

Vertical Relations Under Credit Constraints*

Volker Nocke[†]

John Thanassoulis[‡]

University of Mannheim

University of Oxford

19 October 2009

Abstract

We model the impact credit constraints and market risk have on the vertical relationships between firms in the supply chain. As credit-constrained firms become endogenously risk averse, the optimal supply contract involves risk sharing, thereby inducing double marginalization and higher retail prices. The model gives rise to a new motive for outsourcing supply (or distribution). We identify an intrinsic complementarity between supply and lending, which can help explain the existence of finance arms of major suppliers, and a novel monetary transmission mechanism linking interest rates with short-run pricing that can help explain the *price puzzle* in macroeconomics.

Keywords risk aversion; vertical contracting; double marginalization; outsourcing; market risk; risk sharing; slotting fees; financial companies; finance arms; monetary transmission mechanism; price puzzle

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.¹ It is also well known that firms are subject to substantial market risk – whether on the demand side or supply side. Incorporating

*We would like to thank Meg Meyer, Jean Tirole, Mike Whinston, and seminar participants in Barcelona, Oxford, Leuven, Strathclyde, Warwick, the 2009 CEPR IO Conference in Mannheim, and the 2009 EARIE Conference in Ljubljana. We are also grateful to the ESRC for financial support (grant RES-000-22-3468).

[†]Other affiliations: University of Oxford, CEPR, and CESifo. Email: nocke@uni-mannheim.de.

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

¹See for example the Bank of England Credit Conditions survey available at: <http://www.bankofengland.co.uk/publications/other/monetary/creditconditions.htm>.

insights from the corporate finance literature 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, creating scope for finance arms, risk sharing, slotting fees, and outsourcing. Further, we identify a new monetary transmission mechanism from interest rates to the real economy which acts via firms which are at risk of becoming credit constrained.

Consider a firm exposed to demand-side risk which has some investment opportunities, and yet is credit constrained. To fund future investment assets must be accrued which will serve as collateral. These assets are amassed by business activities in the short run. Hence, as evidence corroborates, credit constrained firms' investment level is closely related to their cash flow (Gertler and Gilchrist 1994). Now suppose investment is subject to diminishing marginal returns. We show that this causes a risk neutral firm to be endogenously risk averse in its business activities in the short run. Low demand realizations limit the collateral which the firm can use for investment and so would result in very low investment returns. Thus risk aversion becomes an inherent feature of the firm's short run objective function. Moreover, the extent of risk aversion is endogenously determined by market parameters such as the interest rate and quality of corporate governance.

The endogenously risk-averse firm will seek some insurance from its vertical partners. Consider how this process operates between a credit-constrained firm and its input supplier. The downstream firm, exposed to market demand risk and credit constrained, wishes to pass risk back up to her supplier. So she demands a risk-sharing contract in which the supplier bears some loss for poor demand realizations. But for the supplier to recoup these potential losses she requires payments in high demand states to grow at a rate faster than cost. That is, double marginalization is introduced, causing the retail price of the firm to rise. The cost of the insurance made necessary by the credit constraints is in this sense partly paid for by final consumers.

To see why inefficient pricing must arise here note that the optimal supply contract cannot fully insure the downstream firm. Such full insurance would require the downstream firm to be paid the expected profit and pay an input price equal to the monopoly retail price to the supplier. However, if prices (or the quality of the shopping experience) are not contractible, then the downstream firm would be tempted to raise its retail price slightly and so make some variable profit on lower volumes. This is anticipated by the upstream firm, making double marginalization unavoidable.

Double marginalization occurs in many industries and its impact on myopic profits for the chain can be of the order of 10% (Mortimer, 2008). Why such inefficient contracts exist

at all is not very well understood. Some have argued that simplicity is the cause, others that incentives must be maintained to exert effort ('double moral hazard'; see Romano 1994). However, under the latter explanation, it is unclear why repeated interaction could not mitigate the shirking incentives. Here, we demonstrate that credit constraints alter the shape of the firm's payoff function, causing risk sharing and, hence, double marginalization to become inevitable.

The insurance service of vertical contracts we model is, we believe, reflected in at least two common business practices. Firstly, risk-sharing contracts are an apparently direct manifestation of our model in which the firm may receive explicit support for costs incurred which are repaid depending upon realized demand. Such contracts are, for example, common in the airline industry in which aeroplanes are supplied below cost, with the supplier recouping its costs by charging above cost for service.²

Secondly, slotting fees, a common practice in the grocery market as well as in other industries such as software and publishing, are well explained by our model. These fees are fixed payments many retailers require of manufacturers in return for stocking their products. Theoretical explanations for this practice have portrayed the slotting fee as a signalling device (Klein and Wright, 2007, and references therein). Empirical evidence suggests that an important part of the story is, however, the sharing of risk (Sudhir and Rao, forthcoming; White et al. 2000), which accords with our model.

Having established that a credit-constrained firm exposed to market risk derives insurance value from its vertical contracts, it follows that such a firm has an incentive to outsource supply (or sales) to a non-credit-constrained upstream supplier (or downstream buyer). The credit-constrained firm cannot insure itself whereas an upstream supplier (or downstream buyer) can monitor the volumes supplied to (or sold by) the firm and is therefore in a unique position to offer the valuable insurance. Thus we demonstrate a theoretical link between increased market risk and increased outsourcing. Our result is supported by empirical evidence (Harrigan, 1985; Sutcliffe and Zaheer, 1988) which points in this direction. 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).

It is standard to see the lending of loans and the supply of an input as separate practices offered by different firms. However, this need not be. In fact, we demonstrate (in Section 4) that there exists an intrinsic complementarity between the provision of insurance and lending. A supplier with access to funds at the same rate as the banking sector could

²A case study of the case of the Embraer jet is offered by Figuirodo et al. (2008).

actually lend on rates that the independent banking sector would find unprofitable. This result may offer an original insight into the existence and profitability of finance arms of major companies such as GE and Cisco. The comparative advantage of borrowing from non-bank financial companies is under current debate. In Section 4, we report that financial companies lend almost \$1 for every \$2 lent by a mainstream bank. Gaining an insight into what makes financial companies effective competitors for banks therefore seems a first-order issue.

The complementarity we find between insurance and lending arises because the supplier observes the pledgable assets required for the loan, *and* the volumes sold to make the pledgable assets. Thus the supplier can mitigate the firm's moral hazard problem associated with the risk-sharing contract and reduce the amount of double marginalization.³ Hence, linking supply to loans can result in higher industry profits, lower retail prices and greater average investment than achievable with a separate bank.⁴

We now turn to the question of how the optimal contract responds to changes in market parameters such as the interest rate or the quality of corporate governance. Section 5 takes up this question remaining within the fully optimal contracting framework. We demonstrate that the relevant measure of risk aversion is the absolute coefficient of risk aversion of the investment returns function with respect to pledgable assets. If market variables alter so as to increase (decrease) this measure of risk aversion, then retail prices will strictly rise (fall) in the short run in all risk realizations except those for which the optimal contract involves pooling, and except for the single best state ('no distortion at the top').

We are then in a position to demonstrate that a higher interest rate increases the relevant measure of risk aversion and so leads to an increase in short-run retail prices. This new monetary transmission mechanism is distinct from the seminal balance sheet channel of Bernanke and Gertler (1995). The standard interpretation of the balance sheet channel is that higher interest rates will raise firms' fixed costs and cause firms to run down inventories and reduce investment in the medium term (Ireland, 2005; see also Bank of England, 1999). We instead show that a change in the interest rate at which firms borrow alters the relative costs of bad demand realizations which in turn alters the amount of insurance required by the downstream firms. As this insurance demand from

³That suppliers see volumes whereas banks do not has been noted as an important feature of trade credit by Burkart and Ellingsen (2004). We here are not considering trade credit as the investments we consider are more naturally thought of as occurring in the medium term (and not the 30 to 60 days typical of trade credit).

⁴At a general level, it is known that long-term contracts may sometimes improve upon short-term contracts (see, for example, DeMarzo and Fishman, 2008). However, whether they do or not depends on the economic environment.

vertical partners is altered, so too is the retail price in the short run. This result is a new insight into the *price puzzle*: the macroeconomic link that has been noted between increases in the interest rate and increases in retail prices (Christiano et al., 1999).

All of these results are established in a benchmark model of demand side risk. However, this turns out to be inessential. We show that the same intuitions apply with supply-side risk. And neither is the take-it-or-leave-it framework we employ essential, as we demonstrate the same intuitions in bargaining extensions. These analyses are offered in Section 6 and 7, respectively. Section 8 concludes, with all omitted proofs contained in the Appendix.

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 decreasing returns assumption that gives rise to the firm becoming endogenously risk averse as it means that the rate at which the marginal dollar can be leveraged is decreasing. A related point, but in a model with exogenous credit constraints, is made by Froot et al. (1993). However, Froot et al. do not study the implications this insight has for vertical contracting and, hence, for pricing and the real economy. On the industrial organization side, Rey and Tirole (1986) note that if downstream retailers are (exogenously) risk averse, then exclusive territories have some rationale. They show that the best two-part tariff contract under exclusive territories 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, because the firm's risk aversion is an endogenous outcome in our model, we are able to show that there is an intrinsic complementarity between supplier insurance and lending (which may explain the existence of finance arms) and identify a new monetary transmission mechanism linking interest rates with short-run pricing (which can help explain the *price puzzle* in macroeconomics).

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 $p(Q/z)$.⁵ We assume that D is exposed to market risk in that market size z is a random variable with finite support $\{z_1, \dots, z_n\}$. 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 $0 < z_1 < z_2 < \dots < z_n$. The probability of state z_i is g_i , and $\zeta \equiv \sum_{i=1}^n g_i z_i$ is the expected value of z .

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

- (i) *Marginal revenue $d[Qp(Q/z)]/dQ$ is declining in quantity Q .*
- (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 $Q[p(Q/z) - c]$ is strictly concave in quantity Q . Moreover, it implies that, in demand state z , industry profit is maximized at quantity $Q = zq(c)$, where $q(c)$ is the unique solution in q to $p(q) + qp'(q) = c$. The downstream price that maximizes industry profit is $p(q(c))$ in every demand state z .

Before the demand state is realized, D offers U a contract of the form $\{Q(z_i), W(z_i)\}$, where $Q(z_i)$ is the input (and output) volume in state z_i , and $W(z_i)$ the associated transfer payment from D to U ; 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.) Then, D privately learns the realization of the demand state z and reports state \hat{z} to U . U , for its part, cannot verify the state of demand. D then receives $\hat{Q} = Q(\hat{z})$ units of input from U , transforms the input into a final good, and fetches a retail price of $p(\hat{Q}/z)$ per unit. Finally, D pays $W(\hat{z})$ to U . We assume for simplicity that D has no initial assets. D 's asset level by the end of period 0, a , is therefore given by D 's net profit in that period: $a = \hat{Q}p(\hat{Q}/z) - W(\hat{z})$.

Period 1. In period 1, D has to decide how much to invest in a project. Based on the moral hazard formulation offered by Holmstrom and Tirole (1997), we assume that D is endogenously credit constrained. Specifically, after choosing the investment level I , D 's owner-manager can choose whether or not to shirk at the investment stage. If he does not shirk, D makes a gross profit of $\pi(I)$. If he does shirk, instead, the investment project fails and yields a payoff of zero while the owner-manager receives a private benefit proportional to the size of the investment, $B \cdot I$, where $B \leq 1$.

If D wishes to invest more than its pledgable assets, $I > a$, it can choose to verifiably show its asset level a to an external banking sector so as to attempt to secure a loan of

⁵ D can equivalently be thought of as setting price p and facing demand $zQ(p)$.

$I - a$.⁶ 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

$$BI \leq \pi(I) - (I - a) \quad (1)$$

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

Assumption 2 *We make the following assumptions on the gross 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, $\pi'(0) > 1$, and $\pi'(I) < 1$ for I sufficiently large, so that the first-best level of investment, $\hat{I} \equiv \arg \max_I \pi(I) - I$, 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, $a < (B + 1)\hat{I} - \pi(\hat{I})$, so that the no-shirking constraint (1) is always binding in equilibrium.⁷*

3 Equilibrium Analysis

We solve the model by backward induction. Suppose D 's asset level at the beginning of period 1 is given by a . By Assumption 2(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 the unique solution in I to

$$BI = \pi(I) - (I - a). \quad (2)$$

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

$$1 < \pi'(I(a)) < 1 + B. \quad (3)$$

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-

⁶ D can always choose to hide some or all of its assets. As a result, D can only prove that it has *at least* the asset level that it reveals.

⁷The assumption that the no-shirking constraint is always binding is for convenience. What is really needed for our main results is that the constraint is binding in the worst demand state(s).

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 pledgable asset level a .*

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

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

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

This 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 its pledgable asset level a . To get some intuition, suppose first that D were not credit constrained. In period 1, it could therefore borrow \hat{I} so as to capture the first-best profit of $\pi(\hat{I}) - \hat{I}$, independently of the realization of a . D would therefore be risk-neutral with respect to end of period 0 assets a . However D is (endogenously) credit constrained. If D can get a loan from the banking sector, then $I(a) - a$ is positive. The positive marginal returns to investment implies that each extra dollar in pledgable income can be leveraged so that $I(a) - a$ is increasing in a . Since marginal returns are diminishing, the rate at which the marginal dollar can be leveraged is decreasing, implying that $d^2[I(a) - a]/da^2 < 0$.

The risk aversion will affect the agreement D requires from its supplier U . This will in turn affect the retail prices in period 0 (the “short run”) and the expected level of investment 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. We now determine how.

3.1 The Optimal Contract under Symmetric Information

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

at the investment stage. Hence, from equation (2), D 's problem becomes

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

subject to the individual rationality constraint for U ,

$$\sum_{i=1}^n g_i \{W_i - Q_i c\} \geq 0. \quad (4)$$

This program gives rise to the following solution:

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

Proof. Note first that U 's individual rationality constraint (4) must be binding since, otherwise, D could increase its payoff by offering slightly lower W_i 's without violating (4). The Lagrangian, which is to be maximized over $\{Q_i, W_i\}$, is given by

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

The remainder of the proof, which involves solving the set of first-order conditions, is straightforward and is therefore omitted. ■

Hence, the optimal contract has U bearing all the risk and D delivering a quantity which yields the price that maximizes industry profit in every state. Further, whatever the realization of risk, D completes period 0 with assets equal to the ex ante expected industry profit. U makes good any shortfall and confiscates any excess. So, in the full information case, consumers are unaffected by the market risk. The risk aversion created for D by the credit constraints is passed up to U and no inefficiency need be created.

3.2 The Optimal Contract under Asymmetric Information

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

If the state is z_i and D truthfully reports it, then she would receive a payoff of $BI(Q_i p(Q_i/z_i) - W_i)$. 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 $p(Q_j/z_i)$. This yields D pledgable income of $a = Q_j p(Q_j/z_i) - W_j$ at the end of period 0. Invoking the Revelation Principle, the optimal program therefore requires the pledgable income to be maximized when the truth is being told:

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

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

subject to the individual rationality constraint for U ,

$$\sum_{i=1}^n g_i \{W_i - Q_i c\} \geq 0, \quad (5)$$

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

$$Q_i p \left(\frac{Q_i}{z_i} \right) - W_i \geq Q_j p \left(\frac{Q_j}{z_i} \right) - W_j \text{ for all } j \neq i. \quad (6)$$

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 2 (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:*

$$\frac{\partial}{\partial Q} \left[Q_i^* p \left(\frac{Q_i^*}{z_i} \right) \right] > c \text{ for all } i < n. \quad (7)$$

Property 3 *D 's pledgable income increases in the state:*

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

Property 4 U 's payoff increases in the state:

$$W_i - Q_i^*c \geq W_{i-1} - Q_{i-1}^*c \text{ for all } i > 1. \quad (8)$$

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

By exploring a general input into a downstream firm D , we obtain important corollaries of the above proposition:

Corollary 1 *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.*
2. *The optimal contract has the supplier making payments to D which are not recouped in low demand states. Hence, if marginal cost c is sufficiently small, $W(z_i)$ is negative for small realized demand states z_i and positive for large z_i .*

Proof. For part 1, note that equation (7) 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 (8), increasing in i . Hence we must have some state j such that

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

⁸For 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. Here, D is endogenously risk averse (due to the interaction of the diminishing returns with the endogenous credit constraints), whereas in Hart, D is assumed risk averse.

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 quantity $z_i q(c)$ in state z_i , resulting in the retail price $p(q(c))$ that maximizes joint period-0 profit. Proposition 2 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 its supplier. It can be improved by requiring U to share in the risk faced by the downstream firm D . Intuitively, for U to provide such risk sharing, it must earn more in good states than in bad states (Property 4). Since U earns zero profit on average, it must make a loss in the worst state(s). Hence, we can think of U as providing a fixed payment to D , with D then making demand-dependent repayments.

In essence, D is using U to lower the variance of her end-of-period pledgable income by increasing the proportion which is fixed in advance. However, for U to be able to make back this ex ante committed payment 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 demand. As a result, the optimal risk-sharing contract induces retail prices that are (in almost all demand states) strictly higher than $p(q(c))$. Hence, some of the burden of credit constraints and market risk is borne by consumers.

Our general result appears to us to be reflected in at least two standard business practices: risk-sharing contracts and slotting fees.

Risk-sharing contracts. The payment made by U could be a financial transfer directly to D ; or a sharing in some costs with repayments dependent upon realized demand. These are widespread in many industries: the aircraft manufacturing industry being one example. In that industry, capital-intensive products such as aircrafts or their engines are delivered in return for a below-cost payment in addition to servicing restrictions which guarantee further servicing-related payments. This allows demand dependent repayments to be made via the servicing and maintenance charges (see Figueredo et al., 2008, for a case study of the Embraer jet). Indeed, many outsourcing contracts have taken on such risk-sharing features.⁹

Note that the vertical partner is being required to accept some risk which he is powerless to affect. This is optimal for the credit-constrained firm as the credit constraints, or rather the possibility of being credit-constrained after some demand realizations, cause her to be endogenously risk averse. Thus buyers forcing risk onto their suppliers is not

⁹See “Outsourcing” in *The Economist*, September 29, 2008.

necessarily an unwelcome side-effect of downstream dominance. This insight has relevance to antitrust suspicions of risk-shifting contracts demanded by large, or thought to be powerful, firms.¹⁰

Slotting Fees. We contend that our results are reflected in the common practice in the grocery market of slotting fees. Slotting fees are payments many supermarkets require of their suppliers. While a theoretical consensus has arguably emerged that slotting fees are to be thought of as part of a supplier signalling the quality of her product to the retailer (Klein and Wright, 2007), recent survey evidence suggests that risk sharing is a part of the rationale for slotting fees (Sudhir and Rao, forthcoming; Bloom et al., 2000). Our model provides the first, to our knowledge, model of risk-sharing or slotting fee contracts arising naturally outside of an exogenously imposed two-part tariff context.

If U is a supplier to a large supermarket, D , then it is natural to ask how the slotting fee result would be changed if U were herself credit constrained. An important contribution of this paper is to demonstrate that credit constraints manifest themselves as endogenous risk aversion (Lemma 1). As a result, if U were herself credit constrained in future investments then she would endogenously have a concave utility function as regards the payoff from the current period. Let us denote this utility function by $\mathcal{U}(\cdot)$. Now the downstream firm making take it or leave it offers to her supplier U would have to solve Program Bank with a revised participation constraint for U : $\sum_{i=1}^n g_i \mathcal{U}(W_i - Q_i c) \geq \mathcal{U}(0)$. One could then proceed analogously to above and invoke Hart (1983) once more as U 's utility function maps to risk averse workers in the Hart analysis. The result is that U is still obliged to bear risk and so a slotting fee type contract persists. However as U 's (endogenous) risk aversion becomes more extreme the extent of risk sharing diminishes and so the size of the slotting fee declines, vanishing in the limit of U becoming infinitely risk averse.

Next, consider double marginalization. Why double marginalization exists in the supply chain is a hotly debated topic as it is a common justification for antitrust interventions in business-to-business markets. If one imagines firms restrict themselves to linear supply contracts, then inefficient double marginalization is the natural result (Spengler, 1950). Linear contracts seem to be a common feature in some industries (see Crawford and Yurukoglu, 2008, for a cable TV example) – even though it would appear to be straightforward to contract around the double marginalization inefficiency, via franchise fees for example. The inefficiency created by double marginalization has been estimated to be as large as 10% of profits (Mortimer, 2008). A number of theories have been proposed to explain double marginalization (see Katz, 1989, for a survey). Two we would highlight are double moral hazard (Romano, 1994) and risk sharing between risk-averse firms.

¹⁰See Competition Commission (2008) for an example of this suspicion applying to the UK grocery market.

However, critics have noted that repeat interaction is likely to mitigate the moral hazard problem in the first explanation. Moreover, a convincing explanation for risk aversion has been missing with regard to the second explanation. Our work demonstrates that credit constraints (in conjunction with market risk) – widespread in the economy – can induce double marginalization in the short run.

Remark 1 *In our analysis, we have allowed for general contracts between the upstream supplier U and the downstream buyer B . 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,

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

3.2.1 Implications for Outsourcing

In our model, the upstream firm U provides (partial) insurance to its downstream buyer D . An obvious question is whether the insurance can instead be provided by a third party. The answer is, no, if the third party cannot verifiably observe the input supply (while, arguably, U can). Indeed, in this case, 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.) To the extent that the insurance cannot be provided by a third party, we obtain the following result:

Proposition 3 *The credit-constrained downstream firm D strictly prefers to outsource input production to the non-credit constrained supplier U rather than produce in-house at the same cost.*

Proof. Suppose D were to produce the input in-house at marginal cost c . In this case, in effect the supply contract would satisfy $W_i = cQ_i$ for all states i . Hence, for any demand state realization, the integrated firm would maximize 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 $\partial [Q_i p(Q_i/z_i)] / \partial Q = c$ for all z_i . That is, the integrated firm would implement the non-double-marginalized retail price. However, by Proposition 2, 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).¹¹ Our model suggests a force pushing against integration, which is responsive to market risk.

4 Complementarities between Supplier Insurance and Banking

In the model as presented so far, the supplier U offers her downstream buyer D some pledgable income 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 any pledgable 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. The reason is that, by having to return to U for a loan, D can commit to charge a lower price and therefore one which is less double marginalized. This is because if she under-reports 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. 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 deal with U . D would now be proposing the contract $\{Q_i, T_0^i, T_1^i\}$ where Q_i is delivered in period 0 if the state is z_i in return for payment of T_0^i (which is net of any ‘loan’). D then

¹¹Carlton (1979) offers the same conclusion but in a model of unadjustable input volumes.

invests her available assets and, after the investment returns are realized, she pays U an amount T_1^i that is again conditional on the period-0 demand state.

As U is offering the loan, she must ensure that the amount she makes available does not induce D to shirk at the investment stage. To achieve this, U can ask to be verifiably shown a given level of assets before providing the loan via T_0 . This limits the states that D can misreport. Suppose that the state is z_i but D reports z_j . U will expect D to be able to show a gross profit of $Q_j p\left(\frac{Q_j}{z_j}\right)$. However, D will only be able to do this if her actual gross profit, $Q_j p\left(\frac{Q_j}{z_i}\right)$, weakly exceeds this level. This is only possible if $z_j < z_i$. Thus, D can report only that the state is worse than it is – otherwise, she would be found out at the end of period 0. 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

$$\sum_{i=1}^n g_i \{T_0^i + T_1^i - Q_i c\} \geq 0, \quad (9)$$

$$\left[Q_i p \left(\frac{Q_i}{z_i} \right) - T_0^i \right] \cdot B \leq \pi \left(Q_i p \left(\frac{Q_i}{z_i} \right) - T_0^i \right) - T_1^i, \quad (10)$$

$$\pi \left(Q_i p \left(\frac{Q_i}{z_i} \right) - T_0^i \right) - T_1^i \geq \pi \left(Q_j p \left(\frac{Q_j}{z_i} \right) - T_0^j \right) - T_1^j \text{ for all } j < i. \quad (11)$$

Here, (9) is the individual rationality constraint for U , (10) is D 's no-shirking constraint at the investment stage in period 1, and (11) is D 's incentive constraint when reporting the state of demand 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 (T_0^j) is not altered. These extra assets cannot, therefore, be leveraged.¹²

Proposition 4 *Using U as a bank strictly dominates using a separate banking sector.*

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

¹²We assume here that any such extra assets could still be invested, although not leveraged. Assuming otherwise would only strengthen our result.

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

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

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 (9), the individual rationality constraint of U , is satisfied with equality by (5). By construction of T_1^i , the credit constraint is binding in every state so that the no-shirking constraint (10) 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 (6). The final term is the return available to D if her pledgable assets are $Q_j^* p \left(\frac{Q_j^*}{z_i} \right) - W_j^*$ and she borrows to the point at which the credit constraint binds. We wish to show that this level of borrowing is greater than T_1^j for $j < i$. This is true if and only if having assets of $Q_j^* p \left(\frac{Q_j^*}{z_i} \right) - W_j^*$ and borrowing T_1^j (resulting in investment equal to the level in the right hand side of (11)) 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 $\pi' > 1 \geq B$, 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 pledgable 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 (11) is actually slack (satisfied strictly).

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 (11) for small $\varepsilon > 0$. U remains indifferent, thus continuing to satisfy (9) with equality. By definition of T_1 , (10) is satisfied with equality. It therefore remains to note that the objective function has grown. This follows as, by Property 2 of Proposition 2, the marginal revenue at states below n exceeds c . ■

Proposition 4 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 as long as the loan is in part used to purchase goods from the same firm. This is not because the supplier is trying to buy business. But rather the ability to compare collateral with the volumes purchased can limit the extent to which the firm can double marginalize and so results in a Pareto improvement.¹³

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 consumers or for real-estate assets.¹⁴ 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) offer evidence that financial companies are over-represented (as compared to banks) in lending to those leasing equipment. This fact links in closely with the insurance function of contracts with suppliers which has been a key part of our analysis. Carey et al. also note that finance companies are over-represented in loans to higher risk firms. Carey et al. are, however, unable to refute or confirm the leading theories for this: that banks either cannot lend to risky borrowers due to regulation, or will not as they prefer not to damage a reputation for being willing to renegotiate contract

¹³An anecdotal example of this effect is the CEO of GE who is reported as believing that the combination of a finance arm with their other products is superior to rival good suppliers with no finance arm. (*Financial Times*, “GE extends its global reach to the Middle East”, July 23, 2008.)

¹⁴This is drawn from the Federal Reserve G20 statistical release. Available at <http://www.federalreserve.gov/econresdata/releases/statisticsdata.htm>.

terms in the event of covenant default. However, in the Carey et al. data, it is evident that financial companies lend to all risk types - and regular banks lend *less* to risky borrowers and comparatively more to safer borrowers. Interpreted this way, our result matches the empirical finding. 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.

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. In the data on loans analyzed by Carey et al., almost half (44%) of all loans are provided by multiple lenders. While most commonly these consortia comprise only banks, Carey et al. report that the second-most frequent form of consortium is composed of both banks and finance companies. Thus situations in which a supplier such as U teams up with banks to make loans is not uncommon. Our mechanism does require the supplier/lender to be able to commit not to leverage assets gained through excessive double marginalization. To achieve this, the borrowing firm must be limited in its access to further lenders for top-up loans. Covenants could be written to this effect. Indeed, there is evidence that, if lending is undertaken by a consortium, then covenants are more likely to be required (Bradley and Roberts, 2004). The lender must also be able to resist calls to renegotiate.

5 Interest Rates and Short-Run Prices

Above, we have shown that the interaction between credit constraints and market risk causes a risk-neutral firm to become endogenously risk averse with respect to its pledgable income. The endogenous risk aversion causes the firm to seek to push risk on to its vertical partners. How risk averse the firm is will depend upon market parameters. For example, the anticipated interest rate payable on future investment will alter the relationship between pledgable assets and investment levels and so impact on the extent of endogenous risk aversion. Similarly changes in the quality of monitoring or of corporate governance (the ability to shirk) will alter the level of the credit constraint and so impact endogenous risk aversion and, hence, period-0 contracts. In this section, we will demonstrate that the mechanism identified above generates a channel between the interest rate payable on borrowed sums and the real economy through short-run pricing and long-run investment decisions.

As the interest rate payable on borrowed sums rises, we show that – under some conditions – the relevant measure of D 's risk aversion increases. This implies that D seeks to increase the amount of insurance it secures from U . But better insurance exacerbates the double marginalization problem, thereby causing the short-run retail price to rise. Due to the increased double marginalization, the expected level of pledgable income declines. As a result, the expected long-run investment level declines also.¹⁵

This is a new monetary transmission mechanism between interest rates on firm borrowing and short-run retail pricing. The mechanism is distinct from the seminal balance sheet channel of Bernanke and Gertler (1995). In the balance sheet channel, the existing debt position of firms is worsened as their repayments rise and their net worth falls. This leads to the running down of inventories and a reduction in investment in the medium term.¹⁶ In our model, the interest rate change leads to an alteration in the profit-maximizing short-run behavior of the firm in anticipation of the risk associated with achieving the necessary pledgable income for her future finance needs. The mechanism identified here therefore impacts short-run retail pricing decisions.

Explicitly, we 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 is credit constrained, she will borrow as much as her end-of-period-0 assets allow. The maximal investment level given assets a is denoted $I(a, r)$ and implicitly defined by the equation

$$IB = \pi(I) - (I - a)(1 + r). \quad (12)$$

We must now pose the question of how a change in the interest rate r will alter the optimal contract between the credit-constrained D and U in period 0, and thereby have short-run effects on the real economy. We have established that the credit constraints make D risk averse. Thus the challenge is to determine what the relevant measure of risk aversion is and how it depends on r . Lemma 2 demonstrates that the relevant measure for the fully optimal supply contract is the Arrow-Pratt coefficient of absolute risk aversion of the investment returns function with respect to pledgable assets: $-\frac{\partial^2 I}{\partial a^2} / \frac{\partial I}{\partial a}$.

If an increase in the interest rate r , or indeed a change in any other variable such as monitoring/corporate governance (B) causes this coefficient to rise (fall), then the period-0 retail price will rise (fall). As we are working with the optimal period-0 contract, this effect is not an artefact of a restricted contract class (such as linear or two-part

¹⁵Over the course of the Financial Crisis firms have faced historically higher borrowing rates. For two years from July 2007, the spread of corporate debt as compared to US treasuries has climbed to levels far in excess of anything experienced over the previous 3 years. Our model predicts a link between the increased cost of borrowing and higher (than myopically optimal) retail prices.

¹⁶See the references in the Introduction and Tirole (2006) for a model.

tariff contracts), and so cannot be contracted around. Thus the retail price effects are unavoidable.

Lemma 2 *Suppose a shift r of the model parameters causes the absolute coefficient of risk aversion of $I(a; r)$ with respect to pledgable assets $\left(-\frac{\partial^2 I}{\partial a^2} / \frac{\partial I}{\partial a}\right)$ to change. Then:*

1. *Consider 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 moves strictly in the opposite direction to risk aversion. Hence, the short-run retail price in such states changes in the same direction as the coefficient of risk aversion.*
2. *The result holds weakly at state $i < n$ if the optimal contract in that state involves pooling, $Q_i^* \in \{Q_{i-1}^*, Q_{i+1}^*\}$.*

Lemma 2 demonstrates that if any alteration in market parameters can be linked to a change in the absolute coefficient of risk aversion of the investment function $I(\cdot)$, then period 0 prices will respond. An increase in endogenous risk aversion will exacerbate the double marginalization problem as more risk is shifted to the vertical partners. And this necessarily leads to higher prices in the short run in all states except the highest and those at which the optimal contract involves pooling. In this section, we focus on the interest rate.

To study this interest rate channel, recall that the no-shirking constraint at the investment stage is given by

$$h(I, a, r) \equiv BI + (I - a)(1 + r) - \pi(I) \leq 0.$$

Thus, $h(I, a, r)$ measures the “incentive to shirk” at the investment stage. Consider the partial derivative of this term with respect to the investment level:

$$\frac{\partial h(I, a, r)}{\partial I} = B + 1 + r - \pi'(I) \equiv \gamma(I, r).$$

We may call $\gamma(I, r)$, which measures how the incentive to shirk changes with the investment level, the “marginal incentive to shirk”.

By Lemma 2, an increase in the interest rate r will lead to higher retail prices if it increases the coefficient of (absolute) risk aversion of investment returns with respect to pledgable assets. This happens if and only if the induced fractional change in the curvature of I w.r.t. a is larger than the induced fractional change in the slope of I w.r.t.

a , i.e., if and only if

$$\frac{d}{dr} \ln \left(-\frac{\partial^2 I(a, r)}{\partial a^2} \right) > \frac{d}{dr} \ln \left(\frac{\partial I(a, r)}{\partial a} \right),$$

where (using (12))

$$\frac{\partial I(a, r)}{\partial a} = \frac{1+r}{\gamma} \quad \text{and} \quad -\frac{\partial^2 I(a, r)}{\partial a^2} = \frac{(1+r)(d\gamma/da)}{\gamma^2} = \frac{\partial I(a, r)}{\partial a} \frac{d\gamma/da}{\gamma} \quad (13)$$

We thus have

$$\frac{d}{dr} \ln \left(-\frac{\partial^2 I(a, r)}{\partial a^2} \right) = \frac{d}{dr} \ln \left(\frac{\partial I(a, r)}{\partial a} \right) + \frac{d}{dr} \ln \left(\frac{d\gamma/da}{\gamma} \right).$$

Hence,

$$\frac{d}{dr} \ln \left(-\frac{\partial^2 I(a, r)}{\partial a^2} \right) > \frac{d}{dr} \ln \left(\frac{\partial I(a, r)}{\partial a} \right) \quad \text{if and only if} \quad \frac{d}{dr} \ln \left(\frac{d\gamma/da}{\gamma} \right) > 0.$$

The term $(d\gamma/da)/\gamma$ is the fractional change in the marginal incentive to shirk induced by an increase in the pledgable asset level. An increase in the interest rate r will thus raise the coefficient of risk aversion (and, by Lemma 2, retail prices) if it makes the marginal incentive to shirk more sensitive to the pledgable asset level. The following proposition shows that this is indeed the case, provided the investment return function $\pi(\cdot)$ is sufficiently curved (and the curvature does not increase with I).

Proposition 5 *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:*

1. [cf. the price puzzle] retail prices to rise in the short run (period 0);
2. the expected level of investment to decline in the long run.

Part 1 of the proposition holds if an increase in the interest rate makes the marginal incentive to shirk more sensitive to the pledgable asset level:

$$\frac{d}{dr} \ln \left(\frac{d\gamma/da}{\gamma} \right) =_{\text{sign}} \left(\frac{d^2\gamma}{dadr} \right) \gamma - \frac{d\gamma}{da} \frac{d\gamma}{dr} > 0.$$

As the interest rate rises, the marginal incentive to shirk, γ , changes by $\partial[r - \pi'(I)]/\partial r$. On the one hand, increasing r lowers the level of investment for any given level of assets, $\partial I/\partial r < 0$. As the investment return function $\pi(\cdot)$ is concave, this means that the

marginal return to investment rises and so the marginal gain from investing rises. This acts to reduce the marginal incentive to shirk. However, as r rises, a larger fraction of the return must be paid back to the bank which increases the marginal incentive to shirk. If $\pi(\cdot)$ is sufficiently curved, then the first effect dominates so that $d\gamma/dr < 0$.¹⁷ In this case the denominator of the fractional change in the marginal incentive to shirk induced by an increase in the pledgable asset level falls as interest rates rise.

We turn now to the numerator of the fractional change in the marginal incentive to shirk induced by an increase in the pledgable asset level. An increase in the interest rate makes pledgable assets more valuable, thereby increasing the asset-sensitivity of the marginal incentive to shirk: $d^2\gamma/dadr > 0$. To see this note that $d^2\gamma/dadr = \frac{d}{dr} \left[-\pi''(I) \frac{1+r}{\gamma} \right]$ which captures the link between the investment technology and the marginal incentive to shirk. As the interest rate r increases, the marginal incentive to shirk shrinks ($d\gamma/dr < 0$) by the reasoning above whereas the term $1+r$ becomes larger. The term $-\pi''(I)$ increases weakly in r as raising r causes the level of investment I to decline and so the curvature of the technology function becomes weakly more negative.

Taken together, these effects imply that (under the conditions of the proposition), the numerator of the fractional change in the marginal incentive to shirk induced by an increase in the pledgable asset level increases with r , while the denominator shrinks. Hence, the fraction grows with r : $d \ln([d\gamma/da]/\gamma)/dr > 0$. This implies that an increase in the interest rate results in a greater coefficient of risk aversion and thus in higher retail prices.

The fact that investment is lower in the long run now follows as a corollary. For any given realization of assets we have $\partial I/\partial r < 0$ as the increased payback required lowers the level of borrowing which can be sustained. Further, realized assets are lower for any realization of period 0 market demand due to the increased double marginalization established in the first part of the proposition. Hence, the expected investment level must decline.

The first part of Proposition 5 is closely related to the price puzzle. The price puzzle refers to a long-standing observation in macroeconomics that retail prices appear to rise in the short run when interest rates are raised by the central bank. This is contrary to macroeconomic textbook discussions of the Phillips curve. Textbook macroeconomics would suggest that higher policy interest rates should raise the real interest rate for business investment decisions and so lead to a reduction in investment in the economy.

¹⁷A sufficient condition for this is that

$$-\frac{\partial^2 \pi}{\partial I^2} \geq \frac{(B+r)(1+r)}{\pi(0)}$$

This, it is argued, would shrink output below the economy's natural equilibrium level. In the simplest rendition of the theory this output gap puts downwards pressure on wages and hence on prices. Thus higher policy interest rates would be expected to lead to price falls. However, before this macroeconomic effect occurs, prices (aggregated into an economy-wide price level) seem to first rise for a number of months to a year by a statistically significant amount (Christiano et al., 1999). The exact size of the price puzzle is in dispute as it varies depending on the extent to which the empirical estimation seeks to control for the link between interest rates and inflation expectations (Balke and Emery, 1994). But the existence of the price puzzle has become a broadly accepted stylized fact. Proposition 5 provides a novel explanation of the price puzzle grounded in optimizing firm behavior.

It might be argued that simpler mechanisms are likely to link increasing interest rates with higher retail prices. One such argument might be that if firms rent capital each period an increase in the interest rate payable directly raises firms' marginal cost of production, resulting in higher retail prices. This argument is sensitive to whether the capital stock is fixed or variable in the short run. If the capital stock is fixed in the short run, then interest rate changes would only affect the fixed costs of operation and not the retail prices. And further on its face the Phillips curve approach would seem to rely on some unresponsiveness of the capital stock in the short run. We remain agnostic on the degree of flexibility of capital. We merely note that the link between interest rates on borrowing and retail pricing we have determined in our model operates regardless of the flexibility of the capital stock.

The monetary transmission mechanism we have determined is a cost-push explanation for the price puzzle. Barth and Ramey (2001) propose that interest rate rises might increase costs and so explain the price puzzle. Indeed, these authors show that in aggregate data the price puzzle has a number of features which suggest that the effect is driven by a supply side (cost) shock rather than a demand shock. Gaiotti and Secchi (2006) extend the empirical analysis of the relationship between interest rates and prices using firm level data from Italian Manufacturing. They confirm that a statistically significant cost push channel exists. That is increasing interest rates pushes up retail prices. They find that the effect is most pronounced for firms which carry large amounts of working capital. The question of whether the size of the working capital buffer required is related to the volatility of profits is not discussed in this work. Indeed if profits were not volatile then receipts from sales in the previous period would presumably remove the need for a large capital buffer and so cut the link between interest rates and prices. Hence, we conclude by noting that recent research is supportive of a cost push channel for monetary policy.

Our work offers a consistent hypothesis as to why this might be so.

6 Supply-Side as Opposed to Demand-Side Risk

The analysis so far has modeled market or demand risk relating to the profitability of the retail market for the produced goods. In this setting we have shown that market risk and credit constraints interact to produce an endogenously risk averse firm which seeks to share some of the risk, at a cost, with its vertical partners. Thus we have discovered how retail prices are pushed up; and how market risk creates an incentive to outsource supply processes off to a financially unconstrained vertical partner.

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

In this section, we confirm the results derived earlier in this paper. Whether risk is on the supply side or demand side, the party exposed to the risk seeks to push some of it onto her vertical partners. This creates a rationale for outsourcing of input manufacturing or distribution. This comes at the cost of higher retail price and thus lower consumer welfare.¹⁸

Formally, we consider the permutation of our baseline model to place risks on the upstream costs and not on the downstream demand. In state $i \in \{1, 2, \dots, n\}$, which occurs with probability h_i , U 's marginal cost of producing the input is κ_i . As before, D can transform the input into a final good at zero cost. But in contrast to the baseline model, (inverse) market demand $p(Q)$ is state-independent. Let $q(k_i)$ denote the joint period-0 profit-maximizing output level in state i : $q(k_i) \equiv \arg \max_q [p(q) - k_i]q$. We suppose that U is the party who is in a position to conduct an investment in period 1 and who is credit constrained. Thus U is impacted directly by the risks over the input costs. We suppose that U makes take-it-or-leave-it offers to a now risk-neutral D .

It is immediate that Lemma 1 applies to U . Thus the credit constraints in period 1 combined with input cost risk in period 0 cause U to be endogenously risk averse in period 0. If D could verify the input costs faced by U in period 0 then, analogously to Section

¹⁸Carlton (1977, 1979) analyzes input price risk and finds a theoretical rationale for increased vertical integration as market risk grows. Carlton's work is however underpinned by the assumption that the supplier must commit to production levels before the market parameters are revealed. Thus if demand is weak, the input supplier must throw some production away. This risk of loss causes the input supplier to charge above marginal cost prices - an inefficiency which, Carlton argues, can be mitigated by vertical integration.

3, one can show that the optimal contract has U requiring D to bear all of the risk. We therefore consider the asymmetric case in which U knows the true input costs but D does not. That is, suppose that D cannot verify the input costs faced by U . In selecting a contract $\{Q(\kappa_i), W(\kappa_i)\}$ in period 0, U must solve:

Program U Constrained The optimization program when U uses an independent banking sector is given by

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

subject to the individual rationality constraint for D ,

$$\sum_{i=1}^n h_i \{Q_i p(Q_i) - W_i\} \geq 0, \quad (14)$$

and the incentive compatibility constraint for U ,

$$W_i - Q_i \kappa_i \geq W_j - Q_j \kappa_i \text{ for all } j \neq i. \quad (15)$$

The analysis of this model proceeds in a similar, but not identical manner, to that of our benchmark model. The difference arises as D (the party accepting the contract now) is risk neutral in the transfer payment - but not in the quantity delivered. Hence, the proof used by Hart (1983) is not directly applicable. The proof is amenable after an adaptation which is outlined in Appendix A.1:

Proposition 6 *Suppose cost states are ordered so that $\kappa_1 < \kappa_2 < \dots < \kappa_n$. The solution to Program U Constrained has the following properties:*

Property 1 *No distortion in lowest cost state: $Q_1 = q(\kappa_1)$.*

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

$$Q_i \leq q(\kappa_i) \text{ for all } i > 1.$$

Property 3 *U 's pledgable income is higher in lower cost states:*

$$W_i - Q_i \kappa_i > W_{i+1} - Q_{i+1} \kappa_{i+1} \text{ for all } i < n.$$

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

$$Q_i p(Q_i) - W_i \geq Q_{i+1} p(Q_{i+1}) - W_{i+1} \text{ for all } i < n.$$

Proposition 6 indicates that the intuitions in the benchmark model where risk was attached to the demand side apply also when risk is attached to the supply side. The optimal contract involves some risk sharing with the partner in the supply chain and so all firms' ex post profits depend on the realization of the risk (Proposition 6, Properties 3 and 4). Further, in the case of supply-side risk, credit constraints cause the retail price to be overly responsive to increases in the costs of production. Thus retail prices rise faster than an integrated seller would require in the face of rising costs (Proposition 6, Property 2). So, for example, if raw material costs were to be volatile then when they are high the retail price would be pushed even higher than apparently justified by the cost rise due to the credit constraints.

When the risk applies to the supply side, the upstream firm U gains by having the ability to share risk with the retailer. This is achieved by a franchise fee type contract (Proposition 6, Property 4). This is the analogue of the risk sharing contract under demand side risk. Here the downstream firm pays a demand independent amount to the supplier which she only recoups in good demand states.

Once again, the outsourcing of the retail channel to a non-credit constrained firm is valuable to the credit-constrained upstream firm. The insurance that D offers to U would not be possible if U could sell to an alternative retail outlet which bypassed D without D being aware. Such area exclusivity is often a feature of franchise agreements.

7 Bargaining in the Supply Chain

The assumption of take-it-or-leave-it offers forms a polar case of bargaining in which one party has all the bargaining power. It is natural therefore to ask how the supply chain agreements we have studied might be affected if bargaining power were more equally shared. In this section, we construct such a model and show that the qualitative features of the supply chain contracts are unaffected: risk sharing and double marginalization remain (though, of course, the surplus from the interaction is more equitably divided).

In considering bargaining one's first impulse might be to reach for the celebrated Nash Bargaining Solution (NBS). This sets the percentage gain to the parties from a small change in the agreed contract to be equal. However, in the context of our model, this is a difficult construct to work with analytically as the total surplus to be split depends upon the contract agreed. Instead, we opt for a more simple approach to study shared bargaining power. Common with the NBS we assume that bargaining selects a point on the Pareto frontier so that money is not left on the table. We assume that the bargaining power of D as compared to U is captured by the invariant parameter θ where the agreed

solution splits the total rents available θ parts to D and $1 - \theta$ parts to U . If the size of the pie being bargained over were to be invariant to the agreement, this solution would match the weighted NBS. This approach is not new to the literature - it is known as the proportional bargaining solution (Kalai, 1977).¹⁹

The main impact of bargaining modeled in this way is that D receives proportion θ of the whole pie. Hence her incentive to invest is reduced - but the qualitative features displayed in the analysis remain. More formally, consider the following extension to the benchmark model offered in Section 2. We suppose that, in period 1, D requires the input from U to generate the return from investment. This is compatible with the benchmark model: if D could make take-it-or-leave-it offers to U , then she would demand the input at cost as D is risk neutral in period 1 following the investment. If, however, U has some bargaining power, then she will expect some rents from period-1 trade.

We further maintain the assumption of the benchmark model that only short-run contracts can be negotiated. Thus when bargaining in period 0, the parties are aware that there will be bargaining in period 1 and they can anticipate its outcome. Moreover, we assume that when bargaining in period 1, D 's actual investment level is observed by U (but is not verifiable so that the period-0 contract cannot condition on this outcome). This structure is standard in the analysis of bargaining and hold up. Hence to understand the impact of bargaining on our results we must solve for the sequence of bargaining outcomes in periods 0 and 1.

In period 1, the total pie available to the parties is $[\pi(I) - I + a]$, where a are D 's pledgable assets and I the total investment which can be made. In period 1, D will bargain to receive proportion θ of the surplus. The credit constraint equation becomes $BI = \theta [\pi(I) - I + a]$ which implicitly defines $I(a; \theta)$. And U receives a payoff of $(1 - \theta) [\pi(I) - I + a] = \left[\frac{1-\theta}{\theta}\right] B \cdot I$ in period 1. For D to be credit constrained and be willing to invest up to her credit constraint we require Assumption 2 to apply with B replaced by $\frac{B}{\theta}$.

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

Program Bargain The optimization program when D and U bargain and leverage from

¹⁹The key axiom generating the proportional bargaining solution is known as the axiom of step-by-step negotiations (Kalai, 1977; p.1627). The axiom requires that the bargained solution should be invariant to a decomposition of the bargaining process into stages. Thus if the individuals consider first a subset of the set of feasible alternatives, reach an agreement on the subset which is then used as the threat point in a second bargaining stage over the remaining alternatives, then the final outcome should be the same as the outcome reached if bargaining occurred in just one step.

an independent banking sector is given by

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

subject to the requirements of the proportional bargaining solution (according to which $(1 - \theta)$ times D 's total payoff is equal to θ times U 's total payoff),

$$(1 - \theta) \left\{ \sum_{i=1}^n g_i B \cdot I \left(Q_i p \left(\frac{Q_i}{z_i} \right) - W_i; \theta \right) \right\} = \theta \left\{ \begin{array}{l} \lceil \frac{1-\theta}{\theta} \rceil \sum_{i=1}^n g_i B \cdot I \left(Q_i p \left(\frac{Q_i}{z_i} \right) - W_i; \theta \right) \\ + \sum_{i=1}^n g_i \{W_i - Q_i c\} \end{array} \right\}, \quad (16)$$

and D 's incentive constraint at the quantity-setting stage in period 0:

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

Note that the proportional bargaining solution condition, equation (16), can only be satisfied if

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

Thus Program Bargain and Program Bank are isomorphic if one alters the shirking parameter from B to $\frac{B}{\theta}$. Hence, the results we developed continue to apply.

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

8 Conclusions

Credit-constrained firms are forced to link the scale of their investments to their pledgable assets. This is because the pledgable assets form the collateral for the borrowing required to fund investment. If the investment technology exhibits diminishing marginal returns, then the marginal return to pledgable assets is also diminishing. As a result, risk-neutral firms become endogenously risk averse with respect to the creation of profits which will form future pledgable income.

A firm exposed to market risk, endogenously risk averse due to its credit constraints on future investment, will seek to push some market risk onto its vertical partners. This will be reflected in the optimal supply contracts. The firm will try to reduce the variance of its ex-post profit - and hence of its pledgable income. This is done optimally via a risk-sharing contract which involves some demand-independent payment to the firm and demand-dependent repayments back to its vertical partner(s). This prediction is evidenced in risk-sharing contracts for example in the aircraft manufacturing industry, and in the common practice of charging slotting fees, prevalent in the grocery industry.

However, these contracts come at a cost to consumers in the form of higher prices. Demand-dependent repayments to the vertical partners raises the firm's effective marginal cost, inducing an increase in consumer prices. Thus, double marginalization is a necessary feature of optimal supply contracts. So our model establishes a causal link between credit constraints and higher real economy prices.

Our model predicts that firms exposed to market risk who face credit constraints on their investments will gain by outsourcing supply or sales. Once outsourced, the firm can enact a value-enhancing supply contract with insurance features. This is possible as the (outsourced) supplier sees the volumes being supplied and so can base payments on this variable. The same insurance cannot be provided by a third party as the firm and its suppliers could then collude against the insurer if supply volumes were not observable and verifiable. This link between risk and outsourcing is new to the theoretical literature and supported by empirical evidence.

We have demonstrated the existence of complementarities between the provision of insurance by the vertical partners and the provision of loans for investment. A vertical partner that is able to borrow and lend at the same rate as a bank can lend profitably to the firm when, in some instances, the separate bank cannot. This is because the vertical partner observes both the level of pledgable income being levered *and* the volume of goods sold to generate that income. This allows the vertical partner to limit, but not eradicate, the double marginalization problem created by the need for insurance. This permits lower

short-run retail prices and so increases the expected profits and the expected consumer surplus. Such lending by a vertical partner is evidenced by the existence of finance arms of major companies (such as GE) and other non-bank financial companies. 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.

This paper has demonstrated a new monetary policy transmission mechanism – distinct from the seminal balance sheet channel. The mechanism of this channel is explained by considering how the interest rates payable on investment alter the degree of endogenous risk aversion of the credit constrained firms. As interest rates rise, the sensitivity of the firm’s investment to pledgable income may rise and so cause the firm to become more risk averse. The increase in risk aversion results in greater insurance being demanded from the vertical partner(s) so as to prevent very low realizations of pledgable assets. But this means more double marginalization is introduced immediately, resulting in higher final goods prices in the short run. This is potentially important in explaining empirically observed price dynamics such as the price puzzle.

A Omitted Proofs

Proof of Lemma 2. We aim to show that:

if $Q_{i-1}^* < Q_i^* < Q_{i+1}^*$, then $\frac{\partial}{\partial r} \left[\frac{-\frac{\partial^2 I}{\partial a^2}}{\frac{\partial I}{\partial a}} \right] =_{\text{sign}} -\frac{\partial}{\partial r} Q_i^*(r)$ for all $i < n$;

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

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:

$$Q_i \geq Q_{i-1} \text{ for all } i \in \{2, \dots, n\}, \quad (17)$$

$$Q_i p \left(\frac{Q_i}{z_i} \right) - W_i \geq Q_{i-1} p \left(\frac{Q_{i-1}}{z_i} \right) - W_{i-1} \text{ for all } i \in \{2, \dots, n\}. \quad (18)$$

We now show that (18) must be satisfied with equality. Suppose not at some state i . Consider increasing W_i to $W_i + \varepsilon$ and lowering W_{i-1} to $W_{i-1} - \varepsilon(g_i/g_{i-1})$. (18) remains satisfied if $\varepsilon > 0$ is small. The individual rationality constraint of U , equation (5), is

unaffected by construction. D 's objective function changes by

$$\begin{aligned} & B\varepsilon g_i \left\{ -I' \left(Q_i p \left(\frac{Q_i}{z_i} \right) - W_i \right) + I' \left(Q_{i-1} p \left(\frac{Q_{i-1}}{z_{i-1}} \right) - W_{i-1} \right) \right\} \\ \geq & B\varepsilon g_i \left\{ -I' \left(Q_i p \left(\frac{Q_i}{z_i} \right) - W_i \right) + I' \left(Q_{i-1} p \left(\frac{Q_{i-1}}{z_i} \right) - W_{i-1} \right) \right\} \\ > & 0, \end{aligned}$$

where the first inequality follows from $z_i > z_{i-1}$ and the concavity of the investment function $I(\cdot)$, while the second equality follows by assumption on (18). But this is a contradiction to the optimality of the contract. Hence, constraint (18) must be satisfied with equality.

We next express the optimal period 0 contract purely in terms of quantities $\{Q_i\}$. From (18),

$$W_i - Q_i c = [W_{i-1} - Q_{i-1} c] + \Delta \Pi_i,$$

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 ,

$$\begin{aligned} 0 &= g_1 [W_1 - Q_1 c] + g_2 \{ \Delta \Pi_2 + [W_1 - Q_1 c] \} + g_3 \{ \Delta \Pi_3 + \Delta \Pi_2 + [W_1 - Q_1 c] \} \\ &\quad + \cdots + g_n \left\{ \sum_{j=2}^n \Delta \Pi_j + [W_1 - Q_1 c] \right\} \\ \Rightarrow & -[W_1 - Q_1 c] = \sum_{k=2}^n \left(g_k \sum_{j=2}^k \Delta \Pi_j \right) = \sum_{k=2}^n \left(\Delta \Pi_k \sum_{j=k}^n g_j \right), \end{aligned} \tag{19}$$

where we have swapped the order of summation in the second expression. Thus, equation (19) gives W_1 . Furthermore, we have

$$W_i - Q_i c = \sum_{j=2}^i \Delta \Pi_j - \sum_{j=2}^n \left(\Delta \Pi_j \sum_{k=j}^n g_k \right) \text{ for } i \geq 2. \tag{20}$$

Note that the second term on the right-hand side of (20) is independent of i .

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

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

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

$$\begin{aligned} a_{i+1} - a_i &= Q_{i+1} \left[p\left(\frac{Q_{i+1}}{z_{i+1}}\right) - c \right] - Q_i \left[p\left(\frac{Q_i}{z_i}\right) - c \right] - \Delta \Pi_{i+1} \\ &= Q_i \left[p\left(\frac{Q_i}{z_{i+1}}\right) - p\left(\frac{Q_i}{z_i}\right) \right] \geq 0 \end{aligned} \quad (21)$$

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:

$$\begin{aligned} \frac{\partial a_{i+1}}{\partial Q_l} - \frac{\partial a_i}{\partial Q_l} &= \frac{\partial}{\partial Q_l} \left\{ Q_i \left[p\left(\frac{Q_i}{z_{i+1}}\right) - p\left(\frac{Q_i}{z_i}\right) \right] \right\} \\ &= \begin{cases} 0 & \text{if } i \neq l, \\ \text{MR}_{i+1}(Q_i) - \text{MR}_i(Q_i) > 0 & \text{if } i = l, \end{cases} \end{aligned} \quad (22)$$

where $\text{MR}_{i+1}(Q_i)$ 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 $E[I(a, r)]$ over $\{Q_i\}$, subject to (17) only, with the transfers being determined by (19) and (20).

Now, we turn to period 1. Suppose that the model parameter is at the level r_1 and the optimal contract is $\{Q_i^*(r_1)\}$. The pledgable assets conditional on the state are given above. Recall that (17) must hold. Consider some state $l < n$ and suppose that $Q_{l-1}^* < Q_l^* < Q_{l+1}^*$. In this case, $E[I(a, r_1)]$ is maximized with respect to Q_l^* . That is $E\left[\frac{\partial I}{\partial a}(a, r_1) \frac{\partial a}{\partial Q_l}\right]_{Q_l^*(r_1)} = 0$. Expanding, using (22), this states:

$$\left[\sum_{j=1}^l g_j \frac{\partial I}{\partial a}(a_j, r_1) \right] \frac{\partial a_l}{\partial Q_l} + \left[\sum_{j=l+1}^n g_j \frac{\partial I}{\partial a}(a_j, r_1) \right] \frac{\partial a_{l+1}}{\partial Q_l} = 0. \quad (23)$$

As the investment returns function $I(\cdot)$ is increasing, and using (22), 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 $r_2 > r_1$. We now derive the optimal

change in Q_l . To this end, we use the Taylor expansion identity that

$$\frac{\partial I}{\partial a}(a, r_2) = \frac{\partial I}{\partial a}(a, r_1) \left[1 + (r_2 - r_1) \frac{\frac{\partial^2 I}{\partial a \partial r}(a, r_1)}{\frac{\partial I}{\partial a}(a, r_1)} \right]$$

We aim to sign $E \left[\frac{\partial I}{\partial a}(a, r_2) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)}$, which can be rewritten as

$$\begin{aligned} E \left[\frac{\partial I}{\partial a}(a, r_2) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} &= E \left[\left\{ \frac{\partial I}{\partial a}(a, r_2) - \frac{\partial I}{\partial a}(a, r_1) \right\} \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} \\ &= (r_2 - r_1) E \left[\frac{\frac{\partial^2 I}{\partial a \partial r}(a, r_1)}{\frac{\partial I}{\partial a}(a, r_1)} \frac{\partial I}{\partial a}(a, r_1) \frac{\partial a}{\partial Q_l} \right]. \end{aligned}$$

Expanding out and using (22), yields

$$\begin{aligned} E \left[\frac{\partial I}{\partial a}(a, r_2) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} &= (r_2 - r_1) \left[\sum_{j=1}^l g_j \frac{\frac{\partial^2 I}{\partial a \partial r}(a_j, r_1)}{\frac{\partial I}{\partial a}(a_j, r_1)} \frac{\partial I}{\partial a}(a_j, r_1) \right] \underbrace{\frac{\partial a_l}{\partial Q_l}}_{<0} \\ &\quad + (r_2 - r_1) \left[\sum_{j=l+1}^n g_j \frac{\frac{\partial^2 I}{\partial a \partial r}(a_j, r_1)}{\frac{\partial I}{\partial a}(a_j, r_1)} \frac{\partial I}{\partial a}(a_j, r_1) \right] \underbrace{\frac{\partial a_{l+1}}{\partial Q_l}}_{>0}. \end{aligned}$$

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

$$\frac{\partial}{\partial a} \left[\frac{\frac{\partial^2 I}{\partial a \partial r}(a, r_1)}{\frac{\partial I}{\partial a}(a, r_1)} \right] < 0. \quad (24)$$

As assets increase in the state (from (21)), we have

$$\begin{aligned} E \left[\frac{\partial I}{\partial a}(a, r_2) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} &< (r_2 - r_1) \frac{\frac{\partial^2 I}{\partial a \partial r}(a_l, r_1)}{\frac{\partial I}{\partial a}(a_l, r_1)} \left[\sum_{j=1}^l g_j \frac{\partial I}{\partial a}(a_j, r_1) \right] \frac{\partial a_l}{\partial Q_l} \\ &\quad + (r_2 - r_1) \frac{\frac{\partial^2 I}{\partial a \partial r}(a_l, r_1)}{\frac{\partial I}{\partial a}(a_l, r_1)} \left[\sum_{j=l+1}^n g_j \frac{\partial I}{\partial a}(a_j, r_1) \right] \frac{\partial a_{l+1}}{\partial Q_l} \\ &= 0, \end{aligned} \quad (25)$$

where the equality follows from (23). This therefore proves that $Q_l^*(r_2) < Q_l^*(r_1)$, and so retail prices would be higher under parameter r_2 than r_1 if (24) 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 (24) is reversed).

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

then

$$\frac{\partial}{\partial r} Q_l^*(r) =_{\text{sign}} \frac{\partial}{\partial a} \left[\frac{\frac{\partial^2 I}{\partial a \partial r}(a, r_1)}{\frac{\partial I}{\partial a}(a, r_1)} \right] =_{\text{sign}} - \frac{\partial}{\partial r} \left[\frac{-\frac{\partial^2 I}{\partial a^2}}{\frac{\partial I}{\partial a}} \right],$$

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.

Finally, we consider the case of pooling. We seek to modify the proof above to show that the pooled quantity falls weakly as we move to r_2 if $\frac{\partial}{\partial a} \left[\frac{\frac{\partial^2 I}{\partial a \partial r}(a, r_1)}{\frac{\partial I}{\partial a}(a, r_1)} \right] < 0$. 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 $Q_{l-1}^*(r_1) = Q_l^*(r_1)$, the optimization over state l is constrained, so that $E \left[\frac{\partial I}{\partial a}(a, r_1) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} \leq 0$. If $E \left[\frac{\partial I}{\partial a}(a, r_1) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} = 0$, then the identical proof to above applies showing that $Q_l^*(r_2) \leq Q_l^*(r_1)$. The inequality is weak as Q_l^* will only be able to fall if Q_{l-1}^* does. Suppose instead that $E \left[\frac{\partial I}{\partial a}(a, r_1) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} < 0$. The Taylor expansion around r_2 is now given by

$$E \left[\frac{\partial I}{\partial a}(a, r_2) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} = E \left[\frac{\partial I}{\partial a}(a, r_1) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)} + (r_2 - r_1) E \left[\frac{\partial^2 I}{\partial a \partial r}(a, r_1) \frac{\partial a}{\partial Q_l} \right]_{Q_l^*(r_1)},$$

which is strictly negative for $r_2 - r_1$ small. Hence, again we have $Q_l^*(r_2) \leq Q_l^*(r_1)$. Finally, we obtain

$$Q_{l-1}^*(r_2) \leq Q_l^*(r_2) \leq Q_l^*(r_1) = Q_{l-1}^*(r_1),$$

where the first inequality follows by (17), the second inequality has just been shown, and the equality follows by assumption.

If, instead, $\frac{\partial}{\partial a} \left[\frac{\frac{\partial^2 I}{\partial a \partial r}(a, r_1)}{\frac{\partial I}{\partial a}(a, r_1)} \right] > 0$, then consider the smallest pooled state and repeat the argument of the paragraph above. ■

Proof of Propostion 5. From the discussion in the main text, an increase in the interest rate r will raise the coefficient of risk aversion (and, by Lemma 2, retail prices) if

$$\frac{d}{dr} \ln \left(\frac{d\gamma/da}{\gamma} \right) > 0. \quad (26)$$

Since $\partial I/\partial a > 0$, $\gamma > 0$ (Assumption 2) and so $d\gamma/da = -\pi''(I)\partial I/\partial a > 0$, equation (26) holds if

$$\frac{d^2\gamma}{dadr} \gamma - \frac{d\gamma}{da} \frac{d\gamma}{dr} > 0,$$

where

$$\frac{d\gamma}{dr} = 1 - \pi''(I) \frac{\partial I}{\partial r} \text{ and } \frac{d^2\gamma}{dadr} = -\pi'''(I) \frac{\partial I}{\partial r} \frac{\partial I}{\partial a} - \pi''(I) \frac{\partial^2 I}{\partial a \partial r}$$

Suppose $-\pi''(I)$ is sufficiently large and $\pi'''(I) \geq 0$, as assumed in the statement of the proposition. Then, noting that $\partial I/\partial r < 0$, we obtain $\partial^2 I/\partial a \partial r > 0$, and

$$\frac{d\gamma}{dr} < 0, \quad \frac{d\gamma}{da} > 0, \quad \text{and} \quad \frac{d^2\gamma}{dadr} > 0.$$

Hence,

$$\frac{d^2\gamma}{dadr} \gamma - \frac{d\gamma}{da} \frac{d\gamma}{dr} > 0,$$

implying that equation (26) holds so that an increase in the interest rate results in a larger coefficient of (absolute) risk aversion. Part 1 of the proposition follows from Lemma 2.

For part 2, note that the investment levels fall for any realization of assets as $\partial I/\partial r < 0$ and, further, realized assets are lower for any realization of period 0 market demand (except the largest), due to the lower equilibrium volumes. ■

A.1 Upstream Risk Analysis

This appendix proves Proposition 6. The proof follows closely that used by Hart (1983, Proposition 2). Here we outline the key steps and any differences. The objective is to find the solution to Program U Constrained.

The first step is to show that the incentive compatibility constraint (15) can be replaced by the local constraints:

$$W_i - Q_i \kappa_i \geq W_{i+1} - Q_{i+1} \kappa_i \text{ for all } i \in \{1, \dots, n-1\} \quad (27)$$

$$Q_i \geq Q_{i+1} \text{ for all } i \in \{1, \dots, n-1\} \quad (28)$$

That (15) implies these two equations follows Hart. The converse remains to be shown. Here Hart 1983 Lemma 1 does not apply. Instead we have

Lemma 3 (27) must hold with equality at an optimal solution to Program U Constrained, with (15) replaced by (27) and (28):

Proof. Identical to the first lines of the proof of Lemma 2 above. ■

Using Lemma 3, we can demonstrate that a solution to program U constrained with (15) replaced by (27) and (28) satisfies the original program. This now follows Hart again.

Proof of Proposition 6. We first show that $Q_i \leq q(\kappa_i)$ for all i . Suppose instead that

there exists a given i such that

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

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

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

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

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

For small ε (28) holds at i and $i-1$ as $Q_i > Q_{i+1}$. Next consider the objective function. This changes by

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

Thus U benefits. Next note that this also implies that (27) holds at i as the left hand side has increased. Finally note that (27) holds at $i-1$ as $W_i - Q_i \kappa_{i-1}$ is reduced for it is altered by the amount:

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

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

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

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

This argument can be extended all along the line and proves result 2 of Proposition 6.

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

As positive quantities are sold result 3 follows as from (15):

$$W_i - Q_i \kappa_i \geq W_{i+1} - Q_{i+1} \kappa_i > W_{i+1} - Q_{i+1} \kappa_{i+1} \quad (29)$$

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

For result 4 we have

$$\begin{aligned} & [Q_i p(Q_i) - W_i] - [Q_{i+1} p(Q_{i+1}) - W_{i+1}] \\ = & \text{Lemma 3 } [Q_i p(Q_i) - Q_i \kappa_i] - [Q_{i+1} p(Q_{i+1}) - Q_{i+1} \kappa_i] \end{aligned}$$

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

References

- [1] Balke and Emery (1994), “Understanding the Price Puzzle”, Federal Reserve Bank of Dallas Economic Review, 4th Quarter, 15-26.
- [2] Barth and Ramey (2001), “The Cost Channel of Monetary Transmission.” NBER Macroeconomics Annual 2001, 199–240.
- [3] Bank of England (1999), “The transmission mechanism of monetary policy”, available at <http://www.bankofengland.co.uk/publications/other/monetary/montrans.pdf>.
- [4] Bernanke and Gertler (1995), “Inside the Black Box: The Credit Channel of Monetary Policy Transmission.” Journal of Economic Perspectives, 27-48.
- [5] Bloom, Gundlach and Cannon (2000), “Slotting Allowances and Fees: Schools of Thought and the Views of Practicing Managers”, Journal of Marketing, 92-108
- [6] Bradley and Roberts (2004), “The Structure and Pricing of Corporate Debt Covenants,” SSRN working paper no. 466240.
- [7] Burkart and Ellingsen (2004), “In-Kind Finance: A Theory Of Trade Credit”, American Economic Review, 569-590.
- [8] Carey, Post and Sharpe, “Does Corporate Lending by Banks and Finance Companies Differ? Evidence on Specialization in Private Debt Contracting,” The Journal of Finance, 845-878.
- [9] Carlton (1979), “Vertical Integration in Competitive Markets Under Uncertainty”, Journal of Industrial Economics, 189-209.
- [10] Carlton (1977), “Uncertainty, Production Lags, and Pricing”, The American Economic Review, 67, 244-249.

- [11] Christiano, Eichenbaum and Evans (1999), "Monetary Policy Shocks: What Have We Learnt And To What End?", Handbook of Macroeconomics. Volume 1A. Taylor and Woodford eds. North Holland Press.
- [12] Competition Commission (2008), "The supply of groceries in the UK market investigation," The Stationary Office, UK.
- [13] Crawford and Yurukoglu (2008), "The Welfare Effects of Bundling in Multi-channel Television Markets," University of Arizona.
- [14] DeMarzo and Fishman (forthcoming), "Agency and Optimal Investment Dynamics", Review of Financial Studies.
- [15] Figueiredo, Silveira and Sbragia (2008), "Risk Sharing Partnerships With Suppliers: The Case of Embraer," Journal of Technology, Management and Innovation, 27-37.
- [16] Froot, Scharfstein and Stein (1993), "Risk Management: Coordinating Corporate Investment and Financing Policies", Journal of Finance, 1629-1658.
- [17] Gaiotti and 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, 2013-2037.
- [18] Gertler and Gilchrist (1994), "Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms", Quarterly Journal of Economics, 309-340.
- [19] Harrigan (1985), "Vertical Integration and Corporate Strategy", Academy of Management Journal, 397-425.
- [20] Hart, Oliver D. (1983), "Labour Contracts under Asymmetric Information: An Introduction," Review of Economic Studies, vol. 50(1), 3-35.
- [21] Holmstrom and Tirole (1997), "Financial Intermediation, Loanable Funds, and the Real Sector", Quarterly Journal of Economics, 663-691.
- [22] Hubbard (1998), "Capital-Market Imperfections and Investment", Journal of Economic Literature, 193-225.
- [23] Ireland (2005), "The Monetary Transmission Mechanism", FRB of Boston Working Paper No. 06-1.
- [24] Kalai (1977), "Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons", Econometrica, 45, 1623-1630.

- [25] Klein and Wright (2007), “The Economics of Slotting Contracts”, *Journal of Law and Economics*, 421-455.
- [26] Mahoney (1992), “The Choice of Organizational Form: Vertical Financial Ownership Versus Other Methods of Vertical Integration” *Strategic Management Journal*, 559-584.
- [27] Mortimer (2008), “Vertical Contracts in the Video Rental Industry,” *Review of Economic Studies*, 165–199.
- [28] Rey and Tirole (1986), “The Logic Of Vertical restraints”, *The American Economic Review*, 921-939.
- [29] Romano (1994), “Double Moral Hazard and Resale Price Maintenance,” *RAND Journal of Economics*, 455-466.
- [30] Shin and Stulz (1998), “Are Internal Capital Markets Efficient?”, *The Quarterly Journal of Economics*, 531-552.
- [31] Spengler (1950), “Vertical integration and antitrust policy,” *Journal of Political Economy*, 347-352.
- [32] Sudhir and Rao (forthcoming), “Are Slotting Allowances Efficiency-Enhancing or Anti-Competitive?”, *Journal of Marketing Research*.
- [33] Sutcliffe and Zaheer (1988), “Uncertainty in the Transaction Environment: An Empirical Test”, *Strategic Management Journal*, 1-23.
- [34] Tirole (2006), “The Theory of Corporate Finance,” Princeton University Press.
- [35] White, Troy and 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, 291-298.