Supply Contracts with Financial Hedging

René Caldentey

Stern School of Business, New York University, New York, NY 10012, rcaldent@stern.nyu.edu.

Martin B. Haugh

Department of IE and OR, Columbia University, New York, NY 10027, mh2078@columbia.edu.

Abstract

We study the performance of a stylized supply chain where two firms, a retailer and a producer, compete in a Stackelberg game. The retailer purchases a single product from the producer and afterwards sells it in the retail market at a stochastic clearance price. The retailer, however, is budget-constrained and is therefore limited in the number of units that he may purchase from the producer. We also assume that the retailer's profit depends in part on the realized path or terminal value of some observable stochastic process. We interpret this process as a financial process such as a foreign exchange rate or interest rate. More generally the process may be interpreted as any relevant economic index. We consider a variation (the flexible contract) of the traditional wholesale price contract that is offered by the producer to the retailer. Under this flexible contract, at $t = 0$ the producer offers a menu of wholesale prices to the retailer, one for each realization of the financial process up to a future time τ. The retailer then commits to purchasing at time τ a variable number of units, with the specific quantity depending on the realization of the process up to time τ. Because of the retailer's budget constraint, the supply chain might be more profitable if the retailer was able to shift some of the budget from states where the constraint is not binding to states where it is binding. We therefore consider a variation of the flexible contract where we assume that the retailer is able to trade dynamically between 0 and τ in the financial market. We refer to this variation as the flexible contract with hedging. We compare the decentralized competitive solution for the two contracts with the solutions obtained by a central planner. We also compare the supply chain's performance across the two contracts. We find, for example, that the producer always prefers the flexible contract with hedging to the flexible contract without hedging. Depending on model parameters, however, the retailer may or may not prefer the flexible contract with hedging. Finally, we study the problem of choosing the optimal timing, τ, of the contract, and formulate this as an optimal stopping problem.

Subject Classifications: Finance: portfolio, management. Optimal control: applications. Production: applications.

Keywords: Procurement contract, financial constraints, supply chain coordination.

1 Introduction

We consider the operation of a stylized supply chain with one producer and one retailer. The producer manufactures a single product which it sells to the retailer. The retailer in turn then sells the product in the retail market at a stochastic clearance price. We consider a non-cooperative mode of operation in which both players maximize their own profit functions. In particular, we consider a Stackelberg game where the producer, acting as leader, proposes a retail price or menu of prices to the retailer who then decides how many units to order. As is customary in the supply chain literature (*e.g.*, Lariviere 1998 and Tsay et al 1998), we are interested in characterizing the solution of the game as well as its efficiency. We measure the efficiency using the so-called *competition penalty*, that is, the ratio of the non-cooperative supply chain profits to the centralized supply chain profits (*e.g.*, Cachon and Zipkin 1991).

Our model differs from previous work in two aspects. First, we assume that the retailer operates under a budget constraint. In particular, a limited amount of cash is available to the retailer for purchasing product units from the producer. Budget constraints are quite common in practice due to a number of reasons. For example, many companies have only limited and / or costly access to credit markets. It is also the case that some companies choose to restrict their managers by imposing budget constraints on their actions. The imposition of budget constraints has for the most part been ignored in the extensive research on supply chain management. A recent exception is the work by Buzacott and Zhang (2004) where the interplay between inventory decisions and asset-based financing is investigated.

The second distinguishing aspect of our model is the existence of a financial market or economic index whose movements are correlated[1] with the supply chain's profits. For example, if the producer sells to a foreign retailer and quotes prices in foreign currency units, then his profits, in units of his domestic currency, will be correlated with exchange rate movements. Similarly, if the retailer pays the producer in arrears, then the producer is exposed to interest rate risk (representing the time value of the delayed payment) as well as possible default risk. It could also be the case that the clearance price for the product in the retail market is influenced in part by the overall state of the economy or the state of particular sectors within the economy. These states might be represented by the value of some well-chosen economic index.

The existence of the financial market affects our framework in two ways. First, the movements of the financial market serve as a public signal that the players can use to negotiate the terms of the procurement contract. Second, the financial market can be used to minimize the impact of the budget constraint. In particular, by trading dynamically in the financial market[2] the retailer can shift resources from states where the budget constraint is not binding to states where it is. This ability to shift resources across different states is only of interest when the two players use the financial market to negotiate the terms of the procurement contract.

In this paper we will consider three different types of contract that are offered by the producer to the retailer. In the case of the simple contract, the producer offers at time $t = 0$ a fixed wholesale price to the retailer who then chooses an order quantity. In the case of the flexible contract, the negotiations are also conducted at $t = 0$ but the physical transaction is deferred to a date $\tau > 0$ when the price and order quantity are contingent upon the history of the financial market up to

[1] We use the term 'correlated' loosely in this paper when referring to any form of statistical dependence.

[2] Hereafter we will use the term 'financial market' even when we have a more general economic index in mind. While it is not possible to trade every economic index, many are tradeable. Moreover, the current 'securitization' trend suggests that ever more economic indices will be tradeable in the future.

time τ. It is assumed that no trading in the financial markets takes place. The flexible contract with hedging is similar to the flexible contract except now the retailer has the ability to trade in the financial markets between $t = 0$ and $t = \tau$.

We assume that both players are risk neutral and maximize the *economic value* of their operations, that is the expected value of their payoffs under an appropriate equivalent martingale measure (EMM). Because some of the uncertainty in our framework will be driven by non-financial noise, the setting of this paper is one of *incomplete*[3] *markets*. A standard result from financial economics then implies that a unique EMM will not exist so an appropriate one would need to be identified using economic principals. We will not concern ourselves with the selection of the appropriate EMM in this paper and will instead assume that it has already been identified. In addition to being economically sound, we will see that using an EMM allows us to model the situation where trading in the financial markets takes place for hedging purposes only, and not for speculative purposes. This is consistent with how the financial markets are typically used in practice by non-financial corporations. Of course the ability to trade in the financial markets can and generally does have an *indirect* impact on the players' profits by expanding the set of feasible order quantities.

The remainder of the paper is organized as follows. Section 2 describes the basic supply chain model and financial market in greater detail. Sections 3 and 4 characterize the solution of the non-cooperative game under the flexible contract and flexible contract with hedging, respectively. To complete the analysis of these contracts, we also compute the centralized solutions and use them to determine the efficiency of the non-cooperative supply chain. While the simple contract is the most commonly occurring in practice, it is a special case of the flexible contract with $\tau = 0$ and so we do not need to analyze it separately from the flexible contract. In Section 5 we consider the case where the transaction time, τ, is no longer given exogenously as a fixed time but is instead a decision variable whose value is determined endogenously as part of our equilibrium solution. We will consider the case where τ is deterministic and the case where τ is permitted to be a more general stopping time. Further extensions to the model are then discussed in Section 6 and we conclude in Section 7.

2 Model Description

We now describe the model in further detail. We focus first on the supply chain and then consider the financial markets. Finally, we describe the three types of contracts that we analyze in this paper.

2.1 The Supply Chain

We model an isolated segment of a competitive supply chain with one producer that produces a single product and one retailer that faces a stochastic clearance price[4] for this product. This clearance price, and the resulting cash-flow to the retailer, is realized at a fixed future time $T > 0$. The retailer and producer, however, negotiate the terms of a procurement contract at time $t = 0$. This contract specifies three quantities:

[3] See, for example, Shreve (2004).

[4] Similar models are discussed in detail in Section 2 of Cachon (2003). See also Lariviere and Porteus (2001).

(*i*) A procurement time τ, with $0 \le \tau \le T$, when the retailer will place a single order. While τ will be fixed for most of our analysis, we will also consider the problem of selecting an optimal τ in Section 5.

(*ii*) A rule that specifies the size of the order, q_τ. Depending on the type of contract under consideration, q_τ may depend upon market information available at time τ.

(*iii*) The payment, $\mathcal{W}(q_\tau)$, that the retailer pays to the producer for fulfilling the order. Again, depending on the type of contract under consideration, $\mathcal{W}(q_\tau)$ may depend upon market information available at time τ. The timing of this payment is not important as we shall assume that interest rates are identically zero in Sections 3 and 4. In Section 6, where we will have non-zero interest rates, it will be necessary to specify exactly when the retailer pays the producer.

We will restrict ourselves to transfer payments that are linear on the ordering quantity, the so-called *wholesale price* contract, with $\mathcal{W}(q) = w\,q$ where w is the per-unit wholesale price charged by the producer. We also assume that during the negotiation of the contract the producer acts as a Stackelberg leader. That is, for a fixed procurement time τ, the producer moves first and proposes a wholesale price[5], w_τ, to which the retailer then reacts by selecting the ordering level q_τ.

We assume that the producer has unlimited production capacity and that if production takes place at time τ then the per-unit production cost is constant and equal to c_τ. This function is assumed to be increasing in τ so that production postponement comes at a cost. The producer's payoff as a function of the procurement time, τ, the wholesale price, w_τ, and the ordering quantity, q_τ, is given by

$$\Pi_{\mathrm{P}} := (w_\tau - c_\tau)\, q_\tau. \tag{1}$$

We assume that the retailer is restricted by a budget constraint that limits his ordering decisions. In particular, we assume that the retailer has an initial budget B that may be used to purchase product units from the producer. Depending on the type of contract under consideration, the retailer may be able to trade in the financial market during the time interval $[0,\tau]$, thereby transferring cash resources from states where they are not needed to states where they are.

For a given order quantity, q_τ, the retailer collects a random revenue at time T. We compute this revenue using a linear *clearance price* model. That is, given an ordering quantity, q_τ, the market price at which the retailer sells (clears) these units is a random function, $A - \xi\, q_\tau$, where A is a non-negative random variable and ξ is a positive constant. The random variable A models the market size that we assume is unknown while the fixed parameter, ξ, captures the demand elasticity that we assume is known. The retailer's payoff, as a function of τ, w_τ, and q_τ, then takes the form

$$\Pi_{\mathrm{R}} := (A - \xi\, q_\tau)\, q_\tau - w_\tau\, q_\tau. \tag{2}$$

We have chosen to use a stochastic clearance price formulation for the following reason. Our goal in this paper is to highlight the benefits of using financial markets in the context of a simple supply chain model. With this objective in mind, we would like to use a formulation that simultaneously captures the stochastic nature of the retailer's payoff and at the same time allows us to clearly isolate the impact that financial markets have on the supply chain performance. A clearance price approach is better suited to achieving this objective than say the newsvendor type of formulation

[5]In the case of the flexible contracts that we consider the producer offers a *menu* of wholesale prices. See Section 2.3.

that is commonly encountered in the supply chain literature[6]. Moreover, it is easily justified since in practice unsold units are generally liquidated using secondary markets at discount prices. Therefore, we can view our clearance price as the average selling price across all units and markets.

As stated earlier, depending on the type of contract under consideration, w_τ and q_τ can depend upon market information available at time τ. Since $\mathcal{W}(q)$, Π_{P} and Π_{R} are functions of w_τ and q_τ, it is also the case that these quantities can depend upon market information available at time τ.

2.2 The Financial Market

The financial market is modelled as follows. Let X_t denote the time t value of a tradeable security and let $\{\mathcal{F}_t\}_{0\leq t\leq T}$ be the filtration generated by X_t on a given probability space, $(\Omega, \mathcal{F}, Q)$. It is not the case that $\mathcal{F}_T = \mathcal{F}$ since we assume that the non-financial random variable, A, is $\mathcal{F}$-measurable but not $\mathcal{F}_T$-measurable. We also assume that there is a risk-less *cash account* available in which cash may be deposited. We assume[7] without loss of generality that the interest rate on the cash account is identically equal to zero. Then the time τ gain (or loss), $G_\tau(\theta)$, that results from following a self-financing[8] $\mathcal{F}_t$-predictable trading strategy, θ_t, can be represented as a stochastic integral with respect to X. For example[9], in a continuous-time setting we have

$$G_\tau(\theta) := \int_0^\tau \theta_s \, dX_s \tag{3}$$

while in a discrete-time setting we have

$$G_\tau(\theta) := \sum_{i=0}^{\tau-1} \theta_i \, (X_{i+1} - X_i). \tag{4}$$

We assume that Q is an equivalent martingale measure (EMM) so that discounted security prices are Q-martingales. Since we have assumed that interest rates are identically zero, however, it is therefore the case that X_t is a Q-martingale. Subject to integrability constraints on the set of feasible trading strategies, we also see that $G_t(\theta)$ is a Q-martingale for every $\mathcal{F}_t$-predictable self-financing trading strategy, θ_t.

Our analysis will be simplified considerably by making a *complete financial markets* assumption. In particular, let G_τ be any suitably integrable contingent claim that is $\mathcal{F}_\tau$-measurable. Then a complete financial markets assumption amounts to assuming the existence of an $\mathcal{F}_t$-predictable self-financing trading strategy, θ_t, such that $G_\tau(\theta) = G_\tau$. That is, G_τ is *attainable.* This assumption is very common in the financial literature. Moreover, many incomplete financial models can be made complete by simply expanding the set of tradeable securities. When this is not practical, we can simply assume the existence of a market maker with a known pricing function or pricing kernel[10]

[6]See Cachon (2003) for a recent review of supply chain contract models.

[7]We will relax this assumption in Section 6 when we consider a specific example with interest rate risk.

[8]In words, a trading strategy is self-financing if cash is neither deposited with or withdrawn from the portfolio during the trading interval, $[0, T]$. In particular, trading gains or losses are due to changes in the values of the traded securities. See Shreve (2004) for a technical definition of the self-financing property.

[9]θ_s represents the number of units of the tradeable security held at time s. The self-financing property then implicitly defines the position at time s in the cash account. Because we have assumed interest rates are identically zero, there is no term in (3) or (4) corresponding to gains or losses from the cash account holdings.

[10]See Duffie (2002). More generally, Duffie may be consulted for further technical assumptions (that we have omitted to specify) regarding the filtration, $\{\mathcal{F}_t\}_{0\leq t\leq T}$, feasible trading-strategies, etc.

who is willing to sell G_τ in the market-place. In this sense, we could then claim that G_τ is indeed attainable.

Regardless of how we choose to justify it, assuming complete financial markets simplifies our analysis considerably because, under this assumption, we will never need to solve for a dynamic trading strategy, θ. Instead, we will only need to solve for a contingent claim, G_τ, safe in the knowledge that any such claim is attainable. For this reason we will drop the dependence of G_τ on θ in the remainder of the paper. The only restriction that we will impose on any such trading gain, G_τ, is that the corresponding trading gain process, $G_s := \mathbb{E}_s^{\mathbb{Q}}[G_\tau]$ be a Q-martingale[11] for $s < \tau$. In particular we will assume that any feasible trading gain, G_τ, satisfies $\mathbb{E}_0^{\mathbb{Q}}[G_\tau] = G_0$ where G_0 is the initial amount of capital that is devoted to trading in the financial market. Without any loss of generality we will typically assume $G_0 = 0$. This assumption will be further clarified in Section 2.3.

A key aspect of our model is the dependence between the payoffs of the supply chain and returns in the financial market. We model this dependence in a parsimonious way by assuming that returns in the financial market and the random variable A are dependent. We will make the following assumption regarding the conditional distribution of A.

Assumption 1 *For all $\tau \in [0, T]$, $\mathbb{E}_\tau^{\mathbb{Q}}[A] \geq c_\tau$.*

This condition ensures that for every time and state there is a production level, $q \geq 0$, for which the expected retailer's market price exceeds the producer's production cost. In particular, this assumption implies that it is possible to profitably operate the supply chain.

2.3 The Three Contracts

The final component of our model is the contractual agreement between the producer and the retailer. We consider three different alternatives. Note that in all three cases the contract itself is negotiated at time $t = 0$ whereas the actual physical transaction takes place at time $\tau \geq 0$.

- **Simple Contract (S-Contract):** In the case of the simple contract, the negotiation and physical transaction both take place at the beginning of the planning horizon so that we have $\tau = 0$. In this case, the financial market is not used in the design of the contract and our model reduces to the traditional wholesale price contract. That is, the producer, acting as a Stackelberg leader, offers a fixed wholesale price, w_0, at time $t = 0$. The retailer, acting as a follower, then determines the quantity, q_0, that he will purchase. The budget constraint in this case takes the form $w_0 \, q_0 \leq B$, where B is the retailer's available budget.
- **Flexible Contract (F-Contract):** In the case of the flexible contract, the physical transaction is postponed to a future date $\tau \in [0, T]$. In this case, the two parties are able to negotiate at time $t = 0$ a contract contingent on the public history, $\mathcal{F}_\tau$, that is available at time τ. Specifically, at time $t = 0$ the producer offers an $\mathcal{F}_\tau$-measurable wholesale price, w_τ, to the retailer. In response to this offer, the retailer decides on an $\mathcal{F}_\tau$-measurable ordering quantity[12], $q_\tau = q(w_\tau)$.

[11] Whenever we write $\mathbb{E}_s^{\mathbb{Q}}[\cdot]$ it should be understood as $\mathbb{E}^{\mathbb{Q}}[\cdot|\mathcal{F}_s]$.

[12] There is a slight abuse of notation here and throughout the paper when we write $q_\tau = q(w_\tau)$. This expression should not be interpreted as implying that q_τ is a function of w_τ as this would imply that q_τ is measurable with respect to the σ-algebra generated by w_τ. We only require, however, that q_τ be $\mathcal{F}_\tau$-measurable and so a more appropriate interpretation is to say that $q_\tau = q(w_\tau)$ is the retailer's response to w_τ.

In this, the flexible contract, we assume that the retailer does not hedge his budget constraint by trading in the financial market. Hence, the financial market acts exclusively as a source of public information used to define the terms of the contract. As a result, the budget constraint takes the form

$$w_\tau \, q_\tau \leq B, \qquad \text{for all } \omega \in \Omega.$$

We note that the S-contract is a special case of the F-contract with $\tau = 0$.

- **Flexible Contract with Hedging (H-Contract):** A flexible contract with hedging is similar to the flexible contract but now the retailer has access to the financial markets. In particular, the retailer can use the financial market to hedge the budget constraint by purchasing at date $t = 0$ a contingent claim, G_τ, that is realized at date τ and that satisfies $\mathbb{E}_0^{\mathbb{Q}}[G_\tau] = 0$. Given an $\mathcal{F}_\tau$-measurable wholesale price, w_τ, the retailer purchases an $\mathcal{F}_\tau$-measurable contingent claim, G_τ, and selects an $\mathcal{F}_\tau$-measurable ordering quantity, $q_\tau = q(w_\tau)$, in order to maximize the economic value of his profits. Because of his access to the financial markets, the retailer can weaken the budget constraint which now becomes

$$w_\tau \, q_\tau \leq B + G_\tau =: B_\tau, \qquad \text{for all } \omega \in \Omega.$$

Since the no-trading strategy with $G_\tau \equiv 0$ is always an option, it is clear that for a given wholesale price, w_τ, the retailer is always better-off by trading in the financial market. Whether or not the retailer will still be better off in equilibrium when he has access to the financial market will be discussed in Section 4.

By using a flexible contract, the parties postpone their transaction to a future time and in the process improve their estimates of the market clearance price. In this respect, our flexible contracts are very much related to the literature on supply chain contracts with demand forecast updating (*e.g.* Donohue 2000). In our case, however, the additional information comes from the financial market and its co-dependence with the market clearance price. This feature differs substantially from previous models that normally relate new market information to marketing research and early order commitments (*e.g.*, Azoury (1985), Eppen and Iyer (1997)). As well as being a source of information upon which a contract can be based, however, financial markets also enable the players to hedge their cashflows. In particular, the difference between the equilibrium solutions of the F-contract and H-contract will help us quantify the impact that financial trading has on the supply chain performance.

Before proceeding to analyze these contracts a number of further clarifying remarks are in order.

1. The model assumes a common knowledge framework in which all parameters of the models are known to both agents. Because of the Stackelberg nature of the game, this assumption implies that the producer knows the retailer's budget, B, and the distribution of the market demand. We also make the implicit assumption that the only information available regarding the random variable, A, is what we can learn from the evolution of X_t in the time interval $[0, \tau]$. If this were not the case, then the trading strategy in the financial market could depend on some non-financial information and so it would not be necessary to restrict the trading gain, G_τ, to be $\mathcal{F}_\tau$-measurable. More generally, if Y_t represented some non-financial noise that was observable at time t, then the trading strategy, θ_t, would only need to be predictable with respect to the filtration generated by X *and* Y. In this case the complete financial market assumption is of benefit and it would be necessary for the retailer to solve the much harder problem of finding the optimal θ in order to find the optimal G_τ.

2. In this model the producer does not trade in the financial markets because, being risk-neutral and not restricted by a budget constraint, he has no incentive to do so. In particular, the Q-martingale property of self-financing trading strategies implies that if the producer devoted an initial capital, F_0, to trading then we would need to include a term $-F_0 + \mathbb{E}_0^{\mathbb{Q}}[F_\tau]$ in his objective function. Here F_τ denotes the time τ value of the producers's financial portfolio that results from adopting some self-financing trading strategy. However, the Q-martingale property of trading gains implies that this term is identically zero for all such strategies[13] and so the financial markets provide no benefit to the producer.

3. A potentially valid criticism of this model is that, in practice, a retailer is often a small entity and may not have the ability to trade in the financial markets. There are a number of responses to this. First, we use the word 'retailer' in a loose sense so that it might in fact represent a large entity. For example, an airline purchasing aircraft is a 'retailer' that certainly does have access to the financial markets. Second, it is becoming ever cheaper and easier for even the smallest 'player' to trade in the financial markets. Finally, even if the retailer does not have access to the financial market, then the producer, assuming he is a big 'player', can offer to trade with the retailer or act as his financial broker. As we shall see in Section 4, it would always be in the producer's interest to do so.

4. We claimed earlier that, without loss of generality, we could assume $G_0 = 0$. This is clear for the following reason. If $G_0 = 0$ then with a finite initial budget, B, the retailer has a terminal budget of $B_\tau = B + G_\tau$ with which he can purchase product units at time τ and where $\mathbb{E}_0^{\mathbb{Q}}[G_\tau] = 0$. If he allocated $a > 0$ to the trading strategy, however, then he would have a terminal budget of $B_\tau = B - a + G_\tau$ at time τ but now with $\mathbb{E}_0^{\mathbb{Q}}[G_\tau] = a$. That the retailer is indifferent between the two approaches follows from the fact any terminal budget, B_τ, that is feasible under one modelling approach is also feasible under the other and vice-versa.

5. Another potentially valid criticism of this framework is that the class of contracts is too complex. In particular, by only insisting that w_τ is $\mathcal{F}_\tau$-measurable we are permitting wholesale price contracts that might be too complicated to implement in practice. If this is the case then we can easily simplify the set of feasible contracts. By using appropriate conditioning arguments, for example, it would be straightforward to impose the tighter restriction that w_τ be $\sigma(X_\tau)$-measurable instead where $\sigma(X_\tau)$ is the σ-algebra generated by X_τ.

We complete this section with a summary of the notation and conventions that will be used throughout the remainder of the paper. The superscripts S, F and H are used to index quantities related to the S-contract, F-contract and H-contract, respectively. The subscripts R, P, and C are used to index quantities related to the retailer, producer and central planner, respectively. The subscript τ is used to denote the value of a quantity conditional on time τ information. For example, $\Pi^{\mathrm{H}}_{\mathrm{P}|\tau}$ is the producer's time τ expected payoff under the H-contract. The expected value, $\mathbb{E}_0^{\mathbb{Q}}[\Pi^{\mathrm{H}}_{\mathrm{P}|\tau}]$, is simply denoted by $\Pi^{\mathrm{H}}_{\mathrm{P}}$ and similar expressions hold for the retailer and central planner. Any other notation will be introduced as necessary.

3 The Flexible Contract

We now study the F-contract in which the producer offers a wholesale price[14], w_τ, to the retailer who then selects a corresponding $q_\tau = q(w_\tau)$. We will assume for now that τ is given exogenously

[13] Subject to technical conditions that we mentioned in the previous subsection.

[14] Hereafter we will drop the qualifier "$\mathcal{F}_\tau$-measurable" when this should be clear from the context.

and defer until Section 5 the problem of selecting it in an optimal manner.

The Decentralized Solution

In response to the wholesale price menu, $w_\mathcal{T}$, the retailer selects a menu of ordering quantities, $q_\mathcal{T} = q(w_\mathcal{T})$, by solving the following optimization problem:

$$\Pi_{\mathrm{R}}^{\mathrm{F}}(w_\mathcal{T}) = \mathbb{E}_0^{\mathbb{Q}}\left[\max_{q_\tau \geq 0}\left\{\mathbb{E}_\mathcal{T}^{\mathbb{Q}}\Big[(A - \xi\, q_\tau - w_\tau)\, q_\tau\Big]\right\}\right]$$
$$\text{subject to} \qquad w_\tau\, q_\tau \leq B, \qquad \text{for all } \omega \in \Omega.$$

Note that the expectation inside the max operator is conditional on $\mathcal{F}_\mathcal{T}$. So for each possible realization of X until time τ, the retailer determines the optimal quantity, q_τ, by solving a procurement problem with wholesale price, w_τ, and budget constraint $w_\tau\, q_\tau \leq B$. The retailer's problem therefore decouples for each such realization of X. Let us define $\bar{A}_\tau := \mathbb{E}_\mathcal{T}^{\mathbb{Q}}[A]$ and $\bar{A} := \mathbb{E}_0^{\mathbb{Q}}[\bar{A}_\tau]$.

Straightforward calculations show that the solution to the conditional optimization problem is given by

$$q(w_\tau) = \min\left\{\left(\frac{\bar{A}_\tau - w_\tau}{2\,\xi}\right)^+ , \ \frac{B}{w_\tau}\right\}. \tag{5}$$

The negative effect of the budget constraint on the optimal ordering quantity is clear from (5). Given this, the retailer's best-response strategy, the producer solves

$$\Pi_{\mathrm{P}}^{\mathrm{F}} = \mathbb{E}_0^{\mathbb{Q}}\left[\max_{w_\mathcal{T} \geq c_\mathcal{T}} \left\{(w_\tau - c_\mathcal{T})\, q_\tau(w_\tau)\right\}\right].$$

As was the case with the retailer's problem, the producer's optimization problem decouples for each realization of X until time τ. We use the notation $\Pi_{\mathrm{P}|\tau}^{\mathrm{F}}$ and $\Pi_{\mathrm{R}|\tau}^{\mathrm{F}}$ to denote the payoffs of the producer and retailer, respectively, conditional on $\mathcal{F}_\mathcal{T}$.

Proposition 1 (Flexible Contract Solution)

Under Assumption 1, the equilibrium solution for the flexible contract is

$$w_\tau^F = \frac{\bar{A}_\tau + \delta_\tau^F}{2} \qquad \text{and} \qquad q_\tau^F = \frac{\bar{A}_\tau - \delta_\tau^F}{4\,\xi}, \tag{6}$$

where

$$\delta_\tau^F := \max\left\{c_\mathcal{T} \, , \ \sqrt{(\bar{A}_\tau^2 - 8\,\xi\, B)^+}\right\}.$$

The equilibrium expected payoffs of the players are then given by

$$\Pi_{P|\tau}^F = \frac{(\bar{A}_\tau + \delta_\tau^F - 2\, c_\mathcal{T})\,(\bar{A}_\tau - \delta_\tau^F)}{8\,\xi} \qquad \text{and} \qquad \Pi_{R|\tau}^F = \frac{(\bar{A}_\tau - \delta_\tau^F)^2}{16\,\xi}. \tag{7}$$

Proof: The proof of this result is straightforward and is therefore omitted.

For notational simplicity, we have not made explicit the dependence of the equilibrium wholesale price, ordering quantity, and players' payoffs on the budget B. We will make this a general rule in this and the following sections.

The auxiliary parameter, δ_τ^{F}, can be interpreted as a modified production cost, greater than or equal to the original cost $c_\mathcal{T}$, that is induced by the budget, B. That is, the state-dependent non-cooperative equilibrium in (6) is the same equilibrium that one would obtain if the producer's production cost were $\delta^{\mathrm{F}}_{\tau}$ and the supplier had an unlimited budget. We can think of this modified cost, $\delta^{\mathrm{F}}_{\tau}$, as a negative (random) externality that a limited budget imposes on the entire supply chain. The following is a direct consequence of the previous result.

Corollary 1 *For every $\omega \in \Omega$ the optimal wholesale price, w^F_τ, and optimal quantity, q^F_τ, are non-increasing and non-decreasing, respectively, as a function of the budget B. Furthermore*

$$\lim_{B\downarrow 0} w^F_\tau = \bar{A}_\tau \qquad \text{and} \qquad \lim_{B\downarrow 0} q^F_\tau = 0.$$

The corresponding payoffs, $\Pi^S_{R|\tau}$ and $\Pi^S_{P|\tau}$, are non-decreasing in B and vanish as $B \downarrow 0$.

Note that the optimal wholesale price, w^{F}_τ, increases as the budget, B, decreases. That is, the more cash constrained the retailer is the higher the wholesale price charged by the producer. In fact the limiting value, $\bar{A}_\tau$, is the maximum price that the producer can charge and still have an operative supply chain; see equation (5). Note also from equations (6) and (7) that when the budget is limited, that is $B < B^{\mathrm{F}}_\tau := \frac{\bar{A}^2_\tau - c^2_\tau}{8\,\xi}$, the wholesale price, ordering quantity, and retailer's payoff are independent of the manufacturing cost c_τ. This threshold, B^{F}_τ, is the budget above which the unconstrained optimal solution is achieved for a given path.

We now compare the equilibrium and the expected profits of the agents as a function of τ. More specifically, we compare the flexible contract where $\tau > 0$ with the simple contract where $\tau = 0$. This comparison is relevant as it reveals the agents' incentives to induce the other party to select one type of contract versus the other. We note that this is not a straightforward comparison because the production costs are different under the two contracts. Let us denote by $\Pi^{\mathrm{F}}_{\mathrm{P}} := \mathbb{E}^{\mathbb{Q}}_0[\Pi^{\mathrm{F}}_{\mathrm{P}|\tau}]$ the producer's expected payoff under a flexible contract. Similar notation is used for the retailer and the superscript 'S' will refer to the equilibrium solution of the simple contract.

Proposition 2 *Suppose that $B \le B^F_\tau$ almost surely and $B \le \frac{A^2 - c_0^2}{8\xi}$. Then*

$$\mathbb{E}^{\mathbb{Q}}_0[w^F_\tau] \le w^S, \quad \mathbb{E}^{\mathbb{Q}}_0[q^F_\tau] \ge q^S, \quad \Pi^F_P \le \Pi^S_P \quad \text{and} \quad \Pi^F_R \ge \Pi^S_R.$$

Furthermore, in the limit

$$\lim_{B\downarrow 0} \frac{\Pi^F_P}{\Pi^S_P} = \frac{1}{\bar{A} - c_\tau}\left(\bar{A} - c_\tau\,\mathbb{E}^{\mathbb{Q}}_0\left[\frac{A}{\bar{A}_\tau}\right]\right) \le 1 \quad \text{and} \quad \lim_{B\downarrow 0} \frac{\Pi^F_R}{\Pi^S_R} = \mathbb{E}^{\mathbb{Q}}\left[\frac{\bar{A}}{\bar{A}_\tau}\right] \ge 1.$$

However, if $B \ge B^F_\tau$ almost surely and $B \ge \frac{A^2 - c_0^2}{8\xi}$ then

$$\mathbb{E}^{\mathbb{Q}}_0[w^F_\tau] = w^S + \frac{c_\tau - c_0}{2}, \quad \text{and} \quad \mathbb{E}^{\mathbb{Q}}_0[q^F_\tau] = q^S - \frac{c_\tau - c_0}{4\,\xi}.$$

In addition,

$$\Pi^F_P \ge \Pi^S_P \quad \text{and} \quad \Pi^F_R \ge \Pi^S_R \qquad \text{if and only if} \qquad \mathrm{Var}(\bar{A}_\tau) + c^2_\tau - c^2_0 \ge 2\,\bar{A}\,(c_\tau - c_0).$$

Proof: See Appendix A.

Proposition 2 compares the supply-chain behavior under the simple and flexible contracts as a function of B. If the retailer's budget is small then the producer is worse off using the F-contract whereas the retailer is better off. However, when the budget is large then the agents' preferences over the contract depend on the additional condition $\mathrm{Var}(\bar{A}_\tau) + c_\tau^2 - c_0^2 \geq 2\,\bar{A}\,(c_\tau - c_0)$. This condition will be satisfied when the variance $\mathrm{Var}(\bar{A}_\tau)$ is large and/or the cost differential $c_\tau - c_0$ is small.

Proposition 2 provides only a partial characterization of the agents' preferences over the two types of contracts. In particular, the result does not cover those cases in which the budget has an intermediate value that can be greater than B_τ^{F} for some realizations of X (up to time τ) and less than B_τ^{F} for other realizations. In this case, the comparison between the contracts depends on the specific value of B and the distribution of $\bar{A}_\tau$, and must be done on a case-by-case basis. The example of Figure 1 assumes a uniform distribution for $\bar{A}_\tau$. In Case 1 (see the upper set of graphs)

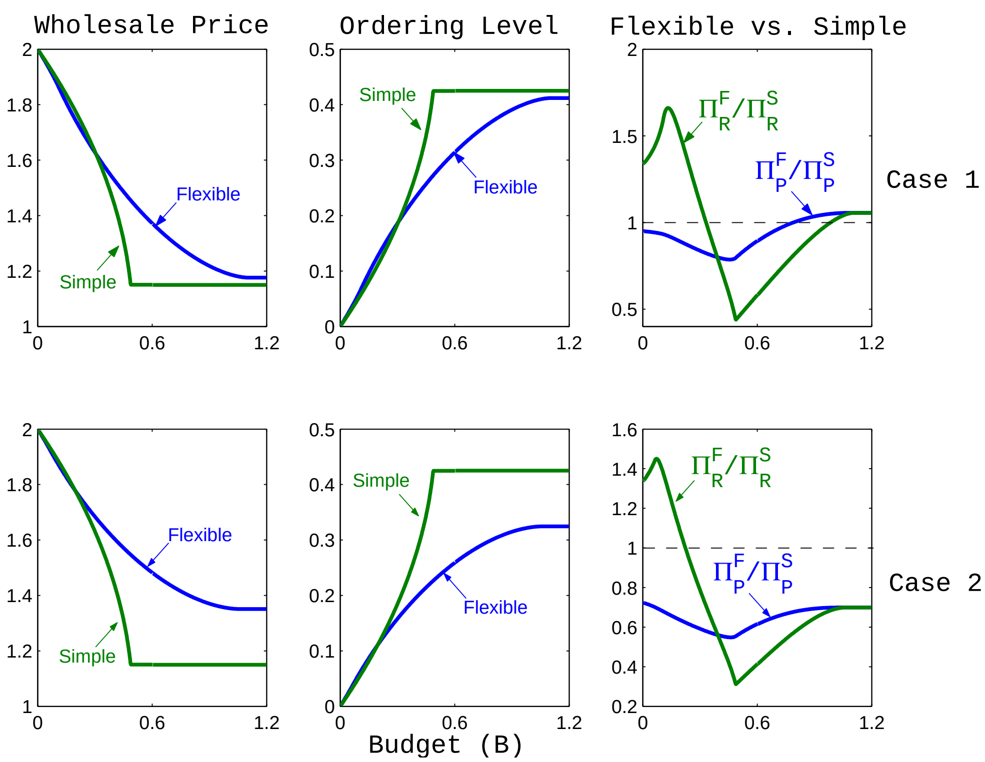

Figure 1: Flexible versus Simple Contract. The conditional demand parameter, $\bar{A}_\tau$, is uniformly distributed in $[1, 3]$. In Case 1 we take $c_\tau = 0.35$ and in Case 2 we take $c_\tau = 0.7$. In both cases $\xi = 1$, $c_0 = 0.3$.

the condition $\mathrm{Var}(\bar{A}_\tau) + c_\tau^2 - c_0^2 \geq 2\,\bar{A}\,(c_\tau - c_0)$ is satisfied while in Case 2 (see the lower set of graphs) the condition is not satisfied. The graphs on the left show the average wholesale price for the flexible and simple contracts. The graphs in the middle compare the ordering levels, while the graphs on the right plot the ratio of the players' payoffs under the two types of contracts. In Case 1 both players prefer the flexible contract when the budget is large and the reverse conclusion holds in Case 2. Furthermore, when the budget is small the retailer prefers the F-contract and the producer prefers the S-contract. These observations are consistent with Proposition 2.

The Centralized Solution

In order to study the efficiency of the non-cooperative or decentralized solution, we first need to compute the centralized solution for the flexible contract model. The centralized solution is found by assuming that a central planner, with the same initial budget B, solves

$$\Pi^{\mathrm{F}}_{\mathrm{C}} = \mathbb{E}^{\mathbb{Q}}_0 \left[\max_{q_\tau \geq 0} \left\{ \mathbb{E}^{\mathbb{Q}}_\tau \Big[(A - \xi\, q_\tau - c_\tau)\, q_\tau \Big] \right\} \right]$$
$$\text{subject to} \qquad c_\tau\, q_\tau \leq B, \qquad \text{for all } \omega \in \Omega.$$

The optimal solution, under Assumption 1, is

$$q^{\mathrm{F}}_{\mathrm{C}|\tau} = \frac{\bar{A}_\tau - \delta^{\mathrm{F}}_{\mathrm{C}|\tau}}{2\,\xi}, \qquad \text{where } \delta^{\mathrm{F}}_{\mathrm{C}|\tau} := \max\left\{ c_\tau\,,\, \bar{A}_\tau - \frac{2\xi B}{c_\tau} \right\}. \tag{8}$$

Defining $B^{\mathrm{F}}_{\mathrm{C}|\tau} := \frac{c_\tau\,(\bar{A}_\tau - c_\tau)}{2\xi}$, we obtain that the central planner's expected payoff is given by

$$\Pi^{\mathrm{F}}_{\mathrm{C}|\tau} = \begin{cases} \frac{B}{c_\tau^2} \Big(c_\tau\,(\bar{A}_\tau - c_\tau) - \xi\, B \Big) & \text{if } \; B \leq B^{\mathrm{F}}_{\mathrm{C}|\tau} \\ \frac{(\bar{A}_\tau - c_\tau)^2}{4\,\xi} & \text{if } \; B \geq B^{\mathrm{F}}_{\mathrm{C}|\tau}. \end{cases}$$

As was the case with the decentralized solution, the optimal quantity for the centralized solution, $q^{\mathrm{F}}_{\mathrm{C}|\tau}$, is non-decreasing in B and goes to zero as $B \downarrow 0$. The threshold, $B^{\mathrm{F}}_{\mathrm{C}|\tau}$, is the limiting budget above which the centralized solution reaches the unconstrained optimal value, $q^{\mathrm{F}}_{\mathrm{C}|\tau} = \frac{\bar{A}_\tau - c_\tau}{2\,\xi}$.

As was the case with Proposition 2, the following result compares the payoff of the central planner under the simple and flexible contracts.

Proposition 3 *Suppose that $B \leq \frac{c_\tau\,(\bar{A}_\tau - c_\tau)}{2\xi}$ almost surely and $B \leq \frac{c_0\,(\bar{A} - c_0)}{2\xi}$. Then*

$$\Pi^F_C \geq \Pi^S_C \qquad \textit{if and only if} \qquad (c_\tau^2 - c_0^2)\,\xi\, B \geq \bar{A}\, c_0\, c_\tau\,(c_\tau - c_0).$$

However, if $B \geq \frac{c_\tau\,(\bar{A}_\tau - c_\tau)}{2\xi}$ almost surely then

$$\Pi^F_C \geq \Pi^S_C \qquad \textit{if and only if} \qquad \mathrm{Var}(\bar{A}_\tau) + c_\tau^2 - c_0^2 \geq 2\,\bar{A}\,(c_\tau - c_0).$$

The proof of Proposition 3 is very similar to the proof of Proposition 2 and is therefore omitted. We see from the first part of the proposition that as $B \downarrow 0$ the central planner prefers the flexible contract. Note that the second part of the proposition is based on the same condition that we derived for the non-cooperative game. Therefore, for B sufficiently large, the retailer, the producer, and the central planner either all prefer the flexible contract or all prefer the simple contract.

Efficiency of The Centralized Solution

Let us now look at the efficiency of the decentralized solution by comparing it to the centralized solution. We first characterize the pathwise efficiency of the F-contract, that is the efficiency for a given outcome in $\mathcal{F}_\tau$. We will then examine the unconditional efficiency of the contract as perceived at time $t = 0$.

We introduce the following ratios:

$$\mathcal{Q}^{\mathrm{F}}_\tau := \frac{q^{\mathrm{F}}_\tau}{q^{\mathrm{F}}_{\mathrm{C}|\tau}} \qquad \text{and} \qquad \mathcal{W}^{\mathrm{F}}_\tau := \frac{w^{\mathrm{F}}_\tau}{c_\tau}.$$

The first ratio, $\mathcal{Q}^{\mathrm{F}}_\tau$, measures the degree of inefficiency of the decentralized solution in terms of production output. The second ratio, $\mathcal{W}^{\mathrm{F}}_\tau$, captures the margin over and above the production cost charged by the producer. Naturally, $\mathcal{W}_\tau^{\mathrm{F}} \geq 1$ and so it follows that $\mathcal{Q}_\tau^{\mathrm{F}} \leq 1$. This inefficiency of the decentralized solution has been long recognized in the economics literature and goes under the name of *double marginalization* (*e.g.*, Spengler 1950). We characterize these performance ratios here in the context of a budget constraint.

By Corollary 1, the double marginalization ratio, $\mathcal{W}_\tau^{\mathrm{F}}$, is a non-increasing function of B and satisfies $\lim_{B\downarrow 0} \mathcal{W}_\tau^{\mathrm{F}} = \frac{\bar{A}_\tau}{c_\tau}$. The ratio, $\mathcal{Q}_\tau^{\mathrm{F}}$, satisfies

$$\mathcal{Q}_\tau^{\mathrm{F}} = \begin{cases} \frac{c_\tau\ \ \bar{A}_\tau - \sqrt{\bar{A}_\tau^2 - 8\,\xi\,B}}{4\,\xi\,B} & \text{if } B \leq B_{\mathrm{C}|\tau}^{\mathrm{F}} \wedge B_\tau^{\mathrm{F}} \\ \frac{\bar{A}_\tau - \sqrt{\bar{A}_\tau^2 - 8\,\xi\,B}}{2\,(\bar{A}_\tau - c_\tau)} & \text{if } B_{\mathrm{C}|\tau}^{\mathrm{F}} \leq B \leq B_\tau^{\mathrm{F}} \\ \frac{c_\tau\,(\bar{A}_\tau - c_\tau)}{4\,\xi\,B} & \text{if } B_{\mathrm{C}|\tau}^{\mathrm{F}} \leq B \leq B_\tau^{\mathrm{F}} \\ \frac{1}{2} & \text{if } B \geq B_{\mathrm{C}|\tau}^{\mathrm{F}} \vee B_\tau^{\mathrm{F}} \end{cases}$$

where $x \vee y := \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$.

Depending on the values of the average market size, $\bar{A}_\tau$, and production cost, c_τ, either $B_{\mathrm{C}|\tau}^{\mathrm{F}} \geq B_\tau^{\mathrm{F}}$ or $B_{\mathrm{C}|\tau}^{\mathrm{F}} \leq B_\tau^{\mathrm{F}}$. For this reason we have to distinguish four possible cases in the computation of $\mathcal{Q}_\tau^{\mathrm{F}}$ as above. It is straightforward to show that $B_{\mathrm{C}|\tau}^{\mathrm{F}} \leq B_\tau^{\mathrm{F}}$ if and only if $\bar{A}_\tau \leq 3\,c_\tau$.

The monotonicity of $\mathcal{W}_\tau^{\mathrm{F}}$ implies that $\mathcal{Q}_\tau^{\mathrm{F}}$ increases in B in the range $B \in [0, B_{\mathrm{C}|\tau}^{\mathrm{F}} \wedge B_\tau^{\mathrm{F}}]$. Within this range, smaller budgets therefore hurt the efficiency of the supply chain with respect to the centralized solution more than larger budgets. In the limit we obtain

$$\lim_{B\downarrow 0} \mathcal{Q}_\tau^{\mathrm{F}} = \frac{c_\tau}{\bar{A}_\tau}.$$

For $B \geq B_{\mathrm{C}|\tau}^{\mathrm{F}} \vee B_\tau^{\mathrm{F}}$, however, the ratio $\mathcal{Q}_\tau^{\mathrm{F}}$ remains constant at $\frac{1}{2}$.

In the range $B_{\mathrm{C}|\tau}^{\mathrm{F}} \wedge B_\tau^{\mathrm{F}} \leq B \leq B_{\mathrm{C}|\tau}^{\mathrm{F}} \vee B_\tau^{\mathrm{F}}$, the behavior of $\mathcal{Q}_\tau^{\mathrm{F}}$ is different depending on the relationship between $B_{\mathrm{C}|\tau}^{\mathrm{F}}$ and B_τ^{F}. If $B_{\mathrm{C}|\tau}^{\mathrm{F}} \leq B_\tau^{\mathrm{F}}$ then $\mathcal{Q}_\tau^{\mathrm{F}}$ is increasing in B. If $B_\tau^{\mathrm{F}} \leq B_{\mathrm{C}|\tau}^{\mathrm{F}}$ then $\mathcal{Q}_\tau^{\mathrm{F}}$ is decreasing in B. In both cases, however, the double marginalization inefficiency is minimized at $B = B_\tau^{\mathrm{F}}$.

To analyze the overall efficiency of the F-contract we look at the *competition penalty*, $\mathcal{P}_\tau^{\mathrm{F}}$, (*e.g.*, Cachon and Zipkin 1999) which is defined as

$$\mathcal{P}_\tau^{\mathrm{F}} := 1 - \left(\frac{\Pi_{\mathrm{R}|\tau}^{\mathrm{F}} + \Pi_{\mathrm{P}|\tau}^{\mathrm{F}}}{\Pi_{\mathrm{C}|\tau}^{\mathrm{F}}} \right).$$

It is clear that $\mathcal{P}_\tau^{\mathrm{F}} \in [0, 1]$ with $\mathcal{P}_\tau^{\mathrm{F}} = 0$ implying that the decentralized chain is perfectly coordinated and achieving the same expected profit as the centralized system. When $\mathcal{P}_\tau^{\mathrm{F}} = 1$, however, the system is completely inefficient. In our setting, we can write the competition penalty as follows:

$$\mathcal{P}_\tau^{\mathrm{F}} = 1 - \left(\frac{\bar{A}_\tau - c_\tau - \xi\, q_\tau^{\mathrm{F}}}{\bar{A}_\tau - c_\tau - \xi\, q_{\mathrm{C}|\tau}^{\mathrm{F}}} \right) \mathcal{Q}_\tau^{\mathrm{F}}.$$

Proposition 4 *The competition penalty, as a function of B, is characterized as follows:*

$$\mathcal{P}_\tau^{F} = \begin{cases} \textit{decreases in } B & \textit{if } B \leq B_{C|\tau}^{F} \wedge B_\tau^{F} \\ \textit{decreases in } B & \textit{if } B_{C|\tau}^{F} \leq B \leq B_\tau^{F} \\ \textit{increases in } B & \textit{if } B_\tau^{F} \leq B \leq B_{C|\tau}^{F} \\ \frac{1}{4} & \textit{if } B \geq B_{C|\tau}^{F} \vee B_\tau^{F}. \end{cases}$$

Proof: The proof is straightforward and is therefore omitted.

Figure 2 summarizes the solution for the F-contract for a given realization in $\mathcal{F}_\tau$. The graphs on the top row correspond to the case $B^{\mathrm{F}}_{\mathrm{C}|\tau} \le B^{\mathrm{F}}_\tau$ while those on the bottom row correspond to $B^{\mathrm{F}}_\tau \le B^{\mathrm{F}}_{\mathrm{C}|\tau}$. The graphs on the left plot the quantity ratio, $\mathcal{Q}^{\mathrm{F}}_\tau$, the graphs in the middle plot the double marginalization ratio, $\mathcal{W}^{\mathrm{F}}_\tau$, and the graphs on the right plot the competition penalty, $\mathcal{P}^{\mathrm{F}}_\tau$. In the case $B^{\mathrm{F}}_\tau \le B^{\mathrm{F}}_{\mathrm{C}|\tau}$, or equivalently $\bar{A}_\tau \le 3\, c_\tau$, the competition penalty is minimized at $B = B^{\mathrm{F}}_\tau$

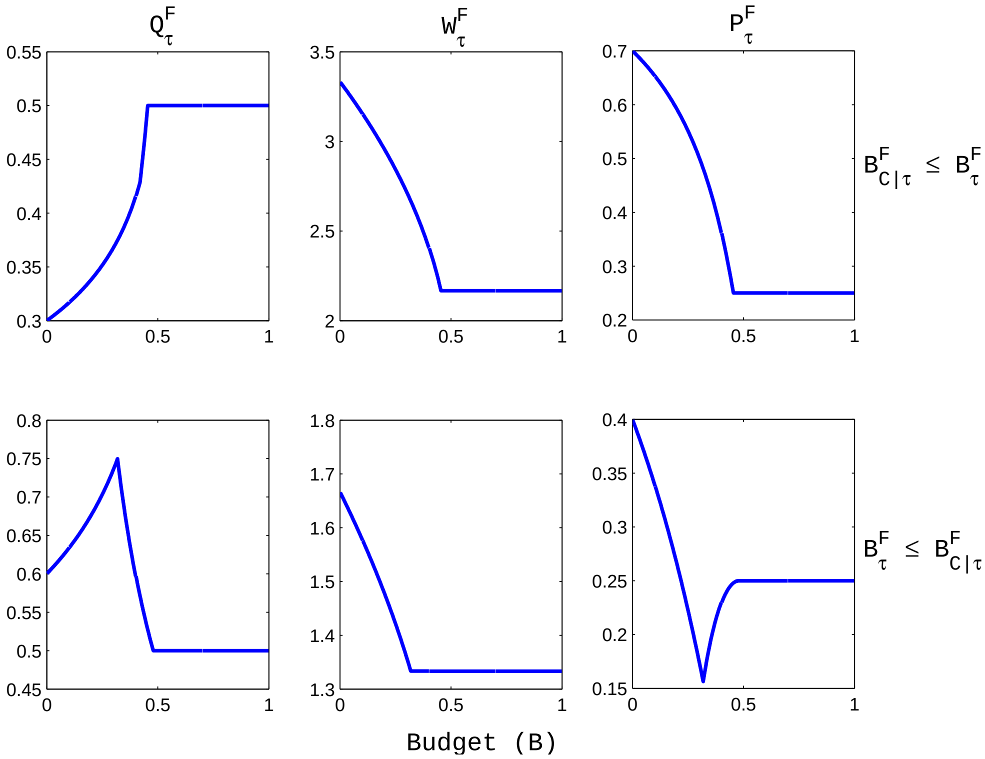

Figure 2: $\mathcal{Q}^{\mathrm{F}}_\tau$, $\mathcal{W}^{\mathrm{F}}_\tau$ and $\mathcal{P}^{\mathrm{F}}_\tau$ are plotted against B for the flexible contract. The demand model is such that $\bar{A}_\tau = 2$ and $\xi = 1$. The production cost is $c_\tau = 0.6$ for the top row and $c_\tau = 1.2$ for the bottom row.

and takes the value

$$\mathcal{P}^{\mathrm{F}}_{\mathrm{min}|\mathrm{X}} = \frac{(5\, c_\tau - \bar{A}_\tau)\,(\bar{A}_\tau - c_\tau)}{(\bar{A}_\tau + c_\tau)\,(7\, c_\tau - \bar{A}_\tau)} \le \frac{1}{4}.$$

If $\bar{A}_\tau = c_\tau$ note that the competition penalty vanishes but this is only due to the fact that $q = 0$ for both the decentralized and centralized supply chains.

Thus far, the efficiency of the F-contract has been discussed in a pathwise fashion, that is conditional on $\mathcal{F}_\tau$. We now consider the unconditional efficiency. In particular, we are interested in characterizing the expected production efficiency, $\mathcal{Q}^{\mathrm{F}} := \mathbb{E}^{\mathbb{Q}}_0[\mathcal{Q}^{\mathrm{F}}_\tau]$, the expected double marginalization, $\mathcal{W}^{\mathrm{F}} := \mathbb{E}^{\mathbb{Q}}_0[\mathcal{W}^{\mathrm{F}}_\tau]$, and the expected competition penalty, $\mathcal{P}^{\mathrm{F}} := \mathbb{E}^{\mathbb{Q}}_0[\mathcal{P}^{\mathrm{F}}_\tau]$.

The computation of these quantities follows directly from our previous analysis though the computations are rather tedious due to the number of different cases that arise in terms of B, B^{F}_τ, and $B^{\mathrm{F}}_{\mathrm{C}|\tau}$. The following proposition summarizes the unconditional efficiency of the F-contract in the limiting cases $B \downarrow 0$ and $B \uparrow \infty$.

Proposition 5 *In the limit as the budget, B, goes to 0 we obtain*

$$\lim_{B\downarrow 0} \mathcal{Q}^F = \mathbb{E}^{\mathbb{Q}}[\frac{c_\mathcal{T}}{\bar{A}_\tau}] \geq \frac{c_\mathcal{T}}{\bar{A}}, \qquad \lim_{B\downarrow 0} \mathcal{W}^F = \frac{\bar{A}}{c_\mathcal{T}}, \qquad \textit{and} \qquad \lim_{B\downarrow 0} \mathcal{P}^F = 1 - \mathbb{E}^{\mathbb{Q}}[\frac{c_\mathcal{T}}{\bar{A}_\tau}] \leq \frac{\bar{A} - c_\mathcal{T}}{\bar{A}}.$$

As $B \to \infty$ we obtain

$$\lim_{B\uparrow\infty} \mathcal{Q}^F = \frac{1}{2}, \qquad \lim_{B\uparrow\infty} \mathcal{W}^F = \frac{\bar{A} + c_\mathcal{T}}{2c_\mathcal{T}}, \qquad \textit{and} \qquad \lim_{B\uparrow\infty} \mathcal{P}^F = \frac{1}{4}.$$

Proof: The proof follows from the nonnegativity of $\bar{A}_\tau$, the bounded convergence theorem, and Jensen's inequality.

Proposition 5 implies that for $B \downarrow 0$ or $B \uparrow \infty$ the expected double marginalization, $\mathcal{W}^{\mathrm{F}}$, decreases with τ. That is, production postponement reduces, on average, the producer's margin. On the other hand, the competition penalty is maximized at $\tau = 0$ for B small and it is constant, independent of τ, for B large.

4 Flexible Contract with Financial Hedging

We now consider the H-contract, that is the flexible contract but where the retailer now has access to the financial markets. The complete financial markets assumption implies that the retailer can modify his budget by purchasing any $\mathcal{F}_\mathcal{T}$-measurable financial claim, $G_\mathcal{T}$, where, as usual, $\{\mathcal{F}_t\}_{0\leq t\leq T}$ is the filtration generated by the financial noise, X_t. Assuming without loss of generality[15] that an initial capital of 0 is devoted to the financial hedging strategy, we then have $\mathbb{E}_0^{\mathbb{Q}}[G_\mathcal{T}] = 0$. The retailer's budget at time τ is then given by $B_\mathcal{T} = B + G_\mathcal{T}$. By optimizing over $G_\mathcal{T}$, the retailer can transfer cash resources from states where the budget constraint is not binding to states where it is. In a partial equilibrium setting, that is for a fixed $w_\mathcal{T}$, it is clear that the retailer will prefer the H-contract to the F-contract. In our competitive setting, however, this is no longer clear. In fact we shall see that on some occasions the retailer will prefer the H-contract but on other occasions he will prefer the F-contract. We shall see that the producer, however, will always prefer the H-contract to the F-contract.

The Decentralized Solution

The sequence of events in the H-contract setting is as follows. At time $t = 0$, the producer offers a menu of wholesale prices, $w_\mathcal{T}$. In response, the retailer selects a menu of ordering quantities, $q_\mathcal{T} = q(w_\mathcal{T})$, as well as an $\mathcal{F}_\mathcal{T}$-measurable financial claim, $G_\mathcal{T}$, that satisfies $\mathbb{E}_0^{\mathbb{Q}}[G_\mathcal{T}] = 0$. At time τ the outcome is observed and the producer immediately manufactures $q_\mathcal{T}$ product units which he then sells to the retailer at a per unit price of w_τ. By construction, the retailer's budget, B_τ, is sufficient to pay the producer for these units. Finally, the retailer sells all the units in the retail market at time T at the stochastic per-unit clearance price, $A - \xi\, q_\tau$.

The distinguishing feature of the H-contract is that the budget constraint is now a path-wise constraint of the form

$$w_\tau\, q_\tau \leq B_\tau, \qquad \text{for all } \omega \in \Omega$$

[15]See Section 2.3.

where $\mathbb{E}_0^{\mathbb{Q}}[B_\mathcal{T}] = B$. The retailer's problem is then given by

$$\Pi_{\mathrm{R}}^{\mathrm{H}}(w_\mathcal{T}) = \max_{q_\mathcal{T} \geq 0,\, B_\mathcal{T}} \; \mathbb{E}_0^{\mathbb{Q}}\left[(\bar{A}_\tau - \xi\, q_\tau - w_\tau)\, q_\tau\right] \tag{9}$$

$$\text{subject to} \quad w_\tau\, q_\tau \leq B_\mathcal{T}, \qquad \text{for all } \omega \in \Omega \tag{10}$$

$$\mathbb{E}_0^{\mathbb{Q}}[B_\mathcal{T}] = B. \tag{11}$$

Note that it is no longer possible to decouple the problem and solve it separately for every realization of X (up to time τ) as we did with the F-contract. This is because the new constraint, $\mathbb{E}_0^{\mathbb{Q}}[B_\mathcal{T}] = B$, binds the entire problem together. We have the following solution to the retailer's problem.

Proposition 6 (Retailer's Optimal Strategy)

Let $w_\mathcal{T}$ be the menu of wholesale prices offered by the producer and let $\mathcal{Q}_\tau$, $\mathcal{X}$ and $\mathcal{X}^c$ be defined as follows

$$\mathcal{Q}_\tau := \left(\frac{\bar{A}_\tau - w_\tau}{2\,\xi}\right)^+, \qquad \mathcal{X} := \{\omega \in \Omega \;:\; B \geq \mathcal{Q}_\tau\, w_\tau\}, \qquad \textit{and} \qquad \mathcal{X}^c := \Omega - \mathcal{X}.$$

The following two cases arise in the computation of the optimal ordering quantity, $q(w_\tau)$, and the financial claim, $G_\mathcal{T}$.

Case 1: *Suppose that $\mathbb{E}_0^{Q}\left[\mathcal{Q}_\tau\, w_\tau\right] \leq B$. Then $q(w_\tau) = \mathcal{Q}_\tau$ and there are infinitely many choices of the optimal claim, $G_\mathcal{T}$. One natural choice is to take*

$$G_\mathcal{T} = \left[\mathcal{Q}_\tau\, w_\tau - B\right] \cdot \begin{cases} \delta & \text{if } \omega \in \mathcal{X} \\ 1 & \text{if } \omega \in \mathcal{X}^c \end{cases}$$

$$\delta := \frac{\int_{\mathcal{X}^c} \left[\mathcal{Q}_\tau\, w_\tau - B\right] \mathrm{d}\mathbb{Q}}{\int_{\mathcal{X}} \left[B - \mathcal{Q}_\tau\, w_\tau\right] \mathrm{d}\mathbb{Q}}.$$

In this case (possibly due to the ability to trade in the financial market), the budget constraint is not binding.

Case 2: *Suppose that $B < \mathbb{E}_0^{\mathbb{Q}}\left[\mathcal{Q}_\tau\, w_\tau\right]$. Then*

$$q_\tau(w_\tau) = \left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi}\right)^+ \qquad \textit{and} \qquad G_\mathcal{T} = q_\tau(w_\tau) w_\tau - B,$$

where $\lambda \geq 0$ solves

$$\mathbb{E}_0^{\mathbb{Q}}\left[w_\tau\, \left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi}\right)^+\right] = B.$$

Proof: It is straightforward to see that $\mathcal{Q}_\tau$ is the retailer's optimal ordering level given the wholesale price menu, $w_\mathcal{T}$, in the absence of a budget constraint. In order to implement this solution, the retailer would need a budget $\mathcal{Q}_\tau\, w_\tau$ for all $\omega \in \Omega$. Therefore, if the retailer can generate a financial gain, $G_\mathcal{T}$, such that $\mathcal{Q}_\tau\, w_\tau \leq B + G_\mathcal{T}$ for all $\omega \in \Omega$ then he would be able to achieve his unconstrained optimal solution.

By definition, $\mathcal{X}$ contains all those states for which $B \geq \mathcal{Q}_\tau w_\tau$. That is, the original budget B is large enough to cover the optimal purchasing cost for all $\omega \in \mathcal{X}$. However, for $\omega \in \mathcal{X}^c$, the initial budget is not sufficient. The financial gain, $G_\mathcal{T}$, then allows the retailer to transfer resources from $\mathcal{X}$ to $\mathcal{X}^c$.

Suppose the condition in Case 1 holds so that $\mathbb{E}_0^{\mathbb{Q}}[\mathcal{Q}_\tau w_\tau] \leq B$. Note that according to the definition of $G_\mathcal{T}$ in this case, we see that $B + G_\mathcal{T} = \mathcal{Q}_\tau w_\tau$ for all $\omega \in \mathcal{X}^c$. For $\omega \in \mathcal{X}$, however, $B + G_\mathcal{T} = (1-\delta)\, B + \delta\, \mathcal{Q}_\tau w_\tau \geq \mathcal{Q}_\tau w_\tau$. The inequality follows since $\delta \leq 1$. $G_\mathcal{T}$ therefore allows the retailer to implement the unconstrained optimal solution. The only point that remains to check is that $G_\mathcal{T}$ satisfies $\mathbb{E}_0^{\mathbb{Q}}[G_\mathcal{T}] = 0$. This follows directly from the definition of δ.

Suppose now that the condition specified in Case 2 holds. We solve the retailer's optimization problem in (9) by relaxing the gain constraint (11) with a Lagrange multiplier, λ. We also relax the budget constraint in (10) for each realization of X up to time τ. The corresponding multiplier for each such realization is denoted by $\beta_\tau\, \mathrm{d}\mathbb{Q}$ where β_τ plays the role of a Radon-Nikodym derivative of a positive measure that is absolutely continuous with respect to $\mathbb{Q}$. The first-order optimality conditions for the relaxed version of the retailer's problem are then given by

$$q_\tau = \frac{(\bar{A}_\tau - w_\tau\,(1+\beta_\tau))^+}{2\,\xi}$$

$$\beta_\tau = \lambda, \qquad \beta_\tau\left(w_\tau\, q_\tau - B + G_\mathcal{T}\right) = 0, \qquad \beta_\tau \geq 0, \qquad \text{and } \mathbb{E}_0^{\mathbb{Q}}[G_\mathcal{T}] = 0.$$

It is straightforward to show that the solution given in Case 2 of the proposition satisfies these optimality conditions; only the non-negativity of β_τ needs to be checked separately. To prove this, note that $\beta_\tau = \lambda$, therefore it suffices to show that $\lambda \geq 0$. This follows from three observations

(a) Since $0 \leq w_\tau$ the function $\mathbb{E}_0^{\mathbb{Q}}\left[w_\tau\,\left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi}\right)^+\right]$ is decreasing in λ.

(b) In Case 2, by hypothesis, we have

$$\mathbb{E}_0^{\mathbb{Q}}\left[w_\tau\,\left(\frac{\bar{A}_\tau - w_\tau}{2\,\xi}\right)^+\right] = \mathbb{E}_0^{\mathbb{Q}}\left[\mathcal{Q}_\tau\, w_\tau\right] > B$$

(c) Finally, we know that λ solves

$$\mathbb{E}_0^{\mathbb{Q}}\left[w_\tau\,\left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi}\right)^+\right] = B.$$

(a) and (b) therefore imply that we must have $\lambda \geq 0$. □

Case 1 of Proposition 6 describes the circumstances when trading in the financial market allows the retailer to completely remove the budget constraint from his optimization problem. When these circumstances are not satisfied as in Case 2, the retailer cannot completely remove the budget constraint. He can, however, mitigate the effects of the budget constraint somewhat so that for a fixed menu of wholesale prices, w_τ, he prefers the H-contract to the F-contract.

Based on the retailer's best-response strategy derived in Proposition 6, the producer's problem can be formulated[16] as

$$\Pi_{\mathrm{P}}^{\mathrm{H}} = \max_{w_\tau,\, \lambda \geq 0} \; \mathbb{E}_0^{\mathbb{Q}} \left[(w_\tau - c_\mathcal{T}) \left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi} \right)^+ \right] \tag{12}$$

$$\text{subject to} \qquad \mathbb{E}_0^{\mathbb{Q}} \left[w_\tau \left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi} \right)^+ \right] \leq B. \tag{13}$$

The following result characterizes the solution of this problem and the corresponding solution of the Stackelberg game.

Proposition 7 (Producer's Optimal Strategy and the Stackelberg Solution)
Let ϕ^H be the minimum $\phi \geq 1$ that solves

$$\mathbb{E}_0^{\mathbb{Q}} \left[\left(\frac{\bar{A}_\tau^2 - (\phi\, c_\mathcal{T})^2}{8\,\xi} \right)^+ \right] \leq B.$$

Define $\delta^H := \phi^H\, c_\mathcal{T}$, then the optimal wholesale price and ordering level satisfy

$$w_\tau^H = \frac{\bar{A}_\tau + \delta^H}{2} \qquad \textit{and} \qquad q_\tau^H = \left(\frac{\bar{A}_\tau - \delta^H}{4\,\xi} \right)^+. \tag{14}$$

The players' expected payoffs satisfy

$$\Pi_{P|\tau}^H = \frac{(\bar{A}_\tau + \delta^H - 2c_\mathcal{T})\,(\bar{A}_\tau - \delta^H)^+}{8\,\xi} \qquad \textit{and} \qquad \Pi_{R|\tau}^H = \frac{((\bar{A}_\tau - \delta^H)^+)^2}{16\,\xi}. \tag{15}$$

Proof: See Appendix A.

As before, we interpret δ^{H} as a modified production cost, greater than or equal to the the original cost, $c_\mathcal{T}$, that is imposed in the supply chain because of the limited budget. Unlike the setting of the F-contract, however, the modified cost in this setting is not stochastic. Note that δ^{H} is nondecreasing in B. Hence, as in the F-contract, the more cash constrained the retailer is the higher the wholesale price charged by the producer.

Suppose now that the budget is limited so that $\delta^{\mathrm{H}} > c_\mathcal{T}$. Then, depending on the value of δ^{H}, Proposition 7 implies that it is possible for $w_\tau^{\mathrm{H}} \geq \bar{A}_\tau$ and $q_\tau^{\mathrm{H}} = 0$ for some outcomes $\omega \in \Omega$. That is, in some cases the producer decides to overcharge the retailer and therefore make the supply chain nonoperative. Because of Assumption 1, this behavior was never optimal in the setting of the F-contract. It occurs in the H-contract setting, however, because the retailer can allocate his limited budget among different states $\omega \in \Omega$. In particular, if the retailer knows that for some outcomes, ω, he will not be purchasing any units then he can transfer the entire budget B from these (non-operative) states to states in which there is a need for cash. It is in the producer's interest, then, to select those states in which he wants to do business with the retailer and those in which he does not. Note that $q_\tau^{\mathrm{H}} = 0$ if and only if $\bar{A}_\tau \leq \delta^{\mathrm{H}}$. Hence the producer "closes" the supply chain when the forecasted demand is low.

We now compare the F-contract with the H-contract in terms of the players expected payoffs under the Nash equilibrium. First we define

$$\widehat{\mathcal{X}} := \{\omega \in \Omega \,:\, \delta_\tau^{\mathrm{F}} = c_\mathcal{T}\}$$

[16]Note that at the optimal solution, the constraint in (13) will be tight if the optimal λ is greater than zero. This will only occur when the budget constraint is binding.

where δ_τ^{F} was defined in proposition 1 in Section 3. The set $\widehat{\mathcal{X}}$ characterizes those states, ω, for which the flexible contract achieves the unconstrained optimal solution, $w_\tau^{\text{F}} = \frac{\bar{A}_\tau + c_\mathcal{T}}{2}$ and $q_\tau^{\text{F}} = \frac{\bar{A}_\tau - c_\mathcal{T}}{4\xi}$. We also recall that the equilibrium wholesale prices and ordering levels for the F-contract and H-contract are

$$w_\tau^{\text{F}} = \frac{\bar{A}_\tau + \delta_\tau^{\text{F}}}{2} \quad q_\tau^{\text{F}} = \frac{\bar{A}_\tau - \delta_\tau^{\text{F}}}{4\,\xi} \quad \text{and} \quad w_\tau^{\text{H}} = \frac{\bar{A}_\tau + \delta^{\text{H}}}{2} \quad q_\tau^{\text{H}} = \frac{(\bar{A}_\tau - \delta^{\text{H}})^+}{4\,\xi},$$

respectively. The difference between the expected payoffs of these two contracts depends on the difference between δ_τ^{F} and δ^{H}, which in turn depends on the set $\widehat{\mathcal{X}}$. We now show that the producer's expected payoff under the H-contract is always greater than his expected payoff under the F-contract.

Proposition 8 *The producer is always better-off if the retailer is able to hedge the budget constraint.*

Proof: See Appendix A.

According to this result, it is in the producer's interest to promote the retailer's ability to trade in the financial market. If the retailer is a small player with limited access to the financial markets, then it would be in the producer's interest to serve as an intermediary between the retailer and the financial markets.

¿From the retailer's perspective, the comparison between the F-contract and H-contract is not so straightforward. We identify three cases.

- **Case 1:** Suppose that $\widehat{\mathcal{X}} = \Omega$. In this case, B is sufficiently large so that $\delta_\tau^{\text{F}} = \delta^{\text{H}} = c_\mathcal{T}$ for all $\omega \in \Omega$ and the two contracts produce the same output. This is not surprising since for large budgets financial trading does not offer any advantage.

- **Case 2:** Suppose that $\widehat{\mathcal{X}} \neq \Omega$ and $\delta^{\text{H}} = c_\mathcal{T}$. In this case, $\delta_\tau^{\text{F}} > c_\mathcal{T}$ for all $\omega \in \widehat{\mathcal{X}}^c$. Therefore, $w_\tau^{\text{H}} \leq w_\tau^{\text{F}}$ and $q_\tau^{\text{H}} \geq q_\tau^{\text{F}}$ for all $\omega \in \Omega$ with strict inequalities in $\widehat{\mathcal{X}}^c$. With regards to the payoffs, using equations (7) and (15) we can conclude that for all $\omega \in \Omega$

$$\Pi_{\text{R}|\tau}^{\text{H}} \geq \Pi_{\text{R}|\tau}^{\text{F}},$$

  with strict inequality in $\widehat{\mathcal{X}}^c$. Note that this case summarizes well the advantages of using financial trading: the ability to trade has increased the output of the supply chain, reduced the wholesale price, reduced the double marginalization inefficiency and has increased the payoff of both agents. These conclusions hold for all $\omega \in \Omega$ in this case. Therefore, they hold in expectation, so that $\mathbb{E}_0^{\mathbb{Q}}[\Pi_{\text{R}|\tau}^{\text{H}}] \geq \mathbb{E}_0^{\mathbb{Q}}[\Pi_{\text{R}|\tau}^{\text{F}}]$.

- **Case 3:** Suppose that $\widehat{\mathcal{X}} \neq \Omega$ and $\delta^{\text{H}} > c_\mathcal{T}$. In this case, $\delta_\tau^{\text{F}} < \delta^{\text{H}}$ for $\omega \in \widehat{\mathcal{X}}$ and the wholesale price (ordering quantity) is smaller (higher) under the F-contract than under the H-contract. In terms of payoffs, the retailer (and the producer as well) therefore prefers the F-contract to the H-contract for $\omega \in \widehat{\mathcal{X}}$. Of course, the choice of the contract has to be made at $t = 0$ when the realization of ω is still unknown. Therefore, the appropriate comparison between the contracts should be based on their time $t = 0$ expected payoffs. As the following example shows, however, the retailer can be better-off or worse-off under the H-contract.

Example 1 Consider the special case in which $\bar{A}_\tau$ takes only the values $\{5, 10\}$ with equal probability and $8\,\xi = 1$ and $B = 9.5$.

If $c_\mathcal{T} = 1$ then we can show that $\delta^{\mathsf{H}} = 9 > c_\mathcal{T}$ and $\mathbb{E}_0^{\mathbb{Q}}[\Pi^{\mathsf{H}}_{\mathsf{R}|\tau}] = 0.25$ and $\mathbb{E}_0^{\mathbb{Q}}[\Pi^{\mathsf{F}}_{\mathsf{R}|\tau}] = 0.342$.

If $c_\mathcal{T} = 4.5$ then $\delta^{\mathsf{H}} = 9 > c_\mathcal{T}$ and $\mathbb{E}_0^{\mathbb{Q}}[\Pi^{\mathsf{H}}_{\mathsf{R}|\tau}] = 0.25$ and $\mathbb{E}_0^{\mathbb{Q}}[\Pi^{\mathsf{F}}_{\mathsf{R}|\tau}] = 0.122$. $\square$

Nevertheless, under some additional conditions we can show that for sufficiently small B the retailer is always better-off under the H-contract.

Proposition 9 *Suppose the random variable $\bar{A}_\tau$ has a bounded support and admits a smooth density bounded away from zero. Furthermore, assume that $\bar{A}_\tau > c_\mathcal{T}$ for all $\omega \in \Omega$. Then, as $B \downarrow 0$ we obtain*

$$\mathbb{E}_0^{\mathbb{Q}}[\Pi^{F}_{R|\tau}] = \xi\, B^2\, \mathbb{E}_0^{\mathbb{Q}}\left[\frac{1}{\bar{A}_\tau^2}\right] + O\big(B^3\big) \qquad \textit{and} \qquad \mathbb{E}_0^{\mathbb{Q}}[\Pi^{H}_{R|\tau}] \geq K\, B^{\frac{3}{2}}$$

for some constant $K > 0$. Hence, for B sufficiently small $\mathbb{E}_0^{\mathbb{Q}}[\Pi^{F}_{R|\tau}] \leq \mathbb{E}_0^{\mathbb{Q}}[\Pi^{H}_{R|\tau}]$.

Proof: See Appendix A.

According to the previous discussion, if $\delta^{\mathrm{H}} = c_\mathcal{T}$ then both players are better-off using the H-contract and so it follows the entire supply chain is also better-off. For the case $\delta^{\mathrm{H}} > c_\mathcal{T}$, it is possible that the retailer prefers the F-contract and so it is not clear which contract has a higher total expected payoff, i.e. the sum of the retailer's and producer's expected profits.

Figure 3 shows the performance of the F-contract and H-contract, in terms of expected wholesale price, ordering level and players' payoffs, as a function of the budget, B. It may be seen that if the budget is small then, on average, the wholesale price is smaller and the ordering level is higher for the F-contract than for the H-contract. This situation is reversed as the budget increases. In terms of the payoffs, both agents prefer the H-contract to the F-contract for all levels of B in this particular example. Furthermore, the benefits of the H-contract with respect to the F-contract are most pronounced for intermediate values of B.

The Centralized Solution

We now solve the centralized solution when hedging by the retailer is permitted. The central planner's problem is similar to the retailer's problem in (9)-(11) and is given by

$$\begin{aligned} \Pi^{\mathrm{H}}_{\mathrm{C}} = \max_{q_\mathcal{T},\, B_\mathcal{T}} \;& \mathbb{E}_0^{\mathbb{Q}}\left[\left(\bar{A}_\tau - \xi\, q_\tau - c_\mathcal{T}\right) q_\tau\right] && (16)\\ \text{subject to} \quad & w_\mathcal{T}\, q_\mathcal{T} \leq B_\mathcal{T}, \qquad \text{for all } \omega \in \Omega && (17)\\ & \mathbb{E}_0^{\mathbb{Q}}[B_\mathcal{T}] = B. && (18) \end{aligned}$$

Proposition 10 summarizes the optimal solution for the central planner. The proof is almost identical to the proof of Proposition 6 and is therefore omitted.

Proposition 10 (Central Planner's Optimal Strategy)

The optimal production strategy, $q^{H}_{C|\tau}$, is given by

$$q^{H}_{C|\tau} = \left(\frac{\bar{A}_\tau - \delta^{H}_{C}}{2\,\xi}\right)^{+} \tag{19}$$

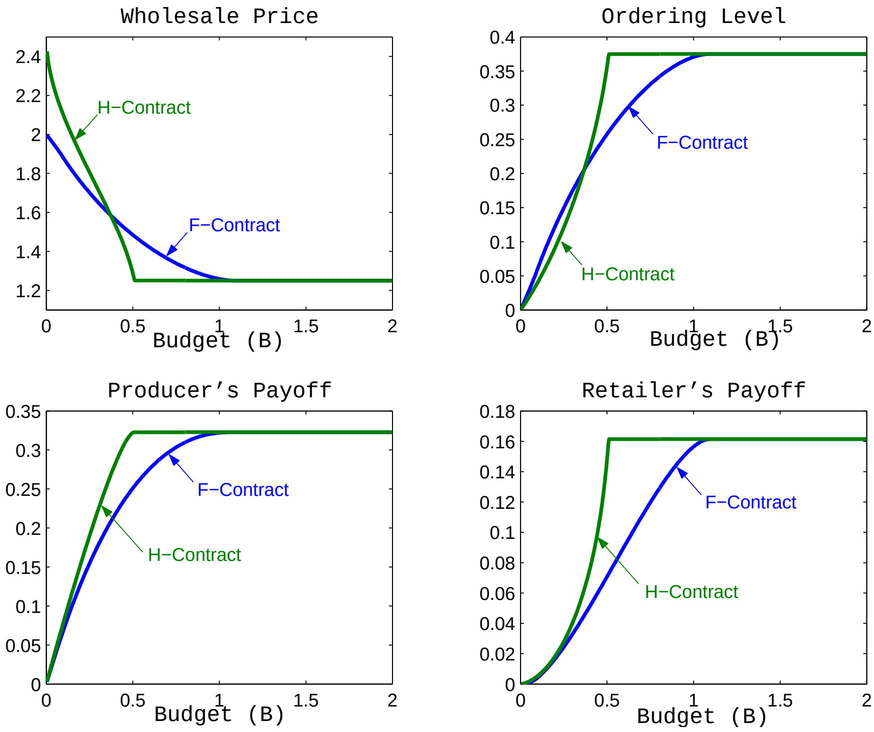

Figure 3: Performance of F-contract and H-contract as a function of the budget. The demand parameter $\bar{A}_\tau$ is uniformly distributed in $[1, 3]$, $\xi = 1$, and $c_\tau = 0.5$.

where δ_C^H is the minimum $\delta \geq c_\tau$ that solves

$$\mathbb{E}^{\mathbb{Q}}\left[c_\tau \left(\frac{\bar{A}_\tau - \delta_C^H}{2\,\xi}\right)^+\right] \leq B.$$

The central planner's optimal payoff given the information available at time τ is

$$\Pi_{C|\tau}^H = \frac{(\bar{A}_\tau + \delta_C^H - 2c_\tau)\,(\bar{A}_\tau - \delta_C^H)^+}{4\,\xi}. \tag{20}$$

Once again, we interpret $\delta_{\mathrm{C}}^{\mathrm{H}}$ as a modified production cost induced by the budget constraint.

Efficiency of The Centralized Solution

With this modified production cost structure in mind, one would expect the centralized solution to be more efficient than the decentralized solution in the sense that $\delta_{\mathrm{C}}^{\mathrm{H}} \leq \delta^{\mathrm{H}}$. This is not always the case, however, as the following example demonstrates.

Example 2 Consider the following instance of the problem with $B = 0.45$, $\xi = c_\tau = 1$, and $\bar{A}_\tau$ uniformly distributed in $[1, 3]$. Since

$$\mathbb{E}_0^{\mathbb{Q}}\left[\left(\frac{\bar{A}_\tau^2 - c_\tau^2}{8\xi}\right)^+\right] = \frac{5}{12} < B \quad \text{and} \quad \mathbb{E}_0^{\mathbb{Q}}\left[c_\tau\left(\frac{\bar{A}_\tau - c_\tau}{2\xi}\right)\right] = \frac{1}{2} > B,$$

it follows that $c_\tau = \delta^{\mathsf{H}} < \delta^{\mathsf{H}}_{\mathsf{C}}$. Furthermore, we can shown that $\delta^{\mathsf{H}}_{\mathsf{C}} \approx 1.103$. Therefore, for values of $\bar{A}_\tau$ in $[1, \delta^{\mathsf{H}}_{\mathsf{C}})$, the central planner does not produce, i.e. $q^{\mathsf{H}}_{\mathsf{C}|\tau} = 0$, while the decentralized supply chain does operate, i.e. $q^{\mathsf{H}}_\tau > 0$. Since

$$\mathbb{E}^{\mathbb{Q}}_0[q^{\mathsf{H}}_\tau] = \mathbb{E}^{\mathbb{Q}}_0\left[\frac{\bar{A}_\tau - c_\tau}{4\,\xi}\right] = \frac{1}{4} \qquad \text{and} \qquad \mathbb{E}^{\mathbb{Q}}_0[q^{\mathsf{H}}_{\mathsf{C}|\tau}] = \frac{B}{c_\tau} = 0.45$$

the central planner, on average, produces more than the decentralized supply chain. □

The previous example highlights an interesting feature of the H-contract: contingent on the outcome ω, the centralized supply chain can produce less than the decentralized solution. This was never the case under the F-contract (or S-contract). On average, however, the central planner always produces more than the decentralized supply chain. To see this, first note that if $\delta^{\mathrm{H}}_{\mathrm{C}} = c_\tau$ then (14) and (19) imply that $q^{\mathrm{H}}_{\mathrm{C}|\tau} \geq q^{\mathrm{H}}_\tau$ for all ω. However, if $\delta^{\mathrm{H}}_{\mathrm{C}} > c_\tau$ then Proposition 10 implies that $c_\tau\, \mathbb{E}^{\mathbb{Q}}_0[q^{\mathrm{H}}_{\mathrm{C}|\tau}] = B$. Then Proposition 7, together with Assumption 1, imply that

$$B \geq \mathbb{E}^{\mathbb{Q}}_0\left[w^{\mathrm{H}}_\tau q^{\mathrm{H}}_\tau\right] = \mathbb{E}^{\mathbb{Q}}_0\left[\left(\frac{\bar{A}_\tau + \delta^{\mathrm{H}}}{2}\right)\left(\frac{\bar{A}_\tau - \delta^{\mathrm{H}}}{4\,\xi}\right)^+\right] \geq c_\tau\, \mathbb{E}^{\mathbb{Q}}_0\left[\left(\frac{\bar{A}_\tau - \delta^{\mathrm{H}}}{4\,\xi}\right)^+\right] = c_\tau\, \mathbb{E}^{\mathbb{Q}}_0[q^{\mathrm{H}}_\tau]$$

implying, in particular, that $\mathbb{E}^{\mathbb{Q}}_0[q^{\mathrm{H}}_{\mathrm{C}|\tau}] \leq \mathbb{E}^{\mathbb{Q}}_0[q^{\mathrm{H}}_\tau]$.

We conclude this section by examining the efficiency[17] of the H-contract in terms of production levels, double marginalization, and the competition penalty. Towards this end, we define the following performance measures that are conditional on the information available at time τ.

$$\mathcal{Q}^{\mathrm{H}}_\tau := \frac{q^{\mathrm{H}}_\tau}{q^{\mathrm{H}}_{\mathrm{C}|\tau}} = \frac{(\bar{A}_\tau - \delta^{\mathrm{H}})^+}{2(\bar{A}_\tau - \delta^{\mathrm{H}}_{\mathrm{C}})^+}, \qquad \mathcal{W}^{\mathrm{H}}_\tau := \frac{w^{\mathrm{H}}_\tau}{c_\tau} = \frac{\bar{A}_\tau + \delta^{\mathrm{H}}}{2c_\tau}, \qquad \text{and}$$

$$\mathcal{P}^{\mathrm{H}}_\tau := 1 - \frac{\mathbb{E}^{\mathbb{Q}}_0[\Pi^{\mathrm{H}}_{\mathrm{P}|\tau}] + \mathbb{E}^{\mathbb{Q}}_0[\Pi^{\mathrm{H}}_{\mathrm{R}|\tau}]}{\mathbb{E}^{\mathbb{Q}}_0[\Pi^{\mathrm{H}}_{\mathrm{C}|\tau}]} = 1 - \frac{(3\bar{A}_\tau + \delta^{\mathrm{H}} - 4c_\tau)\,(\bar{A}_\tau - \delta^{\mathrm{H}})^+}{4\,(\bar{A}_\tau + \delta^{\mathrm{H}}_{\mathrm{C}} - 2c_\tau)(\bar{A}_\tau - \delta^{\mathrm{H}}_{\mathrm{C}})^+}.$$

It is interesting to note that, conditional on $\mathcal{F}_\tau$, the centralized supply chain is not necessarily more efficient than the decentralized operation. For instance, we know that in some cases $\delta^{\mathrm{H}} < \delta^{\mathrm{H}}_{\mathrm{C}}$ (as in Example 2 above) and so for all those ω with $\delta^{\mathrm{H}} < \bar{A}_\tau < \delta^{\mathrm{H}}_{\mathrm{C}}$, $q^{\mathrm{H}}_{\mathrm{C}|\tau} = 0$ and $q^{\mathrm{H}}_\tau > 0$ and the competition penalty is arbitrarily negative. This never occurs under the F-contract. If $\delta^{\mathrm{H}} \geq \delta^{\mathrm{H}}_{\mathrm{C}}$, however, then it is easy to see that the centralized solution is always more efficient than the decentralized supply chain so that $\mathcal{Q}^{\mathrm{H}}_\tau \leq 1$ and $\mathcal{P}^{\mathrm{H}}_\tau \geq 0$.

We also note that if the budget is large enough so that both the decentralized and centralized operations can hedge away the budget constraint then $\delta^{\mathrm{H}} = \delta^{\mathrm{H}}_{\mathrm{C}} = c_\tau$ and

$$\mathcal{Q}^{\mathrm{H}}_\tau = \frac{1}{2} \quad \text{and} \quad \mathcal{P}^{\mathrm{H}}_\tau = \frac{1}{4}.$$

5 Optimal Production Postponement

We now extend the contracts of the previous sections by allowing τ, the time at which the physical transaction takes place, to be an endogenous decision variable that is determined as part of the

[17]It is clear that the central planner will always prefer the H-contract to the F-contract because the ability to hedge the budget constraint increases the set of feasible ordering quantities.

solution to the Nash equilibrium. We discuss this problem initially in the context of the H-contract but later we will assume that B is sufficiently large so that the analysis of the F-contract is the same as that of the H-contract.

We consider two alternatives formulations. In the first alternative, τ is restricted to be a deterministic time in $[0,T]$ that is selected at time $t=0$. Motivated by the terminology of dynamic programming, we refer to this alternative as the optimal *open-loop* production postponement model. In the second alternative, we permit τ to be an $\mathcal{F}_t$-stopping time that is bounded above by T. We call this alternative the optimal *closed-loop* production postponement model. In both cases, the procurement contract offered by the producer takes the form of a pair, (τ, w_τ), where the wholesale price menu, w_τ, is required to be $\mathcal{F}_\tau$-measurable. We note that the producer always prefers the closed-loop model though from a practical standpoint the open-loop model may be easier to implement in practice.

Independently of whether τ is a deterministic time or a stopping time, the optimal ordering level for the retailer, given a contract (τ, w_τ), is an $\mathcal{F}_\tau$-measurable menu, q_τ, that satisfies[18] the conditions in Proposition 6. The producer's problem is therefore given by (12) and (13) but now with τ as an extra decision variable. Furthermore, since the proof of Proposition 7 (see the Appendix) extends to the case of a stopping time, we conclude that the optimal wholesale price menu, w_τ, as a function of τ is still given by equation (14). In summary, the producer's problem of selecting the optimal time τ is given by

$$\Pi_{\mathrm{P}}^{\mathrm{H}} = \max_{\tau, \phi \geq 1} \mathbb{E}_0^{\mathbb{Q}} \left[\frac{(\bar{A}_\tau + \phi\, c_\tau - 2c_\tau)\,(\bar{A}_\tau - \phi\, c_\tau)^+}{8\,\xi} \right] \tag{21}$$

$$\text{subject to} \quad \mathbb{E}_0^{\mathbb{Q}} \left[\left(\frac{\bar{A}_\tau^2 - \phi^2\, c_\tau^2}{8\,\xi} \right)^+ \right] \leq B. \tag{22}$$

Of course τ should be restricted to either a deterministic time or a stopping time depending on which model (open-loop or closed-loop) is under consideration. For a given τ, the objective in (21) is decreasing in ϕ so that the producer's problem reduces to

$$\Pi_{\mathrm{P}}^{\mathrm{H}} = \max_{\tau} \mathbb{E}_0^{\mathbb{Q}} \left[\frac{(\bar{A}_\tau + \phi\, c_\tau - 2c_\tau)\,(\bar{A}_\tau - \phi\, c_\tau)^+}{8\,\xi} \right] \tag{23}$$

$$\text{subject to} \quad \phi = \inf \left\{ \psi \geq 1 \;:\; \mathbb{E}_0^{\mathbb{Q}} \left[\left(\frac{\bar{A}_\tau^2 - \psi^2\, c_\tau^2}{8\,\xi} \right)^+ \right] \leq B \right\}. \tag{24}$$

To solve this optimization problem we would first need to specify the functional forms of $\bar{A}_\tau$ and c_τ and depending on these specifications, the solution may or may not be easy to find. For the remainder of this section, however, we will show how this problem may be solved when additional assumptions are made. In particular, we make the following three assumptions:

1. X_t is a diffusion process with dynamics satisfying

$$\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}W_t, \tag{25}$$

   where W_t a $\mathbb{Q}$-Brownian motion. Note that we have not included a drift term in the dynamics of X_t since it must be the case that X_t is a $\mathbb{Q}$-martingale. This is not a significant assumption and we could easily consider alternative processes for X_t.

[18] It is easy to check that the proof of Proposition 6 remains unchanged if τ is allowed to be a stopping time.

2. We adopt a specific functional form to model the dependence between the market clearance price and the financial market. In particular, we assume that there behaves a well-behaved[19] function, $F(x)$, and a random variable, ε, such that one of the following two models holds.

$$\text{Additive Model:} \qquad A = F(X_T) + \varepsilon, \qquad \text{with } \mathbb{E}^{\mathbb{Q}}[\varepsilon] = 0, \text{ or} \tag{26}$$
$$\text{Multiplicative Model:} \qquad A = \varepsilon\, F(X_T), \qquad \text{with } \varepsilon \geq 0 \text{ and } \mathbb{E}^{\mathbb{Q}}[\varepsilon] = 1. \tag{27}$$

The random perturbation ε captures the non-financial component of the market price uncertainty and is assumed to be independent of X_t. Note that if $F(x) = \bar{A}$, we recover a model for which demand is independent of the financial market.

3. We assume that the initial budget, B, is sufficiently large so that the retailer is able to hedge away the budget constraint for every stopping time, τ. That is, $\phi = 1$ for every $\tau \in \mathcal{T}$. For example, if $\tau \equiv 0$ then there is no time for hedging to take place and so it is necessary that B is at least sufficiently large so that the budget constraint is not binding for the simple contract. This is a significant assumption[20] and effectively reduces the problem to one of finding the optimal (random) timing of the flexible contract when there is no budget constraint.

5.1 Optimal Open-Loop Production Postponement

We now restrict τ to be a deterministic time in $[0, T]$. Based on the third assumption above, the producer's optimization problem in (23) reduces to

$$\max_{\tau \in [0,T]} \mathbb{E}_0^{\mathbb{Q}}\left[(\bar{A}_\tau - c_\tau)^2\right] = \max_{\tau \in [0,T]} \operatorname{Var}(\bar{A}_\tau) + (\bar{A} - c_\tau)^2. \tag{28}$$

We note that in this optimization problem there is a trade-off between demand learning as represented by the variance term, $\operatorname{Var}(\bar{A}_\tau)$, and production costs as represented by $(\bar{A} - c_\tau)^2$. The first term is increasing in τ while the second term is decreasing in τ so that, in general, the optimization problem in (28) does not admit a trivial solution and depends on the particular form of the functions $\operatorname{Var}(\mathbb{E}^{\mathbb{Q}}[A \,|\, X_\tau])$ and c_τ.

The Itô Representation Theorem (*e.g.* Øksendal 1998) implies the existence of an $\mathcal{F}_t$-adapted process, $\{\theta_t : t \in [0, T]\}$, such that

$$A = \bar{A} + \int_0^T \theta_t \, \mathrm{d}X_t + \varepsilon \qquad \text{or} \qquad A = \varepsilon \left(\bar{A} + \int_0^T \theta_t \, \mathrm{d}X_t\right)$$

for the additive or multiplicative model, respectively. In both cases the Q-martingale property of X_t implies

$$\bar{A}_\tau = \bar{A} + \int_0^\tau \theta_t \, \mathrm{d}X_t. \tag{29}$$

In order to compute the variance of $\bar{A}_\tau$ we use the Q-martingale property of the stochastic integral and invoke Itô's isometry (*e.g.* Øksendal 1998) to obtain

$$\operatorname{Var}(\bar{A}_\tau) = \mathbb{E}_0^{\mathbb{Q}}\left[\left(\int_0^\tau \theta_t \, \mathrm{d}X_t\right)^2\right] = \mathbb{E}_0^{\mathbb{Q}}\left[\int_0^\tau \theta_t^2 \, \mathrm{d}[X]_t\right],$$

[19] It is necessary, for example, that $F(\cdot)$ satisfy certain integrability conditions so that the stochastic integral in (29) be a Q-martingale. In order to apply Itô's Lemma it is also necessary to assume that $F(\cdot)$ is twice differentiable. Because this section is intended to be brief, we omit the various technical conditions that are required to make our arguments completely rigorous.

[20] If we only wanted to solve for the open-loop policy it would not be necessary to make this assumption. In that case we could solve for the optimal τ and ϕ in (23) and (24) numerically.

where the process $[X]_t$ is the quadratic variation of X_t with dynamics

$$\mathrm{d}[X]_t = \sigma^2(X_t)\,\mathrm{d}t.$$

It follows that

$$\mathrm{Var}(\bar{A}_\tau) = \int_0^\tau \mathbb{E}_0^{\mathbb{Q}}[(\theta_t\,\sigma(X_t))^2]\,\mathrm{d}t.$$

The open-loop optimal problem therefore reduces to solving

$$\max_{\tau\in[0,T]} \left\{ \int_0^\tau \mathbb{E}_0^{\mathbb{Q}}[(\theta_t\,\sigma(X_t))^2]\,\mathrm{d}t + (\bar{A} - c_\tau)^2 \right\}. \tag{30}$$

If there is an interior solution to this problem ($i.e.$, $\tau^* \in (0,T)$), then it must satisfy the first-order optimality condition

$$\mathbb{E}_0^{\mathbb{Q}}[(\theta_\tau\,\sigma(X_\tau))^2] - 2\,(\bar{A} - c_\tau)\,\dot{c}_\tau = 0, \qquad \text{where } \dot{c}_\tau := \frac{\mathrm{d}c_\tau}{\mathrm{d}\tau}.$$

Example 3 Consider the case in which the security price, X_t, follows a geometric Brownian motion with dynamics

$$\mathrm{d}X_t = \sigma\,X_t\,\mathrm{d}W_t$$

where $\sigma \neq 0$ and W_t is a Q-Brownian motion. The quadratic variation process then satisfies $\mathrm{d}[X]_t = \sigma^2\,X_t^2\,\mathrm{d}t$. To model the dependence between the market clearance price and the process, X_t, we assume a linear model for $F(\cdot)$ so that $F(X) = A_0 + A_1\,X$ where A_0 and A_1 are positive constants. Therefore, depending on whether we consider the additive or multiplicative model, we have

$$A = A_0 + A_1\,X_T + \varepsilon \qquad \text{or} \qquad A = \varepsilon\,(A_0 + A_1\,X_T),$$

where ε is a zero-mean or unit-mean random perturbation, respectively, that is independent of the process X_t. It follows that $\bar{A}_\tau = A_0 + A_1\,X_\tau$ and $\bar{A} = \mathbb{E}_0^{\mathbb{Q}}[A] = A_0 + A_1\,X_0$. In addition, it is clear that θ_t is identically equal to A_1 for all $t \in [0,T]$. We assume that the per unit production cost increases with time and is given by

$$c_\tau = c_0 + \alpha\,\tau^\kappa, \qquad \text{for all } \tau \in [0,T],$$

where α and κ are positive constants.

To impose the additional constraint that $\bar{A}_\tau \geq c_\tau$ for all τ (Assumption 1), we restrict our choice of the parameters A_0, T, c_0, κ, and α so that $A_0 \geq c_0 + \alpha\,T^\kappa$. Since $\mathbb{E}_0^{\mathbb{Q}}[X_t^2] = X_0^2\,\exp(\sigma^2\,t)$ the optimization problem in (30) reduces to

$$\max_{\tau\in[0,T]} \left\{ (A_1\,X_0)^2\,(\exp(\sigma^2\,\tau) - 1) + (\bar{A} - c_0 - \alpha\,\tau^\kappa)^2] \right\}.$$

In general, a closed form solution is not available unless $\kappa = 0$. This is a trivial case in which c_τ is constant and the optimal strategy is to postpone production until time T so that $\tau^* = T$. Figure 4 shows the value of the objective function as a function of τ for four different values of κ. The cost functions are such that it becomes cheaper to produce as κ increases. Note that for $\kappa \in \{4, 8\}$, it is convenient to postpone production using a flexible contract. For the more expensive production cost functions that occur when $\kappa \in \{0.25, 1\}$, production postponement is not profitable and the simple contract is preferred. $\square$

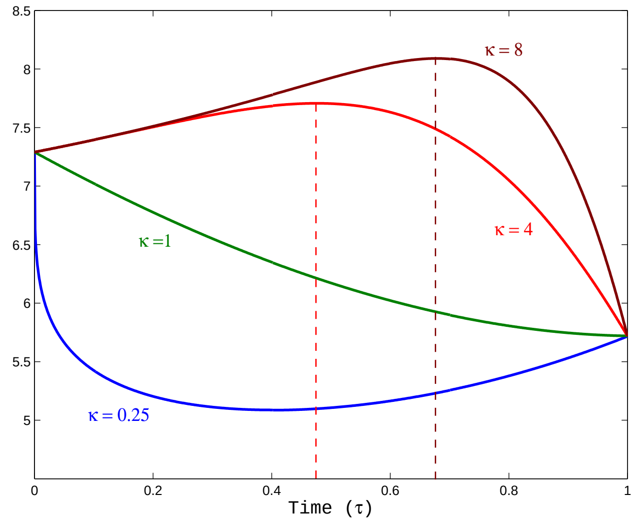

Figure 4: Optimal open-loop production postponement for four different production cost functions parameterized by κ. The other parameters are $X_0 = A_1 = \sigma = T = 1$, $A_0 = 2$, $c_0 = 0.3$ and $\alpha = 0.7$.

5.2 Optimal Closed-Loop Production Postponement

Instead of selecting a fixed transaction time, τ, at $t = 0$, the producer now optimizes over the set of stopping times bounded above by T. In this case, the optimization problem in (23) reduces to the following optimal stopping problem

$$\max_{\tau \in \mathcal{T}} \mathbb{E}_0^{\mathbb{Q}} \left[(\bar{A}_\tau - c_\tau)^2 \right], \tag{31}$$

where $\mathcal{T}$ is the set of $\mathcal{F}_t$-adapted stopping times bounded above by T. Again, the third assumption above has resulted in this simplified form of the objective function. According to the modeling of A in (26) or (27), it follows that $v(\tau, X_\tau) := \bar{A}_\tau = \mathbb{E}_\tau^{\mathbb{Q}}[F(X_T)]$ is a $\mathbb{Q}$-martingale that satisfies

$$\frac{v(t,x)}{\partial t} + \frac{1}{2}\,\sigma^2(x)\,\frac{\partial^2 v(t,x)}{\partial x^2} = 0, \qquad v(T,x) = F(x).$$

We define U to be the set $\{(t,x) \,:\, \mathcal{G}g(t,x) > 0\}$ where $g(t,x) := (v(t,x) - c_t)^2$ is the payoff function and $\mathcal{G}$ is the generator

$$\mathcal{G} := \frac{\partial}{\partial t} + \frac{1}{2}\sigma^2(x)\,\frac{\partial^2}{\partial x^2}.$$

We then obtain

$$U = \left\{(t,x) \,:\, (\sigma(x)\,v_x(t,x))^2 > 2(v(t,x) - c_t)\,\dot{c}_t\right\},$$

where v_x is the first partial derivative of v with respect to x. In general, the set U is a proper subset of the optimal continuation region for the stopping problem in (31). Computing the optimal stopping time analytically is a difficult task and is usually done numerically. However, if U turns out to equal the entire state space then it is clear that it is always optimal to continue.

Example 3: (Continued)

Consider the setting of Example 3 but where now τ is a stopping time instead of a deterministic time. For the linear function $F(X) = A_0 + A_1\,X$, the auxiliary function v satisfies $v(t,x) = A_0 + A_1\,x$, and the region U is given by

$$U = \{(t,x) : (\sigma\, x\, A_1)^2 > 2(A_0 + A_1\, x - c_t)\, \dot{c}_t\}.$$

Straightforward calculations allow us to rewrite U as

$$U = \left\{(t,x) : x > \frac{\dot{c}_t + \sqrt{\dot{c}_t^2 + 2\sigma^2\,(A_0 - c_t)\,\dot{c}_t}}{\sigma^2\, A_1}\right\}.$$

Let us define the auxiliary function

$$\rho(t) := \frac{\dot{c}_t + \sqrt{\dot{c}_t^2 + 2\sigma^2\,(A_0 - c_t)\,\dot{c}_t}}{\sigma^2\, A_1}.$$

Since U is a subset of the optimal continuation region, we know that it is never optimal to stop if $X_t > \rho(t)$. Of course, it is possible that $X_t < \rho(t)$ and yet still be optimal to continue.

We solved for the optimal continuation region numerically by using a binomial model to approximate the dynamics of X_t. In so doing, we can assess the quality of the (suboptimal) strategy that uses $\rho(t)$ to define the continuation region. Figure 5 shows the optimal continuation region and the threshold $\rho(t)$ for four different cost functions. These cost function are given by $c_\tau = c_0 + \alpha\,\tau^\kappa$ with $\kappa = 0.25, 1, 4,$ and 8. When $X(\tau)$ is above the optimal threshold it is optimal to continue. The vertical dashed line corresponds to the optimal open-loop deterministic time computed in Figure 4. For $\kappa = 0.25$ or $\kappa = 1$ this optimal deterministic time equals 0 since X_0 lies below the optimal threshold. For $\kappa = 4$ it equals 0.476, and for $\kappa = 8$ it equals 0.678.

Interestingly, for high values of κ the auxiliary threshold $\rho(t)$ is a good approximation for the optimal solution. However, as κ decreases the quality of the approximation deteriorates rapidly. Except for the case where $\kappa = 0.25$, the optimal threshold increases with time. This reflects the fact that the producer becomes more likely to stop and exercise the procurement contract as the end of the horizon approaches.

We conclude this example by computing the optimal expected payoff for the producer under both the optimal open-loop policy and the optimal closed-loop policy.

κ	Open-Loop Payoff	Closed-Loop Payoff	% Increase
0.25	7.29	7.29	0.0%
1	7.29	7.305	0.2%
4	7.71	7.99	3.7%
8	8.09	8.33	3.8%

Producer's expected payoff for four different production cost functions parameterized by κ. The other parameters are $A_1 = X_0 = \sigma = T = 1$, $A_0 = 2$, $c_0 = 0.3$, $\alpha = 0.7$ and $\xi = 1/8$.

Naturally, the optimal stopping time (closed-loop) policy produces a higher expected payoff than the optimal deterministic time (open-loop) policy. The improvement, however, is only a few percentage points which might suggest that a simpler contract based on a deterministic time captures most of the benefits of allowing τ to be a decision variable. In practice, of course, it would be necessary to model the operations and financial markets more accurately and to calibrate the resulting model correctly before such conclusions could be drawn. $\square$

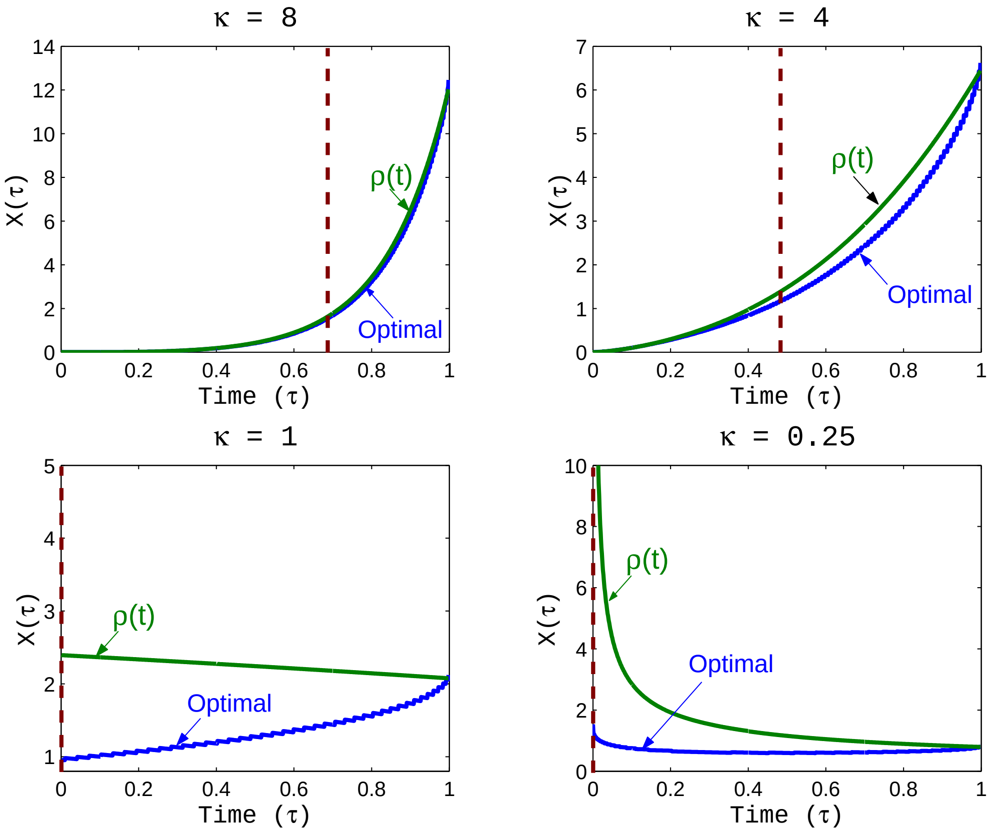

Figure 5: Optimal continuation region for four different manufacturing cost functions parameterized by κ. The other parameters are $A_1 = X_0 = \sigma = T = 1$, $A_0 = 2$, $c_0 = 0.3$ and $\alpha = 0.7$.

6 Extensions

It is easy to extend the basic model to more accurately reflect the manner in which financial markets impact the profitability of operations. In this section we describe some of these extensions. In particular, we focus on extensions[21] where foreign exchange rates, interest rates and the possibility of default influence the profitability of the competitive supply chain. We will concentrate only on the setting of the H-contract as this is a more interesting and challenging setting than that of the F-contract. We will also assume that the transaction time, τ, is deterministic and is given exogenously as in Sections 3 and 4.

Extension 1: Stochastic Interest Rates

We now assume that interest rates are stochastic and that the Q-dynamics of the short rate are given by the Vasicek[22] model so that

$$dr_t \;=\; \alpha(\mu - r_t)\, dt + \sigma dW_t \tag{32}$$

[21] It is straightforward to construct other interesting variations of the basic model.

[22] See, for example, Shreve (2004) for a description of the Vasicek model and other related results that we use in the sequel. While it is not necessary to restrict ourselves to a particular term structure model here, we have done so in order to simplify the exposition.

where α, μ and σ are all positive constants and W_t is a Q-Brownian motion. The short-rate, r_t, is the instantaneous continuously compounded risk-free interest rate that is earned at time t by the 'cash account', *i.e.*, cash placed in a deposit account. In particular, if \$1 is placed in the cash account at time t then it will be worth $\exp\left(\int_t^T r_s\, ds\right)$ at time $T > t$. It may be shown that the time τ value of a zero-coupon-bond with face value \$1 that matures at time $T > \tau$ satisfies

$$Z_\tau^T \;:=\; e^{a(T-\tau)+b(T-\tau)r_\tau} \tag{33}$$

where $a(\cdot)$ and $b(\cdot)$ are known deterministic functions. In particular, Z_τ^T is the appropriate discount factor for discounting a known deterministic cash flow from time T to time $\tau < T$.

Returning to our competitive supply chain, we assume as before that the retailer's profits are realized at time $T \geq \tau$ and that the budget B is only available at that time. However, we also assume that the producer demands payment from the retailer at time τ when the transaction takes place. This means that if $\tau < T$, then the retailer will be forced to borrow against the capital B that would only be available at time T. As a result, the retailer's effective budget at time τ is given by

$$B(r_\tau) \;:=\; BZ_\tau^T = Be^{a(T-\tau)+b(T-\tau)r_\tau}.$$

As before, we assume that the stochastic clearance price, $A - \xi q_\tau$,, depends on the financial market through the co-dependence of the random variable A, and the financial process, X_t. To simplify the exposition, we could assume that $X_t \equiv r_t$ but this is not necessary. If X_t is a financial process other than r_t, we simply need to redefine our definition of $\{\mathcal{F}_t\}_{0\leq t\leq T}$ so that it represents the filtration generated by X_t *and* r_t. Before formulating the optimization problems of the retailer and the producer we must adapt our definition of feasible $\mathcal{F}_\tau$-measurable financial gains, G_τ. Until this point we have insisted that any such G_τ must satisfy $\mathbb{E}_0^{\mathbb{Q}}[G_\tau] = 0$, assuming as before that zero initial capital is devoted to the financial hedging strategy. This was correct when interest rates were identically zero but now we must replace that condition with the new condition[23]

$$\mathbb{E}_0^{\mathbb{Q}}\left[D_\tau G_\tau\right] \;=\; 0 \tag{34}$$

where $D_\tau := \exp\left(-\int_0^\tau r_s\, ds\right)$. The retailer's problem for a given $\mathcal{F}_\tau$-measurable wholesale price, w_τ, is therefore given[24] by

$$\begin{aligned}
\Pi_{\mathrm{R}}^{\mathrm{H}}(w_\tau) = \max_{q_\tau\geq 0,\, G_\tau} \;& \mathbb{E}_0^{\mathbb{Q}}\left[D_T\left(A_T - \xi\, q_\tau\right) q_\tau - D_\tau w_\tau q_\tau\right] && (35)\\
\text{subject to} \quad & w_\tau\, q_\tau \leq B(r_\tau) + G_\tau, \qquad \text{for all } \omega \in \Omega && (36)\\
& \mathbb{E}_0^{\mathbb{Q}}\left[D_\tau G_\tau\right] = 0 && (37)\\
& \text{and } \mathcal{F}_\tau - \text{measurability of } q_\tau. && (38)
\end{aligned}$$

Note that both D_T and D_τ appear in the objective function (35) reflecting the times at which the retailer makes and receives payments. We also had to explicitly impose the constraint that q_τ be

[23]Martingale pricing theory states that the time 0 value, G_0, of the security that is worth G_τ at time τ (and does not pay any intermediate cash-flows) satisfies $G_0/N_0 = \mathbb{E}_0^{\mathbb{Q}}[G_\tau/N_\tau]$ where N_t is the time t price of the *numeraire* security. It is common to take the cash account as the numeraire security and this is the approach we have followed in this paper. Until now, however, the value of the cash account at time t was always \$1 since we assumed interest rates were identically zero. We therefore had $G_0 = \mathbb{E}_0^{\mathbb{Q}}[G_\tau]$ and since we insisted $G_0 = 0$ we obtained $\mathbb{E}_0^{\mathbb{Q}}[G_\tau] = 0$. When interest rates are non-zero we still have $N_0 = 1$ but now $N_\tau = \exp\ \int_0^\tau r_s\, ds$ and so we obtain (34). In the main text we take $D_t = N_t^{-1}$. See Shreve (2004) for a development of martingale pricing theory (or arbitrage pricing theory as it more commonly known).

[24]We write A_T for A to emphasize the timing of the cash-flow.

$\mathcal{F}_\tau$-measurable. This was necessary[25] because of the appearance of D_T in the objective function. We can easily impose the $\mathcal{F}_\tau$-measurability of q_τ by conditioning with respect to $\mathcal{F}_\tau$ inside the expectation appearing in (35). We then obtain

$$\mathbb{E}_0^{\mathbb{Q}}\left[D_\tau\left(\bar{A}_\tau^{\mathrm{D}} - \xi_\tau\, q_\tau - w_\tau\right)\, q_\tau\right] \tag{39}$$

as our new objective function where $\bar{A}_\tau^{\mathrm{D}} := \mathbb{E}_\tau^{\mathbb{Q}}[D_T A_T]/D_\tau$ and $\xi_\tau := \xi Z_\tau^T$. With this new objective function it is no longer necessary to explicitly impose the $\mathcal{F}_\tau$-measurability of q_τ.

While this problem (and the producer's problem that follows) might appear to be more complicated than the corresponding problems of Section 4 they are in fact no more difficult to solve. We have the following solution to the retailer's problem. We omit the proof as it is very similar to the proof of Proposition 6.

Proposition 11 (Retailer's Optimal Strategy)
Let w_τ be an $\mathcal{F}_\tau$-measurable wholesale price offered by the producer and define $\mathcal{Q}_\tau := \left(\frac{\bar{A}_\tau^D - w_\tau}{2\,\xi_\tau}\right)^+$. This is the optimal ordering quantity in the absence of a budget constraint. The following two cases arise:

> **Case 1**: *Suppose $\mathbb{E}_0^{\mathbb{Q}}\left[D_\tau \mathcal{Q}_\tau w_\tau\right] \leq \mathbb{E}_0^{\mathbb{Q}}\left[D_\tau B(r_\tau)\right] = Z_0^T B$. Then $q(w_\tau) = \mathcal{Q}_\tau$ and (possibly due to the ability to trade in the financial market) the budget constraint is not binding.*
>
> **Case 2**: *Suppose $\mathbb{E}_0^{\mathbb{Q}}\left[D_\tau \mathcal{Q}_\tau w_\tau\right] > Z_0^T B$. Then*
>
> $$q_\tau \;=\; \left(\frac{\bar{A}_\tau^D - w_\tau\,(1+\lambda)}{2\xi_\tau}\right)^+ \tag{40}$$
>
> *where $\lambda \geq 0$ solves*
>
> $$\mathbb{E}_0^{\mathbb{Q}}\left[D_\tau w_\tau q_\tau\right] \;=\; \mathbb{E}_0^{\mathbb{Q}}\left[B(r_\tau) D_\tau\right] \;=\; Z_0^T B. \tag{41}$$

Given the retailer's best response, the producer's problem may now be formulated[26] as

$$\Pi_{\mathrm{P}}^{\mathrm{H}} = \max_{w_\tau,\,\lambda\geq 0}\; \mathbb{E}_0^{\mathbb{Q}}\left[D_\tau(w_\tau - c_\tau)\left(\frac{\bar{A}_\tau^{\mathrm{D}} - w_\tau\,(1+\lambda)}{2\,\xi_\tau}\right)^+\right] \tag{42}$$

$$\text{subject to}\qquad \mathbb{E}_0^{\mathbb{Q}}\left[D_\tau w_\tau\left(\frac{\bar{A}_\tau^{\mathrm{D}} - w_\tau\,(1+\lambda)}{2\,\xi_\tau}\right)^+\right] \leq Z_0^T B. \tag{43}$$

The Nash equilibrium and solution of the producer's problem is given by the following proposition. We again omit the proof of this proposition as it it very similar to the proof of Proposition 7.

Proposition 12 (The Equilibrium Solution)
Let γ^H be the minimum $\gamma \geq 0$ that satisfies

$$\mathbb{E}_0^{\mathbb{Q}}\left[\frac{D_\tau}{8\xi_\tau}\;\left((\bar{A}_\tau^D)^2 - \frac{c_\tau^2}{(1-\gamma)^2}\right)^+\right] \;\leq\; Z_0^T B.$$

[25] To be precise, terms of the form $D_T(A_T - \xi q_\tau)$ should also have appeared in the problem formulations of Sections 3 and 4. In those sections, however, $D_t \equiv 1$ for all t and so the conditioning argument we use above allows us to replace A_T with $\bar{A}_\tau$ in those sections.

[26] We assume here and in Extension 2 that the production costs, c_τ, are paid at time τ.

Then the optimal wholesale price and ordering level satisfy

$$w_\tau^H = \frac{c_\tau}{2(1-\gamma^H)} + \frac{\bar{A}_\tau^D}{2} \quad \textit{and} \quad q_\tau^H = \left(\frac{\bar{A}_\tau^D - c_\tau/(1-\gamma^H)}{4\xi_\tau} \right)^+ .$$

Extension 2: Stochastic Interest Rates and Credit Risk

In this extension we assume that the retailer pays the producer at time $T+I$ where $I>0$ so that the producer is therefore paid *in arrears*. The retailer, however, sells as usual in the retail market at time $T \geq \tau$. In this case the retailer's problem may be formulated[27] as

$$\Pi_{\mathrm{R}}^{\mathrm{H}}(w_\tau) = \max_{q_\tau \geq 0,\, G_\tau} \; \mathbb{E}_0^{\mathbb{Q}} \left[D_\tau \left(\bar{A}_\tau^{\mathrm{D}} - \xi_\tau\, q_\tau - w_\tau Z_\tau^{T+I} \right) q_\tau \right] \tag{44}$$

$$\text{subject to} \qquad w_\tau\, q_\tau \leq B + \frac{G_\tau}{Z_\tau^{T+I}}, \qquad \text{for all } \omega \in \Omega \tag{45}$$

$$\mathbb{E}_0^{\mathbb{Q}} \left[D_\tau G_\tau \right] = 0 \tag{46}$$

It is easy to analyze this model as well as the producer's equilibrium problem and obtain analogous results to those of Extension 1. It may also be extended further, however, by allowing for the possibility that the retailer might default on his payment to the producer at time $T+I$. For example, we could use a standard credit risk model and assume that the occurrence of a default coincides with the first arrival of a given point process. This would enable[28] us to write the producer's objective function in the form

$$\mathbb{E}_0^{\mathbb{Q}} \left[D_{T+I} w_\tau q_\tau \, e^{-\int_\tau^{T+I} \nu_s \, ds} \; - D_\tau c_\tau \right] \tag{47}$$

where ν_s is the time s (possibly stochastic) intensity of the given point process. Moreover, if the transaction is a major transaction for the retailer, it might be reasonable to assume that the probability of default depends in part on the realized outcome of A_T. In that case, we should and could assume that the intensity process, $\{\nu_s\}_{T \leq s \leq T+I}$, depends (either deterministically or stochastically) on A_T.

Note that the integral in (47) runs from τ to $T+I$, suggesting that we are only concerned about default occurring in that interval. This is appropriate because if the retailer defaulted on his general obligations before time τ, then the producer would not yet have produced and transferred the production units to him. As a result, his losses due to the retailer's default should be minimized.

While it will not be quite as straightforward[29] to analyze this model where we can explicitly model the possibility of default, it should be still possible to reduce the problem to solving (possibly numerically) for at most two scalar Lagrange multipliers.

[27] $\bar{A}_\tau^{\mathrm{D}}$ and ξ_τ are defined as in Extension 1. The modified objective function in (44) arises from the same conditioning argument we used in Extension 1. The budget constraint in this formulation assumes that B is available only at time $T+I$ and that trading in the financial market is completed by time τ. This is a reasonable interpretation since it is known by time τ exactly how much will need to be paid at time $T+I$. The trading gain or loss, G_τ, is then invested at time τ in a zero-coupon bond that matures at time $T+I$. The realized gain from trading at time $T+I$ will then be G_τ / Z_τ^{T+I}.

[28] See, for example, Lando (2004).

[29] Unless we assume that the process, ν_s, is deterministic.

Extension 3: Foreign Exchange Rates

In our final extension we assume that the retailer and producer are located in countries R and P, respectively, and that these countries do not share a common currency. Let X_t denote the time t value of one unit of currency R in units of currency P. When the producer proposes a contract, w_τ, he does so in terms of currency R so that the retailer pays $q_\tau w_\tau$ units of his domestic currency, *i.e.* currency R, to the producer. The retailer's problem is therefore unchanged from his problem in Section 4.

The producer, however, must convert this payment into currency P and he therefore earns a per-unit profit of $w_\tau X_\tau - c_\tau$. Another minor complication arises because the martingale measure used by the producer, $\hat{Q}$ say, is different[30] to the martingale measure used by the retailer. This occurs because the producer and retailer use domestic currencies. It is still straightforward to formulate the producer's problem and reduce it to solving (possibly numerically) for one or two scalar Lagrange multipliers. It is also possible to assume that the financial market, X_t, is correlated with A_T so that is has both a direct and an indirect impact (via A_T) on the producer's objective function. Many variations of this problem formulation, where exchange rates impact the players' objective functions, are clearly possible.

7 Conclusions and Further Research

In this paper we have studied the performance of a stylized supply chain where two firms, a retailer and a producer, compete in a Stackelberg game. The retailer purchases a single product from the manufacturer and then sells it in the retail market at a stochastic clearance price. The retailer, however, is budget-constrained and is therefore limited in the number of units that he may purchase from the producer. We consider three types of contracts that govern the operation of the supply chain. In the case of the simple and flexible contracts, the retailer does not have access to the financial markets. In the case of the flexible contract with hedging, however, the retailer does have access to the financial markets and so he can, at least in part, mitigate the effects of the budget constraint. For each contract we compare the decentralized competitive solution with the solution obtained by a central planner. We also compare the supply chain's performance across the different contracts. We also examined the problem of choosing the optimal timing, τ, of the contract, and formulated this problem as an optimal stopping problem.

Our model and results extend the existing literature on supply chain contracts by considering a budget-constrained retailer and by including financial markets as (i) a source of public information upon which procurement contracts can be written, and (ii) a means for financial hedging to mitigate the effects of the budget constraint.

We find that in general the more cash constrained the retailer is the higher the wholesale price charged by the producer. We also find that the producer always prefers the flexible contract with hedging to the flexible contract without hedging. Depending on model parameters, however, the retailer may or may not prefer the flexible contract with hedging. One of our main results corresponds to Case 1 in proposition 6. Here we establish that if the budget is large enough, in an average sense, the ability to trade in the financial market allows the retailer to completely remove the budget constraint. This is not possible without financial trading unless the initial budget is so large that, regardless of the demand forecast $\bar{A}_\tau$, the budget constraint is never binding, .

[30] Equivalently, we could assume that they use the same martingale measure but that they use different numeraires.

Another interesting feature of the solution of the H-contract is that when the forecasted demand is low (*i.e.*, $\bar{A}_\tau$ is small) the producer chooses to shut down the supply chain by overcharging the retailer. By doing this, the producer can induce the retailer to transfer its limited budget from low demand states to more profitable high demand states. This is only possible if the retailer has access to the financial market to hedge the budget constraint. Under the F-contract, when access to the financial markets is not available, the producer never chooses to shut down the supply chain operations.

There are many directions in which this research could be extended. First, it would be interesting to consider models where the non-financial noise evolved as an observable stochastic process. In this case it would no longer be necessary for the trading gain, $G_\mathcal{T}$, to be $\mathcal{F}_\mathcal{T}$-measurable. Indeed the trading strategy would now depend on the evolution of both the financial and non-financial noise. Solving for the optimal trading strategy is then an incomplete-markets problem and would require mathematical techniques that are still being developed in the mathematical finance literature. Applying these techniques to our competitive Nash-equilibrium setting where a budget constraint induces the desire to hedge would be particularly interesting and challenging.

A related direction for future research is to build and solve models where the need for hedging is induced by the presence of risk averse agent(s) rather than the presence of a budget constraint. Caldentey and Haugh (2003, 2005) consider such problems in a non-competitive setting where risk aversion is modelled by imposing explicit risk management constraints or by assuming the agent is risk averse with a quadratic utility function.

Third, it would be interesting to explore principal-agent problems in the setting where the risk-averse (or budget constrained) agent has access to financial markets and the principal has imperfect information regarding the actions taken by the agent. Because the agent could use the financial market to smooth his income, it would presumably cost the principal agent less to ensure that the agent behaved optimally. This problem is of course related to the literature regarding executive compensation in corporate finance. In this literature it is often the case that the agent or executive is *not* permitted to trade in his company's stock. However, there is no reason why the agent should not be free to trade in other financial markets that impact his company's performance. There are clearly many variations on this problem that could be explored.

A fourth direction would be to consider other types of contracts that the producer could offer to the retailer. In this paper we have only considered linear price contracts but other contracts could also be used. They include, for example, quantity discount, buy-back and quantity flexibility contracts (*e.g.* Pasternack (1985) and Lovejoy (1999)). A contract that might be of particular interest in our hedging framework is an *affine* contract where the producer offers a contract of the form $(w_\mathcal{T}, v_\mathcal{T})$ to the retailer. In response, the retailer (assuming he accepts the contract) orders the random quantity, $q_\mathcal{T}$, and pays the producer $q_\mathcal{T} w_\mathcal{T} - v_\mathcal{T}$ where $v_\mathcal{T}$ is an $\mathcal{F}_\mathcal{T}$-measurable random variable. If the retailer cannot trade, then this contract is very similar to our H-contract where $v_\mathcal{T}$ may be interpreted as a trading gain that is chosen by the *producer*. Obviously, this would result in an equilibrium that would differ from the equilibrium of the H-contract where it is the retailer who chooses the trading gain. If the retailer did have access to the financial market, however, then this affine contract could be replaced by a contract of the form $(w_\mathcal{T}, V)$ where V is now a *constant* transfer payment. This follows because the retailer could use the financial markets to capitalize the random gain, $v_\mathcal{T}$, obtaining instead $V := \mathbb{E}_0^{\mathbb{Q}}[v_\mathcal{T}]$.

A particularly important direction for future research is to calibrate these models and operations-financial market models more generally. This is not an easy task but it will be necessary to do so if any of these models (competitive or non-competitive) are to be implemented in practice. Accurate calibration would also enable us to determine what types of financial risks are worth hedging and what the resulting economic savings would be.

In Section 6 we described some model extensions that could incorporate foreign exchange risk, interest rate risk and credit risk, among others. It is necessary for future research to further develop these models so that they accurately describe the financial and credit risks encountered in practice. Otherwise, there is little chance that they will be implemented successfully. It is perhaps also worth mentioning that it will be necessary to use numerical methods to solve many of these more realistic models. That should not deter researchers, however, from exploring these research directions.

We are pursuing some of these extensions in our current research.

References

Azoury, K. S. 1985. Bayes Solution to Dynamic Inventory Models under Unknown Demand Distribution. *Management Science* **31**. 1150-1160.

Buzacott, J.A. and R.Q. Zhang. 2004. Inventory Management with Asset-Based Financing. *Management Science*. **50** 1274-1292.

Cachon, G. 2003. Supply Chain Coordination with Contracts. Handbooks in Operations Research and Management Science: Supply Chain Management. Edited by Steve Graves and Ton de Kok. North Holland.

Cachon, G. and P. Zipkin. 1999. Competitive and Cooperative Inventory Policies in a Two-Stage Supply Chain. *Management Science*. **45** 936-953.

Caldentey, R. and M.B. Haugh. 2003. Optimal Control and Hedging of Operations in the Presence of Financial Markets. Working paper, Columbia University.

Caldentey, R. and M.B. Haugh. 2005. The Martingale Approach to Operational and Financial Hedging. Working paper, Columbia University.

Duffie, D. 2001. *Dynamic Asset Pricing Theory*, third edition. Princeton University Press.

Donohue, K.L. 2000. Efficient Supply Contracts for Fashion Goods with Forecast Updating and Two Production Modes. *Management Science*. **46** 1397-1411.

Eppen, G. D. and A. V. Iyer. 1997. Improved Fashion Buying with Bayesian Updates. *Operations Research* **45**. 805-819.

Lando, D. 2004. *Credit Risk Modeling.* Princeton University Press, new Jersey.

Lariviere, M. 1998. Supply Chain Contracting and Coordination with Stochastic Demand. Chapter 8 in Tayur S., M. Magazine, R. Ganeshan (Eds.), *Quantitative Models for Supply Chain Management.* Kluwer Academic Publishers.

Lariviere, M, and E.L. Porteus. 2001. Selling to the Newsvendor: An Analysis of Price-Only Contracts. *Manufac- turing & Service Operations Management.* **3**, 293-305.

Lovejoy, W.S. 1999. Quantity Flexibility Contracts and Supply Chain Performance. *Manufacturing & Service Operations Management.* **1** 89-111.

Øksendal, B. 1998. *Stochastic Differential Equations.* Springer Verlag, New York.

Pasternack, B. 1985. Optimal Pricing and Returns Policies for Perishable Commodities. *Marketing Science.* **4** 166-176.

Shreve, S.E. 2004. *Stochastic Calculus for Finance II: Continuous-Time Models.* Springer Verlag, New York.

Spengler, J.J. 1950. Vertical Integration and Antitrust Policy. *Journal of Political Economy.* **58** 347-352.

Tsay, A. A., S. Nahmias and N. Agrawal. 1998. Modelling Supply Chain Contracts: A Review. Chapter 10 in Tayur S., M. Magazine, R. Ganeshan (Eds.), *Quantitative Models for Supply Chain Management.* Kluwer Academic Publishers.

APPENDIX A: Proofs

Proof of Proposition 2: When $B \leq B_\tau^{\mathrm{F}}$ for all $\omega \in \Omega$, the inequalities follow from Jensen's inequality, the concavity of the function $f(x) = x + \sqrt{x^2 - 8\xi B}$ and the convexity of the functions $g(x) = x - \sqrt{x^2 - 8\xi B}$ and $h(x) = (x - \sqrt{x^2 - 8\xi B})^2$ in the region $x \geq 8\xi B$.

Let us now look at the retailer's payoff ratio when $B \downarrow 0$.

$$\lim_{B\downarrow 0} \frac{\Pi_{\mathrm{R}}^{\mathrm{F}}}{\Pi_{\mathrm{R}}^{\mathrm{S}}} \;=\; \lim_{B\downarrow 0} \mathbb{E}_0^{\mathbb{Q}} \left[\left(\frac{\bar{A}_\tau - \sqrt{\bar{A}_\tau^2 - 8\,\xi\,B}}{\bar{A} - \sqrt{\bar{A}^2 - 8\,\xi\,B}} \right)^2 \right] = \mathbb{E}_0^{\mathbb{Q}} \left[\lim_{B\downarrow 0} \left(\frac{\bar{A}_\tau - \sqrt{\bar{A}_\tau^2 - 8\,\xi\,B}}{\bar{A} - \sqrt{\bar{A}^2 - 8\,\xi\,B}} \right)^2 \right] = \mathbb{E}_0^{\mathbb{Q}} \left[\frac{\bar{A}}{\bar{A}_\tau} \right]$$

The second equality follows from the Bounded Convergence theorem and the third equality uses L'Hôpital's rule. A similar approach can be used to compute the limiting value of the producer's payoff ratio.

For the case $B \geq B_\tau^{\mathrm{F}}$ for all $\omega \in \Omega$, the equalities for w_τ^{F} and q_τ^{F} are straightforward. To verify the inequalities for the producer and retailer's payoff note that

$$\begin{aligned} \mathbb{E}_0^{\mathbb{Q}}[(\bar{A}_\tau - c)^2] \;&=\; (\bar{A} - c_0)^2 + \mathbb{E}_0^{\mathbb{Q}}[\bar{A}_\tau^2] - A^2 + 2\,\bar{A}\,c_0 - 2\,\mathbb{E}_0^{\mathbb{Q}}[\bar{A}_\tau]\,c + c^2 - c_0^2 \\ &=\; (\bar{A} - c_0)^2 + \mathrm{Var}(\bar{A}_\tau) - 2\,\bar{A}(c - c_0) + c^2 - c_0^2. \end{aligned}$$

Therefore,

$$\Pi_{\mathrm{M}}^{\mathrm{F}} \geq \Pi_{\mathrm{M}}^{\mathrm{S}} \quad \Longleftrightarrow \quad \mathbb{E}_0^{\mathbb{Q}} \left[\frac{(\bar{A}_\tau - c)^2}{8\,\xi} \right] \geq \frac{(\bar{A} - c_0)^2}{8\,\xi} \quad \Longleftrightarrow \quad \mathrm{Var}(\bar{A}_\tau) + c^2 - c_0^2 \geq 2\,\bar{A}(c - c_0).$$

The proof for the retailer's payoff is similar. □

Proof of Proposition 7: We will prove a slightly more general result in which τ is permitted to be a stopping time. This result will be used in Section 5. The result in proposition 7 is the special case when τ is deterministic.

Consider an arbitrary $\mathcal{F}_t$-stopping time $\tau \leq T$ and the producer's optimization problem

$$\begin{aligned} \Pi_{\mathrm{M}}^{\mathrm{H}} = \max_{w_\tau,\,\lambda\geq 0} \;& \mathbb{E}_0^{\mathbb{Q}} \left[(w_\tau - c_\tau) \left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi} \right)^+ \right] \\ \text{subject to} \quad & \mathbb{E}_0^{\mathbb{Q}} \left[w_\tau \left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi} \right)^+ \right] \leq B. \end{aligned}$$

To solve this problem, we first relax the budget constraint using a multiplier $\beta \geq 0$. After relaxing the constraint, the new objective function becomes

$$\mathcal{L}(w_\tau, \lambda, \beta) := \mathbb{E}_0^{\mathbb{Q}} \Big[(w_\tau(1-\beta) - c_\tau) \left(\frac{\bar{A}_\tau - w_\tau\,(1+\lambda)}{2\,\xi} \right)^+ \Big]$$

and it is clear that the optimal value of β will satisfy $\beta \leq 1$. In particular, we can restrict $\beta \in [0, 1]$. We introduce the following change of variables:

$$y_\tau := w_\tau\,(1+\lambda) \qquad \text{and} \qquad \phi := \frac{1+\lambda}{1-\beta}.$$

Note that $\phi \geq 1+\lambda$ since $\beta \in [0,1]$ and $\lambda \geq 0$. We can now rewrite the objective function as

$$\mathcal{L}(y_\tau, \phi) = \frac{1}{\phi} \mathbb{E}_0^{\mathbb{Q}}\Big[(y_\tau - c_\mathcal{T}\,\phi)\,\left(\frac{\bar{A}_\tau - y_\tau}{2\,\xi}\right)^+\Big].$$

Let us fix ϕ and optimize $\mathcal{L}(y_\tau, \phi)$ over y_τ. That is, we maximize $\mathcal{L}(y_\tau, \delta)$ point-wise for each y_τ. If $\bar{A}_\tau \geq c_\mathcal{T}\,\phi$ then $y_\tau = (\bar{A}_\tau + c\,\phi)/2$ is optimal. If $\bar{A}_\tau \leq c_\mathcal{T}\,\phi$ then any $y_\tau \geq \bar{A}_\tau$ is optimal. In particular, we can again take $y_\tau = (\bar{A}_\tau + c_\mathcal{T}\,\phi)/2$ as the optimal solution. The corresponding optimal ordering quantity is then given by

$$q_\tau = \left(\frac{\bar{A}_\tau - c_\mathcal{T}\,\phi}{4\,\xi}\right)^+.$$

It only remains now to find the optimal values of ϕ and λ. Given the previous solution, the producer's problem may be formulated as

$$\begin{aligned} &\max_{\lambda\geq 0,\, \phi \geq 1+\lambda} \quad \mathbb{E}_0^{\mathbb{Q}}\left[\left(\frac{\bar{A}_\tau^2 - (c_\mathcal{T}\,\phi)^2}{8\,\xi\,(1+\lambda)}\right)^+ - c_\mathcal{T}\,\left(\frac{\bar{A}_\tau - c_\mathcal{T}\,\phi}{4\,\xi}\right)^+\right] \\ &\text{subject to} \quad \mathbb{E}_0^{\mathbb{Q}}\left[\left(\frac{\bar{A}_\tau^2 - c_\mathcal{T}^2\,\phi^2}{8\,\xi(1+\lambda)}\right)^+\right] \leq B. \end{aligned}$$

We can solve this problem as follows. Suppose the optimal λ is strictly greater than 0. Then the constraint must be binding since the objective function increases as λ decreases. But the first term in the objective function then equals B. Now note that it is possible to increase the objective function by increasing ϕ and maintaining the tightness of the constraint by simultaneously reducing λ. (It is possible to do this since by assumption $\lambda > 0$.) Clearly then we can continue increasing the objective function until $\lambda = 0$. In particular, we can conclude that the optimal value of λ is 0. The optimization problem may be now formulated as

$$\begin{aligned} &\max_{\phi \geq 1} \mathbb{E}_0^{\mathbb{Q}}\left[\left(\frac{\bar{A}_\tau - c_\mathcal{T}\,\phi}{4\,\xi}\right)^+\left(\frac{\bar{A}_\tau + c_\mathcal{T}\,\phi}{2} - c_\mathcal{T}\right)\right] \\ &\text{subject to} \quad \mathbb{E}_0^{\mathbb{Q}}\left[\left(\frac{\bar{A}_\tau^2 - c_\mathcal{T}^2\,\phi^2}{8\,\xi}\right)^+\right] \leq B. \end{aligned}$$

By inspection it is clear that the optimal solution, ϕ^*, satisfies $\phi^* = \max(1, \hat{\phi})$ where $\hat{\phi}$ is the value of ϕ that makes the constraint binding. (If such a ϕ does not exist then we take $\phi^* = 1$.)

The statement of proposition 7 is complete once we identify δ^H with $c_\mathcal{T}\,\phi^*$. □

Proof of Proposition 8: Suppose first that $\delta^{\mathrm{H}} = c_\mathcal{T}$. The producer's expected payoff under the F-contract can be written as

$$\mathbb{E}_0^{\mathbb{Q}}[\Pi_{\mathrm{P}|\tau}^{\mathrm{F}}] = \frac{1}{8\xi}\,\mathbb{E}_0^{\mathbb{Q}}[(\bar{A}_\tau + \delta_\tau^{\mathrm{F}} - 2c_\mathcal{T})\,(\bar{A}_\tau - \delta_\tau^{\mathrm{F}})], \quad \text{where } \delta_\tau^{\mathrm{F}} = \max\left\{c_\mathcal{T}, \sqrt{(\bar{A}_\tau^2 - 8\xi B)^+}\right\}.$$

When $\delta^{\mathrm{H}} = c_\mathcal{T}$ straightforward calculations show

$$\frac{1}{8\xi}\,\mathbb{E}_0^{\mathbb{Q}}[(\bar{A}_\tau + \delta_\tau^{\mathrm{F}} - 2c_\mathcal{T})\,(\bar{A}_\tau - \delta_\tau^{\mathrm{F}})] \leq \frac{1}{8\xi}\,\mathbb{E}_0^{\mathbb{Q}}[(\bar{A}_\tau + \delta^{\mathrm{H}} - 2c_\mathcal{T})\,(\bar{A}_\tau - \delta^{\mathrm{H}})^+] = \mathbb{E}_0^{\mathbb{Q}}[\Pi_{\mathrm{P}|\tau}^{\mathrm{H}}]$$

implying the producer is better off under the H-contract.

Suppose now that $\delta^{\mathrm{H}} > c_\mathcal{T}$ and consider the following optimization problem over $\mathcal{F}_\mathcal{T}$-measurable random variables:

$$\max_{\delta_\tau} \quad \frac{1}{8\xi}\,\mathbb{E}_0^{\mathbb{Q}}[(\bar{A}_\tau + \delta_\tau - 2c_\mathcal{T})\,(\bar{A}_\tau - \delta_\tau)^+] \tag{A1}$$

$$\text{subject to} \quad \frac{1}{8\xi}\,\mathbb{E}_0^{\mathbb{Q}}[(\bar{A}_\tau^2 - \delta_\tau^2)^+] \le B \tag{A2}$$

$$\delta_\tau \ge c_\mathcal{T}, \quad \text{for all } \omega \in \Omega. \tag{A3}$$

Note that δ_τ^{F} is a feasible solution for this problem and that when $\delta_\tau = \delta_\tau^{\mathrm{F}}$ the objective function is equal to the expected payoff of the producer under the F-contract. We will show that the optimal solution to the problem has $\delta_\tau = \delta^{\mathrm{H}}$ so that the optimal objective function is equal to the expected payoff of the producer under the H-contract. In particular, this will imply the producer prefers the H-contract to the F-contract.

Therefore it is enough to show that $\delta_\tau^* = \delta^{\mathrm{H}}$ is an optimal solution to (A1)-(A3). To prove this, first note that because $\delta^{\mathrm{H}} > c_\mathcal{T}$ constraint (A2) must be bidding at optimality. This follows from the fact that $(\bar{A}_\tau + \delta_\tau - 2c_\mathcal{T})\,(\bar{A}_\tau - \delta_\tau)^+$ is decreasing in δ_τ in the range $\delta_\tau \in [c_\mathcal{T}, \bar{A}_\tau]$. Therefore, since the left-hand-side of constraint (A2) can be factored out of the objective function as a constant, B, a solution to (A1)-(A3) also solves

$$\min_{\delta_\tau} \quad \mathbb{E}_0^{\mathbb{Q}}[(\bar{A}_\tau - \delta_\tau)^+] \tag{A4}$$

$$\text{subject to} \quad \frac{1}{8\xi}\,\mathbb{E}_0^{\mathbb{Q}}[(\bar{A}_\tau^2 - \delta_\tau^2)^+] = B \tag{A5}$$

$$\delta_\tau \ge c_\mathcal{T}, \quad \text{for all } \omega \in \Omega. \tag{A6}$$

To solve this problem we relax constraint (A5). The corresponding lagrangian function is

$$\mathcal{L}(\delta, \lambda) := \mathbb{E}_0^{\mathbb{Q}}[\mathcal{L}_\mathcal{T}(\delta_\tau, \lambda)], \qquad \text{where } \mathcal{L}_\mathcal{T}(\delta_\tau, \lambda) := (\bar{A}_\tau - \delta_\tau)^+\,(1 + \lambda\,(\bar{A}_\tau + \delta_\tau)).$$

If $\lambda \ge 0$ then $\delta_\tau \ge \bar{A}_\tau$ for all $\omega \in \Omega$ minimizes $\mathcal{L}(\delta, \lambda)$ subject to $\delta_\tau \ge c_\mathcal{T}$. However, this solution does not satisfy constraint (A5) so it must be that $\lambda < 0$. In this case, the problem of minimizing $\mathcal{L}(\delta, \lambda)$ is solved with $\delta_\tau = \max\left(-\frac{1}{2\lambda},\ c_\mathcal{T}\right)$ for all $\omega \in \Omega$, that is, a constant value. To pick this fixed value of δ_τ we need to impose constraint (A5). By the definition of δ^{H} (and the statements at the end of the proof of Proposition 7) we conclude that $\delta_\tau = \delta^{\mathrm{H}}$ for all $\omega \in \Omega$. $\square$

Proof of Proposition 9: Let $[A_l, A_u]$ be the support of $\bar{A}_\tau$ and f_A be its density.

For B sufficiently small, Proposition 1 implies $\delta_\tau^{\mathrm{F}} = \sqrt{\bar{A}_\tau^2 - 8\,\xi\,B} = \bar{A}_\tau - \frac{4\xi B}{\bar{A}_\tau} + O(B^2)$. This and Proposition 1 again imply that as $B \downarrow 0$, the retailer's payoff satisfies

$$\mathbb{E}_0^{\mathbb{Q}}[\Pi_{\mathrm{R}|\mathcal{T}}^{\mathrm{F}}] = \frac{1}{16\,\xi}\,\mathbb{E}_0^{\mathbb{Q}}\left[\left(\frac{4\xi B}{\bar{A}_\tau} + O(B^2)\right)^2\right] = \xi\,B^2\,\mathbb{E}_0^{\mathbb{Q}}\left[\frac{1}{\bar{A}_\tau^2}\right] + O\big(B^3\big). \tag{A7}$$

In the case of the H-contract, Proposition 7 implies that if B is sufficiently small then $\delta^{\mathrm{H}} > c_\mathcal{T}$ and solves

$$\int_{\delta^{\mathrm{H}}}^{A_u} \left(\frac{z^2 - (\delta^{\mathrm{H}})^2}{8\xi}\right) f_A(z)\,\mathrm{d}z = B.$$

The mean-value theorem then implies the existence of an $\tilde{A} \in [\delta^{\mathrm{H}}, A_u]$ such that

$$f_A(\tilde{A}) \int_{\delta^{\mathrm{H}}}^{A_u} \left(\frac{z^2 - (\delta^{\mathrm{H}})^2}{8\xi}\right) \mathrm{d}z = B.$$

After integrating and simplifying we obtain

$$(A_u - \delta^{\mathrm{H}})^2 = \frac{24\xi B}{(A_u + 2\,\delta^{\mathrm{H}})\, f_A(\tilde{A})}. \tag{A8}$$

By Proposition 7 and the mean-value theorem, the retailer's payoff under the H-contract satisfies

$$\mathbb{E}_0^{\mathbb{Q}}[\Pi^{\mathrm{H}}_{\mathrm{R}|\tau}] = \frac{1}{16\xi} \int_{\delta^{\mathrm{H}}}^{A_u} (z - \delta^{\mathrm{H}})^2\, f_A(z)\,\mathrm{d}z = \frac{f_A(\hat{A})}{16\xi} \int_{\delta^{\mathrm{H}}}^{A_u} (z - \delta^{\mathrm{H}})^2\,\mathrm{d}z = \frac{f_A(\hat{A})\,(A_u - \delta^{\mathrm{H}})^3}{48\,\xi},$$

for some $\hat{A} \in [\delta^{\mathrm{H}}, A_u]$. We can then combine this identity with condition (A8) to obtain

$$\mathbb{E}_0^{\mathbb{Q}}[\Pi^{\mathrm{H}}_{\mathrm{R}|\tau}] = \frac{f_A(\hat{A})}{48\xi} \left(\frac{24\xi}{(A_u + 2\,\delta^{\mathrm{H}})\, f_A(\tilde{A})} \right)^{\frac{3}{2}} B^{\frac{3}{2}} \geq K\ B^{\frac{3}{2}}, \tag{A9}$$

for constant $K > 0$. (Since f_A is bounded away from zero it is easy to check that $K > 0$.) The statement of the proposition now follows immediately from (A7) and (A9). $\square$