983) is not applicable. We can however adapt our previous work to solve this model variant.

The full information benchmark

If the cost state is revealed to all then U's problem in designing the tariff is to solve

$$\max_{\{Q_i,W_i\}} \sum_{i=1}^{n} h_i B \cdot I \left(W_i - Q_i \kappa_i\right) \tag{22}$$

$$\text{subject to D accepting} \quad : \quad \sum_{i=1}^{n} h_i \left\{Q_i p\left(Q_i\right) - W_i\right\} = 0 \tag{23}$$

where the probability of cost κ_i is h_i.

Given a realised demand state κ_i, the parties prefer efficient retail pricing and so volumes $q(\kappa_i)$ will be sold ($q(\cdot)$ defined after assumption 1). U, with all the bargaining power, requires a payment from D equal to the expected industry profit; the expectation taken over al possible input cost realisations. This arrangement is efficient as U is endogenously risk averse due to the credit constraints and the contract allows her to pass all the risk on to D.

Thus the full information solution has U being fully insured, and D bearing all the risk. The retail price responds fully to the input cost state and moves to the optimal level given the realized costs.[11] However this cannot be the outcome when only U can observe the cost state. She would then have an incentive to claim costs are higher than they really are with a view to extracting some extra rents from D's variable payment.

The asymmetric information case

If only U observes the cost state then she is exposed to adverse selection. She must therefore structure the contract so that she doesn't have an incentive to misreport the cost state. Thus U's problem is given by the maximisation of (22) subject to D accepting (23) and also to an incentive compatibility constraint:

$$W_i - Q_i \kappa_i \geq W_j - Q_j \kappa_i \text{ for all } j \neq i \tag{24}$$

[11]The proof of this parallels that used for Proposition 1 and is therefore omitted.

We denote this program U. Adapting the methodology offered by Hart allows us to establish the following key result which is derived in Appendix B:

Proposition 13 *The solution to program U is*

1. *No distortion in lowest cost states:*

$$Q_1 = q(\kappa_1)$$

2. *Too little supply, in higher input cost states*

$$Q_i \le q(\kappa_i) \text{ for all } i \ge 1$$

3. *U's pledgeable income is higher in lower cost states:*

$$W_i - Q_i\kappa_i > W_{i+1} - Q_{i+1}\kappa_{i+1}$$

4. *D's payoff is also higher in lower cost states:*

$$Q_i p(Q_i) - W_i \ge Q_{i+1} p(Q_{i+1}) - W_{i+1}$$

Proposition 13 indicates that the intuitions in the benchmark model where risk was attached to the demand side apply also when risk is attached to the supply side. The optimal contract involves some risk sharing with the partner in the supply chain and so all firms' ex post profits depend on the realization of the risk. Further, in the case of supply side risk, credit constraints cause the retail price to be overly responsive to increases in the costs of production. Thus retail prices rise faster than an integrated seller would require in the face of rising costs. So, for example, if raw material costs were to be volatile then when they are high the retail price would be pushed even higher than apparently justified by the cost rise due to the credit constraints.

When the risk applies to the supply side the upstream firm U gains by having the ability to share the risk with the retailer. Thus the outsourcing of the retail channel to a non-credit constrained firm is valuable to the credit-constrained upstream firm.[12]

[12] The insurance that D offers to U would not be possible if U could open up an alternative retail outlet which bypassed D without D being aware. This therefore provides a rationale for upstream firms being strict in not opening up competition with their existing retail channel.

These results and the intuitions underlying them reinforce the results from the benchmark model.

7 A Bargaining Extension

Our analysis so far has been simplified by assuming that one party in the supply chain is exogenously endowed with the ability to make take it or leave it offers. Here we demonstrate that this is not an essential assumption. Our results and intuitions extend to a setting in which U and D bargain with each other; neither one being able to offer take it or leave it contracts to her vertical partner.

In considering bargaining one's first impulse might be to reach for the celebrated Nash Bargaining Solution (NBS). This sets the percentage gain to the parties from a small change in the agreed contract to be equal. However this is a difficult construct to work in this setting as the total surplus to be split depends upon the contract agreed. Analytically the percentage expected utility change is not easy to calculate and less easy to work with. Instead we offer a more simple approach. Common with the NBS we assume that bargaining selects a point on the Pareto frontier so that money is not left on the table. We assume that the bargaining power of D as compared to U is captured by the invariant parameter γ where the agreed solution splits the total rents available γ parts to D and $1-\gamma$ parts to U. If the size of the pie being bargained over were to be invariant to the agreement this solution would match the weighted NBS. This approach is not new to the literature - it is known as the proportional bargaining solution (Kalai 1977). The key axiom generating the proportional bargaining solution is known as the axiom of step-by-step negotiations (Kalai 1977 p1627). The axiom requires that the bargained solution should be invariant to a decomposition of the bargaining process into stages. Thus if the individuals consider first a subset of the set of feasible alternatives, reach an agreement on the subset which is then used as the threat point in a second bargaining stage over the remaining alternatives, then the final outcome should be the same as the outcome reached if bargaining occurred in just one step.

The main impact of bargaining modelled in this way is that D receives proportion γ of the whole pie. Hence her incentive to invest is reduced - but the qualitative features displayed in the analysis remain.

More formally, consider the benchmark model offered in Section 2. In period 1 the total pie available to the parties is $[\pi(I) - I + a]$ where a are D's pledgeable assets and I the total investment which can be made. D will bargain to receive proportion γ of this.

Hence the credit constraint equation becomes

$$BI = \gamma \left[\pi(I) - I + a\right]$$

which implicitly defines $I(a;\gamma)$. And U receives a payoff of $(1-\gamma)\left[\pi(I) - I + a\right] = \left[\frac{1-\gamma}{\gamma}\right] B \cdot I$. For D to be credit constrained and be willing to invest up to her credit constraint we require assumption 2 to apply with B replaced by $\frac{B}{\gamma}$.

Now consider period 0. The bargained agreement, $\{Q(z_i), W(z_i)\}$ must satisfy the following program:

Program Bargain The optimization program when D and U bargain and leverage from an independent banking sector is given by

$$\max_{\{Q_i, W_i\}} \sum_{i=1}^{n} g_i B \cdot I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i; \gamma\right)$$

subject to the requirements of the proportional bargaining solution:

$$\begin{aligned} (1-\gamma)\{D\text{'s total payoff}\} &= \gamma\{U\text{'s total payoff}\} \\ (1-\gamma)\left\{\sum_{i=1}^{n} g_i B \cdot I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i; \gamma\right)\right\} &= \gamma \left\{\begin{array}{c} \left[\frac{1-\gamma}{\gamma}\right] \sum_{i=1}^{n} g_i B \cdot I\left(Q_i p\left(\frac{Q_i}{z_i}\right) - W_i; \gamma\right) \\ + \sum_{i=1}^{n} g_i \{W_i - Q_i c\} \end{array}\right\} \end{aligned} \tag{25}$$

and the incentive constraint at the quantity setting stage for D,

$$Q_i p\left(\frac{Q_i}{z_i}\right) - W_i \geq Q_j p\left(\frac{Q_j}{z_i}\right) - W_j \text{ for all } j \neq i$$

However note that the proportional bargaining solution condition (25) can only be satisfied if

$$\sum_{i=1}^{n} g_i \{W_i - Q_i c\} = 0$$

Thus Program Bargaining and Program Bank are isomorphic if one alters the corporate governance parameter from B to $\frac{B}{\gamma}$. Hence the results we developed continue to apply.

Some reflection reveals the following intuition. After the investment stage the parties ill bargain to split the rents from trade in period 1 proportion γ to D and the remainder to U. At the end of period 0, D will optimally invest all her pledgeable income to secure the maximum possible leverage and so the greatest return from period 1 trade. Thus $D's$ return arises only from the profits secured at the end of period 1. If at period 0 U should bargain for some positive rents then overall she will secure more than $1-\gamma$ of the total surplus: she secures $1-\gamma$ proportion in period 1 and more in period 0. This is not possible according to the proportional bargaining solution. Hence U also defers her extraction of rents until after the investment has occurred. But then the risk and contracting results we have discussed all apply. The only change is that the incentive to invest is reduced for D – or analogously her shirking parameter rises to $\frac{B}{\gamma}$.

8 Conclusions

Credit constrained firms are forced to link the scale of their investments to their pledgeable assets. This causes otherwise risk neutral firms to behave as if they were endogenously risk averse when trying to amass pledgeable assets to fund an investment program. Therefore a credit constrained firm, with investment opportunities, which is exposed to market risk will alter her supply contracts with her vertical partners. The contracts will be altered to share profit risk throughout the vertical chain. This however comes at a cost to both the firm and ultimately to consumers in the form of higher prices..

The benchmark model analysed is the optimal contract between a credit constrained downstream firm exposed to demand risk and her upstream supplier. The downstream firm is shown to require the supplier to enter into a risk sharing or insurance type contract. The downstream firm wishes to convert an ex post risky income stream into a more certain one. The upstream supplier is therefore required to make a payment (or equivalently share development costs) with the downstream firm irrespective of realised demand. To win agreement to this the upstream firm will receive an input price per unit which is greater than the cost of production. It is therefore optimal, in the face of the market risk, to introduce double marginalization into the supply chain. As a result retail prices are pushed up by the fact that the downstream firm is credit constrained.

The observable result of this contract is that the upstream firm will make an actual loss if low demand is realised. Such payments from the supplier to the buyer are common in the grocery industry where they are known as slotting fees. In other industries we have described a risk sharing contract which are common in industries such as aerospace.

Our analysis indicates that there exist complementarities between insurance, which is provided by the supplier, and the provision of loans. The supplier, by linking the level of the loan to the actual volumes sold, can lower the incentive the downstream firm has to double marginalize. This allows retail prices to be lower and so increases the expected profits and the expected consumer surplus. This therefore provides a new rationale for

	Short Run Retail price movement:	**Long Run** Expected investment levels:
Tighter credit markets (increased interest payable on loans)	↑	↓
Worsening corporate governance of investing firm	↑	↓
Positive shock to investment technology	↓	↑
Exogenous increase in firm assets	↓	↑
Positive shock to market size	↓ if some large demand realisations are not made less likely	↑
Decrease in market risk	↓ if risk reduction sufficient	↑
Diversification across symmetric markets (conglomorate benefit)	↓ if present in sufficient markets	↑
Expansion into risky markets	↑ if new market sufficiently risky	↓

Table 1: Price And Investment Implications Of Credit Constraints

trade credit. However, for trade credit to be superior to external finance requires strategies which are not renegotiation proof.

Whether a supplier finances all of the investment, or just offers insurance, our model demonstrates a motivation for a credit constrained firm to outsource supply to a non-credit constrained supplier. This link between risk and outsourcing is new to the theoretical literature, and supported by empirical evidence.

The optimal supply contract has the supplier making a loss at low demand realisations and recouping those losses if demand is high. We can approximate this by a two part tariff. This allows us to consider how changes in the market structure alter consumer surplus in the short run (through prices) and in the long run (through investment). Pursuing this allows us to establish the new insights when investment shows sufficiently decreasing marginal returns. These are summarised in Table 1.

Our benchmark model analyses the case of demand risk. However the results are unaltered if there is instead supply side risk (volatile materials costs for example). In this setting it is still the case that credit constrained firms are risk averse and contract with their vertical partners to share risk resulting in retail prices which are to high (and so overly responsive to rises in materials costs for example). Further the motivation for outsourcing remains: a credit constrained upstream firm exposed to supply cost risk has a motivation to outsource the sales channel to a separate firm so as to benefit from the risk sharing possibilities.

A Omitted Proofs

Proof of Lemma 3. Dropping the scalar B and differentiating the objective function (15) twice with respect to w, and using (16), yields:

$$\begin{aligned} \frac{d^2 E_z I(a(z,w))}{dw} &= \int_{\underline{z}}^{\overline{z}} \frac{d^2 I(a(z,w))}{da^2} \left[\frac{\partial a(z,w)}{\partial w}\right]^2 dG(z) \\ &+ \int_{\underline{z}}^{\overline{z}} \frac{dI(a(z,w))}{da} \frac{\partial^2 a(z,w)}{\partial w^2} dG(z), \end{aligned} \tag{26}$$

where

$$\frac{\partial a(z,w)}{\partial w} = q(w) + (w-c)\frac{dq(w)}{dw} + z\frac{d}{dw}\{q(w)[p(q(w)) - w]\}$$

and

$$\frac{\partial^2 a(z,w)}{\partial w^2} = 2\frac{dq(w)}{dw} + (w-c)\frac{d^2 q(w)}{dw^2} + z\frac{d^2}{dw^2}\{q(w)[p(q(w)) - w]\}$$

The first term on the RHS of (26) is strictly negative because $I(a)$ is strictly concave in a (Lemma 1). We now show that the second term is negative as well, proving that $E_z I$ is strictly concave in w. To see this, recall from Lemma 1 that $dI(a)/da > 1$ for all a. Second, note that the strict concavity of the joint profits of the vertical chain with respect to w (Assumption 3) implies that $\frac{\partial^2 a(1,w)}{\partial w^2} < 0$. Strict convexity of $\pi(w)$ in w (Assumption 3) implies

$$\frac{\partial^3 a(z,w)}{\partial w^2 \partial z} = \frac{d^2}{dw^2}\{q(w)[p(q(w)) - w]\} > 0,$$

and so $\partial^2 a(z,w)/\partial w^2 > 0$ for z sufficiently large. Hence, there exists a unique $\widetilde{z} > 1$ such that $\partial^2 a(w,z)/\partial w^2 < 0$ if $z < \widetilde{z}$ and $\partial^2 a(w,z)/\partial w^2 > 0$ if $z > \widetilde{z}$. (But note that $\widetilde{z}$ may be smaller or larger than $\overline{z}$.) The second term in (26) can be rewritten as

$$
\begin{aligned}
& \int_{\underline{z}}^{\overline{z}} \frac{dI(a(z,w))}{da} \frac{\partial^2 a(z,w)}{\partial w^2} dG(z) \\
= & \int_{\underline{z}}^{\widetilde{z}} \frac{dI(a(z,w))}{da} \frac{\partial^2 a(z,w)}{\partial w^2} dG(z) + \int_{\widetilde{z}}^{\overline{z}} \frac{dI(a(z,w))}{da} \frac{\partial^2 a(z,w)}{\partial w^2} dG(z) \\
< & \left[\frac{dI(a(\widetilde{z},w))}{da}\right] \cdot \int_{\underline{z}}^{\widetilde{z}} \frac{\partial^2 a(z,w)}{\partial w^2} dG(z) + \left[\frac{dI(a(\widetilde{z},w))}{da}\right] \cdot \int_{\widetilde{z}}^{\overline{z}} \frac{\partial^2 a(z,w)}{\partial w^2} dG(z) \\
= & \left[\frac{dI(a(\widetilde{z},w))}{da}\right] \cdot E_z \left[\frac{\partial^2 a(z,w)}{\partial w^2}\right] = \frac{dI(a(\widetilde{z},w))}{da} \frac{\partial^2 a(1,w)}{\partial w^2} \\
< & \ 0
\end{aligned}
$$

where the first inequality follow from the fact that $a(z,w)$ is increasing in z and from the fact that $dI(a)/da$ is positive but declining in a. The last inequality follows from the strict concavity of industry profits in w. ■

Proof of Proposition 6.

The proof consists in showing that some double marginalization is strictly preferred to no double marginalization and then invoking the concavity of the objective function shown in Lemma 3. Pledgeable assets at the end of period 0, (16) are a function of the realized demand and the agreed per unit wholesale price w. Dropping the scalar B and differentiating the objective function (15) with respect to w and evaluating at $w = c$, yields

$$E_x \left[\frac{dI(a(z,c))}{dw}\right] = \int_z \frac{dI(a(z,c))}{da} \frac{\partial a(z,c)}{\partial w} dG(z)$$

Consider the second term in the integral. Evaluating the term at the mean demand state, $z = 1$, yields

$$\frac{\partial a(1,c)}{\partial w} = \left\{\frac{d}{dw} q(w) \left[p(q(w)) - c\right]\right\}_{w=c} = 0,$$

where the second equality follows from the fact that industry profits are maximized when there is no double marginalization, $w = c$. The term $\partial a(z,c)/\partial w$ is strictly decreasing in z as

$$\frac{\partial a(z,c)}{\partial w \partial z} = \left\{\frac{d}{dw} \left[q(w) \left[p(q(w)) - w\right]\right]\right\}_{w=c} < 0$$

where the inequality follows as D's profits decline as the wholesale price increases. Hence,

$$\frac{\partial a(z,c)}{\partial w}\begin{cases} >0 & \text{if } z<1 \\ <0 & \text{if } z>1 \end{cases}$$

Next, we consider the first term in the integral. Recall that I increases at a rate greater than unity and is also concave in assets (Lemma 1). Thus $\frac{dI(a(z,c))}{da}$ is declining in z. We thus have

$$\begin{aligned} \frac{dI(a(z,c))}{dw} &= \int_{\underline{z}}^{1}\left\{\frac{dI(a(z,c))}{da}\frac{\partial a(z,c)}{\partial w}\right\}dG(z)+\int_{1}^{\bar{z}}\left\{\frac{dI(a(z,c))}{da}\frac{\partial a(z,c)}{\partial w}\right\}dG(z) \\ &> \int_{\underline{z}}^{1}\left\{\frac{dI(a(1,c))}{da}\frac{\partial a(z,c)}{\partial w}\right\}dG(z)+\int_{1}^{\bar{z}}\left\{\frac{dI(a(1,c))}{da}\frac{\partial a(z,c)}{\partial w}\right\}dG(z) \\ &= \frac{dI(a(1,c))}{da}E_z\left[\frac{\partial a(z,c)}{\partial w}\right] \\ &= 0 \end{aligned}$$

where the last equality follows from $E_z\left[\frac{\partial a(z,c)}{\partial w}\right]=\frac{\partial a(1,c)}{\partial w}=0$.This shows that, starting from $w=c$, the downstream firm can increase its profit by offering a larger wholesale price. As the objective function is concave (Lemma 3) we thus have $w^*>c$. The slotting fee follows from (14). ■

Proof of Lemma 4. Suppose w is set at $w^*(\theta_1)$ so that $E_z\left[\frac{dI}{da}(a,\theta_1)\cdot\frac{da}{dw}\right]=0$. Suppose also that $\frac{da}{dw}|_{w^*(\theta_1)}$ changes sign at $\tilde{z}$. Then, in the case of 1, we have

$$\begin{aligned} & E_z\left[\frac{dI}{da}(a,\theta_2)\cdot\frac{da}{dw}\right] \\ = & \int_{z=\underline{z}}^{\tilde{z}} x(a,\theta_2)\cdot\frac{dI}{da}(a,\theta_1)\underbrace{\frac{da}{dw}}_{+ve}dG(z)+\int_{z=\tilde{z}}^{\bar{z}} x(a,\theta_2)\cdot\frac{dI}{da}(a,\theta_1)\underbrace{\frac{da}{dw}}_{-ve}dG(z) \\ > & \int_{z=\underline{z}}^{\tilde{z}} x(a|_{\tilde{z}},\theta_2)\cdot\frac{dI}{da}(a,\theta_1)\frac{da}{dw}dG(z)+\int_{z=\tilde{z}}^{\bar{z}} x(a|_{\tilde{z}},\theta_2)\cdot\frac{dI}{da}(a,\theta_1)\frac{da}{dw}dG(z) \\ = & x(a|_{\tilde{z}},\theta_2)\cdot E_z\left[\frac{dI}{da}(a,\theta_1)\cdot\frac{da}{dw}\right]=0 \end{aligned}$$

The inequality follows as $\frac{dx}{da}<0$ yielding the result. Part 2 follows similarly. ■

Proof of Lemma 6. The investment function I is given by (18). The derivative of this with respect to r yields:

$$a=I+\left[B+1+r-\frac{\partial\pi}{\partial I}\right]\frac{dI}{dr} \tag{27}$$

We then have

$$0 \leq \frac{d}{dr}\left[\frac{\partial \pi}{\partial I} - r\right] = \frac{\partial^2 \pi}{\partial I^2}\frac{dI}{dr} - 1 \Leftrightarrow 1 \leq -\frac{\partial^2 \pi}{\partial I^2}\frac{I-a}{\left[B+1+r-\frac{\partial \pi}{\partial I}\right]}$$

Using (18) the requirement becomes

$$1+r \leq \left[-\frac{\partial^2 \pi (I)}{\partial I^2}\right] \cdot \left[\frac{\pi (I) - IB}{B+1+r-\frac{\partial \pi}{\partial I}}\right] \tag{28}$$

The condition that investment is always worthwhile and yet credit constrained implies $1 < \frac{\partial \pi}{\partial I} < 1+r+B$. Also

$$\pi (I) - IB \geq \pi (I) - I > (\pi (0) + I) - I = \pi (0)$$

Hence a sufficient condition for (28) to be satisfied is if

$$1+r \leq \left[-\frac{\partial^2 \pi (I)}{\partial I^2}\right] \cdot \frac{\pi (0)}{r+B} \Leftrightarrow \left[-\frac{\partial^2 \pi (I)}{\partial I^2}\right] \geq \frac{(B+r)(1+r)}{\pi (0)}$$

Yielding the result. ■

Proof of Proposition 7. Taking derivatives of the investment function I given in (18) allows us to establish the following identities, in addition to (27):

$$1+r = \left[B+1+r-\frac{\partial \pi}{\partial I}\right]\frac{dI}{da} \tag{29}$$

$$\left[B+1+r-\frac{\partial \pi}{\partial I}\right]^3 \frac{d^2 I}{dadr} = \left[B+1+r-\frac{\partial \pi}{\partial I}\right]^2 - (1+r)\left\{B+1+r-\frac{\partial \pi}{\partial I} + \frac{\partial^2 \pi}{\partial I^2}(I-a)\right\} \tag{30}$$

$$\left[B+1+r-\frac{\partial \pi}{\partial I}\right]\frac{d^2 I}{da^2} = \frac{\partial^2 \pi}{\partial I^2}\left[\frac{dI}{da}\right]^2 \tag{31}$$

Differentiating (30) again with respect to a yields

$$\begin{aligned} & \left[B+1+r-\frac{\partial \pi}{\partial I}\right]^3 \frac{d^3 I}{da^2 dr} - 3\left[B+1+r-\frac{\partial \pi}{\partial I}\right]^2 \frac{\partial^2 \pi}{\partial I^2}\frac{dI}{da}\frac{d^2 I}{dadr} \\ = & -2\left[B+1+r-\frac{\partial \pi}{\partial I}\right]\frac{\partial^2 \pi}{\partial I^2}\frac{dI}{da} - (1+r)\left\{-\frac{\partial^2 \pi}{\partial I^2}\frac{dI}{da} + \frac{\partial^3 \pi}{\partial I^3}\frac{dI}{da}(I-a) + \frac{\partial^2 \pi}{\partial I^2}\left(\frac{dI}{da}-1\right)\right\} \\ = & -(1+r)\left\{\frac{\partial^2 \pi}{\partial I^2} + \frac{\partial^3 \pi}{\partial I^3}\frac{dI}{da}(I-a)\right\} \end{aligned} \tag{32}$$

We now seek to apply Lemma 5. Therefore

$$\frac{dx}{da} =_{\text{sign}} \frac{dI}{da}\frac{d^3I}{da^2dr} - \frac{d^2I}{dadr}\frac{d^2I}{da^2}$$

Hence, using (31)

$$\begin{aligned}
\frac{dx}{da} &=_{\text{sign}} \left[B+1+r-\frac{\partial\pi}{\partial I}\right]\frac{d^3I}{da^2dr} - \frac{d^2I}{dadr}\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da} \\
&=_{\text{sign}} \left[B+1+r-\frac{\partial\pi}{\partial I}\right]^3\frac{d^3I}{da^2dr} - \left[B+1+r-\frac{\partial\pi}{\partial I}\right]^2\frac{d^2I}{dadr}\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da} \\
&= 2\left[B+1+r-\frac{\partial\pi}{\partial I}\right]^2\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da}\frac{d^2I}{dadr} - (1+r)\left\{\frac{\partial^2\pi}{\partial I^2}+\frac{\partial^3\pi}{\partial I^3}\frac{dI}{da}(I-a)\right\} \text{ from (32)}
\end{aligned}$$

We now apply (30) to yield:

$$\begin{aligned}
\frac{dx}{da} &=_{\text{sign}} 2\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da}\left\{\left[B+1+r-\frac{\partial\pi}{\partial I}\right]^2 - (1+r)\left\{B+1+r-\frac{\partial\pi}{\partial I}+\frac{\partial^2\pi}{\partial I^2}(I-a)\right\}\right\} \\
&\quad -(1+r)\left[B+1+r-\frac{\partial\pi}{\partial I}\right]\left\{\frac{\partial^2\pi}{\partial I^2}+\frac{\partial^3\pi}{\partial I^3}\frac{dI}{da}(I-a)\right\}
\end{aligned}$$

Using (29) we can simplify this to

$$\begin{aligned}
\frac{dx}{da} &=_{\text{sign}} \frac{\partial^2\pi}{\partial I^2}\left[B+1+r-\frac{\partial\pi}{\partial I}\right](1+r) - (1+r)^2\frac{\partial^3\pi}{\partial I^3}(I-a) \\
&\quad -2\frac{\partial^2\pi}{\partial I^2}(1+r)\left\{1+r+\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da}(I-a)\right\}
\end{aligned}$$

Note that the first and second terms are negative as $\frac{\partial^2\pi}{\partial I^2} < 0$ and $\frac{\partial^3\pi}{\partial I^3} \geq 0$.

We will have $\frac{dx}{da} < 0$ if

$$0 \geq 1+r+\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da}(I-a) = (1+r)\left[1-\frac{\partial^2\pi}{\partial I^2}\frac{dI}{dr}\right] = (1+r)\frac{d}{dr}\left[r-\frac{\partial\pi}{\partial I}\right]$$

which follows from Lemma 6. The full result then follows from Lemma 5.

Note that it also follows that at higher interest rates the expected level of investment declines. This is because investment levels fall for any realization of assets as $\frac{dI}{dr} < 0$, and further realised assets are lower for any realization of period 0 market demand due to the larger per unit input prices. ■

Proof of Lemma 7. The proof parallels that for Lemma 6 and therefore an abridged version is offered here. Note that $\frac{\partial}{\partial B}\left[\frac{\partial \pi}{\partial I}-B\right]=\pi''(I)\frac{\partial I}{\partial B}-1$. From (2)

$$\frac{\partial I}{\partial B}=\frac{-I}{1+B-\pi'(I)} \tag{33}$$

Therefore the result holds if $[-\pi''(I)]>\frac{-1}{\frac{\partial I}{\partial B}}=\frac{1+B-\pi'(I)}{I}$. As $\pi'(I)>1$ and $I=\frac{\pi(I)-I+a}{B}>\frac{\pi(0)+I-I+a}{B}>\frac{\pi(0)}{B}$ we have the result. ■

Proof of Proposition 8. We proceed as with the proof of Proposition 7 and so establish the following identities, in addition to (33):

$$\left[1+B-\frac{\partial \pi}{\partial I}\right]\frac{\partial I}{\partial a} = 1 \tag{34}$$

$$\left[1+B-\frac{\partial \pi}{\partial I}\right]^3\frac{\partial^2 I}{\partial a\partial B} = -\left[1+B-\frac{\partial \pi}{\partial I}+I\frac{\partial^2 \pi}{\partial I^2}\right] \tag{35}$$

$$\left[1+B-\frac{\partial \pi}{\partial I}\right]^3\frac{\partial^3 I}{\partial a^2\partial B}-3\left[1+B-\frac{\partial \pi}{\partial I}\right]^2\frac{\partial^2 \pi}{\partial I^2}\frac{\partial I}{\partial a}\frac{\partial^2 I}{\partial a\partial B} = -\frac{\partial I}{\partial a}I\frac{\partial^3 \pi}{\partial I^3} \tag{36}$$

$$\left[1+B-\frac{\partial \pi}{\partial I}\right]\frac{\partial^2 I}{\partial a^2} = \frac{\partial^2 \pi}{\partial I^2}\left[\frac{\partial I}{\partial a}\right]^2 \tag{37}$$

We now apply Lemma 5 by noting that

$$\begin{aligned}
&\frac{dI}{da}\frac{d^3 I}{da^2 dB}-\frac{d^2 I}{da dB}\frac{d^2 I}{da^2}\\
&\underset{\substack{\text{sign}\\(34)\text{ and }(37)}}{=}\left[1+B-\frac{\partial \pi}{\partial I}\right]^3\frac{\partial^3 I}{\partial a^2\partial B}-\left[1+B-\frac{\partial \pi}{\partial I}\right]^2\frac{\partial^2 \pi}{\partial I^2}\frac{\partial I}{\partial a}\frac{\partial^2 I}{\partial a\partial B}\\
&\underset{\substack{\text{sign}\\(35)\text{ and }(36)}}{=}2\frac{\partial^2 \pi}{\partial I^2}\left\{-1+\frac{\partial^2 \pi}{\partial I^2}\frac{\partial I}{\partial B}\right\}-I\frac{\partial^3 \pi}{\partial I^3}
\end{aligned}$$

This is negative overall as $\frac{\partial^2 \pi}{\partial I^2}<0$, $\frac{\partial^3 \pi}{\partial I^3}\geq 0$ and Lemma 7 applies. The result then follows from Lemma 5. Finally consider the investment level. D's objective function is given by maximising $E_z B\cdot I$. Dropping the scalar B, and setting $B_2>B_1$ we have

$$E_z I\left(a\left(w^*\left(B_2\right),z\right);B_2\right)\leq E_z I\left(a\left(w^*\left(B_1\right),z\right);B_2\right)\leq E_z I\left(a\left(w^*\left(B_1\right),z\right);B_1\right)$$

The first inequality follows as $w^*(B_2)$ is the optimal per unit price to use when corporate governance is characterised by B_2 and this is lower if corporate governance is characterised by B_1. The second inequality follows as $\frac{dI}{dB}<0$ from (33). ■

Proof of Proposition 9. The investment function is defined by

$$\pi\left(I;\theta\right)-\left(I-a\right)=BI$$

The proof proceeds by seeking to sign the derivatives of I and then invoking Lemma 5. We will repeatedly use the fact that $B+1-\frac{\partial\pi}{\partial I}>0$ which is a corollary of the requirement that $\frac{dI}{da}>1$ and is given in (3). The first derivatives of the I function are

$$1 = \left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{dI}{da} \tag{38}$$

$$\frac{\partial\pi}{\partial\theta} = \left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{dI}{d\theta} \tag{39}$$

which implies that $\frac{dI}{d\theta}>0$. Differentiating (38) with respect to θ yields

$$\left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{d^2I}{dad\theta}=\frac{dI}{da}\left[\frac{\partial^2\pi}{\partial I\partial\theta}+\frac{\partial^2\pi}{\partial I^2}\frac{dI}{d\theta}\right] \tag{40}$$

Using (39) we can rewrite the above equation as

$$\left[B+1-\frac{\partial\pi}{\partial I}\right]^3\frac{d^2I}{dad\theta}=\left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{\partial^2\pi}{\partial I\partial\theta}+\frac{\partial^2\pi}{\partial I^2}\frac{\partial\pi}{\partial\theta}$$

And now differentiating with respect to a yields

$$\begin{aligned}&\left[B+1-\frac{\partial\pi}{\partial I}\right]^3\frac{d^3I}{da^2d\theta}-3\left[B+1-\frac{\partial\pi}{\partial I}\right]^2\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da}\frac{d^2I}{dad\theta}\\ =\;& -\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da}\frac{\partial^2\pi}{\partial I\partial\theta}+\left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{\partial^3\pi}{\partial I^2\partial\theta}\frac{dI}{da}+\frac{\partial^3\pi}{\partial I^3}\frac{dI}{da}\frac{\partial\pi}{\partial\theta}+\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da}\frac{\partial^2\pi}{\partial I\partial\theta}\end{aligned} \tag{41}$$

We now seek to apply Lemma 5. Therefore

$$\frac{dx}{da}=_{\text{sign}}\frac{dI}{da}\frac{d^3I}{da^2d\theta}-\frac{d^2I}{dad\theta}\frac{d^2I}{da^2}$$

We can simplify this by differentiating (38) with respect to a to derive

$$\left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{d^2I}{da^2}=\frac{\partial^2\pi}{\partial I^2}\left[\frac{dI}{da}\right]^2$$

Hence

$$\begin{aligned}
\frac{dx}{da} &=_{\text{sign}} \left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{d^3 I}{da^2 d\theta}-\frac{d^2 I}{dad\theta}\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da} \\
&=_{\text{sign}} \left[B+1-\frac{\partial\pi}{\partial I}\right]^3\frac{d^3 I}{da^2 d\theta}-\left[B+1-\frac{\partial\pi}{\partial I}\right]^2\frac{d^2 I}{dad\theta}\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da} \\
&= 2\left[B+1-\frac{\partial\pi}{\partial I}\right]^2\frac{\partial^2\pi}{\partial I^2}\frac{dI}{da}\frac{d^2 I}{dad\theta}+\left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{\partial^3\pi}{\partial I^2\partial\theta}\frac{dI}{da}+\frac{\partial^3\pi}{\partial I^3}\frac{dI}{da}\frac{\partial\pi}{\partial\theta}\text{ from (41)} \\
&=_{\text{sign}} 2\frac{\partial^2\pi}{\partial I^2}\left\{\frac{\partial^2\pi}{\partial I\partial\theta}+\frac{\frac{\partial^2\pi}{\partial I^2}\frac{\partial\pi}{\partial\theta}}{\left[B+1-\frac{\partial\pi}{\partial I}\right]}\right\}+\left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{\partial^3\pi}{\partial I^2\partial\theta}+\frac{\partial^3\pi}{\partial I^3}\frac{\partial\pi}{\partial\theta}\text{ from (40)} \\
&= \frac{\partial\pi}{\partial\theta}\left\{\frac{2\left[\frac{\partial^2\pi}{\partial I^2}\right]^2}{\left[B+1-\frac{\partial\pi}{\partial I}\right]}+\frac{\partial^3\pi}{\partial I^3}\right\}+2\frac{\partial^2\pi}{\partial I^2}\frac{\partial^2\pi}{\partial I\partial\theta}+\left[B+1-\frac{\partial\pi}{\partial I}\right]\frac{\partial^3\pi}{\partial I^2\partial\theta}
\end{aligned}$$

As $\frac{\partial\pi}{\partial\theta}>0$ and $\frac{\partial^3\pi}{\partial I^3}\geq 0$ and $\frac{\partial^2\pi}{\partial I\partial\theta}\leq 0$ and $\frac{\partial^3\pi}{\partial I^2\partial\theta}\geq 0$ then $\frac{dx}{da}>0$ and so optimal w decreases by Lemma 5 yielding the result.

Finally consider the investment level. D's objective function is given by maximising $E_z B\cdot I$. Dropping the scalar B we have

$$E_z I\left(a\left(w^*\left(\theta_2\right),z\right);\theta_2\right)\geq E_z I\left(a\left(w^*\left(\theta_1\right),z\right);\theta_2\right)\geq E_z I\left(a\left(w^*\left(\theta_1\right),z\right);\theta_1\right)$$

The first inequality follows as $w^*\left(\theta_2\right)$ is the optimal per unit price to use when the investment technology is parametrised by θ_2. The second inequality follows as $\frac{dI}{d\theta}>0$ from (39). ■

Proof of Proposition 11. For part 1 note that the downstream firm optimizes B times by the investment level as she is credit constrained (15). Therefore

$$E_H\left[I\left(a\left(z,w_H^*\right)\right)\right]\geq E_H\left[I\left(a\left(z,w_G^*\right)\right)\right] \tag{42}$$

We now seek to compare the expectation of $I\left(a\left(z,w_G^*\right)\right)$ under distribution H and G.

Note that for any y

$$\begin{aligned}
\Pr\left[I\left(a\left(z_H, w_G^*\right)\right) > y\right] &= \Pr\left[a\left(z_H, w_G^*\right) > I^{-1}\left(y\right)\right] \\
&= \Pr\left[z_H > \frac{I^{-1}\left(y\right) - E\left[z_H\right] \cdot q\left(w_G^*\right)\left(w_G^* - c\right)}{q\left(w_G^*\right)\left[p\left(q\left(w_G^*\right)\right) - w_G^*\right]}\right] \text{ using (16)} \\
&\geq \Pr\left[z_H > \frac{I^{-1}\left(y\right) - q\left(w_G^*\right)\left(w_G^* - c\right)}{q\left(w_G^*\right)\left[p\left(q\left(w_G^*\right)\right) - w_G^*\right]}\right] \text{ as } E\left[z_H\right] > 1 \\
&\geq \Pr\left[z_G > \frac{I^{-1}\left(y\right) - q\left(w_G^*\right)\left(w_G^* - c\right)}{q\left(w_G^*\right)\left[p\left(q\left(w_G^*\right)\right) - w_G^*\right]}\right] \text{ as } H \succ_{\text{FOSD}} G \\
&= \Pr\left[I\left(a\left(z_G, w_G^*\right)\right) > y\right]
\end{aligned}$$

As this is true for any y the variable $I\left(a\left(z, w_G^*\right)\right)$ under H first order stochastically dominates $I\left(a\left(z, w_G^*\right)\right)$ under G and so

$$E_H\left[I\left(a\left(z, w_G^*\right)\right)\right] \geq E_G\left[I\left(a\left(z, w_G^*\right)\right)\right] \tag{43}$$

Combining (42) and (43) yields the result.

For part 2 we now suppose that H can be formed from G by repeatedly taking ε probability mass from some point $z^\dagger \in [\underline{z}, \tilde{z}]$ and moving it to a higher realization (to some $z > z^\dagger$). Such sweeps of the density function will satisfy (21). The proof consists in showing that $E_z\left[\frac{dI(a(z,w))}{da}\frac{\partial a(z,w)}{\partial w}\right]$ declines under such a movement of probability mass.

Suppose first that G is altered by taking an ε of mass from point $z^\dagger \in [\underline{z}, \tilde{z}]$ and moving it to some $z \in (z^\dagger, \tilde{z}]$. We have

$$\frac{\partial}{\partial z}\left[\frac{dI\left(a\left(z,w\right)\right)}{da}\frac{\partial a\left(z,w\right)}{\partial w}\right] = \frac{d^2 I}{da^2}\frac{\partial a}{\partial z}\frac{\partial a}{\partial w} + \frac{dI}{da}\frac{\partial^2 a}{\partial w \partial z}$$

The second term is negative as $\frac{\partial^2 a}{\partial w \partial z} < 0$ as D's profits decline in the wholesale price. The first term is negative as $\frac{\partial a}{\partial w} > 0$ for $z \leq \tilde{z}$ by definition. So the integrand in $E_z\left[\frac{dI(a(z,w))}{da}\frac{\partial a(z,w)}{\partial w}\right]$ declines in z for $z \leq \tilde{z}$ and so this movement of mass must lower the expectation.

Suppose instead we move mass from point $z^\dagger \in [\underline{z}, \tilde{z}]$ to some $z > \tilde{z}$. Now note that

$$E_z\left[\frac{dI}{da}\frac{\partial a}{\partial w}\right] = \underbrace{\int_{\underline{z}}^{\tilde{z}} \frac{dI}{da}\frac{\partial a}{\partial w} dH\left(z\right)}_{(i)} + \underbrace{\int_{\tilde{z}}^{\bar{z}} \frac{dI}{da}\frac{\partial a}{\partial w} dH\left(z\right)}_{(ii)}$$

Term (i) was positive as $\left.\frac{\partial a}{\partial w}\right|_{[\underline{z},\tilde{z}]} > 0$ and the weight here is being reduced by the movement of mass. Term (ii) is negative as $\frac{\partial a}{\partial w}\Big|_{(\tilde{z},\bar{z}]} < 0$ and the weight here is being increased by the movement of mass. Therefore combining the movement of mass lowers $E_z\left[\frac{dI(a(z,w))}{da}\frac{\partial a(z,w)}{\partial w}\right]$.

Hence we must have

$$E_{H(z)}\left[\frac{dI}{da}\frac{\partial a}{\partial w}\right]_{w_G^*} < E_{G(z)}\left[\frac{dI}{da}\frac{\partial a}{\partial w}\right]_{w_G^*} = 0$$

and so we have $w_H^* < w_G^*$ as required. ■

Proof of Proposition 12. For the first result note that the downstream firm optimizes B times by the investment level as she is credit constrained (15). Therefore

$$E_{K'}\left[I\left(a\left(z, w_{K'}^*\right)\right)\right] \geq E_{K'}\left[I\left(a\left(z, w_K^*\right)\right)\right]$$

As $G_{K'}$ is a mean preserving contraction of G_K we will have

$$E_{K'}\left[I\left(a\left(z, w_K^*\right)\right)\right] \geq E_K\left[I\left(a\left(z, w_K^*\right)\right)\right]$$

if $I\left(a\left(z, w_K^*\right)\right)$ is a concave function of z. This is so as

$$\frac{d^2 I\left(a\left(z,w\right)\right)}{dz^2} = \underbrace{\frac{d^2 I}{da^2}}_{<0 \text{ from Lemma 1}}\left[\frac{da}{dz}\right]^2 + \frac{dI}{da}\underbrace{\frac{d^2 a}{dz^2}}_{=0 \text{ from (16)}} < 0$$

■

We now turn to the second result. This follows if $\lim_{K\to\infty} w_K^* = c$. Proposition **??** confirms that From Proposition 6, we know that, for any D, the equilibrium contract involves double marginalization. Recall that the equilibrium wholesale price satisfies

$$w^* = \arg\max_w E_{g_K}\left[I\left(a\left(z,w\right)\right)\right]$$

Since $I\left(a\left(z,w\right)\right)$ is strictly concave in w (Lemma 3), w^* is uniquely determined by the first-order condition, and so

$$\frac{dE_z I(a\left(z,w\right))}{dw} = \int_{\underline{z}}^{\overline{z}} \frac{dI(a\left(z,w\right))}{da}\frac{da\left(z,w\right)}{dw} dG_K\left(z\right) < 0 \text{ if and only if } w > w_K^*$$

To prove that $\lim_{K\to\infty} w_K^* = c$ it thus suffices to show that $\left[dE_z I/dw\right]_{\tilde{w}} < 0$ for any fixed $\tilde{w} > c$ and K sufficiently large. Since industry profits are strictly concave in w, attaining a maximum at $w = c$ (Assumption 2), we have $\frac{\partial a(1,\tilde{w})}{\partial w} < 0$. Moreover, since D's profits before the fixed fee decline in the wholesale price, $\frac{\partial^2 a(z,\tilde{w})}{\partial z \partial w} < 0$ the $\frac{da(z,w)}{dw}$ term is decreasing in z. Hence, either $\frac{da(z,\tilde{w})}{dw} < 0$ for all $z > \underline{z}$, in which case $dE_z I(a(z, \widetilde{w}))/dw < 0$, or else there exists a unique $\widehat{z} \in (\underline{z}, 1)$ such that $\frac{da(\hat{z},\tilde{w})}{dw} = 0$. In the latter case,

$$\frac{da(z, \tilde{w})}{dw} \begin{cases} > 0 & \text{if } z < \widehat{z}, \\ < 0 & \text{if } z > \widehat{z}. \end{cases}$$

We can then split the derivative $dE_z I(a(z, \widetilde{w}))/dw$ into the sum of positive and negative terms:

$$\begin{aligned} & \frac{dE_z I(a(z, \tilde{w}))}{dw} \\ = & \int_{\underline{z}}^{\widehat{z}} \frac{dI(a(z, \tilde{w}))}{da} \frac{da(z, \tilde{w})}{dw} dG_K(z) + \int_{\widehat{z}}^{\bar{z}} \frac{dI(a(z, \tilde{w}))}{da} \frac{da(z, \tilde{w})}{dw} dG_K(z) \\ < & \left\{ \frac{dI(a(z, \tilde{w}))}{da} \frac{da(z, \tilde{w})}{dw} \right\}_{\underline{z}} G_K(\widehat{z}) + \int_{\widehat{z}}^{\bar{z}} \frac{da(z, \tilde{w})}{dw} dG_K(z) \\ < & \left\{ \frac{dI(a(z, \tilde{w}))}{da} \frac{da(z, \tilde{w})}{dw} \right\}_{\underline{z}} G_K(\widehat{z}) + \frac{da(1, \tilde{w})}{dw} [\Pr(z_K \geq 1)] \end{aligned}$$

where the first inequality follows as by the strict concavity of $I(a)$ and the fact that $\frac{\partial^2 a(z,\tilde{w})}{\partial z \partial w} < 0$ the brace is maximized at the smallest possible demand realization $\underline{z}$. For the second term the fact that $dI/da > 1$ (Lemma 1) has been used. The second inequality follows from the fact that in the region $[\hat{z}, \bar{z}]$, $\frac{da(z,\tilde{w})}{dw}$ is negative and declining in z. As K increases, the density of z undergoes a mean-preserving contraction around its mean of $1 > \widehat{z}$, and so the first term on the RHS declines and vanishes in the limit as $K \to \infty$. As to the second term, it is negative and converges to $\left[\frac{d}{dw} q(w) \left[p(q(w)) - c\right]\right]_{\tilde{w}} < 0$. Hence, for sufficiently large K, $dE_z I(a(z, \widetilde{w}))/dw < 0$.

B Upstream Risk Analysis

This appendix proves Proposition 13. Is is helpful to confirm that the optimal contract can be found by only considering local incentive compatibility constraints. Hence we note that U's program is equivalent to the following:

Program U'

$$\max_{\{Q_i,W_i\}} \sum_{i=1}^{n} h_i B \cdot I\left(W_i - Q_i\kappa_i\right)$$

$$\text{subject to D accepting} \ : \ \sum_{i=1}^{n} h_i \left\{Q_i p\left(Q_i\right) - W_i\right\} = 0$$

and

$$W_i - Q_i\kappa_i \geq W_{i+1} - Q_{i+1}\kappa_i \text{ for all } i \in \{1, \ldots, n-1\} \tag{44}$$

$$Q_i \geq Q_{i+1} \text{ for all } i \in \{1, \ldots, n-1\} \tag{45}$$

This program differs from program U by the reduction of the global incentive compatibility conditions to two local ones. It is clear that (24) implies (44). It also implies (45) as we have, from (24)

$$W_i - Q_i\kappa_i \geq W_{i+1} - Q_{i+1}\kappa_i \text{ and } W_{i+1} - Q_{i+1}\kappa_{i+1} \geq W_i - Q_i\kappa_{i+1}$$

Summing and simplifying yields (45).

Therefore a solution of the original program (U) is also a solution of program U'. The converse remains to be shown.

Lemma 8 *(44) must hold with equality at an optimal solution to program* U'.

Proof. Suppose not so that $W_i - Q_i\kappa_i > W_{i+1} - Q_{i+1}\kappa_i$. Consider lowering W_i to $W_i - a$ and raising W_{i+1} to $W_{i+1} + b$. (44) remains satisfied if a and b are small. (45) is unaffected. D's IR constraint is unaffected if $ah_i = bh_{i+1}$But U's objective function changes by

$$\begin{aligned} & B\left\{-h_i I'\left(W_i - Q_i\kappa_i\right) a + h_{i+1} I'\left(W_{i+1} - Q_{i+1}\kappa_{i+1}\right) b\right\} \\ =_{\text{sign}} & I'\left(W_{i+1} - Q_{i+1}\kappa_{i+1}\right) - I'\left(W_i - Q_i\kappa_i\right) > 0 \end{aligned}$$

as I is concave increasing and $W_i - Q_i\kappa_i > W_{i+1} - Q_{i+1}\kappa_i \geq W_{i+1} - Q_{i+1}\kappa_{i+1}$. A contradiction to the optimality of the contract. ■

Using Lemma 8 we can demonstrate that a solution to program U' satisfies program U. This follows if (44) and (45) imply (24).

To show this consider the downwards inequality:

$$\left[W_{i+1} - Q_{i+1}\kappa_{i+1}\right] - \left[W_i - Q_i\kappa_{i+1}\right] =_{\text{Lemma 8}} \left(Q_i - Q_{i+1}\right)\kappa_{i+1} - \left(Q_i - Q_{i+1}\right)\kappa_i \geq_{\text{Eq (45)}} 0$$

as required. To extend the inequalities along successive states consider the derivative with respect to κ of (44) and, noting (45), we have:

$$[W_i - Q_i\kappa_{i-1}] - [W_{i+1} - Q_{i+1}\kappa_{i-1}] \geq 0$$

(44) implies that $W_{i-1} - Q_{i-1}\kappa_{i-1} \geq W_i - Q_i\kappa_{i-1}$. Combining yields that when the truth is $i-1$ truthful reporting dominates reporting i or $i+1$. Similar arguments extend the inequality throughout the line. Hence we have shown that Program U and Program U' are equivalent.

Proof of Proposition 13. We first show that $Q_i \leq q(\kappa_i)$ for all i. Suppose instead that there exists a given i such that

$$q(\kappa_i) < Q_i \Rightarrow \frac{d}{dQ}\{Q_i p(Q_i)\} < \kappa_i$$

and suppose further that $Q_i > Q_{i+1}$. Consider lowering Q_i by ε and lowering W_i by

$$\varepsilon \cdot \max\left(\frac{d}{dQ}\{Q_i p(Q_i)\}, \kappa_{i-1}\right)$$

which is less than $\varepsilon\kappa_i$. D will accept this change as her expected payoff changes by

$$h_i\left\{-\varepsilon\frac{d}{dQ}\{Q_i p(Q_i)\} + \varepsilon \cdot \max\left(\frac{d}{dQ}\{Q_i p(Q_i)\}, \kappa_{i-1}\right)\right\} \geq 0$$

For small ε (45) holds at i and $i-1$ as $Q_i > Q_{i+1}$. Next consider the objective function. This changes by

$$h_i BI'(W_i - Q_i\kappa_i) \cdot \left\{\varepsilon\kappa_i - \varepsilon \cdot \max\left(\frac{d}{dQ}\{Q_i p(Q_i)\}, \kappa_{i-1}\right)\right\} > 0$$

Thus U benefits. Next note that this also implies that (44) holds at i as the left hand side has increased. Finally note that (44) holds at $i-1$ as $W_i - Q_i\kappa_{i-1}$ is reduced for it is altered by the amount:

$$-\varepsilon \cdot \max\left(\frac{d}{dQ}\{Q_i p(Q_i)\}, \kappa_{i-1}\right) + \varepsilon\kappa_{i-1}$$

This is a contradiction and so if $Q_i > Q_{i+1}$ then we have $Q_i \leq q(\kappa_i)$.

If $Q_i = Q_{i+1} > Q_{i+2}$ then we have

$$\frac{d}{dQ}\{Q_i p(Q_i)\} = \frac{d}{dQ}\{Q_{i+1} p(Q_{i+1})\} \geq \kappa_{i+1} > \kappa_i$$

This argument can be extended all along the line and proves result 2 of Proposition 13.

At κ_1 we have $\frac{d}{dQ}\{Q_1 p(Q_1)\} = \kappa_1$ as the argument above can be used to raise Q_1, altering W_1 to keep D indifferent, and maximising the profit in this state. Due to the direction of the inequalities in Program U' this doesn't alter any of the other relevant constraints.

As positive quantities are sold result 3 follows as from (24):

$$W_i - Q_i\kappa_i \geq W_{i+1} - Q_{i+1}\kappa_i > W_{i+1} - Q_{i+1}\kappa_{i+1} \tag{46}$$

using the fact that $\kappa_{i+1} > \kappa_i$.

For result 4 we have

$$\begin{aligned} & [Q_i p(Q_i) - W_i] - [Q_{i+1} p(Q_{i+1}) - W_{i+1}] \\ =_{\text{Lemma 8}} & [Q_i p(Q_i) - Q_i\kappa_i] - [Q_{i+1} p(Q_{i+1}) - Q_{i+1}\kappa_i] \end{aligned}$$

And $q(\kappa_i) \geq Q_i > Q_{i+1}$ which implies that if costs are κ_i then profits are higher with volume Q_i than Q_{i+1} so the above expression is positive as required. ■

References

[1] Balke and Emery (1994), "Understanding the Price Puzzle", Federal Reserve Bank of Dallas Economic Review, 4th Quarter, 15-26

[2] Bloom, Gundlach and Cannon (2000), "Slotting Allowances and Fees: Schools of Thought and the Views of Practicing Managers", Journal of Marketing, 92-108

[3] Carlton (1979), "Vertical Integration in Competitive Markets Under Uncertainty", Journal of Industrial Economics, 189-209.

[4] Carlton (1977), "Uncertainty, Production Lags, and Pricing", The American Economic Review, 67, 244-249.

[5] Christiano, Eichenbaum and Evans (1999), "Monetary Policy Shocks: What Have We Learnt And To What End?", Handbook of Macroeconomics. Volume 1A. Taylor and Woodford eds. North Holland Press.

[6] Froot, Scharfstein and Stein (1993), "Risk Management: Coordinating Corporate Investment and Financing Policies", Journal of Finance, 1629-1658.

[7] Gertler and Gilchrist (1994), "Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms", Quarterly Journal of Economics, 309-340

[8] Harrigan (1985), "Vertical Integration and Corporate Strategy", Academy of Management Journal, 397-425.

[9] Hubbard (1998), "Capital-Market Imperfections and Investment", Journal of Economic Literature, 193-225.

[10] Hart, Oliver D. (1983), "Labour Contracts under Asymmetric Information: An Introduction," Review of Economic Studies, vol. 50(1), 3-35.

[11] Holmstrom and Tirole (1997), "Financial Intermediation, Loanable Funds, and the Real Sector", Quarterly Journal of Economics, 663-691.

[12] Hubbard and Palia (1999), "A Reexamination of the Conglomerate Merger Wave in the 1960s: An Internal Capital Markets View", Journal of Finance, 1131-1152

[13] Kalai (1977), "Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons", Econometrica, 45, 1623-1630.

[14] Klein and Wright (2007), "The Economics of Slotting Contracts", Journal of Law and Economics, 421-455.

[15] Mahoney (1992), "The Choice of Organizational Form: Vertical Financial Ownership Versus Other Methods of Vertical Integration" Strategic Management Journal, 559-584.

[16] Peterson and Rajan (1997), "Trade Credit: Theories and Evidence", The Review of Financial Studies, 661-691.

[17] Shin and Stulz (1998), "Are Internal Capital Markets Efficient?", The Quarterly Journal of Economics, 531-552

[18] Sudhir and Rao (forthcoming), "Are Slotting Allowances Efficiency-Enhancing or Anti-Competitive?", Journal of Marketing Research.

[19] Sutcliffe and Zaheer (1988), "Uncertainty in the Transaction Environment: An Empirical Test", Strategic Management Journal, 1-23