Supply Chain Postponement Contracts: A Game-Theoretical Model
Abstract #004-0571

Chester Chambers
Cox School of Business,
Southern Methodist University
PO Box 750333
Dallas, TX 75275-0333
cchamber@cox.smu.edu

Eli M. Snir
Cox School of Business,
Southern Methodist University
PO Box 750333
Dallas, Texas 75275-0333
esnir@cox.smu.edu
Abstract: We consider a system with a monopolist manufacturer and multiple buyers in which contracts are written at time 0 that include a fee structure and a lead-time schedule. As lead-times shorten, buyer’s forecast accuracy improves, but the manufacturer’s cost increases. The key differentiator from previous work is that each buyer has a private assessment of his demand uncertainty, unknown to the producer. We produce a profit maximizing menu of contracts that the manufacturer should offer that includes a price-lead-time schedule with a two-part tariff. We explain how price discrimination should be used by the manufacturer and characterize the social inefficiency that is introduced. The contribution of the work arises form the pricing of lead-times with imperfect information, which facilitates the inclusion of lead-times as a decision variable for the manufacturer.

1.0 Introduction

The growing literature on Supply Chain Management has often touted the value of postponing production as part of an effort to better match supply and demand. The objective is typically to provide a better product selection, and to reduce inventory levels, stockouts, and shortages. Facilitating this postponement requires a high level of cooperation between supply chain partners. For example, it may be necessary for manufacturers to expend resources to accommodate shorter lead times and buyers may be implicitly asking manufacturers to reject orders from other customers that include longer lead times. Consequently, supply chain partners have incentives to enter into contracts which reserve production capacity but defer order details. In such settings, the partners may negotiate fulfillment contracts far in advance of product delivery. This type of arrangement is of particular relevance in settings where the partners are willing to agree early on, on an aggregate level of production but would like to postpone the specification of order quantities at the SKU level. This structure is reasonable when considering many seasonal items where neither the load on the manufacturer, nor the space used by the retailer differs widely across the items in question. Various industries including fashion apparel, cosmetics, and toys exhibit these characteristics.

From the manufacturer’s point of view, the key to facilitating such an arrangement is a shorter lead time between finalizing order specifications and delivery. To allow for postponement, generic modules or common components may be purchased or produced well in advance, and generic processes may be initiated early. Postponement requires that these modules be finalized or assembled into finished goods adjacent to their delivery to the retail channel. Contracts which specify the terms of such arrangements can become quite complex for at least three reasons. (1) Decision makers may gain information about demand over time and it is generally easier to predict what demand will be at some future time rather than today. (2) The contract involves multiple SKU’s with varying demand characteristics. Consequently, the ability to predict demand may differ dramatically from one SKU to another. (3) We must account for the fact that the decision makers behave strategically. Each player leverages his position when possible, and neither player will accept a ‘bad deal’ when other options exist. The objective of this analysis is to evaluate decisions regarding postponement in a game-theoretic context where supply chain partners behave strategically in designing contracts for postponement. In the process, we develop several insights regarding contract structure and supply chain efficiency.
We model a 2-player supply chain with an upstream monopolist manufacturer (alternately a brand-owner/distributor, e.g., Sports Obermeyer, Mattel toys, Kenebo cosmetics, Levi Strauss clothing, etc.) and a competitive downstream retail market. We assume that buyers purchase multiple items (SKUs) from the manufacturer. We also assume that both parties seek a contract that reserves capacity at time $T$, but allows for the specification of order quantities at the SKU level up to a pre-specified time before the retail season begins.

Demand $x$ is realized at the item level at time 0. At this time all merchandise is delivered to retailers. To assure timely delivery, contracts require that retailers finalize orders with a positive lead time ($t$). Each contract is for a specific item (SKU). Thus, each contract must specify three parameters (i) the lead time for finalizing product specifications ($t$), (ii) a non-linear price schedule, and (iii) order quantity ($Q$). The non-linear price schedule is of the form of a “two-part tariff”, which includes a fixed payment $R$ and a per-unit price $p$. We allow each of these components to be functions of $t$.

We focus upon the value of this arrangement that derives from demand uncertainty at the item level. The lead time $t \in [0, T]$ is measured relative to the retail delivery date of 0. A shorter lead time allows the retailer to learn more about demand and to order higher quantities of products (SKUs) expected to be more popular and lower quantities of less promising items. With less uncertainty regarding demand, rational buyers order less safety stock, and thus place smaller orders on average. However, offering shorter lead time is costly for the manufacturer. He has to invest in manufacturing and transportation technology that enable postponement.

Specifically, we model the situation of a monopolist manufacturer selling to a retail buyer. We examine the case when there are two types of products, one with low variance and the other with high variance. The manufacturer offers two production lead times; a long lead time and a short lead time. We concentrate on the contract structure offered by the manufacturer where low variance items are produced with long lead times and high variance items are produced with short lead times. When reducing lead time is costly, this correspondence is consistent with the decisions made by a central planner. In developing the optimal solution we begin by analyzing the case of centralized supply chain, which maximizes social welfare, in section 2.

In section 3, we determine the terms of the contracts that the manufacturer must offer (lead times, quantities, and prices) so that retailers find it in their best interest to adopt the lead time and quantity for various products as desired by the manufacturer. The rationality and strategic behavior of the retailers result in self-selection constraints on the manufacturer’s profit maximization problem. We show that the optimal solution to this problem involves offering a lead time and order quantity that are equal to those of a centralized supply chain only for the low variance items. For the high variance items, lead time is set shorter and order quantity is set larger, than would be chosen by a centralized supply chain. In setting the fixed payments, $R$, the manufacturer extracts all profit from the high variance items, while allowing the retailer to maintain some of the profit on the low variance items. This is required to assure that the self-selection constraints are met.

2.0 Selected Literature

This work builds upon the framework of Maskin and Riley (1984). That work deals with the pricing of assets with at least two customer types and an exogenous parameter that will fall within some specified range. The key point is that one customer type will evaluate the asset at a
higher value than the other given any realization of this exogenous parameter but the customer cannot be forced to expose his type before the seller presents his offerings. This forces the seller to configure a menu of contracts that gives the “high” type customer a positive return while simultaneously providing the “low” type customer a non-negative return. The complication is that the contract offered to the “low” type customer must be stated in such a way that it is no more attractive to the “high” type customer than the contract that we want her to accept. Our work extends the logic of Maskin and Riley in that we are dealing with three contract attributes simultaneously (lead time, price, and quantity.)

Our work is also related to Donohue (2000), which considers a single producer that delivers a product to a single retailer where the customer gains imperfect information about demand after an opportunity to place an initial order with a long lead time. The buyer can then place a second order with a short lead time that reflects this information. Both orders arrive at the buyer simultaneously. The paper focuses upon parameter values, including prices and salvage values, which result in efficient contracts. Our work differs in that we consider two products with different levels of demand uncertainty. This is important because it allows strategic behavior on the part of the buyer. We also calculate both optimal prices and lead times which maximize social welfare, while Donohue (2000) takes lead times as given. This is important because the lead times are negotiated values which alter both the manufacturer’s cost and the demand uncertainty.

3.0 The Decision Setting

It is assumed that the demand for individual items is uncorrelated although this assumption can be relaxed. At any time associated with a lead time \( t \) a buyer can make decisions using only what information is available. We assume that the buyer’s understanding of demand for an item \( x \) is such that he is willing to assume that it is normally distributed with a mean value of \( \mu \) and a finite variance. Since the level of uncertainty in his prediction rises with the lead time, it is appropriate to state the variance as a function of \( t \). To make this notion more concrete we assume that predicting demand for a specific item is akin to predicting the position of a random variable that obeys geometric Brownian motion.

This assumption implies that the variance associated with our prediction increases linearly in \( t \) (See Dixit & Pindyck (1992) for coverage of the characteristics of Brownian motion.) We denote by \( D(Q \mid \theta, t) \) the CDF of demand faced for a specific item of type \( \theta \) at lead time \( t \). Similarly, we denote the pdf of demand as \( d(Q \mid \theta, t) \) [We frequently drop the subscript when no confusion is introduced.] We note that, even if multiple items share a common value of \( \mu \), the items may differ significantly in the demand variability they face, and consequently in the value of postponement. For example, the producer may have historical sales data on items in one group, and no such data on newer items in another group. We may think of these groups as two product types. We assume that the variability in an item’s demand is linear in its type. An item of type \( i \) is associated with the parameter \( \theta \) which relates to its demand uncertainty. Based on this description, ordering decisions are made assuming that

\[
x_i(t) \sim D(Q \mid \theta, t) = N\left(\mu, \theta t \sigma^2\right).
\]

We define \( w \) as the retailer’s selling price for the final product in a competitive market. We let \( c \geq \frac{w}{2} \) represent the producer’s per unit production cost. We also define \( L(t) \), \( R(t) \), and
$p(t)$ as the producer’s cost of offering a lead time of $t$, the fixed price paid by the retailer to the manufacturer for placing an order with a lead time of $t$, and the per unit price paid by retailers to the manufacturer, when the lead time is $t$. We also label the retailer’s order quantity $Q(\theta_i, t)$.

We further specify properties of $L(t)$ to include: (i) $L_t < 0$; (ii) $L_{tt} > 0$; and (iii) $L(T) > 0$. For tractability we assume that $L(t) = \frac{L_0}{t}$, Thus we state that the manufacturer’s cost is higher for a shorter lead time,\(^1\) with the marginal cost of lower lead time increasing, and there is always some fixed cost of placing an order.

### 3.1 Centralized Planning (Social Welfare) and Order Size

We begin the analysis by calculating the maximum expected value of revenue minus production costs that can be derived from this relationship. We refer to this as the social welfare $S(\theta_i, t, Q)$ for the joint problem of the retailer and the manufacturer. For each item (SKU) of a given type $\theta_i$ and lead time $t$, the socially efficient order size corresponds to the newsvendor solution. The parameters of the problem depend only on the demand variability faced by the retailer and the costs faced by the manufacturer. For a given type, lead time, and quantity ordered we have:

$$E[S(\theta_i, t, Q)] = w\{E[x \mid x \leq Q] + Q \cdot \Pr\{x > Q\}\} - cQ - L(t)$$

(1)

It can be shown that:

$$E[x \mid x \leq Q] = \mu D(Q \mid \theta_i, t) - \sigma^2 t d(Q \mid \theta_i, t)$$

(2)

Where, $D(Q \mid \theta_i, t)$ is the probability that the demand for items of type $\theta_i$, with a lead time of $t$ is no greater than $Q$. We may restate (1) as:

$$E[S(\theta_i, t, Q)] = w\{\mu D(Q \mid \theta_i, t) - \sigma^2 t d(Q \mid \theta_i, t) + Q(1 - D(Q \mid \theta_i, t))\} - cQ - L(t)$$

(3)

Given our setting description, $D(Q \mid \theta_i, t) = \Phi\left(\frac{Q - \mu}{\sigma \sqrt{\theta_i t}}\right)$, and $d(Q \mid \theta_i, t) = \frac{1}{\sigma \sqrt{\theta_i t}} \phi\left(\frac{Q - \mu}{\sigma \sqrt{\theta_i t}}\right)$, where, $\Phi(\bullet)$ and $\phi(\bullet)$ are the CDF and pdf of the Normal distribution, respectively.

We define $Q^*(\theta_i, t)$ as the solution to the newsvendor problem defined by (1). Note that this value defines the socially efficient purchasing quantity, as a function of $\theta_i$ and $t$. The derivative of (1) w.r.t. $Q$ evaluated at $Q^*$ can be stated as:

$$\frac{\partial E[S(\theta_i, t, Q)]}{\partial Q} = w\{Qd(Q \mid \theta_i, t) + [1 - D(Q \mid \theta_i, t)] - Qd(Q \mid \theta_i, t)\} - c$$

(4)

Equivalently:

\(^1\) Property (i) and the assumption that $c \geq w/2$ also imply that, in the newsvendor solution presented below, the optimal purchasing quantity $Q^*(\theta_i, t) > \mu$. 

---

1. Property (i) and the assumption that $c \geq w/2$ also imply that, in the newsvendor solution presented below, the optimal purchasing quantity $Q^*(\theta_i, t) > \mu$. 

---

4
Postponement Contracts

\[
\frac{\partial E[S(\theta, t, Q)]}{\partial Q} = w\left[1 - D(Q | \theta, t)\right] - c. \tag{5}
\]

Noting that \( D(Q | \theta, t) = \mu + (\theta t)^{1/2} \), we can state the socially optimal order quantity for each item type \( \theta \) at time \( t \) from the F.O.C. as:

\[
Q^\star(\theta, t) = \mu + (\theta t)^{1/2} \phi^{-1}\left(\frac{w-c}{w}\right) \tag{6}
\]

This shows that the efficient purchased quantity increases with lead time \((t)\), and type \((\theta)\). The increase in the efficient purchase quantity is a direct result of variance increasing in type and time. In a centralized supply chain the quantity of each product type would be purchased in accordance with \( Q^\star(\theta, t) \).

It is important to identify the derivative of the expected social welfare function w.r.t. \( Q \) for other points as well. It is clear that since \( Q^\star \) maximizes expected social welfare we have \( \frac{\partial E[S(\theta, t, Q)]}{\partial Q} = 0 \) at \( Q = Q^\star(\theta, t) \). For \( Q < Q^\star(\theta, t) \) we have: \( \frac{\partial E[S(\theta, t, Q)]}{\partial Q} > 0 \) and \( \frac{\partial^2 E[S(\theta, t, Q)]}{\partial Q^2} < 0 \). Similarly, for \( Q > Q^\star(\theta, t) \) we have: \( \frac{\partial E[S(\theta, t, Q)]}{\partial Q(t)} < 0 \) and \( \frac{\partial^2 E[S(\theta, t, Q)]}{\partial Q^2} < 0 \).

The following Lemma analyzes important characteristics of the social welfare function when the socially efficient quantity is purchased.

**Lemma 1**: When the socially efficient quantity is purchased, social welfare is linear in \( \mu \), and decreasing in variance, with:

\[
E[S(\theta, t, Q^\star)] = \mu(w-c) - w(t\theta)^{1/2} \phi^{-1}\left(\frac{w-c}{w}\right) - L(t). \tag{7}
\]

**Proof**: Noting that \( D(Q | \theta, t) = \Phi\left(\frac{Q-\mu}{(t\theta)^{1/2}\sigma}\right) \), substitution into (3) yields,

\[
E[S(\theta, t, Q^\star)] = w\phi^{-1}\left(\frac{w-c}{w}\right) - Q^\star - L(t).
\]

Rearranging terms yields:
Postponement Contracts

\[
E[S(\theta,t,Q)] = \mu(w-c) - w(\theta t)^{\frac{1}{2}} \sigma \phi \left[ \Phi^{-1}\left(\frac{w-c}{w}\right) \right] + Q^* \left[ w - w + w \left(\frac{c}{w}\right) - c \right] - L(t) \quad \text{QED.}
\]

We note that \(E[S(\theta,t,Q)]\) is linear in \(\mu\). Hence, when evaluating the expected social welfare and the socially optimal quantity is purchased, the distribution of \(\mu\), can be replaced by its expected value.

Another important conclusion is that social welfare is strictly decreasing in \(\sigma\). In other words, lowering variance increases profit, and low type products, meaning those with lower values of \(\theta\), are more profitable. When comparing different demand distributions, variability only impacts stock-out costs. With lower variability the probability of stocking out is reduced, and profit increases. This can also be shown by inspecting an alternate specification of the profit function: \(\text{Profit} = w^*\text{Demand} - w^*E[x|x>Q] - c^*Q\). Only the second term changes with variability hence lower variability increases profit.

3.2 Social Welfare and Lead Time

We now turn to the socially optimal lead time, for a given item (SKU). To identify this we hold the purchase quantity at \(Q^*(\theta,t)\). Under appropriate conditions there is an interior solution to the socially optimal lead time, denoted by \(t^*\), i.e. \(0 < t^* < T\). This socially optimal lead time must vary by type because the type defines the rate at which variance is increasing in \(t\) while the producer’s cost is decreasing in \(t\). Hence, we seek \(t^*(\theta)\). The logic for an interior solution is that at \(t = 0\) increasing lead time lowers cost substantially, but has only a small impact on demand variability, so at \(t = 0\) increasing lead time is beneficial. At the other extreme, when \(t = T\), the cost savings from increasing lead time is marginal \((L_0 > 0)\), but the increased cost from demand variability is substantial. Alternatively, we may say that it is assumed that \(T\) is large enough, so that \(\forall \theta, t^*(\theta) < T\).

**Lemma 2**: When \(L(t) = \frac{L_0}{t}\), the socially optimal lead time is given by

\[
t^*(\theta) = \left( \frac{2L_0}{w(\theta)^{\frac{1}{2}}\sigma \phi \left[ \Phi^{-1}\left(\frac{w-c}{w}\right) \right]} \right)^{\frac{1}{3}}
\]

**Proof**: The F.O.C. for the social welfare function (1) is \(\frac{dE[S(\theta,t,Q)]}{dt} = 0\). We rewrite this as

\[
\frac{dE[S(\theta,t,Q)]}{dt} = \frac{\partial E[S(\theta,t,Q)]}{\partial t} + \frac{\partial E[S(\theta,t,Q)]}{\partial Q} \frac{\partial Q}{\partial t}. \quad \text{When maximizing social welfare } Q(t)
\]

\(= Q^*(t)\), and this becomes:

\[
\frac{dE[S(\theta,t,Q)]}{dt} = \frac{\partial E[S(\theta,t,Q^*)]}{\partial t} + \frac{\partial E[S(\theta,t,Q^*)]}{\partial Q^*} \frac{\partial Q^*}{\partial t}. \quad \text{From the}
\]

\[
\end{proof}
F.O.C. that defines $Q^*(t)$ we have $\frac{\partial E[S(\theta,t,Q)]}{\partial Q} = 0$. (This is simply the Envelope Theorem.)

Therefore, the third term equals 0 and $\frac{dE[S(\theta,t,Q)]}{dt} = \frac{\partial E[S(\theta,t,Q^*)]}{\partial t}$. Using (7) we can rewrite this equality as, $\frac{\partial E[S(\theta,t,Q^*)]}{\partial t} = -w(\theta)^{1/2}\sigma\Phi^{-1}\left(\frac{w-c}{w}\right) - L_t$. Using the F.O.C., setting $c_t = 0$, and $L(t) = \frac{L_0}{t}$, we get the result QED.

By substitution and simplification we may also state the following, additional result.

**Lemma 3:** The social welfare at the efficient level of $Q^*(\theta,t^*(\theta))$ and $t^*(\theta)$ is decreasing in $\theta$ and $\sigma^2$ and is: $E[S(\theta,t^*(\theta),Q^*(\theta,t^*(\theta)))] = \mu(w-c) - \frac{3L_0}{t^*(\theta)}$.

Again, this indicates that low variance product types are more profitable.

Figures 1 and 2 exemplify these key results. Here we show the social welfare for two different types of products, as a function of order quantity and lead time. In both cases we use the following parameters: $w=$20; $c=$5; $\mu=50$; $\sigma=15$; and $L_0=50$. For the low type we choose $\theta_L=1$ and for the high type we choose $\theta_H=2$. Using the results of Lemmas 1 and 2, we can show that for the low type the optimal lead time is $t^*(\theta_L)=1.03$ and the optimal order quantity is $Q^*(\theta_L,t^*)=60.28$, yielding social welfare of 604.7. Similarly, for the high type: $t^*(\theta_H)=0.819$ and the optimal order quantity is $Q^*(\theta_H,t^*)=62.95$, yielding social welfare of 566.94. As can be seen from the figures, the social welfare for the low type is higher than for the high type, for any fixed lead time and order quantity. In addition, as discussed previously for the low type the optimal quantity is smaller and the optimal lead time is longer than for the high type.
Figure 1: Social welfare for type $\theta_L$ product

Figure 2: Social welfare for type $\theta_H$ product
3.2 The Manufacturer’s Problem
The manufacturer is a monopolist selling various items into the retail channel. In general terms, he offers a “menu of contracts” to induce retailers to purchase items with appropriate lead times. High variability items should be purchased with lower lead times, while low variability items correspond to longer lead times.

For the sake of exposition, assume two types of items, those with high variability (θ_H) and those with low variability (θ_L); i.e. i ⇔ H or L. The “high” type may be items that are new offerings (a new shade of lipstick for Kenebo, a new cut of jeans for Levi’s, or a new line of action figures for Mattel for instance.) These items are not likely to require dramatically new or additional process steps, but their demand is much harder to predict when compared to older products with a long sales history. We have already shown that the item type affects social welfare and the optimal order quantity. The frequency of type θ_i items in the population is: f_i.

The manufacturer offers two lead time options for these items t_1 and t_2 with t_1 < t_2. Therefore, we develop two contracts, to populate his “menu” corresponding to the different lead times. Each contract specifies lead time (t_j), a non-linear price including a fee related to lead time R(t_j), a price per unit p(t_j), and order quantity Q(t_j), j=1,2. The manufacturer’s problem is to offer these contracts to maximize his profit, when the decision variables are the prices and lead times for each item type.

Lemma 4: It is optimal for the manufacturer to set the unit price at marginal cost: p(t_j) = c(t_j) = c.
Proof: By setting price in such a way, the retailer’s optimal order quantity is identical to the socially optimal decision. Clearly, the newsvendor solution is decreasing in p, thus there is no benefit in charging a higher price if there are no restrictions on R(t). QED.

The rationale here is similar to setting a fixed tax instead of a tax rate. Given the price schedule, offered quantity, and lead time, the retailer’s profit from an item of type θ_i is denoted by π(θ_i,t_j,Q_j). Given the assumptions stated earlier we have,

\[ E[\pi(\theta_i,t_j,Q_j)] = w \{ E[x | x \leq Q_j] + Q_j \Pr\{x > Q_j\}\} - p(t_j)Q_j - R(t_j) \]  

From Lemma 4 we deduce that it is optimal to set the unit price such that: p(t_i) = c(t_i). Therefore, from (1) we get,

\[ E[\pi(\theta_i,t_j,Q_j)] = E[S(\theta_i,t_j,Q_j)] + L(t_j) - R(t_j) \]  

Equation (9) implies that the optimal order quantity for type θ_i at lead time t_j is identical to the socially efficient quantity, from (6). Thus the manufacturer’s decision is clear if he is a monopolist considering a single product; produce a quantity at a lead time that maximizes social welfare, and charge the buyer a price per unit equal to variable cost, along with a fixed fee equal to the total contribution of the product.

\[ This\ assumption\ can\ be\ relaxed\ substantially! \]
However, if the manufacturer offers multiple items, we need to devise a schedule of quantities, lead times and fixed fees for each item type. To maximize profit, the manufacturer chooses lead times for different items, with a quantity and fixed fee at each lead time. The manufacturer’s profit function depends on the revenue from each item, and the relative frequency of each item type. The manufacturer faces two types of constraints in developing this menu. Specifically, he must respect the Individual Rationality (IR) of the buyer, for each item type. This refers to the fact that the retailer’s expected profit from an item must be non-negative, or else he does not place the order. He must also provide Incentive Compatibility (IC) for each item type. Formally, we state these constraints below.

$$ (IR1) \quad E[\pi(\theta_H, t_1, Q(t_1))] \geq 0,$$

$$ (IR2) \quad E[\pi(\theta_L, t_2, Q(t_2))] \geq 0,$$

$$ (IC1) \quad E[\pi(\theta_H, t_1, Q(t_1))] \geq E[\pi(\theta_H, t_2, Q(t_2))], \text{ and}$$

$$ (IC2) \quad E[\pi(\theta_L, t_2, Q(t_2))] \geq E[\pi(\theta_L, t_1, Q(t_1))].$$

Formalizing the manufacturer’s problem with \( p(t) = c \) we have:

$$ \max_{t_1, t_2, R(t_1), R(t_2), Q(t_1), Q(t_2)} \left[ \Pi(t_1, t_2, R(t_1), R(t_2)) \right] \equiv \begin{cases} f_H \left[ R(t_1) - L(t_1) \right] \\ f_L \left[ R(t_2) - L(t_2) \right] \end{cases} $$

Subject to (10). Note that from Lemma 4, the manufacturer’s profit margin is zero. He sells items at marginal cost. This suggests that quantity decisions do not directly affect the manufacturer’s profit. However their importance is still substantial. Choosing quantities impacts retailers’ IR and IC constraints, and enables self-selection for the different types. Hence, strategically setting quantities maximizes manufacturer profit. To solve the manufacturer’s problem we follow the literature on vertical differentiation and second-degree price discrimination to identify a mechanism that assures self-selection. (See Tirole 1988, Maskin and Riley 1984, and Leng 2005 for background and examples.) By self selection we mean, given the alternatives offered by the manufacturer, retailers find it in their best-interest to purchase items at the appropriate lead times and quantities.

### 3.3 Setting Quantities and Lead Times:

The IR and IC constraints imply that it may not be in the manufacturer’s best interest to offer socially efficient quantities for each item type. The reasoning is that a contract with a fixed fee that extracts all surplus from a buyer of one type may allow the buyer of another type to hold onto much of the surplus created and the fixed fee for this contract (which would be paid by both players) results in revenues that are strictly less than the best that the manufacturer can achieve. Since the low type items generate more surplus, the manufacturer would like to extract this amount as \( R(t_1) \). However, he must make sure that the terms of this contract are at least as attractive to this buyer as the terms of the contract targeting the customer for the high type product. Simultaneously, he must ensure that the contract targeting the buyer of high type products do not yield negative profits for customers interested in these items. This is facilitated by the application of the following result.
**Proposition 1:** The optimal fixed payments set by the manufacturer are characterized by:

\[
R^* (t_i) = E \left[ S \left( \theta_H, t_i, Q(t_i) \right) \right] + L(t_i) \tag{12}
\]

\[
R^* (t_2) = E \left[ S(\theta_L, t_2, Q(t_2)) \right] + L(t_2) - E \left[ S(\theta_L, t_1, Q(t_1)) \right] + E \left[ S(\theta_H, t_1, Q(t_1)) \right] \tag{13}
\]

**Proof:** The manufacturer creates contracts such that IR1 and IC2 are binding. Thus, the contracts offered reflect,

\[
\text{(IR1)} \quad E \left[ \pi \left( \theta_H, t_1, Q(t_1) \right) \right] \geq 0.
\]

From (9) we get,

\[
R^* (t_i) = E \left[ S \left( \theta_H, t_i, Q(t_i) \right) \right] + L(t_i)
\]

Setting \( R(t_2) \) so that IC2 binds yields:

\[
\text{(IC2)} \quad E \left[ \pi \left( \theta_L, t_2, Q(t_2) \right) \right] = E \left[ \pi \left( \theta_L, t_1, Q(t_1) \right) \right]
\]

Again, from (9) we get,

\[
E \left[ S(\theta_L, t_2, Q(t_2)) \right] + L(t_2) - R(t_2) = E \left[ S(\theta_L, t_1, Q(t_1)) \right] + L(t_1) - R(t_1), \text{ thus}
\]

\[
R^* (t_2) = E \left[ S(\theta_L, t_2, Q(t_2)) \right] + L(t_2) - E \left[ S(\theta_L, t_1, Q(t_1)) \right] - L(t_1) + R(t_1)
\]

Inserting \( R^*(t_1) \) we get,

\[
R^* (t_2) = E \left[ S(\theta_L, t_2, Q(t_2)) \right] + L(t_2) - E \left[ S(\theta_L, t_1, Q(t_1)) \right] + E \left[ S(\theta_H, t_1, Q(t_1)) \right]
\]

To show that under (12) and (13) (IR2) holds, note that, because profit is decreasing in variance:

\[
E \left[ \pi \left( \theta_L, t_2, Q(t_2) \right) \right] > E \left[ \pi \left( \theta_H, t_2, Q(t_2) \right) \right] = 0
\]

To show that under (12) and (13) (IC1) holds,

\[
E \left[ \pi \left( \theta_H, t_1, Q(t_1) \right) \right] = E \left[ S(\theta_H, t_2, Q(t_2)) \right] - E \left[ S(\theta_L, t_2, Q(t_2)) \right] + E \left[ S(\theta_L, t_1, Q(t_1)) \right] - E \left[ S(\theta_H, t_1, Q(t_1)) \right] - E \left[ S(\theta_H, t_1, Q(t_1)) \right] - E \left[ S(\theta_H, t_1, Q(t_1)) \right]
\]

which is negative because the difference in profit reduction, between \( t_1, Q(t_1) \) and \( t_2, Q(t_2) \) is larger for type \( \theta_L \) than for type \( \theta_H \). QED

Inspecting the results of Proposition 1 indicates the manufacturer’s ability to extract surplus. Note that low variance items generate more surplus than high variance items, hence the fixed fee for type \( \theta_L \) can be set such that type \( \theta_H \) could not profit under the contract terms offered for \( \theta_L \). For high variance items, which are less profitable, the manufacturer can extract all surplus. This is true because retailers do not have alternatives for purchasing these items and it is
not viable for them to purchase them under the terms of the contract for low variance items. Conversely, if the manufacturer attempts to extract too much of the surplus on the low variance items the retailer could purchase these under the contract for the high variance items. This limits the ability of manufacturer to extract all surplus as his profit from low variance items. At most the fixed fee can be set such that when purchasing the low variance items, the retailer is indifferent between the low variance contract and the high variance contract. In this case the retailer chooses the “appropriate” contract; i.e. the low variance contract for low variance items. We note that Proposition 1 also implies that the low type products are more profitable for both the manufacturer and the retailer.

Once \( R(t_1) \) and \( R(t_2) \) are set in (12) and (13), the manufacturer’s problem becomes unconstrained, and we may state his objective as,

\[
\begin{align*}
\text{Max}_{t_1, t_2, R(t_1), R(t_2), Q(t_1), Q(t_2)} & \left\{ \Pi(t_1, t_2, R(t_1), R(t_2)) \right\} \\
&= \left\{ f_H \left[ R(t_1) - L(t_1) \right] \right\} \\
&+ f_L \left[ R(t_2) - L(t_2) \right]
\end{align*}
\]

(14)

After rearranging, and noting that \( f_H + f_L = 1 \), we get:

\[
\begin{align*}
\text{Max}_{t_1, t_2, Q(t_1), Q(t_2)} & \left\{ \Pi(t_1, t_2, Q(t_1), Q(t_2)) \right\} = E \left[ S(\theta_H, t_1, Q(t_1)) \right] + f_L \left[ E \left[ S(\theta_L, t_2, Q(t_2)) \right] \right] \\
&- E \left[ S(\theta_L, t_1, Q(t_1)) \right]
\end{align*}
\]

(15)

**Proposition 2:** The manufacturer’s offer for the low variance type is socially efficient, i.e.,

\[ t^*_L = t^c(\theta_L) \text{ and } Q^c(t_2) = Q^c(\theta_L, t^c(\theta_L)). \]

**Proof:** Evaluating the manufacturer’s profit from the order quantity offered at lead time \( t_2, Q(t_2) \):

\[
\frac{\partial \Pi(t_1, t_2, Q(t_1), Q(t_2))}{\partial Q(t_2)} = f_L \frac{\partial E \left[ S(\theta_L, t_2, Q(t_2)) \right]}{\partial Q(t_2)}.
\]

Evaluating the manufacturer’s profit at the lead time for the low variance type, \( t^*_L \):

\[
\frac{\partial \Pi(t_1, t_2, Q(t_1), Q(t_2))}{\partial t_2} = f_L \frac{\partial E \left[ S(\theta_L, t_2, Q(t_2)) \right]}{\partial Q(t_2)}.
\]

Hence, when the manufacturer chooses the quantity and lead time for a type \( \theta_L \) item, his profit-maximizing choice is identical to maximizing social welfare. \( \Box \)

The intuition behind Proposition 2 is largely based on the fact that profit is decreasing in variance. As shown in Proposition 1 the terms of the low variance contract are unattractive for high variance items. This is achieved by setting a fixed fee sufficiently high. Given this fact, the manufacturer is then free to set the remaining terms of this contract to maximize social welfare. Thus for low variance items it is optimal to offer the socially efficient lead time and quantity. Thus, for these goods the monopolist manufacturer’s decision is aligned with that of a centralized supply chain.
**Proposition 3:** The manufacturer’s offer for the high variance type includes a lead time that is shorter than is socially efficient and a quantity that is larger than is socially efficient. i.e.: 

\[ t^*_1 < t^e(\theta_H) \] and \[ Q^*(t^*_1) > Q^e(\theta_H, t^*_1) \].

**Proof:** Consider the two parts of the proposition in turn. For part 1, evaluating the manufacturer’s profit from the lead time \( t \) offered to \( \theta_H \) we have,

\[
\frac{\partial \Pi(t_1, t_2, Q(t_1), Q(t_2))}{\partial t_1} = \frac{\partial E\left[S(\theta_H, t_1, Q(t_1))\right]}{\partial t_1} - f_L \frac{\partial E\left[S(\theta_L, t_1, Q(t_1))\right]}{\partial t_1}. \tag{16}
\]

To characterize the optimal solution for the short lead time, \( t^*_1 \), we evaluate the F.O.C. at the efficient lead time, \( t^e(\theta_H) \). From the F.O.C. for the social welfare function we have:

\[
\frac{\partial E\left[S(\theta_H, t_1, Q)\right]}{\partial t_1} \bigg|_{t_1 = t^e(\theta_H)} = 0. \]

Since optimal lead times decrease in variance, we know that \( t^e(\theta_L) > t^e(\theta_H) \). Therefore, \( \frac{\partial E\left[S(\theta_L, t_1, Q)\right]}{\partial t_1} \bigg|_{t_1 = t^e(\theta_H)} > 0. \) Since the F.O.C. must hold at the optimal solution to (15) we know that \( t^*_1 < t^e(\theta_H) \). For part two, we evaluate the optimal order quantity at \( t^*_1 \). Evaluating the manufacturer’s profit at \( Q(t^*_1) \) we have:

\[
\frac{\partial \Pi(t_1, t_2, Q(t_1), Q(t_2))}{\partial Q(t_1)} = \frac{\partial E\left[S(\theta_H, t_1, Q(t_1))\right]}{\partial Q(t_1)} - f_L \frac{\partial E\left[S(\theta_L, t_1, Q(t_1))\right]}{\partial Q(t_1)}. \tag{17}
\]

To characterize the optimal solution for the short lead time, \( Q^*(t^*_1) \), we evaluate this F.O.C. at the efficient quantity. From the F.O.C. for maximizing social welfare we know that at \( Q(t^*_1) = Q^e(\theta_H, t^*_1) \), we have \( \frac{\partial E\left[S(\theta_H, t^*_1, Q)\right]}{\partial Q} = 0. \) Since, the optimal quantity increases in variance we know that \( Q^e(\theta_L, t^*_1) < Q^e(\theta_H, t^*_1) \), thus \( \frac{\partial E\left[S(\theta_L, t^*_1, Q)\right]}{\partial Q} \bigg|_Q = Q^e(\theta_H, t^*_1) < 0. \)

The F.O.C. for (15) w.r.t. \( Q(t^*_1) \), requires that: \( Q^*(t^*_1) > Q^e(\theta_H, t^*_1) \) and the manufacturer’s optimal quantity to offer at lead time \( t^*_1 \) is higher than the efficient quantity for type \( \theta_H \). The justification for this is that when increasing the offered quantity (from \( Q^e \)) the derivative of \( E[S(\theta_H, Q(t^*_1), t^*_1)] \) is negative. \( QED \)

To understand the intuition behind Proposition 3, note that at the socially optimal solution, high variance items have shorter lead times and larger order quantities than low variance items. This follows from Lemmas 1 and 2. Inspection of the optimal lead time and quantity for the high variance items indicates that it is in the manufacturer’s best interest to deviate from the socially optimal values. The reasoning here is that while a contract could be
structured, with these efficient values \((t_i = t^*(\theta_H)\) and \(Q(t_i) = Q^*(\theta_H, t_i)\)), these values specify a contract that is quite attractive to the buyer of the low variance items. Therefore, to maximize his profits, the manufacturer must choose different values, which make the high variance contract less attractive for the buyer of low variance items. Thus the specified lead time and order quantity are moved away from the efficient levels. When this is done surplus decreases for both high and low variance items. The impact of these deviations from socially optimal levels have a more pronounced impact on the valuation of this contract by the customer for low type items. Consequently, the manufacturer can offer a contract that is not “too” inefficient but still separate these two groups.

The optimal order quantity \(Q^*(t_i)\) can also be deduced at this time. Using (17) and (5) we can state the F.O.C for quantity at \(t_1\) as,

\[
\frac{\partial \Pi(t_1, t_2, Q(t_1), Q(t_2))}{\partial Q(t_1)} = \left\{ w\left[1-D(Q(t_i)\mid \theta_H, t_i)\right]-c \right\} - f_L \left\{ w\left[1-D(Q(t_i)\mid \theta_L, t_i)\right]-c \right\} \tag{18}
\]

**Lemma 5:** The optimal quantity at the short lead time, \(Q(t_1)\), exists, and is the solution to:

\[
\Phi \left( \frac{Q^*(t_i^*)-\mu}{(t_i^*\theta_H)^{\frac{\mu}{\sigma}}} \right) - f_L \Phi \left( \frac{Q^*(t_i^*)-\mu}{(t_i^*\theta_L)^{\frac{\mu}{\sigma}}} \right) = (1-f_L) \left( \frac{w-c}{w} \right) \tag{19}
\]

**Proof:** Substituting \(D(Q \mid \theta, t) = \Phi \left( \frac{Q-\mu}{(t\theta)^{\frac{\mu}{\sigma}}} \right)\), into (18) and equating to zero we get (19).

The LHS of (19) is the difference between two CDF’s, hence it attains a value of 0 at \(Q = -\infty\), and \((1-f_L)\) at \(Q = +\infty\). The RHS is constant and less than \((1-f_L)\forall c > 0\). Thus, there exists a solution to (19). If the LHS is monotonic, this solution is unique. Taking the derivative of the LHS w.r.t. \(Q\) generates the condition for monotonicity:

\[
\frac{1}{(t_i\theta_H)^{\frac{\mu}{\sigma}}} \phi \left( \frac{Q-\mu}{(t_i\theta_H)^{\frac{\mu}{\sigma}}} \right) - f_L \frac{1}{(t_i\theta_L)^{\frac{\mu}{\sigma}}} \phi \left( \frac{Q-\mu}{(t_i\theta_L)^{\frac{\mu}{\sigma}}} \right) > 0. \tag{20}
\]

This is equivalent to: 
\(d(Q \mid \theta_H, t_i) - f_L d(Q \mid \theta_L, t_i) > 0\). In other words, the LHS is monotonic whenever \(f_L\) is sufficiently high. \(QED.\)

To understand these results we work with the same data used to generate Figures 1 & 2, but we hold lead times constant, and set \(f_L=0.6\). This allows us to identify the impact of changes in quantity decisions. We choose to hold lead times at the socially efficient levels; i.e. for both products \(t = t'(\theta)\). As stated previously these values are \(t'(\theta_L)=1.03\) and \(t'(\theta_H)=0.819\). Based on the results of Proposition 3, the manufacturer finds it optimal to offer a quantity greater than is
socially efficient, and $Q^*=71.08$. The surplus from the high type products in this case would be: 577.11.

**Figure 3: Social welfare as a function of Quantity with fixed lead times**

![Diagram showing social welfare as a function of quantity with fixed lead times.](image)

We are now positioned to deduce the optimal lead time, $t_1^*$.

**Lemma 6:** The optimal short lead time exists, and is the solution to:

$$
(t_1^*)^{3/2} \left[ \frac{Q^* - \mu}{\sigma \sqrt{t_1^*}} \right] - f_L \phi \left( \frac{Q^* - \mu}{\sigma \sqrt{t_1^*}} \right) = \frac{2L_0 (1 - f_L)}{w \sigma}.
$$

**Proof:** Taking the derivative of $E[S(\theta_H, Q(t_1), t_1)]$ w.r.t. $t_1$ we get (we can take only partials b/c of the envelope theorem):

$$
\frac{\partial E[S(\theta_H, Q, t_1)]}{\partial t_1} = -\frac{1}{2} w \sigma^2 \theta d(Q | \theta, t) - \frac{\partial L}{\partial t_1}.
$$

So (16) becomes:

$$
-\frac{1}{2} w \sigma^2 \theta_H d(Q | \theta_H, t_1) - \frac{\partial L}{\partial t_1} + f_L \left[ \frac{1}{2} w \sigma^2 \theta_L d(Q | \theta_L, t_1) + \frac{\partial L}{\partial t_1} \right].
$$

Note that in this case we are holding lead times constant, hence this optimal quantity is different than when lead times are optimized, as well.
Equating to zero, we get:

\[
\frac{w\sigma}{2} \left[ \frac{\theta_H \phi \left( \frac{Q - \mu}{\sigma\sqrt{\theta_H t_1}} \right)}{\sqrt{t_1}} - f_L \frac{\theta_L \phi \left( \frac{Q - \mu}{\sigma\sqrt{\theta_L t_1}} \right)}{\sqrt{t_1}} \right] = -\frac{\partial L}{\partial t_1} (1 - f_L),
\]

and

\[
\frac{w\sigma}{2} \left[ \theta_H \phi \left( \frac{Q - \mu}{\sigma\sqrt{\theta_H t_1}} \right) - f_L \theta_L \phi \left( \frac{Q - \mu}{\sigma\sqrt{\theta_L t_1}} \right) \right] = \frac{L_0}{t_1^s} (1 - f_L) \sqrt{t_1},
\]

therefore

\[
(t_1)^{1/2} \frac{w\sigma}{2} \left[ \theta_H \phi \left( \frac{Q - \mu}{\sigma\sqrt{\theta_H t_1}} \right) - f_L \theta_L \phi \left( \frac{Q - \mu}{\sigma\sqrt{\theta_L t_1}} \right) \right] = L_0 (1 - f_L)
\]

which is consistent with \( t' \). At \( t_1 = 0 \) the LHS of (15) is zero; while it is unbound when \( t_1 \) increases. The RHS is a constant, so an interior solution must exist. \( \text{QED} \)

We can depict the impact of the manufacturer’s optimization over lead times for the two product types. We maintain the previous numerical values and describe the optimal behavior in Figure 4. Here we hold purchased quantities at their socially efficient values, which are lead time dependent, \( Q^*(t_1) \). The socially efficient values for the low variance and high variance items are 1.03 and 0.82 respectively. These would generate surplus of 604.7 and 566.94, as discussed earlier. When the manufacturer optimally sets the lead time for the high variance item, he would offer this product type at a shorter lead time of 0.655, to make it unattractive to customers of low variance items.
From (19) we can see that,

$$\Phi\left(\frac{Q^* - \mu}{(t_1 \theta_H)^{1/2} \sigma}\right) - \Phi\left(\frac{w-c}{w}\right) = f_L$$

When $Q^* = \mu$, the numerator of the LHS of (22) equals the denominator, corresponding to $f_L = 1$. When $Q^* = Q^*(\theta_L, t_2)$ the numerator equals 0, corresponding to $f_L = 0$. Using the definition of $Q^*(\theta_L, t_2)$, and the facts that $\theta_L < \theta_H$ and $\Phi(x)$ is increasing in $x$ we note that for $Q^* \in (\mu, Q^*(\theta_L, t_2))$:

$$\Phi\left(\frac{Q^* - \mu}{(t_2 \theta_H)^{1/2} \sigma}\right) < \Phi\left(\frac{Q^* - \mu}{(t_2 \theta_L)^{1/2} \sigma}\right) < \left(\frac{w-c(t_2)}{w}\right)$$

Hence, the numerator and denominator are negative for $Q^* < Q^*(\theta_L, t_2)$. Also note that $Q^*(\theta_L, t_2) < Q^*(\theta_H, t_2)$, because the variance for type $\theta_H$ is larger. Similarly, the optimal lead time for the $t_1$ is set so that $t_1^* < t^*(\theta_H)$. This reduces the value for the high-variance items, to enable the manufacturer to capture more surplus from the low-variance items.

Based on the previous results and the derivation of the optimal lead time and order quantity set by the manufacturer, we can specify the implications from this model. Figure 2a
Postponement Contracts

replicates Figure 2, which shows the surplus for type $\theta_H$ products as a function of lead time and order quantity. The difference here is that the range of values for $Q$ and the scale have been changed to clarify the important issues. As discussed earlier the socially efficient values for type $\theta_H$ are $t^e(\theta_H)=0.819$; $Q^e(\theta_H, t^e)=62.95$, yielding social welfare of 566.94, as shown in Figure 2a. For type $\theta_L$ we have $t^e(\theta_L)=1.03$ and $Q^e(\theta_L, t^e)=60.28$, yielding social welfare of 604.7.

It is possible to specify a contract using these socially efficient values, which would meet the Individual Rationality and Incentive Compatibility constraints of both types, as specified in (10). Using the results of Proposition 1, the manufacturer would set the lead times and order quantities at their efficient levels, and choose the fixed payment at the short lead time as the gross profit generated by high variance products: $R(t_1) = 566.94+50/0.819 = 627.96$. The fixed payment for the long lead time items would be $R(t_2) = (604.7+50/1.03) – 599.54+566.94 = 620.54$. Since $f_L=0.6$, the manufacturer’s expected profit would be, from (14), 570.04.

However, the manufacturer can increase his profit, by offering low variance items at their socially efficient values and making the contract for high variance products less attractive, as depicted in Figure 2a. At the socially efficient point high variance products are offered at shorter lead time and larger quantity than low variance items. At this point the surplus function for high variance items is flat, while it is increasing for low variance items. By shortening the lead time and raising the quantity, the surplus function for high variance changes modestly, while the change in the surplus for low variance items is more pronounced. The optimal solution to this problem is $t^*(\theta_H)=0.681$ and $Q^*(\theta_H, t^*)=69.22$, yielding social welfare of 556.37. The fixed payments now become: at the short lead time $R(t_1) = 556.37+50/0.681 = 629.77$, and at the long lead time $R(t_2) = (604.7+50/1.03) – 574.08+556.37 = 635.43$. Since $f_L=0.6$, the manufacturer’s expected profit rises to 574.75, from (14).
Postponement Contracts

Figure 2a: Surplus of type $\theta_H$

![Graph showing surplus and lead time relationship]

Socially efficient point
Profit Maximizing point

4.0 Discussion and Conclusion

In this paper we develop the menu of contracts offered by a monopolist vendor across the supply chain for products with different levels of demand uncertainty. Retailers may face demand uncertainty for a variety of reasons, including product novelty, technological changes, changing tastes, and other factors. When different products within the same product line face different demand variability, the optimal response suggests placing different orders for them. It is shown, in the context of a simple newsvendor model, that optimal orders vary both in lead time and order quantity. We use these results as a benchmark for a monopolist manufacturer’s decisions when offering multiple lead times. Furthermore, it is shown that at these optimal decisions, higher variance items are ordered earlier, and in larger quantities than lower variance items, and that profit is decreasing in demand variability. Hence, a low variance item generates more total surplus than a higher variance item.

We model a scenario with 2 product types, which vary in demand variability. It is assumed that a monopolist manufacturer is knowledgeable about the two possible values of demand uncertainty, but cannot discern the demand uncertainty for a specific product at a specific retailer. This may be due to the importance of local information, which the retailer does not provide the manufacturer. The manufacturer offers retailers a menu of contracts for these items in advance of the selling season using a multi-dimensional contract. The dimensions of the contract are lead time, order quantity and a two-part price, which includes a fixed-fee per order.
and a per unit price. In maximizing profit, one of the objectives of these contracts is that retailers self-select and choose the shorter lead time for the higher variance items.

Using a standard newsvendor model we show that such a menu of contracts exists, with different products being offered at different lead times. The key insights from our model are that low variance items are offered at their optimal lead time and quantity, while high variance items are offered with a shorter lead time and larger quantity than would be set by a centralized supply chain. Contracts are stipulated in this manner to assure that low variance items are procured at the appropriate lead time. When inspecting retailer profit, we see that all profit on high variance items is appropriated by the manufacturer, while retailers maintain some profit on low variance items. This is consistent with the economics literature on “information rents” that are required when monopolist sellers cannot distinguish among types of buyers.

References


