Vertical Relations Under Credit Constraints*

Volker Nocke† John Thanassoulis‡
University of Mannheim University of Oxford

December 31, 2011

Abstract

We model the impact credit constraints and market risk have on the vertical relationships between firms in the supply chain. Firms which might face credit constraints in future investments become endogenously risk averse when accumulating pledgeable assets. In the short run, the optimal supply contract therefore involves risk sharing, so inducing double marginalization. Credit constraints thus result in higher retail prices. This is true whether the firm is debt or equity funded. Further, there is an intrinsic complementarity between supply and lending which reduces financing inefficiencies created by informational asymmetries. This provides a new theory of finance arms of major suppliers. Finally, the model offers a concise explanation for several empirical regularities of firm behavior: a theory of countervailing power based on credit constraints; a monetary transmission mechanism linking the cost of borrowing with short-run retail prices; and a motive for outsourcing supply (or distribution) in the face of market risk.

Keywords: risk aversion; vertical contracting; double marginalization; outsourcing; market risk; risk sharing; financial companies; finance arms; monetary transmission mechanism; countervailing power.

JEL Classification: L14, L16.

---

*We would like to thank the editor George-Marios Angeletos and three anonymous referees who have allowed us to greatly improve this paper. We would also like to thank Tracy Lewis, Meg Meyer, Patrick Rey, Jean Tirole, Mike Whinston, and seminar participants at CEMFI, Columbia, Cologne, DIW Berlin, Frankfurt, Leuven, Naples, Northwestern, Oxford, Paris School of Economics, Strathclyde, Texas-Austin, UCLA, UPF, Warwick, the 2009 CEPR Applied IO Conference in Mannheim, the 2009 EARIE Conference in Ljubljana, and the 2009 Duke-Northwestern-Texas IO Theory Conference in Evanston. We are also grateful to the ESRC for financial support (grant RES-000-22-3468).

†Other affiliations: CEPR and CESifo. Email: nocke@uni-mannheim.de.

‡Email: john.thanassoulis@economics.ox.ac.uk.

1 Introduction

Credit constraints have been known to be a part of corporate reality for decades (Hubbard, 1998, and references therein). Massively reduced access to credit has been a feature of the major financial crisis of recent years. It is also well known that firms are subject to substantial market risk – whether on the demand side or supply side. Incorporating corporate finance aspects into an industrial organization model of the vertical supply chain, we study the interaction between credit constraints and market risk, and their effects on short-run retail pricing, long-run investment, and welfare. We show that credit constraints and market risk impact optimal vertical contracting, resulting in higher prices through double marginalization and slotting fees. We show that combining supply with lending can reduce financing frictions and so provide a new theory of finance arms. Further, the model gives rise to a novel theory of countervailing power based on credit constraints, and provides a rationale for outsourcing. Finally we identify a new transmission mechanism from interest rates to the real economy which acts via firms which are at risk of becoming credit constrained.

Consider a vertical supply chain consisting of a single upstream firm ("he") supplying a single downstream firm ("she"), and exposed to demand-side risk. The joint-profit maximizing supply contract would involve per unit input prices at the upstream firm's marginal cost, irrespective of any demand-side risk. But now suppose the downstream firm has some future investment opportunities. The size of the loan she is able to raise to fund the investment, and therefore the actual investment level, depend on the size of the pledgeable assets the firm owns. Under the standard assumption that investment is subject to diminishing marginal returns, when pledgeable assets are low, the induced investment level is low as well. This implies that the return on a marginal dollar of investment would be high, and so an extra marginal dollar of pledgeable assets can be greatly levered through the banking sector. Hence we show that the profit-maximizing firm becomes endogenously risk averse when accumulating pledgeable assets.

As a result, the optimal contract between the downstream firm and its upstream supplier involves risk sharing and double marginalization. The endogenously risk-averse downstream firm wants to insure her level of pledgeable assets. So she demands a risk-sharing contract in which the supplier bears some loss for poor demand realizations. For the supplier to recoup these potential losses, he requires payments in high demand states to grow at a rate faster than cost. Hence, double marginalization is introduced, causing the retail price of the downstream firm to rise. The cost of the insurance made necessary by the credit constraints is in this sense partly paid for by final consumers.

The optimal supply contract can be thought of as involving a fixed payment from the upstream to the downstream firm and demand-dependent repayments. This may help explain the increasingly common use of "slotting fees" in the grocery market as well as in other industries such as software and publishing. These fees are fixed payments many retailers require of manufacturers in return for stocking their products. Empirical evidence suggests that an important part of the story is the sharing of risk (Sudhir and Rao, 2006; White et al., 2000), which accords with our model.[1]

These results apply whether or not the firm carries debt. The presence of debt can make equity more tolerant of risk as losses fall disproportionately on debt holders (see Jensen and Meckling, 1976, and a long literature which follows). However, conditional on the firm surviving the short term risk and being able to make investments for the future the endogenous risk aversion remains. The marginal value of extra pledgeable assets is greatest in the worst states – conditional on the firm surviving. The optimal supply contract involves risk sharing in which the downstream firm seeks to increase her profits conditional on survival: thus high retail prices remain.

It is standard to see the input suppliers and the banking sector as two completely separate industries. However, if the input supplier also provides profit insurance, as in our model, it is no longer clear whether such a separation is indeed optimal. In fact, we demonstrate that there exists an intrinsic complementarity between the provision of insurance and lending. An input supplier with access to funds at the same rate as the banking sector could actually lend on rates that the independent banking sector would find unprofitable. This result offers an original insight into the existence and profitability of finance arms of major companies such as GE and Cisco. As financial companies lend almost \$1 for every \$2 lent by a mainstream bank, gaining an insight into what makes financial companies effective competitors to banks therefore seems a first-order issue.

The complementarity we find between supply insurance and lending arises because of the countervailing incentives the downstream firm faces when dealing with the insurer and the lender. By pooling insurance and lending, the downstream firm can effectively reduce her temptation to misreport the demand state. This reduces the informational inefficiencies which hinder financing and so allows for larger investments, for less double marginalization, and therefore for higher profits.

As the pre-investment degree of risk aversion of the downstream firm is endogenous, it is a function of market-level and firm-level parameters such as the interest rate, the quality of corporate governance, the firm's asset endowment and her bargaining power in the vertical chain. We demonstrate that if parameters change so as to increase the coefficient of absolute risk aversion with respect to pledgeable assets, then retail prices will rise in the short run (and vice-versa). We use this result to demonstrate that firms with greater exogenous assets ("asset-rich firms") are less risk averse generating the prediction that there should be less double-marginalization amongst asset-rich downstream firms.

Relaxing our assumption that the downstream firm has all of the bargaining power, we

[1] Theoretical explanations for this practice have portrayed the slotting fee as a signalling device (Klein and Wright, 2007, and references therein).

show that an increase in the downstream firm's bargaining power *vis-à-vis* her supplier makes the firm less risk averse when accumulating pledgeable assets. Hence, a more powerful downstream firm charges lower retail prices in the short run. The model therefore gives rise to a new theory of *countervailing power* (Galbraith, 1952) that is based on credit constraints. We also show that an increase in the interest rate that the downstream firm has to pay to finance her investment makes the firm more risk averse when accumulating pledgeable assets. Hence, a higher interest rate leads to higher retail prices in the short run.

We finally demonstrate a link between market risk and outsourcing. A credit-constrained downstream firm cannot insure herself. By outsourcing input supply, however, the downstream firm can purchase insurance as the upstream supplier is in a unique position to monitor the volumes supplied to the downstream firm. Our result is supported by empirical evidence (Harrigan, 1985; Sutcliffe and Zaheer, 1988) which points in this direction.[2]

Related Literature.

Our paper builds on some existing insights from the industrial organization and corporate finance literatures. On the corporate finance side, we build on Holmstrom and Tirole (1997) in modelling credit constraints as an endogenous outcome, caused by a moral hazard problem associated with the firm's investment project. In contrast to Holmstrom and Tirole, however, we assume that the firm's investment project has decreasing returns. It is this (standard) decreasing-returns assumption that implies that the rate at which the marginal dollar can be leveraged is decreasing. A related issue is studied by Froot et al. (1993). Froot et al.'s core model supposes that external finance is more costly than internally generated funds. If so then a firm would seek to avoid needing to access external funds in the face of market risk and so would hedge financial variables correlated with this risk. Thus capital market imperfections can generate risk-averse behavior. Our contribution here is to demonstrate that firm risk aversion arises from standard assumptions of diminishing returns to investment. Further, we explore the implications credit constraints have on final prices, vertical contracting and so on consumers – Froot et al. do not consider these real economy issues.

Firms might borrow from their suppliers either for investment purposes (usually through finance arms) or by implication of making late repayments for inputs received (known as "trade credit"). We demonstrate that a supplier has complementarities in providing insurance and also credit for future investment projects. As trade credit is usually very short term (measured in tens of days) our work offers insight to the understanding of finance arms. While finance arms are arguably not well understood, there is a literature studying trade credit. Burkart and Ellingsen (2004) note that goods are less divertable

[2] The main theoretical arguments in the extant literature have had difficulty with this empirical evidence as they work in the opposite direction. These theories commonly cite problems of incomplete contracting, which mandate integration in the face of risk to save on contracting costs (Mahoney, 1992).

to private benefits than money and so they argue that trade credit and bank lending are complementary. Cuñat (2007) suggests that a supplier can enforce repayment as he has a long-term relationship with a downstream buyer over inputs which cannot be supplied by another. Cuñat therefore argues that this essential relationship can make a supplier able to provide liquidity insurance to a firm which a bank would not. We show that even without any exogenous advantage over banks in terms of the longevity of the relationship, or the divertability of the loan, there is a complementarity between supplier insurance and lending for investment purposes.

On the industrial organization side, it is probably fair to say that the literature has been skeptical about the assumption that firms are risk averse. It is perhaps more accepted that small owner-managed firms might inherit the risk aversion of the owner; but, in general, it has proved harder to justify why risk aversion should significantly alter the behavior of firms with dispersed ownership. Such explanations would typically require that managers' interests cannot be fully aligned with those of the owners. Here, we demonstrate that risk aversion can result even without such a separation of goals between owners and managers, and that it has real impacts on prices and investments.

Assuming exogenously risk-averse downstream firms, Rey and Tirole (1986) show that the best two-part tariff contract involves double marginalization. Our work demonstrates that risk aversion would be expected if the firms are credit constrained and that double marginalization results not only from two-part tariff contracts but even from the fully optimal contract. More importantly, however, we demonstrate that credit constraints lead to an intrinsic complementarity between supplier insurance and lending (which may explain the existence of finance arms); create a new transmission mechanism linking interest rates with short-run pricing; and provide a theory of countervailing power which predicts that more powerful retailers will charge lower prices.

Plan of the Paper.

The model is introduced in Section 2. We consider a credit-constrained downstream firm facing demand-side risk.[3] The model is solved and the optimal supply contract characterized in Section 3. Debt, amongst other robustness checks, are studied in Section 4. The complementarity between lending and insurance is analyzed in Section 5. Comparative statics methods are used to study changes in firm wealth, countervailing power, and interest rates in Section 6. The incentive to outsource production due to market risk is demonstrated in Section 7. Finally, Section 8 concludes, with all omitted proofs contained in the Appendices.

[3]The paper could easily be rewritten to consider a credit-constrained upstream firm facing supply-side risk.

2 The Model

We consider a model of a vertically related industry with two firms, a downstream firm D and an upstream firm U. There are two periods: period 0 and period 1.

Period 0. In period 0, U can produce an intermediate input at marginal cost $c \geq 0$. U supplies the input to D which D transforms into a final good on a one-to-one basis at zero cost, and then sells on. When choosing output Q and facing market size z, D faces inverse demand $p(Q/z)$.[4] We assume that D is exposed to market risk in that market size z is a random variable with finite support $\{z_1, ..., z_n\}$ and state z_i occurs with probability g_i. 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 $0 < z_1 < z_2 < \cdots < z_n$.

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

(i) *Marginal revenue $d\left[Qp\left(Q/z\right)\right]/dQ$ is declining in quantity Q for all Q such that $P(Q/z) > 0$.*

(ii) *The reservation price exceeds marginal cost at $Q = 0$, $p(0) > c$, and falls below marginal cost, $P(Q) < c$, for Q sufficiently large.*

Assumption 1 implies that, in any demand state z, industry profit $Q\left[p\left(Q/z\right) - c\right]$ is strictly concave in quantity Q. Moreover, it implies that, in demand state z, industry profit is maximized at quantity $Q = zq\left(c\right)$, where $q(c)$ is the unique solution in q to $p\left(q\right) + qp'\left(q\right) = c$. The downstream price that maximizes industry profit is $p\left(q\left(c\right)\right)$ in every demand state z.

Before the demand state is realized, D offers U a menu of contracts of the form $\{Q_i, W_i\}_{i=1}^n$, where $Q_i = Q(z_i)$ is the input (and output) volume in state z_i, and $W(z_i)$ the associated transfer payment from D to U. If U rejects D's offer, both firms make zero profit. (That is, we assume for now that D has all of the bargaining power.) If U accepts D's offer, then D privately learns the realization of the demand state z and reports state $\hat{z}$ to U. As U and D cannot contract on the state of demand z (nor on the retail price), D can equivalently be thought of picking $\{\hat{Q} = Q\left(\hat{z}\right), \hat{W} = W\left(\hat{z}\right)\}$ from the menu.[5] D then receives $\hat{Q} = Q\left(\hat{z}\right)$ 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 $W\left(\hat{z}\right)$ to U. For notational simplicity, we assume in the baseline model that D has no initial assets. We relax this assumption to explore the comparative statics with respect to firm wealth in Section 6. D's asset level by the end of period 0, a, is therefore given by D's net profit in that period: $a = \hat{Q}p\left(\hat{Q}/z\right) - W(\hat{z})$.

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

[5] Note that contractibility of the quantity of input that U ships to D does not imply contractibility of the state of demand z as D can choose to alter quantities requested from U given z.

Period 1. Overnight between periods 0 and 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 D is endogenously credit constrained. Specifically, after choosing the investment level I, D can choose whether or not to shirk at the investment stage. If she does not shirk and invests amount I then in period 1, D makes a gross profit of $\pi(I)$. If instead she does shirk on the investment, then the investment project fails and yields a payoff of zero in period 1, while D receives a benefit proportional to the size of the investment, $B \cdot I$, where $B \leq 1$.[6]

If D wishes to invest more than her pledgeable assets, $I > a$, she can choose to verifiably show her asset level a to an external banking sector so as to attempt to secure a loan of $I - a$.[7] 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

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

since, otherwise, D would decide to shirk and so would be unable to pay back her loan.

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

(i) *The marginal gross return of investment is positive but diminishing:* $\pi(I)$ *is strictly increasing and strictly concave in* I*. Further,* $\pi(0) > 0$, $\pi'(0) > 1+B$, *and* $\pi'(I) < 1$ *for* I *sufficiently large. This implies that the first-best level of investment,* $\widehat{I} \equiv \arg\max_I \pi(I) - I$, *is strictly positive.*

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

Assumption 2(i) contains the standard assumption that investment opportunities exhibit diminishing marginal returns. Assumption 2(ii), which states that the no-shirking constraint is always binding, is for convenience. What is needed for our results is that D would find herself constrained in the level of her investment if she should experience the worst demand state(s) in period 0.

3 Equilibrium Analysis

We solve the model by backward induction.

[6] This reduced-form approach to period 1 allows us to simplify the working of the model while demonstrating our results and intuition cleanly. Section 4 discusses numerous extensions to the model.

[7] D can always choose to hide some or all of her assets. As a result, D can only prove that she has *at least* the asset level that she reveals.

3.1 The Investment Decision and Period-1 Profits

Suppose D's asset level at the end of period 0 is given by a. By Assumption 2(ii), 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 given by a solution in I to

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

Given pledgeable assets $a \geq 0$, we confirm that equation (2) has a unique positive solution $I(a)$ with $0 < I(a) < \hat{I}$ where $\hat{I}$ is the first best investment level. To see this, consider the function $\Psi(I) \equiv \pi(I) - (1 + B) I$. Positive roots to $\Psi(I) + a = 0$ satisfy (2). By Assumption 2(i), $\Psi(0) + a = \pi(0) + a > 0$ for $a > 0$; and $\Psi'(0) = \pi'(0) - (1 + B) > 0$. By assumption 2(ii), $\Psi\left(\hat{I}\right) + a < 0$. As $\Psi(\cdot)$ is concave there is a unique positive investment level $I(a) \in (0, \hat{I})$ at which $\Psi(I(a)) + a = 0$ as stated.

Note that Assumption 2 also ensures that at $I(a)$ the marginal gross return satisfies

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

The first inequality follows as the investment level is below the first-best level. The second inequality is an implication of credit being constrained at $I(a)$. Since the no-shirking constraint is binding, D's net payoff at the end of the second period is $\pi(I(a)) - [I(a) - a] \equiv BI(a)$. The following lemma holds:

Lemma 1 *D's net payoff, $BI(a) \equiv \pi(I(a)) - [I(a) - a]$, is (i) increasing at a rate greater than B and (ii) strictly concave in the pledgeable asset level a.*

Proof. Implicitly differentiating $I(a)$ in equation (2) yields

$$\frac{dI}{da} = \frac{1}{1 + B - \pi'(I)} > 1 \quad \text{and} \quad \frac{d^2 I(a)}{da^2} = \frac{\pi''(I)\left[\frac{dI}{da}\right]^2}{1 + B - \pi'(I)} < 0,$$

where the inequalities follow from equation (3) and Assumption 2(i). ■

The fact that firm D's objective function in the asset accumulation stage becomes concave is a key preliminary result. It shows that the interaction of credit constraints *and* diminishing marginal returns to investment make firm D endogenously risk averse with respect to changes in her pledgeable asset level a. To get some intuition, suppose that D were not credit constrained. After period 0, firm D would therefore borrow $\hat{I} - a$ to invest at the first-best level $\hat{I}$, resulting in end-of-period-1 wealth of $\pi\left(\hat{I}\right) - \hat{I} + a$, for every realization of a. D would therefore be risk-neutral with respect to end-of-period-0 asset level a. Thus diminishing returns to investment alone do not yield the endogenous risk aversion result.

Now consider the fact that D is assumed credit constrained, and the extent of the constraint is endogenous as it depends on D's pledgeable assets. As D has an investment opportunity she can secure a loan of some size from the banking sector. But as D is credit constrained the size of this loan grows with her pledgeable assets. However, as marginal returns to investment are diminishing, the rate at which the marginal dollar of pledgeable assets can be leveraged is itself diminishing. This says that the returns to extra pledgeable assets decline in the level of assets – and so D becomes endogenously risk averse.

3.2 Period-0 Vertical Contracting

The risk aversion identified in Lemma 1 will affect the agreement D requires from her supplier U. This will in turn affect the retail prices in period 0 (the "short run") and the expected level of investment reached in period 1 (the "long run"). Thus credit constraints will – via the supply-chain relationship – affect consumer welfare both in the short and long run.

Let us now analyze period-0 contracting. If the state is z_i and D truthfully reports it, then she would receive a payoff of $BI\left(Q_i p\left(Q_i/z_i\right) - W_i\right)$. Suppose instead D were to lie and claim that the state is z_j, thereby requesting volume 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 assets of $a = Q_j p\left(Q_j/z_i\right) - W_j$ at the end of period 0. Invoking the Revelation Principle, we restrict attention to contracts that maximize the value of D's pledgeable assets 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,

$$\sum_{i=1}^{n} g_i \left\{W_i - Q_i c\right\} \geq 0, \tag{4}$$

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

$$Q_i p\left(\frac{Q_i}{z_i}\right) - W_i \geq Q_j p\left(\frac{Q_j}{z_i}\right) - W_j \text{ for all } j \neq i. \tag{5}$$

This problem is isomorphic to one explored by Hart (1983) in the context of optimal labor contracts. U here maps to workers (the marginal cost c corresponding to workers' reservation wage) in Hart's analysis and D maps to a firm demanding labor specifically. The following proposition then follows:

Proposition 1 *(Hart, 1983, Proposition 2) The solution to Program Bank, $\{Q_i^*, W_i^*\}_{i=1}^n$, has the following properties:*

Property 1 *There is no distortion at the top:* $\frac{\partial}{\partial Q}\left[Q_n^* p\left(\frac{Q_n^*}{z_n}\right)\right] = c.$

Property 2 *There is inefficiently low quantity demanded in all other states:*

$$\frac{\partial}{\partial Q}\left[Q_i^* p\left(\frac{Q_i^*}{z_i}\right)\right] > c \text{ for all } i < n. \tag{6}$$

Property 3 *D's pledgeable assets increase in the state:*

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

Property 4 *U's payoff increases in the state:*

$$W_i - Q_i^* c \geq W_{i-1} - Q_{i-1}^* c \text{ for all } i > 1. \tag{7}$$

Proof. Hart (1983) yields all four conditions.[8] We have a strict inequality in his second condition as U is risk neutral here. ■

Note that the optimal contract involves risk-sharing: both U and D are better off in better demand states (Properties 3 and 4).[9] By exploring a general input into a downstream firm D, we obtain important corollaries of the above proposition.

Corollary 1 *[Risk Sharing and Prices] 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 in all except the best demand state. That is, the optimal contract induces double marginalization.*

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

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

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

$$\frac{\partial^2 f}{\partial Q \partial z} = \left\{\frac{\partial\left(\frac{\partial f}{\partial Q}\right)}{\partial\left(\frac{Q}{z}\right)}\right\} \frac{\partial\left(\frac{Q}{z}\right)}{\partial z} =_{\text{sign}} -\left[-\frac{Q}{z^2}\right] > 0,$$

where we have used the fact that the term in curly brackets is negative (as marginal revenue is declining). The other assumptions follow as U is assumed risk neutral and $I(\cdot)$ has been shown to be concave.

[9] As U's income increases in the state the optimal contract is robust to Innes' (1990) critique. Had the monotonicity not held then D would have an incentive to misreport the state as being better than it was and so save on repayments to U. Such a contract is sub-optimal here as it increases the variance of D's pledgeable assets rather than reducing it.

2. *The optimal contract has the supplier making a net loss in low demand states. Hence, if marginal cost c is sufficiently small, the transfer from D to U, $W(z_i)$ is negative for small realized demand states z_i and positive for large z_i.*

Proof. For part 1, note that equation (6) 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 (and, thus, retail prices being above) the industry-profit maximizing levels.

For part 2, note that U's individual rationality constraint is binding, $\sum_{i=1}^{n} g_i \{W_i^* - Q_i^* c\} = 0$, while $\{W_i^* - Q_i^* c\}$ is, by equation (7), increasing in i. Hence we must have some state j such that

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

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

In the absence of either credit constraints or market risk, or both, the optimal supply contract would stipulate the myopic first best quantity $z_i q(c)$ in state z_i, resulting in the retail price $p(q(c))$ that maximizes joint period-0 profit. Proposition 1 and Corollary 1 show that the interaction of credit constraints and market risk imply that this (joint period-0 profit maximizing) contract is not an optimal one for the endogenously risk-averse firm D to demand of her supplier. It can be improved by requiring U to share in the risk faced by the downstream firm D. Intuitively, such risk sharing implies that U is made to provide D with profit insurance. Since U earns zero profit on average,[10] he must make a loss in the worst state(s) and profits in the best yielding Part 2 of Corollary 1.[11] We will provide empirical evidence that such risk sharing contracts in which U can suffer losses are widespread in the following subsection. If a third party could verifiably observe the quantity U supplies, then she could, in principle, provide the insurance to D. However, as we discuss in Section 7, this would have no effect on the retail price implications explained here.

Why should the desire to have U supply insurance result in high retail prices (double marginalization)? In essence, D is using U to lower the variance of the value of her end-of-period pledgeable assets. This is done, in effect, by having U make a fixed payment to D which D repays according to the state of realized demand. However, for U to be able to make back this ex ante committed payment in expectation, the variable payments made to U must increase in volumes by more than the marginal cost of supply. Hence, double marginalization is created. This double marginalization is optimally spread across (almost) all demand states to reduce the temptation D has to misreport the state of

[10]More formally, U is held to his participation constraint. That his payoff from the contract equals zero is a normalization.

[11]Note that this implies that $a \geq 0$ as in state 1, $Q_1^* < z_1 q(c)$ and so $Q_1^* p(Q_1^*/z_1) > c Q_1^* \geq W_1^*$, and D's pledgeable income increases in the state.

demand. As a result, the optimal risk-sharing contract induces retail prices that are (in almost all demand states) strictly higher than the myopic industry profit maximizing level, $p(q(c))$. Hence, some of the burden of credit constraints and market risk is borne by consumers.

Remark 1 *In our analysis, we have allowed for general contracts between the upstream supplier U and the downstream firm D. Suppose instead that firms were restricted to two-part tariff contracts of the form*

$$W(Q) = f + wQ,$$

where f is a fixed fee and w the per-unit input price. It can be shown that, in this case, the equilibrium contract in period 0, (f^*, w^*), involves double marginalization (in all demand states), $w^* > c$, and payment of a slotting fee from the upstream firm to the downstream firm, $f^* = -E[z] \cdot q(w^*)(w^* - c) < 0$.

3.3 Evidence on Retail Pricing and Contract Form

We have presented a model in which credit constrained firms have high retail prices as their supply contracts are altered to share risk, thereby increasing marginal costs. Empirical evidence of this effect is provided by Chevalier and Scharfstein (1996). These authors study retail prices in the U.S. supermarket industry. In each city studied they compare the prices charged by local supermarkets against those of national chains over a period when some cities faced a bad demand shock (recession) while others did not. Arguing that local stores are more likely to be credit constrained, they offer the empirical finding that firms facing credit constraints post higher prices than non-constrained firms during bad demand outcomes.

Chevalier and Scharfstein offer an alternative explanation through a model of constant and fixed input prices with the presence of consumer switching costs giving supermarkets a rationale for altering the price level. The mechanism we offer also fits the empirical findings without recourse to switching costs. Chevalier and Scharfstein rule out an explanation involving credit constrained supermarkets affecting their marginal input cost by arguing that "*it is hard to think of any reason why* [such an] *interpretation ... would be true.*" We offer such a reason by demonstrating that the credit constraints themselves create the desire to alter input costs and share risks.

There is also substantial evidence that the form of contract we have explored is in widespread use. We have noted that the optimal contract can be thought of as involving a fixed payment from U to D (and demand-dependent repayments). In the marketing literature, this fixed payment is known as a '*slotting fee.*' Slotting fees are commonly used in the grocery industry as well as in software and publishing industries. While it has been noted that slotting fees can be rationalized by, for example, suppliers signalling the quality of their products to retailers (Klein and Wright, 2007), recent survey evidence suggests that risk sharing is a part of the rationale for slotting fees (Sudhir and Rao, 2006; Bloom et al., 2000). Further note that the very largest retailers, such as Wal-Mart and Costco in the US, do not use slotting fees.[12] One might suspect that these very large retailers are unlikely to be credit constrained and so this observation is in line with the predictions of our model.

There are also many examples of firms signing risk-sharing contracts with their vertical partners. One celebrated example are the "*Risk and Revenue Sharing Partnerships*" used by Rolls Royce.[13] Case studies exist exploring such arrangements in multiple industries.[14]

4 Firm Debt, Risk Aversion and Model Robustness

A debate exists within the Corporate Finance literature as to whether external financial frictions cause firms to be effectively risk averse, or instead risk loving. We have studied an asset accumulation stage in which the downstream firm is trying to secure funds to allow borrowing to take place. As the amount which can be borrowed increases in the level of pledgeable assets and the return to investment has diminishing marginal returns (both standard assumptions), the firm is endogenously risk averse. Other theoretical mechanisms exist which would also result in a firm acting as if she is effectively risk averse. Managers who are risk averse and hold a large portion of their firm's stock which they cannot sell on will take decisions under risk aversion (Stulz, 1984). Managers may in the future be judged by their performance and so they wish to avoid being outliers on the bad side as they will then lose their jobs; this causes the managers to become risk averse (Breeden and Viswanathan, 1990; DeMarzo and Duffie, 1995). If taxes are a convex function of earnings, then risk aversion is immediate (Smith and Stulz, 1985).

However, theory is not unambiguous in pointing to financial constraints leading to risk aversion. Jensen and Meckling's (1976) seminal work explained that the presence of debt can make equity more tolerant of risk as losses fall disproportionately on debt holders. This influential idea of risk-shifting led to a large literature. Building on this work, Brander and Lewis (1986) argue that the risk-shifting rationale, combined with limited liability, will make firms act in a risk-loving way in their production decisions.

Further insight can be gleaned from a number of empirical studies. Panousi and Papanikolaou (forthcoming) show strong evidence for risk aversion, rather than risk-loving

[12] See Federal Trade Commission (2001), "Report on the Federal Trade Commission Workshop on Slotting Allowances and Other Marketing Practices in the Grocery Industry", available at: http://www.ftc.gov/os/2001/02/slottingallowancesreportfinal.pdf.

[13] See the 2008 Annual Report for example, available at http://www.rolls-royce.com/reports/2008/finance-directors-review.html.

[14] See for example Figueiredo et al. (2008) for aircraft in Brazil and Camuffo et al. (2007) for air-conditioning in Italy.

behavior, playing an active role in firms' investment strategies. Panousi and Papanikolaou show that firms exposed to greater uncertainty (as evidenced by a more volatile stock return) invest less; and the distribution of the effect implies, they argue, that risk aversion is the cause. That firms' objective function is concave (implying risk aversion) is similarly the empirical conclusion of Leahy and Whited (1996).[15] Finally, Chevalier and Scharfstein (1996) has already been cited as evidence supporting our link between credit constraints and retail price rises.

There is therefore substantial evidence in favor of firms acting in a risk averse way. In fact, our core results are entirely robust to the presence of debt contracts as modelled, for example, by Brander and Lewis (1986). To study this, let us extend our core model slightly and suppose that D has a debt contract at the beginning of period 0 which we model as D commencing with negative assets of $-d$. After period 0-trade, firm D will generate a net asset position of a. If sufficient profit is made in period 0 to pay off the debt, then a will be positive. If, however, profits in period 0 are less than d, then a will be negative. In that case, D has outstanding debt. D is not necessarily bankrupt, however, as D has future ($t = 1$) investment opportunities. Thus D may be able to borrow funds to both pay off her end-of-period-0 debt and invest the remainder between periods 0 and 1.

Lemma 2 *There is a minimum end-of-period-*0 *asset level* $\hat{\kappa} < 0$ *below which* D *is unable to borrow funds. If the end-of-period-*0 *asset level* a *satisfies* $a \geq \hat{\kappa}$*, then* D *can borrow to make some investment and pay off her existing debt.*

Proof. Let $\tilde{I} = \max_I \Psi(I) = \max_I \pi(I) - (1 + B) I$. By Assumption 2 $\tilde{I} \in (0, I(0))$. Set $\hat{\kappa} = -\Psi\left(\tilde{I}\right) < 0$. If end of period $t = 0$ assets are $a \geq \hat{\kappa}$ then $\Psi(I) + a = 0$ has a unique positive root such that $I(a) \geq \tilde{I}$. This follows from the same reasoning as in the core model which defined the implicit function $I(a)$. Hence D can invest at level $I(a)$ before period $t = 1$. If $a < \hat{\kappa}$, then $\Psi(I) + a < 0$ at all investment levels I, and so D will be unable to secure a loan. ■

If $a < \hat{\kappa}$, then D is unable to access any loans as her moral hazard problem is too severe. Thus she carries her existing debt into period $t = 1$ and can make no extra investment. Therefore if end of period 0 assets are a, then D's final payoff is given by $V(a)$ where:[16]

$$V(a) = \begin{cases} BI(a) > 0 & \text{if } a \geq \hat{\kappa}, \\ a + \pi(0) < 0 & \text{if } a < \hat{\kappa}. \end{cases} \tag{8}$$

Note that there is a jump in the value function if realized end-of-period-0 assets are equal to $\hat{\kappa} < 0$.

[15] Not all empirical studies have generated clear evidence of risk aversion. See for example Brainard, Shoven, Weiss, Cagan and Hall (1980).

[16] Note that if $a < \hat{\kappa}$ then $V(a) < 0$ as $\hat{\kappa} = -\max_I \pi(I) - (1 + B) I < -\pi(0)$ by Assumption 2.

In our core model, once D observes the state of demand she will realize whether or not she will be able to survive into period 1. As any firm always has the option of shutting down, we must extend our model to allow for this possibility. We suppose that after learning the demand state, and before receiving inputs from U, D may shut down and default on any outstanding debt and secure a private payoff of b. We allow for any payoff b which lies within the jump range given in (8): $\hat{\kappa} + \pi(0) \leq b < BI(\hat{\kappa})$. If the payoff parameter is equal to zero, $b = 0$, then we have a literal model of limited liability. If $b < 0$, then the model captures some shut down costs; whereas if $b > 0$, then the model captures some management perks. With this limited liability condition, D's net payoff allowing for debt is a function of the end of period 0 assets, a, and is given by $V^{\text{debt}}(a)$:

$$V^{\text{debt}}(a) = \begin{cases} BI(a) > 0 & \text{if } a \geq \hat{\kappa}, \\ b & \text{otherwise.} \end{cases} \tag{9}$$

There are at least three motivations for this modelling choice. The first is that in many markets it is illegal to trade when the firm knows it is insolvent. Here we are exactly assuming that as soon as D realizes that she has a negative NPV, she shuts down. A second motivation is that D has some (arbitrarily small) time discounting. In this interpretation, if D can secure b either at the end or at the start of $t = 0$, because of time discounting D prefers to secure b sooner rather than later. The final interpretation is that D's existing debt is senior to any supply contract with U. Thus U appreciates that if a bad demand state is reported then he will not receive full payment from D. In such a situation U would not be willing to supply D as per the terms of the contract.

The jump in the payoff function creates the potential for risk-loving behavior. The downstream firm, D, can now gain disproportionately if she can push her end of period 0 assets to be above $\hat{\kappa}$.

Proposition 2 *Suppose that D has a debt contract so that she begins period $t = 0$ with negative assets $-d$. As long as the grid of probabilities $\{g_i\}$ is sufficiently fine:*

1. *D will shut down if the demand state is bad enough; that is, there exists some state $k \geq 1$ such that D shuts down in state $i < k$.*

2. *There will be double marginalization at all demand states at which D remains in business, except the highest.*

3. *U provides insurance in all states at which D remains in business.*

Proof. See Appendix A. ■

The downstream firm D wishes to manage her supply contract to allow her to remain in business, and subject to this, to invest to maximize her payoff. D therefore wishes to grow her returns in poor states sufficiently to get over the point at which she can borrow if necessary, repay her existing debt and potentially invest for the future. This requirement is entirely compatible with the desire to smooth pledgeable assets conditional on being in business. Both requirements cause D to wish to grow her assets in low states (conditional on remaining in business) and she is willing to sacrifice assets in high states to do this. This follows as, conditional on remaining in business, the marginal payoff to greater pledgeable assets is much larger if the level of assets is low (yet above $\hat{\kappa}$) than if the level of assets is high. If demand is so bad that D will have an asset value which is too negative to permit any further borrowing then D will shut down. Thus U is asked to provide the revenue boost in low demand states which permits D to continue as a going concern. U is willing to participate due to expected profits in higher demand states. It follows that identical reasoning to Corollary 1 guarantees that double marginalization remains.

Our result is compatible with the risk-shifting insight of Jensen and Meckling (1976) as the downstream firm is unable to opportunistically exploit the upstream supplier. In the standard risk-shifting set-up, the debt is present before commercial and risk decisions are taken. The firm therefore has an incentive to exploit the debt holders through the commercial decisions taken. In contrast, the interaction with the supplier is, in our model, part of the commercial decision making process. U therefore agrees to a supply contract in the knowledge of the prior debt level. The supplier is therefore used for profit insurance, the optimal contract protecting the supplier from being exploited as a repository for extra risks. The effect of this income insurance is retail double marginalization.

We conclude that effective risk aversion amongst firms is well founded in the literature; and further standard debt contracts would not affect our retail price results.

4.1 Discussion of Model Assumptions and Extensions

4.1.1 Future Investment Correlated with Period-0 Demand

In our core model, the expected returns from a given investment in period 1 do not vary with the realized period-0 demand. This is a natural assumption; but it is possible to imagine correlation between current (period-0) demand and future (period-1) investment returns. If current demand was informative as to the quality of the product, for example, then low current demand might signal lower future returns from any given investment. On the other hand, if current demand is low due to a macro-economic cycle reining in spending then returns to investment in period 1 might be expected to be high as the cycle moves to a growth phase.

Current (period-0) demand may alter either the level of future expected returns, or their gradient, or both. The impact on our results differs depending on which aspect of future investment returns are affected. Let us first suppose that the level of future investment returns is positively correlated with period-0 demand. To capture this, suppose that the period 0 state (z_i) alters the expected period 1 returns to $\pi(I) + \theta\varphi(z_i)$, where $\theta > 0$ is a parameter measuring the strength of correlation, and φ an increasing function. This additive formulation permits the core model as a limiting case with $\theta = 0$, and implies that though the level of expected profits is correlated across states, the marginal returns to extra investment is not. The higher is demand in period 0, the greater the level of future expected returns to investment. Suppose also that $\pi(0) + \theta\varphi(z_1) > 0$ and that $\sum_i g_i\theta\varphi(z_i) = 0$, implying that the ex ante expected returns are the same as before.

This alteration to equation (2) delivers that D's objective function in state i is $B \cdot I(a + \theta\varphi(z_i))$. That is, the new additive term in period-1 returns acts as a state-dependent change in pledgeable assets. For the same reasons as before, D's objective function is concave in the end-of-period-0 asset level a for any given realization of the demand state. Moreover, as the revenue function $f(Q, z; \theta) \equiv QP(Q/z) + \theta\varphi(z)$ is increasing and concave in Q, increasing in z, and $\partial^2 f(Q, z; \theta)/\partial Q\partial z > 0$, the core results, Proposition 1 and Corollary 1, continue to hold. An increase in θ (the parameter measuring the strength of correlation between period-0 demand and the level of the returns to period-1 investment) tends to increase D's endogenous risk aversion and reinforces the results. Intuitively, this is because, for a given quantity Q, an increase in θ induces a mean-preserving spread of the revenue function $f(Q, z; \theta)$.

Now let us consider the impact of positive correlation between period-0 demand and the gradient of the returns to investment. Intuitively one might have in mind that period-0 demand alters returns multiplicatively rather than additively, so that the gross returns to investment are given by $\pi(I) \cdot \varphi(z_i)$. In this case low period-0 demand has two effects. Firstly, after low period-0 demand, the expected returns in period 1 are shifted down. This effect was analyzed above and acts to reinforce our results. Secondly, the marginal return to extra investment is reduced after low period-0 demand. This latter effect on the marginal returns acts to lower the marginal returns to extra assets in low period-0 states and raise them in high period-0 states – hence this is a force towards making D risk loving and reversing our results. Which of these opposing effects dominates will depend upon the specifics of any given situation. Overall, and as already noted by Froot et al. (1993), D will seek to insulate the marginal value of returns to extra period-0 assets. Unlike in Froot et al., this is achieved by a risk-sharing contract which explicitly moves transfers across states subject to the supplier U participating.[17]

4.1.2 Functional Form of Investment Returns and Shirking Returns

In the core model, D faced a moral hazard problem as she could misuse the funds meant for investment and shirk instead. If she shirked then D secured a payoff which was linear

[17] Unlike in Froot et al. (1993), the optimization cannot be solved simply state by state as the supply contract is state dependent and the participation constraint of U links the states. An implication of this is that explicit randomness in future returns is more difficult to manage.

and proportional to the investment funds I. The linearity assumption is not key to the results. One might instead consider any shirking reward function $\mathcal{B}(I)$ which was convex increasing: $\mathcal{B}'(I) > 0$, $\mathcal{B}''(I) \geq 0$. If D has only a small amount of money to invest then it is difficult to discretely redirect it to unproductive perks; however if the funds D has are large then detailed oversight is more difficult and potentially large sums can be redirected to unproductive perks.

Continuing to restrict attention to end of period-0 assets which leave D credit constrained (Assumption 2(ii)), our results continue to apply. In particular, D would invest $I(a)$ given by the point at which the credit constraint becomes binding: $\mathcal{B}(I) = \pi(I) - (I - a)$. This has a unique solution in I due to the convexity of $\mathcal{B}(\cdot)$ and concavity of $\pi(\cdot)$. The payoff to D will be given by $\mathcal{B}(I(a))$. Lemma 1 can be extended to demonstrate that D is endogenously risk averse as $\mathcal{B}(I(a))$ is concave in a:

Lemma 3 *D's net payoff, $\mathcal{B}(I(a)) \equiv \pi(I(a)) - [I(a) - a]$, is increasing and strictly concave in the pledgeable asset level a.*

Proof. Assumption 2(ii) ensures that at $I(a)$ the marginal gross return satisfies $1 < \pi'(I(a)) < 1 + \mathcal{B}'(I(a))$. The first inequality follows as the investment level is below the first-best level; the second inequality is an implication of credit being constrained at $I(a)$. Implicitly differentiating $I(a)$ yields $d[\mathcal{B}(I(a))]/da = \mathcal{B}'(I(a)) / (1 + \mathcal{B}'(I(a)) - \pi'(I(a))) > 0$. Hence, D's net payoff is increasing in a. Its curvature is given by

$$\frac{d}{da^2}[\mathcal{B}(I(a))] = \left(\frac{dI}{da}\right)^2 \frac{\{\pi''(I(a)) \cdot \mathcal{B}'(I(a)) + \mathcal{B}''(I(a))(1 - \pi'(I(a)))\}}{1 + \mathcal{B}'(I(a)) - \pi'(I(a))} < 0$$

The sign is determined by the sign of the term in curly brackets. This is negative as the technology is concave, $\pi'' < 0$; the investment level is below the first best, so $\pi'(I(a)) > 1$; and the private benefit from shirking is convex in investment, $\mathcal{B}''(I(a)) \geq 0$. ■

If pledgeable assets are low then the investment returns to extra assets would be high. Further, borrowing would be easier as at low investment levels shirking is less valuable to (or less feasible for) D. Together these imply that extra assets could be substantially leveraged. At high pledgeable asset levels the opposite is true. Hence, with the generalized private benefit function it continues to be the case that the marginal return to extra assets is greatest at low asset levels, and so D is endogenously risk averse. Our results on short term retail prices therefore continue to apply.

We now turn to the shape of the returns to investment. The standard assumption concerning future technology is that there are decreasing returns to scale so that $\pi''(I) \leq 0$. However over some ranges it is possible for the technology to exhibit increasing marginal returns to investment. This would come about if some critical scale were required, for example. Under a generalized private benefit function all that would matter (Lemma 3) is the curvature of $\pi(I) - \mathcal{B}(I)$. If the private benefit function is sufficiently convex to render this sum concave, then our results and approach are unchanged. Otherwise, our results would change.

To build intuition, let us consider the case of a convex technology: $\pi''(I) > 0$. In this case the larger the pledgeable asset level at the end of period 0, the greater the amount that can be borrowed and invested. D would be risk loving. We have noted above that though this result can be generated in theory, empirically it appears uncompelling as a first order approximation (Leahy and Whited, 1996; Panousi and Papanikolaou, forthcoming). The logic of our core model would dictate that D would like the opposite of insurance from U. D would like extra pledgeable assets from U in good states of the world, paid for by accepting reduced profits in bad states of the world. As a result the direction of the binding incentive constraints would be reversed: the optimal structure of the contract would ensure that D did not wish to report the state as being better (not worse) than it really was. To ensure truth telling U increases the quantity that D must sell as this increases revenues by more in higher states. Consumers would see prices below (not above) the short-run profit-maximizing level.[18]

4.1.3 Modelling of Period-1 Contracting

Suppose that after investment, in period 1, D must select a supplier and contract upon the same informational basis as in period 0. As this is the final period of the model (there are no future investments), D is risk neutral with respect to the assets she accumulates post investment. Hence, D's optimal contract involves sourcing the input at marginal cost. This is unaffected by the level of investment achieved. Now consider the borrowing phase which takes place overnight between periods 0 and 1. If in all period-1 states of demand D can repay her loan when not shirking, then the model proceeds unaffected as for overnight borrowing decisions all that matters is the expected period-1 profit, not the profit in each state. Suppose instead that D can only repay her loan if period 1 demand is not in the worst state(s). As D would have a default risk in period 1, the bank would require a positive interest rate so as to break even in expectation. It remains the case that the amount that can be borrowed grows in the level of the pledgeable assets. As there are diminishing returns to investment, the rate at which extra assets can be leveraged declines in the assets. Hence, D remains endogenously risk averse in the period-0 asset accumulation phase. Thus the features of the optimal vertical contract summarized in Corollary 1 remain, though the specific optimal contract will depend upon the interaction of the investment return function with the post-investment demand risk.

[18] We have made the assumption that a bank can inspect the pledgeable assets at the end of period 0 before deciding on the loan to make available. If the supplier U can also demand to inspect the pledgeable assets, then D cannot lie by claiming the state is better than it is. If she did lie she would be unable to demonstrate the expected level of pledgeable assets. In this case, the myopic contract of marginal cost supply would be optimal.

4.1.4 Credit Constraints Upstream and Downstream

We now consider the impact if U also had investment opportunities to fund between periods 0 and 1, and U were credit constrained. The analysis of Lemma 1 would deliver that U would also be endogenously risk averse in the asset accumulation phase (period 0). To build intuition, first consider the case of D not being credit constrained while U is. D, who offers the supply contract to U, would be risk neutral. Her preferred supply contract would then be the standard myopically optimal one in which the wholesale price per unit equals the marginal cost. This generates perfect insurance for U: his end-of-period-0 assets are independent of the realized downstream demand. Consider now the case where D and U are both credit constrained. As U is endogenously risk averse, he would value the optimal contract from our baseline model less and so would be unwilling to accept it. U, for his part, would require a contract closer to the standard marginal cost pricing one. The result will be a weighted average of these two contracts, thus preserving the features of the optimal vertical contract summarized in Corollary 1.[19]

5 Complementarities between Supplier Insurance and Banking

In the model as presented so far, the supplier U offers her downstream buyer D some profit insurance. The downstream firm D then goes to the banking sector to borrow to fund the investment. If U could borrow and lend at the same (zero) interest rate as banks can, then U could take the place of the bank, providing the loan for investment as well as period-0 income insurance. In fact, this section shows that borrowing from U and committing not to use a separate banking sector strictly dominates using a banking sector.

Lending from finance companies as opposed to from banks is an important source of corporate debt. Yet the extent to which bank loans and non-bank loans are substitutable is, arguably, poorly understood. Here we propose that by combining supply with lending the inefficiencies in financing created by informational asymmetries can be reduced. Concretely, by having to return to U for a loan, D can commit to charge a lower retail price and therefore one which is less double marginalized. This is because if she misreports the state in period 0 and so makes extra profits, U can commit not to allow them to be leveraged. This permits D to credibly discipline herself, reducing her incentive to misreport and shirk. As a result, this section will offer a novel explanation for the existence of supplier finance arms.

To derive this result, suppose that D committed not to use a banking sector and only

[19] This is of relevance to slotting fees as one might suspect that suppliers to supermarkets are at least as likely to be credit constrained as the supermarkets themselves.

deal with U. D would now be proposing the contract $\{Q_i, T_0^i, T_1^i\}$, where Q_i is quantity of input delivered in period 0 if the state is z_i, T_0^i is a payment from D to U at the end of period 0 (so that D's investment in period 1 is her revenue minus T_0^i), and T_1^i is a payment from D to U at the end of period 1, after the investment returns are realized.

The program to solve with no bank is as follows.

Program No Bank The optimal program when U provides the loan is given by:

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

subject to

$$\begin{aligned} \sum_{i=1}^{n} g_i \left\{ T_0^i + T_1^i - Q_i c \right\} &\geq 0, & (10) \\ \left[Q_i p \left(\frac{Q_i}{z_i} \right) - T_0^i \right] \cdot B &\leq \pi \left(Q_i p \left(\frac{Q_i}{z_i} \right) - T_0^i \right) - T_1^i, & (11) \\ \pi \left(Q_i p \left(\frac{Q_i}{z_i} \right) - T_0^i \right) - T_1^i &\geq \pi \left(Q_j p \left(\frac{Q_j}{z_i} \right) - T_0^j \right) - T_1^j \text{ for all } j < i. & (12) \end{aligned}$$

Here, (10) is the individual rationality constraint for U, (11) is D's no-shirking constraint at the investment stage in period 1, and (12) is D's incentive constraint when reporting the state of demand in period 0. The optimal contract can involve a large penalty if D were unable to show all of the assets that she should have earned in period 0 according to her demand report. This parallels the base line model in that D cannot claim to have pledgeable assets which she does not have as collateral for a loan. This implies that D can only under-report but not over-report the demand state in period 0.

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.[20]

Proposition 3 *[Finance Arms] 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 assets of $Q_i^* p \left(Q_i^*/z_i\right) - W_i^*$ and invests an amount $I \left(Q_i^* p \left(Q_i^*/z_i\right) - W_i^*\right)$, borrowing the difference between these two.

We first show that U can replicate the optimal contract D would set if using a banking

[20] We assume here that any such extra assets could still be invested, although not leveraged. Assuming otherwise would only strengthen our result.

sector. Let

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

where volumes $\{Q_i^*\}$ are as in the contract with the separate banks, and T_1^i is the size of the loan provided. Then, equation (10), the individual rationality constraint of U, is satisfied with equality by (4). By construction of T_1^i, the credit constraint is binding in every state so that the no-shirking constraint (11) always holds with equality. Finally, from the definition of the loan,

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

where the inequality is by the incentive constraint (5). The final term is the return available to D if her pledgeable assets are $Q_j^* p\left(Q_j^*/z_i\right) - W_j^*$ 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 $Q_j^* p\left(Q_j^*/z_i\right) - W_j^*$ and borrowing T_1^j (resulting in investment equal to the level in the right-hand side of (12)) leaves the no-shirking constraint at the investment stage 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 $\pi' > 1 \geq B$ (see equation (3)), we must have

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

The left-hand side is the profit available if D borrows T_1^j to invest a total of $Q_j^* p\left(\frac{Q_j^*}{z_i}\right) - T_0^j$. 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. We thus obtain

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

as required. Hence, the period-0 incentive constraint (12) is actually slack.

But as the incentive constraint on the report of the demand state in period 0 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 payments

T_1^i and T_0^i retain the form given above. This new tariff satisfies (12) for small $\varepsilon > 0$. U remains indifferent, thus continuing to satisfy (10) with equality. 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 Property 2 of Proposition 1, the marginal revenue at states below n exceeds c. ■

The complementarity between supplier insurance and lending results from countervailing incentives being pooled. When applying for a loan, D would like to over-report the size of her assets so as to secure a larger loan. In contrast, in her supply insurance relationship, D would like to under-report the demand state so as to secure a larger insurance payout. By committing to leverage only those assets that are consistent with D's demand report, U can effectively reduce D's temptation to under-report the demand state and thus remove some double marginalization from the supply contract.[21] Note that a third party, such as the bank, would be in a worse position than U to provide both the insurance and the lending unless the third party could verifiably observe the input supply of U to D.[22]

Note that our mechanism does not require that the upstream firm U provides all of the lending to D. Instead, U may cooperate with banks in a consortium of lenders – with the banks providing "inframarginal" lending (the part of the loan that would be provided even in the worst demand state) and U only providing the "marginal" lending that is sensitive to the reported demand state. To enforce this, the borrowing firm must be limited in its access to further lenders for top-up loans. Covenants could be written to this effect.[23]

5.1 Empirical Evidence: Lending via Financial Companies Versus Banks

Proposition 3 provides a rationale for suppliers maintaining finance arms, as indeed many major firms do (e.g., GE, Cisco). The finance arm will be able to offer terms which improve on those from a bank by linking the size of the loan to the quantity of input supplied. That a supplier with the same access to capital markets as an external bank can lend on rates that the independent banking sector would find unprofitable, is a new result. Understanding when such non-bank lenders have a comparative advantage over banks is important. In 2008, U.S. financial companies lent just over 608 billion dollars to business borrowers. This figure does not include financial companies lending to private

[21] The mechanism offered here is related to the literature on countervailing incentives; see, e.g., Lewis and Sappington (1989). These authors show that, with countervailing incentives, the optimal contract may involve pooling in some states. Instead, our focus is to show that by pooling principals (supplier, bank), the agent (buyer) derives a benefit.

[22] See also the discussion on outsourcing in Section 7.

[23] Indeed, there is evidence that, if lending is undertaken by a consortium, then covenants are more likely to be required (Bradley and Roberts, 2004).

consumers or for real-estate assets.[24] This compares with bank lending to businesses of 1.5 trillion dollars (commercial and industrial assets on U.S. bank balance sheets at end 2008). Thus financial companies lend almost \$1 for every \$2 lent by a mainstream bank. Therefore gaining an insight into what makes financial companies effective competitors for banks is arguably a first-order issue.

Carey et al. (1998) note that finance companies are over-represented in loans to higher risk firms. Similarly, Denis and Mihov (2003) report that borrowing firms with a lower credit score, so higher risk, are more likely to borrow from a finance company rather than a bank. Such a distribution of loans could be explained if there is a complementarity between supplying input and lending, implying that such lenders can make a profit even with risky borrowers, whereas banks cannot. Our mechanism is new: Carey et al. report that there is as yet little literature on why finance companies may have a competitive advantage over banks in lending to riskier firms. They speculate that financial regulation might have encouraged banks to avoid these loans to limit risks; or that banks might wish to preserve a reputation for being lenient which they could not be with the riskiest borrowers. However these arguments are not explicitly modelled or tested.

In our model the lending provided by U is identical to that which a bank can provide. The extent to which finance companies provide loans which are a substitute for bank loans has been explored by Billett, Flannery and Garfinkel (1995) and by Preece and Mullineaux (1994). Both of these studies consider the stock market response to the announcement of new loans from either banks or non-banks. Both articles find that there is no statistically significant difference in the market's reaction to new loans whether they originate from a bank or a non-bank. Thus even though finance companies appear to have an advantage in lending to riskier borrowers, the loans they extend carry the same good news which a bank loan would. This finding sits comfortably with the mechanism we have proposed.

In conclusion the evidence available suggests that lending by finance companies is a substitute for bank lending, and that finance companies are over-represented in loans to riskier firms. The mechanism we have proposed is new and can explain why a finance company would have a comparative advantage in lending to the firms that she supplies. James and Smith (2000) report that these bank-type loans from either banks or non-banks commonly contain covenants which link to the borrower firm's commercial performance. Our mechanism uses exactly this route to explain the advantages of supplier lending.

6 Cross-sectional Predictions

We have shown that the interaction between credit constraints and market risk causes a risk-neutral (downstream) firm to become endogenously risk averse with respect to her

[24] This is drawn from the Federal Reserve G20 statistical release. Available at http://www.federalreserve.gov/econresdata/releases/statisticsdata.htm.

pledgeable assets. The endogenous risk aversion causes the firm to seek to push risk on to her vertical partner, and this results in double marginalization so that retail prices rise. How risk averse the firm is, and so how large the effect on retail prices, will depend upon market-level and firm-level parameters. Thus retail prices of potentially credit-constrained firms are affected by alterations in D's bargaining power (Section 6.2); in the interest rates payable on debt (Section 6.3); or by how asset-rich firm D is.

We first study the general question of how a change in any parameter θ will alter the optimal contract between the credit-constrained D and U in period 0. Let $I(a;\theta)$ denote the (endogenous) investment level as a function of the realized asset level a and some parameter θ. The following lemma demonstrates that if a change in θ increases the Arrow-Pratt measure of absolute risk aversion of the investment function with respect to pledgeable assets, then the period-0 retail price will rise in all demand states except the highest and those at which the optimal contract involves pooling. As we are working with the optimal period-0 contract, this effect is not an artifact of a restricted contract class (such as linear or two-part tariff contracts).

Lemma 4 *Suppose a change in model parameter θ causes the coefficient of absolute risk aversion, $-\frac{\partial^2 I}{\partial a^2}\Big/\frac{\partial I}{\partial a}$, to increase (decrease) at all pledgeable asset levels. Then, at all states $i < n$ at which the optimal contract does not involve pooling, $Q_{i-1}^* < Q_i^* < Q_{i+1}^*$, the optimal quantity sold in period 0 decreases (increases). Hence, the short-run retail price in such states increases (decreases). The result holds weakly at state $i < n$ if the optimal contract in that state involves pooling, $Q_i^* \in \{Q_{i-1}^*, Q_{i+1}^*\}$.*

Proof. See Appendix B. ■

6.1 Asset-Rich Firms and Double Marginalization

Suppose that the downstream firm has some exogenous assets equal to $A \geq 0$. We assume these are not so large that Assumption 2 is contravened – though we recall all that is required for our results is that the downstream firm would be credit constrained if realized demand in period 0 were bad enough. Credit constraints would then ensure that the investment level $I(a, A)$ would be given implicitly by

$$B \cdot I = \pi(I) - (I - [a + A]). \tag{13}$$

Thus if the downstream firm has extra assets A then her investment returns as a function of pledgeable assets match those in the core model: $I(a, A) \equiv I(a + A)$. Using the proof of Lemma 1 applied to (13) the coefficient of absolute risk aversion is therefore given by

$$-\frac{\partial^2 I}{\partial a^2}\Big/\frac{\partial I}{\partial a} = -\pi''(I)\left[\frac{\partial I}{\partial a}\right]^2. \tag{14}$$

Hence, invoking Lemma 4, the impact of a change in initial assets (A) can be understood by reference to how the change alters the marginal return to assets, and the curvature of the investment returns function. We derive:

Proposition 4 *[Asset Rich] Suppose that the curvature of the technology function is regular in the sense that it declines in magnitude at higher investment levels, i.e., $\pi'''(I) \geq 0$. Then, an increase in D's initial asset endowment (i.e., an increase in A) results in lower retail prices in the short run (period 0).*

To understand this result suppose an asset-rich downstream firm signed the same period-0 supply contract as an asset-poor firm. If D is asset rich, then she will have more pledgeable assets at the end of period 0 for any given realization of demand. These extra assets allow the amount borrowed, and thus the investment level, to grow. But at higher investment levels, the marginal return to extra investment is lower, and thus the marginal incentive to shirk larger. The size of the marginal incentive to shirk captures the size of the moral hazard problem. Hence, if assets were to increase a little further then the amount of extra borrowing would be modest. Therefore $\partial I/\partial a$ shrinks as the initial asset endowment grows. Further, if the initial asset endowment grows then the level of investment rises, and we have assumed that the investment returns function π has regular curvature in that the curvature is negative and weakly increasing in the investment level I. Overall, therefore, both of the elements of risk aversion given in (14) decline. Hence, Lemma 4 guarantees that the degree of risk aversion felt in period 0 declines. Therefore asset-rich firms seek to transfer less risk to their vertical partners, so less double marginalization is induced and so retail prices fall.

6.2 Countervailing Power and Credit Constraints

Now we relax the assumption that D has all of the bargaining power. This allows us to study how changes in D's bargaining power as compared to her supplier alters the retail prices faced by consumers. We show that an increase in D's bargaining power corresponds to an increase in the expected level of pledgeable assets and this lowers the degree of risk aversion D faces in the pledgeable asset accumulation phase. This therefore means that a more powerful buyer agrees a bargained supply contract which results in lower retail prices. The model thus provides a novel *theory of countervailing power* based on credit constraints.

To model a more equal distribution of bargaining power between U and D, we let U receive an expected payoff of β from the relationship. That is, β is a measure of U's bargaining strength and the bargaining analogue of Program Bank is modified by replacing the individual rationality constraint of U (equation (4)) by

$$\sum_{i=1}^{n} g_i \{W_i - Q_i c\} \geq \beta.$$

A stronger downstream firm D is then one whose expected level of assets is higher as less need be left to U through bargaining. From inspection of D's objective it follows that efficient bargaining does not alter the generic structure of the optimal contract as the new problem is isomorphic to Program Bank – it remains the case that double marginalization is unavoidable. Further:

Proposition 5 *[Countervailing Power] Suppose that the curvature of the technology function is regular in the sense that it declines in magnitude at higher investment levels, i.e., $\pi'''(I) \geq 0$. Then, an increase in D's bargaining power (a smaller value of β) induces lower retail prices in the short run (period 0).*

If D's bargaining power rises, then U secures a lower return. This is equivalent to D gaining extra assets in addition to the income she makes through her normal business dealings in period 0. The extra assets for D then act in the same way as exogenous assets did in Section 6.1. Hence, the intuition explaining why an asset rich D is not very risk averse follows as above.

Proposition 5 provides a novel theory of countervailing power based on credit constraints: consumer prices are lower the larger is the credit-constrained downstream firm's bargaining power *vis-à-vis* her upstream supplier. The term "countervailing power" was coined by Galbraith (1952) but Snyder (2008) notes that formalizing the concept has proved difficult. Several theories of countervailing power (or buyer power) have recently been proposed in which upstream and downstream firms bargain. One strand of the literature builds on Katz (1987) and models bargaining as a supplier matching the price of some outside option. As such the question of bargaining power does not arise. A second influential strand has considered bilateral bargaining (Chipty and Snyder, 1999; Inderst and Wey, 2007). In this setting, the bargained transfer depends upon the expected incremental cost of supply – and this can differ between buyers of different size. However, without mandating inefficient bargaining, there is typically no retail price effect from changes in bargaining power. Here we are able to offer, to our knowledge, the first model of countervailing power based on credit constraints. Our model captures retail price effects within an efficient bargaining paradigm.

6.3 The Cost of Borrowing and Retail Prices

Now we study the effect of changes in the cost of capital, the interest rate payable by D, on the optimal supply contract. We show that an increase in the interest rate can be expected to increase D's endogenous risk aversion and thus lead to higher retail prices in the short run.

Suppose that money borrowed from the external banking sector between periods 0 and 1 needs to be repaid at an interest rate of r. As D has to pay back $(I-a)(1+r)$ to the bank, the no-shirking constraint in period 1 is now given by

$$BI + (I-a)(1+r) - \pi(I) = 0. \tag{15}$$

The equality follows as the downstream firm D is credit constrained and so will invest as much as she is able to. This relationship between investment and assets implies:

$$\frac{\partial I}{\partial a} = \frac{1+r}{B+1+r-\pi'(I)} \text{ and } \frac{\partial^2 I}{\partial a^2} = \pi''(I)\left[\frac{\partial I}{\partial a}\right]^3 \frac{1}{1+r}. \tag{16}$$

Hence, analogously to (14) we have

$$-\frac{\partial^2 I}{\partial a^2}\Big/\frac{\partial I}{\partial a} = -\pi''(I)\left[\frac{\partial I}{\partial a}\right]^2 \cdot \frac{1}{1+r}. \tag{17}$$

Proposition 6 *Suppose the curvature of the technology function, $-\pi''(I)$, is sufficiently large in magnitude and declining at higher investment levels (i.e., $\pi'''(I) \geq 0$). Then, an increase in the interest rate causes retail prices to rise in the short run (period 0).*

Invoking Lemma 4 on (17), the effect of interest rates on risk aversion can be established by assessing the implications of higher interest rates on the sensitivity of investment to assets $[\partial I/\partial a]$ and on the curvature of the investment returns function $[\pi''(I)]$. At high interest rates less can be invested for given pledgeable assets $(\partial I/\partial r < 0)$. As the curvature of the investment returns function declines with the investment level due to its regularity, higher interest rates raise the curvature of the investment returns function. That is $-\pi''(I)$ becomes larger. Next consider the sensitivity of the investment level to pledgeable assets. As interest rates rise the level of investment declines and the returns to investment at the margin are greater at lower investment levels due to the diminishing returns to investment. This acts to lower the incentive to shirk at the margin, so lowering the moral hazard problem and making investment more responsive to extra assets. However, as interest rates increase more needs to be repaid to the banks. This acts to raise the marginal incentive to shirk. Whether the investment level becomes more or less sensitive to extra assets depends on which effect dominates. A sufficient, but not necessary, condition for the first effect to dominate is that the investment returns function, $\pi(I)$, is sufficiently curved. In this case, increasing r would raise the sensitivity of investment to assets $[\partial I/\partial a]$. Finally, raising r shrinks the last term in (17). A sufficient (though not necessary) condition for the first two effects to dominate the third effect is again if the curvature of the investment returns function is sufficiently large. This then yields that increasing the interest rate r raises the coefficient of risk aversion of investment with respect to pledgeable assets. Proposition 6 thus obtains. An increase in the firm's cost of borrowing for investment results in higher retail prices in the pledgeable asset accumulation phase.

We have established a mechanism by which increasing interest rates which a firm must pay on future investments raises its aversion to risk in the asset accumulation phase and so results in price rises in the short term. Evidence in the macro literature exists for this effect at an aggregate level: the '*price puzzle*.' The price puzzle captures the empirical result that 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 after an increase in interest rates (Christiano et al., 1999). This is a puzzle as the standard Phillips curve explanation would dictate that prices should fall, not rise, if interest rates rise. There are other explanations for the price puzzle. For example Sims (1992) argues that the price puzzle is a statistical artifact which arises from not fully capturing inflation expectations.[25] Of relevance to us is that evidence exists that a price puzzle effect operates at more disaggregated industry levels (Gaiotti and Secchi, 2006). Proposition 6 provides a novel explanation, compatible with this evidence, for why retail prices might rise when the cost of borrowing a firm faces rises.[26]

7 Outsourcing

We have studied how an optimal supply contract has the upstream supplier U providing insurance to the downstream buyer D. An obvious question is whether the insurance can instead be provided by a third party. The answer comes in two parts. First, if the third party can verifiably observe the input supply, then D may decide to source the input from U at marginal cost c and separately secure insurance from the third party. However, the retail price implications are unchanged as the insurance would induce double marginalization for the same reason as before.

Second, if the third party cannot verifiably observe the input supply, then U and D would have an incentive to collude and under-report the supply of input from U to D. (Of course, this is not possible when U provides insurance.) This would prevent a third party from providing insurance to D. In this case, we obtain the following result:

Proposition 7 *[Outsourcing] The credit-constrained downstream firm D strictly prefers to outsource input production to U rather than produce in-house at the same cost.*

[25] This is disputed by Hanson (2004).

[26] Note that at the firm level increasing interest rates would only feed into higher short-run prices via an increase in firm costs if firms rent capital each period so that capital is a marginal cost of production. It is more standard to see capital as fixed in the short run. Our link between interest rates and prices operates regardless of the flexibility of the capital stock.

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 its payoff 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 $\partial\left[Q_i p\left(Q_i/z_i\right)\right]/\partial Q = c$ for all z_i. That is, the integrated firm would implement the non-double-marginalized retail price. However, by Proposition 1, Property 2, though implementable, this is not the optimal tariff when D is outsourcing input production to U. Hence, D strictly prefers outsourcing to U. ■

Our model thus provides a new rationale for credit constrained firms exposed to market risk to outsource supply: the suppliers can provide revenue insurance that a third party cannot to the same extent.

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) analyzing executive interviews and Sutcliffe and Zaheer (1988) experimentally find evidence that firms do move more production outside the firm when exposed to demand risk. 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).[27] Our model suggests a force pushing against integration, which is responsive to market risk.

8 Conclusions

In this paper, we analyze a model of vertical relations between a downstream buyer and her upstream supplier. The downstream buyer is endogenously credit constrained which means that the scale of her investment is constrained by the level of her pledgeable assets to be below first best. Assuming that the downstream buyer's investment technology exhibits diminishing marginal returns, the firm becomes endogenously risk averse when accumulating pledgeable assets.

As a result, the optimal contract between the (endogenously) risk-averse downstream firm and her upstream supplier involves risk sharing. This holds even if the downstream firm holds debt ex ante. Conditional on remaining in business, the downstream firm seeks revenue insurance as extra pledgeable assets have the biggest impact when pledgeable asset levels are low. However, such insurance comes at a cost to consumers in the form of higher prices. Demand-dependent repayments to the supplier raise the downstream

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

firm's effective marginal cost, inducing an increase in consumer prices. Thus double marginalization is a necessary feature of optimal supply contracts under credit constraints.

As supplier-insurance and lending for investment are subject to countervailing incentives, their pooling within one principal allows the downstream firm to reduce the double marginalization problem. So our model can explain why finance arms of major companies (such as GE) can lend profitably when banks cannot. Why such non-bank lending arrangements should exist and be thriving is currently not settled in the literature. Our model offers a contribution to this debate.

As the downstream firm's risk aversion is endogenous, it is affected by changes in market-level and firm-level parameters. A result which follows is that exogenously asset-rich firms, or firms with greater bargaining power, will have less need for risk-sharing contracts and so can be expected to set lower retail prices. Our model thus generates potentially testable cross-sectional price predictions. Further, we have a new theory of countervailing power as a more powerful downstream firm will set lower retail prices even with efficient bargaining.

Finally, our model predicts that if input supply is not verifiable, then risk-averse firms exposed to market risk will gain by outsourcing supply (or sales). Once outsourced, the firm can enact a value-enhancing supply contract with insurance features. The same insurance cannot be provided by a third party if input supply is not verifiable by that third party.

These results have all been demonstrated in a model of downstream credit constraints and demand-side risk. However, the results are more general and would apply analogously to a model of upstream credit constraints and supply-side risk. There is, to our knowledge, little current empirical evidence which directly isolates the impact of credit constraints on pricing levels; though evidence of risk sharing and double marginalization is widespread.

While the shape of the optimal contract between the credit-constrained (and thus risk-averse) downstream firm and her upstream supplier does not rely on the cause of the downstream firm's risk aversion, we would have been unable to obtain several of our results without explicitly modelling the interaction between the credit constraints and risk aversion. First, the complementarity between lending and insurance (giving rise to a theory of finance arms) obviously requires a role for credit constraints. Second, the theory of countervailing power is based on bargaining power altering the endogenous degree of risk aversion via the credit constraints. Finally the relationship between interest rates and short-run retail prices relies on the interest rate altering the endogenous degree of risk aversion via the credit constraints. In contrast, the results on double marginalization and slotting fees arise from the risk-sharing motive of the risk-averse downstream firm. The same is true for our results on outsourcing. The industrial organization literature has been skeptical about modelling firms as being risk averse. Our paper shows that firms are expected to be risk averse if they face investment projects with diminishing returns and there is some chance of the firms being credit-constrained.

A Proof of Proposition 2

Define $R(Q,z) = Qp(Q/z)$. We have that $R(Q,z) > 0$; $\frac{\partial R}{\partial Q} > 0$; $\frac{\partial^2 R}{\partial Q^2} < 0$; $\frac{\partial R}{\partial z} > 0$; $\frac{\partial^2 R}{\partial Q \partial z} > 0$. D's problem is given by Program Bank with $B \cdot I(\cdot)$ replaced by $V^{\text{debt}}(R(Q_i,z_i) - W_i - d)$ as given in (9). Following Hart (1983, Result 1) we confirm that Q_i must be weakly increasing in the state i. As D should not have an incentive to claim that the state is $i-1$ when, in fact, it is i, nor the reverse, (5) implies that $R(Q_i,z_i) - R(Q_i,z_{i-1}) \geq R(Q_{i-1},z_i) - R(Q_{i-1},z_{i-1})$. Taking a Taylor expansion we require $(z_i - z_{i-1}) \frac{\partial R}{\partial z}(Q_i,z_{i-1}) \geq (z_i - z_{i-1}) \frac{\partial R}{\partial z}(Q_{i-1},z_{i-1})$. As $z_i > z_{i-1}$, and using the fact that $\partial^2 R/\partial z \partial Q > 0$ we must have $Q_i \geq Q_{i-1}$.

Following Hart (1983, proof of equation 18), note that D's payoff must increase weakly in the state as

$$R(Q_i,z_i) - W_i - d \geq R(Q_{i-1},z_i) - W_{i-1} - d \geq R(Q_{i-1},z_{i-1}) - W_{i-1} - d. \tag{18}$$

The first inequality is (5), whereas the second inequality follows from $z_{i-1} < z_i$. This implies that D pays off her debt if the state of period-0 demand is high enough, and not otherwise. Let the lowest state at which D pays off her debt be k. Thus the optimal contract would have $a_i < \hat{\kappa}$ for states $i < k$, and $a_i \geq \hat{\kappa}$ for states $i \geq k$. Hence k satisfies

$$R(Q_{k-1},z_{k-1}) - W_{k-1} - d < \hat{\kappa} \leq R(Q_k,z_k) - W_k - d. \tag{19}$$

Combined with (9) we have that in any state $i < k$, D prefers her limited liability payoff of b and so shuts down before securing any input from U. For any other state $i \geq k$, we identify the binding local incentive compatibility constraints.

Lemma 5 *If the grid of probabilities is sufficiently fine then in the downstream firm's problem the incentive constraints (5) can be replaced by (19) and*

$$\begin{aligned} Q_i &\geq Q_{i-1} \text{ for all } i; &(20)\\ R(Q_i,z_i) - W_i &\geq R(Q_{i-1},z_i) - W_{i-1} \text{ for all } i > k; &(21)\\ R(Q_k,z_{k-1}) - W_k - d &< \hat{\kappa}. &(22) \end{aligned}$$

Proof. We have shown that (5) implies (20) and (19) follows from (18). (21) is immediate from (5). (22) is the condition for D to shut down in state z_{k-1} rather than report state k.

Now the reverse implication. Suppose we maximize the objective function in Program Bank subject to (4), and constraints (20), (21) and (22). We show the solution satisfies (5).

Using the structure of our model, we go beyond Hart (1983) and show that (21) is binding for all $i \geq k+1$. Suppose that this did not hold for some $i > k+1$ so that $R(Q_i, z_i) - W_i > R(Q_{i-1}, z_i) - W_{i-1}$. Consider D seeking more risk sharing by lowering W_{i-1} to $W_{i-1} - \varepsilon/g_{i-1}$ and raising W_i to $W_i + \varepsilon/g_i$ with ε sufficiently small that (21) for i continues to be satisfied. The constraint on $i-1 > k$ is satisfied as Q_{i-1} is unchanged while the transfer to U is reduced. U's participation constraint, (4), is unaffected. D's objective function changes by

$$\begin{aligned} & \frac{d}{d\varepsilon}\left[g_i V^{\text{debt}}\left(R(Q_i, z_i) - W_i - \frac{\varepsilon}{g_i} - d\right) + g_{i-1}V^{\text{debt}}\left(R(Q_{i-1}, z_{i-1}) - W_{i+1} + \frac{\varepsilon}{g_{i-1}} - d\right)\right]_{\varepsilon=0} \\ = & -V^{\text{debt}\prime}\left(R(Q_i, z_i) - W_i - d\right) + V^{\text{debt}\prime}\left(R(Q_{i-1}, z_{i-1}) - W_{i-1} - d\right) > 0. \end{aligned}$$

The inequality follows from (18) and the concavity of V^{debt}. A contradiction to the optimality of the contract. If (22) is not too tight, $R(Q_k, z_{k-1}) - W_k - d < \hat{\kappa}_-$, then the same proof as above yields the result for $i = k+1$.

Suppose however that such a deviation cannot deliver equality in constraint (21) for $i = k+1$ as (22) is too tight: $R(Q_k, z_{k-1}) - W_k - d = \hat{\kappa}_-$. In this case consider a deviation which transfers payments to U from state k to $k+1$ and allows D pledgeable assets of $\hat{\kappa}$ in state $k-1$ also. Thus we consider a deviation which sets $Q_{k-1} = Q_k$, changes the repayments to $\tilde{W}_{k-1} = \tilde{W}_k = W_k - \varepsilon$ and raises the state $k+1$ payment to $\tilde{W}_{k+1} = W_{k+1} + \eta$. For small ε a firm in state $k-2$ will not deviate to report state $k-1$ as $\partial R/\partial z > 0$. As the payments and quantities are the same in states k and $k-1$, (21) is satisfied for state k. The same is true at state $k+1$ as long as ε and η are small. For U to be indifferent to this change we require that

$$(g_{k-1} + g_k)\left[W_k - \varepsilon - cQ_k\right] + g_{k+1}\left[W_{k+1} + \eta - cQ_{k+1}\right] = g_k\left[W_k - cQ_k\right] + g_{k+1}\left[W_{k+1} - cQ_{k+1}\right],$$

or

$$g_{k-1}\left[W_k - \varepsilon - cQ_k\right] - \varepsilon g_k + g_{k+1}\eta = 0.$$

D will see a change in its objective function given by

$$\begin{aligned} & g_{k-1}V^{\text{debt}}(\hat{\kappa}) + \frac{d}{d\varepsilon}\Big[g_k V^{\text{debt}}\left(R(Q_k, z_k) - W_k + \varepsilon - d\right) \\ & +g_{k+1}V^{\text{debt}}\begin{pmatrix} R(Q_{k+1}, z_{k+1}) - W_{k+1} - d \\ -\frac{\varepsilon g_k}{g_{k+1}} + \frac{g_{k-1}}{g_{k+1}}\left[W_k - \varepsilon - cQ_k\right] \end{pmatrix}\Big]_{\varepsilon=0} \\ = & \; g_{k-1}V^{\text{debt}}(\hat{\kappa}) + g_k V^{\text{debt}\prime}\left(R(Q_k, z_k) - W_k - d\right) \\ & -(g_k + g_{k-1})V^{\text{debt}\prime}\left(R(Q_{k+1}, z_{k+1}) - W_{k+1} - d + \frac{g_{k-1}}{g_{k+1}}\left[W_k - cQ_k\right]\right). \end{aligned}$$

As g_{k-1} tends to zero, this expression is positive as V^{debt} is concave and payments increase in the state (by (18)). Hence for a sufficiently fine grid of probabilities the objective function cannot have been maximized by the given choice of k (equation (19)) and so we have a contradiction.

Using the inductive technique of Hart (1983, Result 1) one can establish that (21) implies that in any state $i \geq k$ the downstream firm would not wish to misreport the state as being any other $j \geq k$. Misreporting the state as $j < k$ would necessitate shutting-down. Hence (5) is established for states $i \geq k$.

Finally we demonstrate truthful behavior for states $i < k$. In such states we require D to shut down and receive the payoff of b. As $\partial R/\partial z > 0$, (22) ensures that misreporting state k will lead to a payoff below b and so is not preferable. Finally we show that deviating to misreport $k+1$ instead of k is even less desirable. As $R(Q_{k+1}, z_{k+1}) - W_{k+1} = R(Q_k, z_{k+1}) - W_k$, we have for $i < k$,

$$\begin{aligned} & [R(Q_k, z_i) - W_k] - [R(Q_{k+1}, z_i) - W_{k+1}] \\ = \; & [R(Q_{k+1}, z_{k+1}) - R(Q_{k+1}, z_i)] - [R(Q_k, z_{k+1}) - R(Q_k, z_i)] \\ = \; & (z_{k+1} - z_i) \left[\frac{\partial R}{\partial z}(Q_{k+1}, z_i) - \frac{\partial R}{\partial z}(Q_k, z_i) \right] \geq 0. \end{aligned}$$

The inequality follows as $\partial^2 R/\partial z \partial Q > 0$. Proceeding inductively ensures that the truth-telling condition (5) is satisfied at any state $i < k$. Hence, we have (5) satisfied in all states. ■

Having established Lemma 5 we can invoke the proof of Hart (1983, Proposition 2) to deliver double marginalization at all states $i \geq k$, and coinsurance with U.

B Proofs Of Section 6

Proof of Lemma 4. We aim to show that:

$$\begin{aligned} \text{if } Q_{i-1}^* \; &< \; Q_i^* < Q_{i+1}^*, \text{ then } \frac{\partial}{\partial \theta}\left[\frac{-\frac{\partial^2 I}{\partial a^2}}{\frac{\partial I}{\partial a}}\right] =_{\text{sign}} -\frac{\partial}{\partial \theta} Q_i^*(\theta) \text{ for all } i < n; \\ \text{if } Q_i^* \; &\in \; \left\{Q_{i-1}^*, Q_{i+1}^*\right\}, \text{ then } \frac{\partial}{\partial \theta} Q_i^*(\theta) = 0 \text{ or } \frac{\partial}{\partial \theta}\left[\frac{-\frac{\partial^2 I}{\partial a^2}}{\frac{\partial I}{\partial a}}\right] =_{\text{sign}} -\frac{\partial}{\partial \theta} Q_i^*(\theta) \text{ for all } i < n. \end{aligned}$$

We first characterize the optimal period 0 contract in some more detail. Result 1 in Hart (1983) shows that the set of incentive constraints in Program Bank can be replaced with the following set of local constraints:

$$Q_i \geq Q_{i-1} \text{ for all } i \in \{2, ..., n\}, \tag{23}$$

$$Q_i p\left(\frac{Q_i}{z_i}\right) - W_i \geq Q_{i-1} p\left(\frac{Q_{i-1}}{z_i}\right) - W_{i-1} \text{ for all } i \in \{2, ..., n\}. \tag{24}$$

(24) is satisfied with equality as shown by the proof of Lemma 5 in Appendix A. Now

express the optimal period 0 contract purely in terms of quantities $\{Q_i\}$. From (24), $W_i - Q_i c = [W_{i-1} - Q_{i-1}c] + \Delta\Pi_i$, where:

$$\Delta\Pi_i = Q_i \left[p\left(\frac{Q_i}{z_i}\right) - c \right] - Q_{i-1} \left[p\left(\frac{Q_{i-1}}{z_i}\right) - c \right].$$

The term $\Delta\Pi_i$ measures the industry profit gain if D does not lie and claim the state is marginally worse than it is (reporting $i-1$ instead of i). Iterating, we obtain $W_i - Q_i c = \sum_{j=2}^{i} \Delta\Pi_j + [W_1 - Q_1 c]$. From the individual rationality constraint for U:

$$\begin{aligned} 0 &= g_1 [W_1 - Q_1 c] + g_2 \{\Delta\Pi_2 + [W_1 - Q_1 c]\} + g_3 \{\Delta\Pi_3 + \Delta\Pi_2 + [W_1 - Q_1 c]\} \\ &\quad + \cdots + g_n \left\{ \sum_{j=2}^{n} \Delta\Pi_j + [W_1 - Q_1 c] \right\} \\ \Rightarrow &\quad -[W_1 - Q_1 c] = \sum_{k=2}^{n} \left(g_k \sum_{j=2}^{k} \Delta\Pi_j \right) = \sum_{k=2}^{n} \left(\Delta\Pi_k \sum_{j=k}^{n} g_j \right). \end{aligned} \tag{25}$$

Equation (25) gives W_1. Furthermore:

$$W_i - Q_i c = \sum_{j=2}^{i} \Delta\Pi_j - \sum_{j=2}^{n} \left(\Delta\Pi_j \sum_{k=j}^{n} g_k \right) \quad \text{for } i \geq 2. \tag{26}$$

Note that the second term on the right-hand side of (26) is independent of i.

We now discuss the pledgeable assets which will be available to D at the end of period 0, given any realization of the state. We have

$$a_i = Q_i p\left(\frac{Q_i}{z_i}\right) - W_i = Q_i \left[p\left(\frac{Q_i}{z_i}\right) - c \right] - \sum_{j=2}^{i} \Delta\Pi_j + \sum_{j=2}^{n} \left(\Delta\Pi_j \sum_{k=j}^{n} g_k \right).$$

First, we show that assets are increasing in the state:

$$\begin{aligned} a_{i+1} - a_i &= Q_{i+1} \left[p\left(\frac{Q_{i+1}}{z_{i+1}}\right) - c \right] - Q_i \left[p\left(\frac{Q_i}{z_i}\right) - c \right] - \Delta\Pi_{i+1} \\ &= Q_i \left[p\left(\frac{Q_i}{z_{i+1}}\right) - p\left(\frac{Q_i}{z_i}\right) \right] \geq 0. \end{aligned} \tag{27}$$

The inequality follows as $z_{i+1} > z_i$. Next, we consider $\partial a_i / \partial Q_l$ for some realization of risk i and some contracted quantity at state l. From the last equation:

$$\begin{aligned} \frac{\partial a_{i+1}}{\partial Q_l} - \frac{\partial a_i}{\partial Q_l} &= \frac{\partial}{\partial Q_l} \left\{ Q_i \left[p\left(\frac{Q_i}{z_{i+1}}\right) - p\left(\frac{Q_i}{z_i}\right) \right] \right\} \\ &= \begin{cases} 0 & \text{if } i \neq l, \\ \mathrm{MR}_{i+1}(Q_i) - \mathrm{MR}_i(Q_i) > 0 & \text{if } i = l, \end{cases} \end{aligned} \tag{28}$$

where $\mathrm{MR}_{i+1}(Q_i)$ is the marginal revenue in state z_{i+1}, evaluated at output Q_i. The final line follows as marginal revenue grows in higher demand states (see Footnote 8).

Hence, we have demonstrated that D's problem can be rewritten as: maximize $E[I(a,\theta)]$ over $\{Q_i\}$, subject to (23) only, with the transfers being determined by (25) and (26).

Now, we turn to period 1. Suppose that the model parameter is at the level θ_1 and the optimal contract is $\{Q_i^*(\theta_1)\}$. Consider some state $l < n$ and suppose that $Q_{l-1}^* < Q_l^* < Q_{l+1}^*$. In this case, $E[I(a,\theta_1)]$ is maximized with respect to Q_l^* as (23) is not binding. That is $E\left[\frac{\partial I}{\partial a}(a,\theta_1)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)} = 0$. Expanding, using (28), yields:

$$\left[\sum_{j=1}^{l} g_j \frac{\partial I}{\partial a}(a_j,\theta_1)\right]\frac{\partial a_l}{\partial Q_l} + \left[\sum_{j=l+1}^{n} g_j \frac{\partial I}{\partial a}(a_j,\theta_1)\right]\frac{\partial a_{l+1}}{\partial Q_l} = 0. \tag{29}$$

As the investment returns function $I(.)$ is increasing, and using (28), we must have $\partial a_l/\partial Q_l < 0 < \partial a_{l+1}/\partial Q_l$. Suppose that the model parameter rises slightly to $\theta_2 > \theta_1$.

$$\frac{\partial I}{\partial a}(a,\theta_2) = \frac{\partial I}{\partial a}(a,\theta_1)\left[1 + (\theta_2-\theta_1)\frac{\frac{\partial^2 I}{\partial a \partial \theta}(a,\theta_1)}{\frac{\partial I}{\partial a}(a,\theta_1)}\right].$$

We aim to sign $E\left[\frac{\partial I}{\partial a}(a,\theta_2)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)}$, which can be rewritten as

$$\begin{aligned} E\left[\frac{\partial I}{\partial a}(a,\theta_2)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)} &= E\left[\left\{\frac{\partial I}{\partial a}(a,\theta_2) - \frac{\partial I}{\partial a}(a,\theta_1)\right\}\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)} \\ &= (\theta_2-\theta_1) E\left[\frac{\frac{\partial^2 I}{\partial a \partial \theta}(a,\theta_1)}{\frac{\partial I}{\partial a}(a,\theta_1)}\frac{\partial I}{\partial a}(a,\theta_1)\frac{\partial a}{\partial Q_l}\right]. \end{aligned}$$

Expanding out and using (28), yields

$$\begin{aligned} E\left[\frac{\partial I}{\partial a}(a,\theta_2)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)} &= (\theta_2-\theta_1)\left[\sum_{j=1}^{l} g_j \frac{\frac{\partial^2 I}{\partial a \partial \theta}(a_j,\theta_1)}{\frac{\partial I}{\partial a}(a_j,\theta_1)}\frac{\partial I}{\partial a}(a_j,\theta_1)\right]\underbrace{\frac{\partial a_l}{\partial Q_l}}_{<0} \\ &+ (\theta_2-\theta_1)\left[\sum_{j=l+1}^{n} g_j \frac{\frac{\partial^2 I}{\partial a \partial \theta}(a_j,\theta_1)}{\frac{\partial I}{\partial a}(a_j,\theta_1)}\frac{\partial I}{\partial a}(a_j,\theta_1)\right]\underbrace{\frac{\partial a_{l+1}}{\partial Q_l}}_{>0}. \end{aligned}$$

Suppose that an increase in assets a reduces the Taylor quotient,

$$\frac{\partial}{\partial a}\left[\frac{\frac{\partial^2 I}{\partial a \partial \theta}(a,\theta_1)}{\frac{\partial I}{\partial a}(a,\theta_1)}\right] < 0. \tag{30}$$

As assets increase in the state (from (27)), we have

$$E\left[\frac{\partial I}{\partial a}(a,\theta_2)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)} < (\theta_2-\theta_1)\frac{\frac{\partial^2 I}{\partial a\partial\theta}(a_l,\theta_1)}{\frac{\partial I}{\partial a}(a_l,\theta_1)}\left[\sum_{j=1}^{l} g_j\frac{\partial I}{\partial a}(a_j,\theta_1)\right]\frac{\partial a_l}{\partial Q_l}$$
$$+(\theta_2-\theta_1)\frac{\frac{\partial^2 I}{\partial a\partial\theta}(a_l,\theta_1)}{\frac{\partial I}{\partial a}(a_l,\theta_1)}\left[\sum_{j=l+1}^{n} g_j\frac{\partial I}{\partial a}(a_j,\theta_1)\right]\frac{\partial a_{l+1}}{\partial Q_l}=0, \quad (31)$$

where the equality follows from (29). This therefore proves that $Q_l^*(\theta_2)<Q_l^*(\theta_1)$, and so retail prices would be higher under parameter θ_2 than θ_1 if (30) holds. The reverse result follows analogously by reversing the inequality if an increase in assets raises the Taylor quotient (i.e., if the inequality in (30) is reversed).

Hence, we have shown that if at state l with model parameter θ, $Q_{l-1}^*<Q_l^*<Q_{l+1}^*$, then

$$\frac{\partial}{\partial\theta}Q_l^*(\theta)=_{\text{sign}}\frac{\partial}{\partial a}\left[\frac{\frac{\partial^2 I}{\partial a\partial\theta}(a,\theta_1)}{\frac{\partial I}{\partial a}(a,\theta_1)}\right]=_{\text{sign}}-\frac{\partial}{\partial\theta}\left[\frac{-\frac{\partial^2 I}{\partial a^2}}{\frac{\partial I}{\partial a}}\right],$$

where the last equality follows algebraically. The last term is the rate of change of the coefficient of risk aversion and so proves result 1 of the lemma.

Now consider the case of pooling. We seek to modify the proof above to show that the pooled quantity falls weakly as we move to θ_2 if $\frac{\partial}{\partial a}\left[\frac{\frac{\partial^2 I}{\partial a\partial\theta}(a,\theta_1)}{\frac{\partial I}{\partial a}(a,\theta_1)}\right]<0$. Consider the largest pooled state z_l, where $Q_{l-1}^*=Q_l^*<Q_{l+1}^*$. Note that $l<n$ as we know that at state n, the efficient quantity $z_n q(c)$ is delivered, while at state $n-1$, there is strictly too little quantity: $Q_{n-1}^*<z_{n-1}q(c)<z_n q(c)=Q_n^*$. As we have $Q_{l-1}^*(\theta_1)=Q_l^*(\theta_1)$, the optimization over state l is constrained, so that $E\left[\frac{\partial I}{\partial a}(a,\theta_1)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)}\le 0$. If $E\left[\frac{\partial I}{\partial a}(a,\theta_1)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)}=0$, then the identical proof to above applies showing that $Q_l^*(\theta_2)\le Q_l^*(\theta_1)$. The inequality is weak as Q_l^* will only be able to fall if Q_{l-1}^* does. Suppose instead that $E\left[\frac{\partial I}{\partial a}(a,\theta_1)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)}<0$. The Taylor expansion around θ_2 is now given by

$$E\left[\frac{\partial I}{\partial a}(a,\theta_2)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)}=E\left[\frac{\partial I}{\partial a}(a,\theta_1)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)}+(\theta_2-\theta_1)E\left[\frac{\partial^2 I}{\partial a\partial\theta}(a,\theta_1)\frac{\partial a}{\partial Q_l}\right]_{Q_l^*(\theta_1)},$$

which is strictly negative for $\theta_2-\theta_1$ small. Hence, again we have $Q_l^*(\theta_2)\le Q_l^*(\theta_1)$. Finally, we obtain $Q_{l-1}^*(\theta_2)\le Q_l^*(\theta_2)\le Q_l^*(\theta_1)=Q_{l-1}^*(\theta_1)$, where the first inequality follows by (23), the second inequality has just been shown, and the equality follows by assumption. If, instead, $\frac{\partial}{\partial a}\left[\frac{\partial^2 I}{\partial a\partial\theta}(a,\theta_1)/\frac{\partial I}{\partial a}(a,\theta_1)\right]>0$, then consider the smallest pooled state and repeat the argument. ■

Proof of Proposition 4. Standard algebraic manipulations using (13) deliver that $\frac{\partial I}{\partial a}=\frac{\partial I}{\partial A}=1/[B+1-\pi'(I)]$, which is strictly positive by (3). Hence, differentiating (14) with respect to A then yields the result as $\pi'''\ge 0$, and it can be shown that $\partial^2 I/\partial A\partial a<0$. ■

Proof of Proposition 5. The bargaining program in which U requires a profit of β can be converted to Program Bank by adding β to the transfers in all states. This implies that the optimal contract can be found by reducing D's assets by β and proceeding as in Section 3. Hence the investment function the downstream faces, $I(a;\beta)$, is defined implicitly by:

$$IB = \pi(I) - (I - [a - \beta]) \tag{32}$$

and so the result follows from Proposition 4. ■

Proof of Propostion 6. Using (16) in (17) we have

$$\begin{aligned} \frac{d}{dr}\left[-\frac{\partial^2 I}{\partial a^2}\Big/\frac{\partial I}{\partial a}\right] &= -\frac{\partial I}{\partial r}\pi'''(I)\left[\frac{\partial I}{\partial a}\right]^2\frac{1}{1+r} \\ &\quad -2\pi''(I)\frac{\partial I}{\partial a}\frac{1}{1+r}\frac{\partial^2 I}{\partial a\partial r} + \pi''(I)\left[\frac{\partial I}{\partial a}\right]^2\frac{1}{(1+r)^2}. \end{aligned} \tag{33}$$

The first line is positive as $\pi''' \geq 0$ and investment levels shrink in the interest rate. Recalling that $\pi(I)$ is concave so that $\pi''(I) < 0$, the second line is equal in sign to $2(1+r)\frac{\partial^2 I}{\partial a\partial r} - \frac{\partial I}{\partial a}$. Using (16) we have

$$\frac{\partial^2 I}{\partial a\partial r} = \frac{1}{1+r}\frac{\partial I}{\partial a} - \frac{1}{1+r}\left(\frac{\partial I}{\partial a}\right)^2\left(1 - \pi''(I)\frac{\partial I}{\partial r}\right).$$

Hence the sign of the second line of (33) is given by

$$2(1+r)\frac{\partial^2 I}{\partial a\partial r} - \frac{\partial I}{\partial a} = \frac{\partial I}{\partial a}\left[1 - 2\frac{\partial I}{\partial a}\left(1 - \pi''(I)\frac{\partial I}{\partial r}\right)\right]$$

which is positive if $\pi''(I) < 0$ is sufficiently large in magnitude. Hence, an increase in the interest rate results in a larger coefficient of (absolute) risk aversion. ■

References

[1] Billett M.T., M.J. Flannery and J.A. Garfinkel, (1995), The Effect of Lender Identity on a Borrowing Firm's Equity Return, *Journal of Finance*, 50(2), 699-718.

[2] Bloom P., G. Gundlach and J. Cannon (2000), "Slotting Allowances and Fees: Schools of Thought and the Views of Practicing Managers," *Journal of Marketing*, 64(2), 92-108

[3] Bradley, M. and M. Roberts (2004), "The Structure and Pricing of Corporate Debt Covenants," SSRN working paper no. 466240.

[4] Brainard, W. C., J. B. Shoven, L. Weiss, P. Cagan, and R. E. Hall, (1980). The financial valuation of the return to capital, *Brookings Papers on Economic Activity* 1980(2): 453-511.

[5] Brander J. and T. Lewis (1986), "Oligopoly and Financial Structure: The Limited Liability Effect", *American Economic Review*, 76(5), 956-970.

[6] Breeden, D. and S. Viswanathan (1990), Why do firms hedge? An asymmetric information model, Working paper, Duke University.

[7] Burkart, M.and T. Ellingsen (2004), "In-Kind Finance: A Theory Of Trade Credit," *American Economic Review*, 94(3), 569-590.

[8] Camuffo A., A. Furlan and E. Rettore (2007), "Risk Sharing In Supplier Relations: An Agency Model For The Italian Air-Conditioning Industry", *Strategic Management Journal*, 28, 1257–1266.

[9] Carey M., M. Post and S. Sharpe (1998), "Does Corporate Lending by Banks and Finance Companies Differ? Evidence on Specialization in Private Debt Contracting," *Journal of Finance*, 53(3), 845-878.

[10] Carlton, D. (1979), "Vertical Integration in Competitive Markets Under Uncertainty," *Journal of Industrial Economics*, 27(3), 189-209.

[11] Chevalier, J. and D. Scharfstein (1996), "Capital-Market Imperfections And Countercyclical Markups: Theory and Evidence," *American Economic Review*, 86(4), 703-725.

[12] Chipty, T. and C. Snyder (1999), "The Role of Firm Size in Bilateral Bargaining: A Study of the Cable Television Industry," *Review of Economics and Statistics*, 81(2), 326-340.

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

[14] Cuñat, V. (2007), "Trade Credit: Suppliers as Debt Collectors and Insurance Providers," *Review of Financial Studies*, 20(2), 491-527.

[15] DeMarzo, P.M. and D. Duffie (1995), Corporate incentives for hedging and hedge accounting, *The Review Of Financial Studies*, 8(3), 743-771.

[16] Denis D.J., and V.T. Mihov, 2003, The choice among bank debt, non-bank private debt, and public debt: evidence from new corporate borrowings, *Journal of Financial Economics*, 70, 3-28.

[17] Figueiredo, P., G. Silveira and R. Sbragia (2008), "Risk Sharing Partnerships With Suppliers: The Case Of Embraer", Journal of Technology, Management and Innovation, 3(1), 27-37.

[18] Froot K., D. Scharfstein and J. Stein (1993), "Risk Management: Coordinating Corporate Investment and Financing Policies," *Journal of Finance*, 48(5), 1629-1658.

[19] Gaiotti, E. and A. Secchi (2006), "Is there a cost channel of monetary policy transmission? An investigation into the pricing behavior of 2,000 firms," *Journal of Money, Credit and Banking*, 38(8), 2013-2037.

[20] Galbraith, J. (1952), *American Capitalism: The Concept of Countervailing Power*, Boston: Houghton Mufflin.

[21] Hanson, M., 2004, The 'Price Puzzle' Reconsidered, *Journal of Monetary Economics*, 51, 1385-1413

[22] Harrigan, K. (1985), "Vertical Integration and Corporate Strategy," *Academy of Management Journal*, 28(2), 397-425.

[23] Hart, O. (1983), "Labour Contracts under Asymmetric Information: An Introduction," *Review of Economic Studies*, 50(1), 3-35.

[24] Holmstrom, B. and J. Tirole (1997), "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 112(3), 663-691.

[25] Hubbard, G. (1998), "Capital-Market Imperfections and Investment," *Journal of Economic Literature*, 36, 193-225.

[26] Inderst, R. and C. Wey (2007), "Buyer power and supplier incentives," *European Economic Review*, 51, 647-667.

[27] Innes, R.D. (1990), Limited liability and incentive contracting with ex-ante action choices, *Journal of Economic Theory*, 52(1), 45-67.

[28] James C. and D.C. Smith, (2000), Are Banks Still Special? New Evidence on Their Role in the Corporate Capital-Raising Process, *Journal of Applied Corporate Finance*, 13(1), 52-63.

[29] Jensen, M.C. and W.H. Meckling (1976), Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, *Journal of Financial Economics*, 3(4), 305-360.

[30] Katz, M. (1987), "The Welfare Effects of Third-Degree Price Discrimination in Intermediate Good Markets," *American Economic Review*, 77(1), 154-167.

[31] Klein, B. and J. Wright (2007), "The Economics of Slotting Contracts," *Journal of Law and Economics*, 50(3), 421-455.

[32] Leahy, J. V. and T. M. Whited (1996). The effect of uncertainty on investment: Some stylized facts, *Journal of Money, Credit and Banking* 28(1): 64-83.

[33] Lewis, T. and D. Sappington (1989), "Inflexible Rules in Incentive Problems," *American Economic Review*, 79(1), 69-84.

[34] Mahoney, J. (1992), "The Choice of Organizational Form: Vertical Financial Ownership Versus Other Methods of Vertical Integration" *Strategic Management Journal*, 13(8), 559-584.

[35] Panousi, P. and D. Papanikolaou (forthcoming), Investment, Idiosyncratic Risk, and Ownership, *Journal of Finance.*

[36] Preece, D.C., and D.J. Mullineaux, 1994, Monitoring by Financial Intermediaries: Banks vs. Nonbanks, *Journal of Financial Services Research*, 8 193-202.

[37] Rey, P. and J. Tirole (1986), "The Logic Of Vertical restraints," *American Economic Review*, 76(5), 921-939.

[38] Shin, H. and R. Stulz (1998), "Are Internal Capital Markets Efficient?" *Quarterly Journal of Economics*, 113(2), 531-552.

[39] Sims, C. A., 1992, Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review* 36(5), 975-1000.

[40] Smith, C.W., and R. Stulz, 1985, The determinants of firms' hedging policies, *Journal of Financial and Quantitative Analysis* 20, 391-405.

[41] Snyder, C. (2008), "Countervailing Power," in: The New Palgrave Dictionary in Economics, Second Edition, Eds. Durlauf and Blume. North Holland Press.

[42] Stulz, R. 1984, Optimal hedging policies, *Journal of Financial and Quantitative Analysis* 19, 127-140.

[43] Sudhir, K. and V. Rao (2006), "Do Slotting Allowances Enhance Efficiency or Hinder Competition?" *Journal of Marketing Research*, 43(2), 137-155.

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

[45] Tirole, J. (2006), *The Theory of Corporate Finance*, Princeton University Press.

[46] White, C., L. Troy and N. Gerlich (2000), "The Role of Slotting Fees and Introductory Allowances in Retail Buyers' New-Product Acceptance Decisions," *Journal of the Academy of Marketing Science*, 28(2), 291-298.