2011 MSOM Special Interest Group Meetings
Ann Arbor, Michigan, June 26, 2011

PROCEEDINGS

Interface of Finance, Operations, and Risk Management
(iFORM) SIG

iFORM SIG Chairs:
Rene Caldentey
Jiri Chod
Lingxiu Dong
Danko Turcic
Trade Credit in Supply Chains: Multiple Creditors and Priority Rules

Song Alex Yang
London Business School, Regent’s Park, London, United Kingdom. sayang@london.edu

John R. Birge
The University of Chicago Booth School of Business, Chicago, IL 60637. john.birge@ChicagoBooth.edu

Priority rules determine the order of repayment when the debtor cannot repay all of his debt. In this paper, we study how different priority rules influence trade credit usage and supply chain efficiency when multiple creditors are present. We find that with only demand risk, when the wholesale price is exogenous, trade credit with high priority can lead to high chain efficiency, yet trade credit with low priority allows more retailers to obtain trade credit and suppliers to gain higher profits. When the supplier has control of the wholesale price, however, we show that the supplier should extend unlimited trade credit with net terms. We also study the case when demand risk mingles with other risks, especially those with longer terms. Under this setting, we show several scenarios when the optimal trade credit policy should change according to different risks and that, in general, trade credit with low priority results in high chain efficiency. Finally, we use empirical data to show that, at an aggregate level, trade credit usage reacts to changes in the law according to our theory.

Key words: supply chain management; newsvendor model; trade credit; priority rules; bankruptcy; financial constraint

History:

1. Introduction

As a type of short-term financing, trade credit allows the upstream firm in a supply chain to finance the downstream firm. Empirical evidence suggests that trade credit is an important source of external financing. For example, the Financial Times reports that in 2007, 90% of world merchandise trade is financed by trade credit, with a value of approximately $25 trillion. Because of its wide usage, researchers in economics and finance have proposed many theories explaining the role of trade credit. Recognizing the intrinsic connection of trade credit and supply chain contracting and inventory management, Yang and Birge (2010) propose a theory suggesting that trade credit serves as a risk-sharing mechanism between parties in a supply chain, hence enhancing supply chain efficiency. In that paper, the authors assume bank debt is senior to trade credit. That is, when the

retailer (the debtor) is unable to meet all of his debt obligations, the trade creditor will not receive anything until the bank debt is fully repaid.

In practice, however, priority rules are more complicated. On the one hand, priority can be determined in the form of a contractual agreement. For example, in a debt covenant, the initial creditor could specify that any future creditors must have low priority related to her claim. On the other hand, there are certain priority principles (see Schwartz 1997) that bankruptcy judges normally obey. Therefore, even if a priority rule is written in a contract, it may be overruled by bankruptcy judges as it may violate some of these general principles.

More interestingly, the legal statutes that govern priorities change constantly over time. Recently, for example, the Bankruptcy Abuse Prevention and Consumer Protection Act of 2005 (BAPCPA) has strengthened the supplier’s right of reclamation. That is, the supplier, when qualified, could receive (part of) the trade credit claim before other unsecured or secured creditors. Intuitively, these changes of law, together with different priority rules mutually agreed upon by the creditor and the debtor, may change the terms and usage of trade credit, as well as supply chain efficiency. In this paper, we try to develop a deeper understanding of this impact. Specifically, we try to answer the following questions: first, under different priority rules, how do the terms of trade credit and institutional funding change? Are suppliers willing to extend as much trade credit as retailers need? How do those changes influence the retailer’s order quantity and other investment decisions?

Among different priority rules, which one is most favorable for the supplier (the trade creditor), the retailer (the debtor), the supply chain, and/or end customers? Are the current legal rules efficient? Finally, given the flexibility judges have in bankruptcy rulings, how should they take different factors into consideration?

We organize this paper into three parts. We first summarize related literature and priority rules governing trade credit in practice. The second part (Sections 3 and 4) studies the influence of priorities on the supply chain under several different settings, all extended from the classical “selling to a newsvendor” model. The differences among these situations mainly comes from the nature of risk exposures and the sequence of events. As in Yang and Birge (2010), we start with the setting in which the retailer faces only demand risk. In this setting, we discuss two scenarios: first, when the wholesale price w is exogenous and, second, when it is endogenous. In the second setting, in addition to the demand risk, the retailer faces another risk when he has the opportunity to invest in a separate (generic) project. At last, we use empirical data to show that, at an aggregate level, trade credit usage reacts to changes in the law according to our theory.

In this paper, terms including priority and seniority are used interchangeably.
1.1. Literature Review

Lying at the interface of operations and finance, our work is closely related to many areas in both subjects. The research on trade credit, viewed mainly as a financing decision, is largely from financial economists. Among various theoretical investigations, Schwartz (1974), Emery (1984), Frank and Maksimovic (2004), Longhofer and Santos (2003), Wilner (2000), and Cuñat (2007) are mostly related to our work. For representative works, see Petersen and Rajan (1994, 1997), Ng et al. (1999), Love et al. (2007), Giannetti et al. (2008), Klappler et al. (2010), and the references therein.

As an additional research stream, the interface of operations and finance has recently received fast-growing attention. Besides Yang and Birge (2010), papers related to supply chain financing include Haley and Higgins (1973), Gupta and Wang (2009), Zhou and Groenevelt (2008), Federgruen and Wang (2010), Dada and Hu (2008), Lai et al. (2009), Kouvelis and Zhao (2009), and Caldentey and Chen (2010).

Another closely related area of research is law and finance. General literature on this area include Franks et al. (1996), La Porta et al. (1997), La Porta et al. (1998), and Haselmann et al. (2010). For theoretical and empirical works that examine priorities, see Adler (1992), Hart and Moore (1995), Barclay and Smith (1995), Schwartz (1997), and Schwartz (1989).

A separate stream of literature that is closely related to this research concerns the supplier’s right of reclamation. This area is almost exclusively contributed by legal scholars and practitioners. Garvin (1996) provides a thorough study of Uniform Commercial Code (U.C.C.) reclamation rights by tracing back the statutory history and summarizing various theories to support the right of reclamation. Kull (1998) and Dagan (2004) also study the reclamation right under U.C.C. Studies by Gilday (2003), Katz and Dion (2005), and Turner (2006) look at the Chapter 11 Critical Vendor Motion. BAPCPA reclamation, given its short history, is mostly studied by practitioners. Works in this area include: Morris (2007), Erens and Friedman (2007), Larsen (2007), and Nathan (2008). Among comparisons to other jurisdictions, Klotz (1992) summarizes the reclamation rights under Canadian laws.

2. Priority Rules: Practice and Definitions

Before going into the model, we briefly summarize what trade credit related priority rules are in practice. Under the current US legal system, trade credit is normally treated as general unsecured claims, which are the class with the lowest priority. Under different circumstances, however, trade credit may be assigned higher priority. For example, trade credit offered to the buyer when the
buyer is under bankruptcy is considered a post-petition claim, which has administrative priority and has to be paid back before other claims. However, in this paper, we are mostly interested in trade credit the seller extends when the buyer is out of bankruptcy. In this case, as a pre-petition claim, trade credit may have higher priority through several statutes.

First, outside bankruptcy, the seller’s reclamation of goods is governed by Section 2-702 of the Uniform Commercial Code (U.C.C). Under this statute, the trade creditor, discovering that the buyer has received goods while insolvent, can reclaim goods that are sold to the buyer within 10 days after the buyer received them.

Note that U.C.C. reclamation is only relevant before the buyer files for bankruptcy. In bankruptcy cases, US Bankruptcy Code Section 546(c) allows Section 2-702 of U.C.C to have the same effect. Further, BAPCPA amended Section 546(c) by expanding the reclamation period from 10 days to 45 days. However, even with this amendment, this reclamation right is weak. Although the statutes place the reclaiming seller ahead of the buyer’s general unsecured creditors, the seller is still behind the buyer’s secured creditors who have security interests in the goods. For example, it is common for a bank loan to use the buyer’s floating inventory as collateral. In this case, the bank holds a secured interest on the inventory and thus the seller cannot use the inventory reclamation right defined under U.C.C. or Section 546(c).

Another statute trade creditors can use to reclaim their goods is the Chapter 11 Critical Vendor Motion. This motion grants a subset of trade creditors high priority for trade credit on all goods sold within 90 days before the debtor files for Chapter 11 bankruptcy. Note that although this motion allows a longer reclamation period than the U.C.C. and Section 546(c) allow, it has two limitations. First, to qualify for this rule, the vendor has to be selected by the purchaser and approved by the judge. Second, this right does not apply to Chapter 7 (liquidation) bankruptcies, which most small firms use. BAPCPA reclaims, however, can be used in both Chapter 7 and Chapter 11 bankruptcies.

Finally, among legal scholars, it is agreed that the strongest reclamation right the supplier could enjoy is Section 503(b)(9) of BAPCPA. According to this statute, trade creditors who provided goods to a distressed debtor within 20 days of the commencement of a bankruptcy case are entitled to an administrative priority equal to the value of the goods sold. In bankruptcy, administrative priority is the highest preference afforded to an unsecured claim. This statute serves the direct

---

3 Bris et al. (2006) report that among corporate bankruptcies that are filed in New York and Arizona from 1995 to 2001, Chapter 11 bankruptcies have median pre-bankruptcy assets of $1.2 million, and Chapter 7 bankruptcies have median pre-bankruptcy assets of $0.11 million.
motivation of this paper, as we are interested in whether trade credit should have high or low priority when the buyer defaults.

Note that Section 503(b)(9) has its own flaws. For example, Morris (2007) points out that Section 503(b)(9) provides no guidance as to the proper standard for determining the value of the subject goods and moreover does not prescribe the date on which the goods should be valued. For example, the amount of such a claim varies significantly depending on whether the goods are valued on a retail, wholesale, or liquidation basis.

After summarizing how priorities are treated in practice, we introduce some notation that we use in this paper and link them to priorities in practice as mentioned above. In this paper, we only consider standard debt contracts. Except for priorities, we assume no covenant in terms of actions the debtor must or must not perform. We consider two types of financial claims: a bank loan and trade credit. A bank loan contract includes the market value $B$ and the face value $L_b$.\footnote{$L$ is short for “liability,” and the subscript $b$ is short for “bank loan.”} A trade credit contract consists of three elements: the wholesale (credit) price $w$, which is due when demand is realized, the trade credit discount $d_{tc}$, which specifies that if the retailer pays cash upon the delivery of the goods, he only needs to pay $w(1 - d_{tc})$ per unit, and the line of trade credit $L_s$.\footnote{$s$ is short for “supplier.”} that is, the maximum amount of trade credit the supplier is willing to extend. As it is common that a supplier extends credit on net-terms (see Giannetti et al. 2008), we assume $d_{tc} = 0$ for most of this paper.

Further, we assume that upon default, the payoffs of different parties (creditors and the debtor) only depend on the total value of the debtor, but not on the value of specific assets held by the debtor.\footnote{For most of this paper, this assumption does not make a difference as the firm are all in cash at the end of time horizon.} Given this assumption, let $y$ be the total value of the retailer before paying any claims; what the creditors receive should only be a function of $y$. Correspondingly, we define default thresholds $\theta_b$ (for bank loan) and $\theta_s$ (for trade credit) as: when $y < \theta_b$, the bank loan defaults; when $y < \theta_s$, trade credit defaults.\footnote{More complicated default rules are possible. For example, when one of the claims is collateralized, it is possible that this claim is paid off in full when $y$ is small yet collateral value is high, but default occurs when $y$ is large yet collateral value is low. For simplicity, we do not cover this scenario in this paper.} An allocation rule ($l_b(y)$ for the bank loan and $l_s(y)$ for trade credit) governs how the bank and the trade creditor are paid off when default occurs. By definition, the allocation rule needs to satisfy: $\forall y < \theta_b$, $l_b(y) < L_b$, and $\forall y \geq \theta_b$, $l_b(y) = L_b$; $\forall y < \theta_s$, $l_s(y) < L_s$, and $\forall y \geq \theta_s$, $l_s(y) = L_s$. 

\[ 4 \] $L$ is short for “liability,” and the subscript $b$ is short for “bank loan.”

\[ 5 \] $s$ is short for “supplier.”

\[ 6 \] More complicated default rules are possible. For example, when one of the claims is collateralized, it is possible that this claim is paid off in full when $y$ is small yet collateral value is high, but default occurs when $y$ is large yet collateral value is low. For simplicity, we do not cover this scenario in this paper.
Given this definition, a bank loan in a perfectly competitive market breaks even; that is,

\[ B = \int_0^{\theta_b} l_b(y) dF(y) + L_b \bar{F}(L_b). \]  

(1)

Clearly, different priority rules are reflected in the allocation rules \( l_b(y) \) and \( l_s(y) \), and lead to different \( L_b \) and \( L_s \). We use \( L_t = L_b + L_s \) to represent the total liability.

Depending on the magnitude of \( \theta_b \) and \( \theta_s \), we define priority (seniority) as: when \( \theta_b > \theta_s \), trade credit is senior to the bank loan; when \( \theta_b < \theta_s \), the bank loan is senior to trade credit; when \( \theta_s = \theta_b \), the bank loan and trade credit have equal seniority. While there are an infinite number of possible allocation rules, in this paper, we mainly focus on three representative cases, whose payoffs are shown in Figures 1.

*Figure 1  Payoffs with Different Priority*

(a) Trade Credit is Strictly Senior  
(b) Trade Credit is Strictly Junior  
(c) Trade Credit and the Bank Loan have the Same Priority

The left panel shows the case when trade credit is strictly senior to the bank loan. This can be seen as an approximation of the case when trade credit is subject to Section 503(b)(9) administrative reclamation. In this case, the bank receives nothing before the trade credit is fully repaid. The allocation rule follows: \( l_s(y) = \min(y, L_s), l_b(y) = \min((y - L_s)^+, L_b) \).

The second scenario is when the bank loan is strictly senior to trade credit, as shown in panel (b) of Figure 1. This is to approximate the scenario where the bank loan uses inventory as collateral, including the case when trade credit is subject to U.C.C. or Section 546(c) rejections. In this case, the trade creditor receives nothing until the bank loan is fully repaid. As illustrated, the allocation rule is: when \( y \leq \theta_b \), \( l_b(y) = y, l_s(y) = 0 \); when \( y \in (\theta_b, \theta_s) \), \( l_b(y) = L_b, l_s(y) = y - L_b \); when \( y \geq L_t \), \( l_b(y) = L_b, l_s(y) = L_s \).

This implicitly assumes the bank also has the revenue generated by the inventory as collateral.
Finally, the third scenario is shown in panel (c) of Figure 1. This describes the case when both trade credit and the bank loan are classified as general unsecured claims and have the same priority. Under this scheme, upon default, both creditors’ payoffs are proportional to the face value of their claim. In this case, \( \theta_b = \theta_s = L_t, l_b(y) = \min\left(\frac{L_b}{L_t} y, L_b\right), \) and \( l_s(y) = \min\left(\frac{L_s}{L_t} y, L_s\right) \).

3. Priority Rules and Demand Risk Sharing

With the above framework, we lay out the basic model, which is similar to Yang and Birge (2010). The sequence of events is shown in Figure 2.

Consider a “selling to a newsvendor” supply chain. The supplier and retailer are both risk-neutral. Both firms’ objective is to maximize their expected profit. The time horizon is divided into two periods. At the beginning of the first period, the supplier, as the Stackelberg game leader, publishes a wholesale price \( w \) and a line of credit \( L_s \). For simplicity, we confine our discussion to net-term trade credit terms. The retailer then places an order quantity \( x \). At the end of the first period, the supplier incurs unit production cost \( c \) and delivers the \( x \) units of goods to the retailer. The retailer has capital \( K \) at the end of the first period. To finance his inventory, the retailer borrows \( B \) from a bank, with a face value \( L_b \), and pays part of the order by cash.

During the second period, the retailer sells the goods on hand at retail price \( p = 1 \). Let \( \xi \) be the demand realized during the second period. Assume \( \xi \) has a cumulative distribution function \( F() \), with density function \( f() \), and failure rate (hazard rate) \( h() \). To simplify the analysis, we assume \( \xi \) has an increasing failure rate; that is, \( h(x) \) increases in \( x \). At the end of the second period, all revenue is realized as cash, and all unsold goods can be salvaged at salvage price \( s \) with total salvage value of \( s(x - \xi)^+ \). When the total amount \( sx + (p - s) \min(x, \xi) \) exceeds the sum of the face value of the bank loan \( L_b \) and the trade credit used by the retailer \( L_s \), the retailer pays off both
claims and keeps the rest. Otherwise, some obligation is partially paid off, the amount depending on the allocation rule. For simplicity, we assume \( s = 0 \); that is, unsold goods have no salvage value.

In this paper, we assume no distress costs. Without distress costs, it is easy to show that the supplier’s capital position is irrelevant to the problem we are interested in.

3.1. The Line of Trade Credit under Exogenous Wholesale Price

In the first scenario, we assume the wholesale price \( w \) offered by the supplier is exogenous; that is, neither the supplier nor the retailer has any control of the magnitude of \( w \). This is of interest for several reasons: first, in practice, the wholesale price and line of credit may be determined by different departments within a company. For example, the wholesale price may be determined by the marketing or production department, while the line of credit is set by a credit manager in the finance department. Second, certain regulations prohibit the supplier from discriminating customers through prices. For example, the Robinson-Patman Act prohibits a seller from selling comparable goods to different buyers at different prices for a transaction that crosses state borders.\(^9\)

Similarly, franchise laws at the state level require auto-makers to sell to all dealers within one state at the same price. Another possible reason for a fixed price is that the supplier may have signed a long-term supply contract that specifies the wholesale price with the retailer. All these constraints make the wholesale price independent of the retailer’s financial health. Therefore, the only decision the supplier can make is to determine the line of trade credit.

Intuitively, when facing different priority rules, the supplier anticipates different responses from the retailer, hence different profits for herself, even given the same wholesale price and credit limit. Focusing on how the supplier should offer different credit limits under different priority rules, we examine several different scenarios.\(^a\)

We start with the case when the retailer can use only a bank loan or trade credit. In these cases, as the retailer has only one creditor, priority rules are irrelevant. Later, we study three scenarios in which the retailer can use both trade credit and a bank loan, but under different priority rules.

As Yang and Birge (2010) point out, by extending trade credit, the supplier can effectively reduce the retailer’s marginal cost, and hence induce the retailer to order a larger quantity. This improves the supplier’s profit and the supply chain’s efficiency at the same time. This sword, however, is double-edged. Intuitively, when the supplier cannot adjust the wholesale price to compensate for the default risk associated with trade credit, the retailer may use more trade credit than the supplier prefers. Bank debt, on the other hand, can use the interest rate as a tool. When both sources of

\(^9\) Among antitrust scholars, the Robinson-Patman Act is universally unpopular. For related literature, see Posner (1976), Ross (1984) and references therein.
financing are available, priority rules should play an important role in not only allocating the profit between both the supplier and the retailer, but also in determining the order quantity and chain efficiency. Therefore, two questions arise naturally: first, under different priority rules, what is the optimal credit extension policy? In reaction, how does the retailer order and finance his inventory differently?

3.1.1. Only a Bank Loan is Available The first scenario we consider is that the retailer has only access to bank financing. Obviously, when his capital $K$ is less than the minimal capital without financing need $\kappa_{sc} = w\bar{F}^{-1}(w)$, the retailer tries to maximize his profit $\pi_r = \int_{L_b}^{x} \bar{F}(\xi) d\xi - K$ subject to his financing constraint $wx - K = \int_{0}^{L_b} \bar{F}(\xi) d\xi$, where the left hand side is the amount the retailer borrows from the bank, and the right hand side is the expected payoff the bank will receive. Solving the retailer’s problem leads to the following condition:

$$F(x) = \bar{F}(L_b) \frac{\partial x}{\partial L_b} = w.$$ 

(2)

Not surprisingly, the retailer orders $\bar{F}^{-1}(w)$, the same amount as when he has sufficient capital, reassuring the validity of the Modigliani-Miller (M-M) theorem under this setting.

3.1.2. Only Trade Credit is Available Now we introduce trade credit into the picture. Suppose the supplier offers a line of credit $\bar{L}_s$ to a retailer with internal capital $K$; that is, the retailer can at most obtain trade credit to the amount of $\bar{L}_s$. Further, we assume the retailer has no access to a bank loan. This setting can be discussed under two scenarios depending on the retailer’s financing need.

First, when the retailer has relatively more capital, the trade credit offered is more than the optimal level the retailer chooses to use, that is, $L_s < \bar{L}_s$, where $L_s$ is the amount of trade credit actually used by the retailer. In this case, the retailer maximizes $\pi_r = \int_{L_s}^{x} \bar{F}(\xi) d\xi - K$ subject to the financing constraint $wx - K = L_s \leq \bar{L}_s$.

Obviously, as $L_s < \bar{L}_s$, the optimal solution is an interior optimum and must satisfy the first order condition:

$$\bar{F}(x) = \bar{F}(L_s) \frac{\partial x}{\partial L_s} = w\bar{F}(L_s).$$

(3)

We can easily show that there exists a unique $(x, L_s)$ pair that satisfies the first order condition, as well as the financing constraint.\footnote{This result is shown in Lemma 4.} Under this condition, we show that the supply chain behaves according to the following lemma.

\footnote{The subscript $sc$ is short for “sufficient capital.” It is easy to show that when $K > \kappa_{sc}$, the retailer has no incentive or need to use external financing, and his order quantity equals the classical price-only quantity $\bar{F}^{-1}(w)$.}
Lemma 1. When the retailer receives an unlimited amount of trade credit, as $K$ increases, $x$ decreases, $L_s$ increases, and the retailer’s profit $\pi_r$ and the chain’s profit both decrease.

This lemma suggests that when offered sufficient trade credit, retailers with less capital place larger orders, which allows them to gain more profits. However, intuitively, when a retailer with extremely limited capital loads up his inventory, the supplier may end up with less profit.

This result naturally leads to the second case, when the retailer has relatively less capital and exhausts his line of trade credit. In this case, the retailer basically orders as much as he can; that is, $x$ satisfies $wx - K = \bar{L}_s$. To determine $\bar{L}_s$, the supplier tries to maximize $\int_0^{\bar{L}_s} \bar{F}(\xi) d\xi + K - cx$ subject to the above constraint. Setting the first order condition to zero, that is:

$$\bar{F}(\bar{L}_s) = c \frac{\partial x}{\partial \bar{L}_s} = \frac{c}{w},$$

we obtain the following result:

Lemma 2. When only trade credit is available, the supplier offers a line of trade credit $\bar{L}_s^*$ that satisfies $\bar{F}(\bar{L}_s^*) = \frac{c}{w}$.

This result is intuitive: the supplier’s marginal cost to produce one extra unit is $c$ and the marginal revenue is $w\bar{F}(\bar{L}_s)$; setting these two equal leads to the optimal solution. Note that the line of trade credit $\bar{L}_s^*$ is only a function of the supplier’s profit margin and is independent of the retailer’s capital $K$. Therefore, it is immune to any criticism of discrimination.

Facing this credit policy, the retailer’s behavior is summarized in the following results.

Proposition 1. $\exists \kappa_u = [w\bar{F}^{-1}(c) - \bar{L}_s^*]^+$ such that:

1. for $K < \kappa_u$, the retailer exhausts his line of trade credit ($L_s = \bar{L}_s^*$). In this interval, as $K$ increases, $x$ increases, $\pi_s$ increases, and $\pi_r$ is concave. If $w^2 > c$; that is, the supplier’s profit margin is greater than the retailer’s, $\pi_r$ decreases in $K$; and otherwise, $\pi_r$ first increases and then decreases in $K$;

2. for $K \in [\kappa_u, \kappa_{sc}]$, the retailer does not use up the line of trade credit the supplier offers, that is, $L_s < \bar{L}_s^*$. Lemma 1 holds. In addition, the supplier’s profit $\pi_s$ decreases in $K$ when $L_s$ is sufficiently small. When $ln[h(x)]$ is concave, $\pi_s$ is concave.

Note that many common distribution families, including the uniform and exponential distributions, satisfy that $ln[h(x)]$ is concave. In the following analysis, we assume $\pi_s$ is concave in $K$ when $K \in [\kappa_u, \kappa_{sc}]$. The
Also, note that under this scheme, when the retailer’s capital reaches the point that the credit he needs is exactly his credit limit, the retailer orders \( x = F^{-1}(c) \), that is, the quantity under the integrated chain.

The above results are illustrated in the following numerical example, as Figure 3 shows. The parameters are: \( \xi \sim \text{Uniform}[0, 1] \), \( c = 0.4 \), \( w = 0.5 \) and 0.9.

As the figure shows, when \( w = 0.5 \), that is, the supplier has a lower profit margin, she sets a credit limit \( \hat{L}_s = 0.2 \). Prior to hitting this limit, when the retailer’s capital decreases, both his order quantity and profit increase. However, once hitting the credit limit, the retailer’s order decreases as \( K \) decreases; when \( K \) becomes extremely small, his profit drops. Interestingly, as the retailer’s profit is maximized at \( \kappa_u \), when he has the freedom to his capital stock \( K \), \( \kappa_u \) is an obvious choice, and it also allows the decentralized chain to achieve coordination. On the supplier side, the supplier’s profit obtains a small increase initially by extending trade credit, but drops after that when the retailer’s capital decreases, as the higher sales cannot compensate for the retailer’s default risk.
The other case is when \( w = 0.9 \). With a higher margin, the supplier is willing to extend effectively unlimited trade credit, regardless of the retailer’s capital position; that is, \( \kappa_u = 0 \). In this regime, when the retailer’s capital decreases, the efficiency of the supply chain keeps improving with both the retailer’s and supplier’s profit increasing.

### 3.1.3. Trade Credit is Senior to the Bank Loan

Starting at this section, we focus on the impact of priority rules by examining cases where the retailer has access to both trade credit and a bank loan. We first study the case when trade credit is strictly senior to the bank loan, as shown in Figure 1 panel (a). Under this scheme, the retailer’s financing constraint is:

\[
\int_{L_s}^{L_t} F(\xi) d\xi = wx - K_0 - L_s,
\]

where \( L_s \) is bounded by \( \bar{L}_s \). The objective is to maximize his profit \( \pi_r = \int_{L_t}^{\bar{L}_t} F(\xi) d\xi - K \). Obviously, as the bank loan always has a positive interest rate, while the trade credit’s implied interest rate is zero, the retailer should first use trade credit and then the bank loan. Thus only one of the following two scenarios are possible.

First, the retailer has relatively more capital and uses only trade credit; that is, \( L_s \leq \bar{L}_s \), and \( L_b = 0 \). Obviously, this degenerates to the last case when the retailer only uses only trade credit.

Second, the retailer exhausts all trade credit extended to him and also uses a bank loan with a strictly positive amount; that is, \( L_s = \bar{L}_s \) and \( L_b > 0 \). In this case, the retailer’s decision is to determine his order quantity \( x \) and the amount of the bank loan \( L_b \) jointly. At an interior optimum, the following first order condition holds:

\[
\bar{F}(x) \frac{\partial x}{\partial L_b} - \bar{F}(L_t) = 0,
\]

where \( \frac{\partial x}{\partial L_b} = \frac{F(L_t)}{w} \). Note that the retailer’s first order condition under this priority rule is \( \bar{F}(x) = w \), which is exactly the same as when the retailer faces no capital constraint.

In this case, the supplier’s problem is to maximize her profit \( \pi_s = -cx + (K + B) + (\int_0^{L_t} \bar{F}(\xi) d\xi - \text{Bank’s Expected Payoff}) \), where the first term is the supplier’s marginal production cost, the second part is how much (cash) she receives upon delivery, and the third part is the expected payoff she receives at the end of the sales horizon. Note that as the bank loan breaks even, Bank’s Expected Payoff equals the face value of the loan \( B \); therefore, \( \pi_s = \int_0^{L_t} \bar{F}(\xi) d\xi - cx + K \), which can also be written as:

\[
\pi_s = (w - c)x - \int_0^{L_s} \bar{F}(\xi) d\xi.
\]

\(^{12}\) It is a coincidence that the supplier’s profits with \( K \geq \kappa_{sc} \) are the same in both cases.
As $x$ is independent of $\bar{L}_s$, it is obvious that $\pi_s$ decreases in $\bar{L}_s$, implying that whenever the retailer uses the bank loan, the supplier should not offer trade credit at all.

Combining these two cases, we find that when trade credit is senior to the bank loan, there is a bifurcation in the amount of trade credit offered (and used). The supplier offers trade credit if and only if both of the following conditions hold: first, her profit is greater than that when the supplier does not offer trade credit at all; that is, $\pi_{ts} > \pi_s(\kappa_{sc}) = (w - c)\bar{F}^{-1}(w)$. Second, the retailer prefers using only trade credit, instead of using both trade credit and a bank loan. Analysis of these two conditions leads to the following proposition.

**Proposition 2.** \( \exists \kappa_{ts}^n \in [0, \kappa_{sc}) \) such that:

1. for $K < \kappa_{ts}^n$, the supplier does not offer trade credit, the retailer uses only the bank loan, and the chain behaves exactly as if there is no capital constraint;
2. for $K \in [\kappa_{ts}^n, \kappa_{sc}]$, the supplier offers a line of trade credit $\bar{L}_s^* = \bar{F}^{-1}(c/w)$, and the retailer uses only trade credit. Proposition 1 holds.

This proposition leads to three distinctive cases in terms of trade credit usage, depending upon the magnitude of $\kappa_{ts}^n$. First, when $\kappa_{ts}^n = 0$, the supplier effectively offers unlimited trade credit; that is, even though she may set a line of trade credit credit, it is never reached. In this case, adding a bank loan becomes irrelevant, and the retailer only uses trade credit. This is the case illustrated in Figure 3 (with $w = 0.9$).

The other two scenarios, both with a strictly positive $\kappa_{ts}^n$ that satisfies $\pi_s(\kappa_{ts}^n) = \pi_s(\kappa_{sc})$ are more interesting. Figure 4 illustrates a numerical example of those two scenarios. The parameters are: $c = 0.6$, $\xi \sim \text{Uniform}[0, 1]$. The dotted line shows the supplier’s profit when only trade credit can be used.

In the second case ($w = 0.8$), when $\pi_s(\kappa_u) > \pi_s(\kappa_{sc})$, $\kappa_{ts}^n \in (0, \kappa_u)$. Depending on $K$, the supplier offers trade credit in three stages. When $K$ is large, the retailer enjoys sufficient trade credit. With medium $K$, the line of trade credit $\bar{L}_s^*$ is reached, as shown by the flat top in trade credit usage. Yet as it still allows the retailer to order more than it would be able to without trade credit, the retailer continues using only trade credit. For small $K$, however, if trade credit is still offered, the retailer starts to use the bank loan. In response, the supplier cuts off trade credit completely.

The third scenario ($w = 0.9$) shows the case with $\kappa_{ts}^n \in (\kappa_u, \kappa_{sc})$. In this scenario, the supplier’s profit margin is so low that he cuts off trade credit even before the credit limit is reached; that is, the retailer with low financing need can enjoy unlimited trade credit, yet the retailer with very limited capital can only borrow a bank loan.

\(^{13}\) The superscript $ts$ is short for “trade credit is senior”.
3.1.4. The Bank Loan is Senior to Trade Credit

In this section, we study another case, that is, when the bank loan is strictly senior to trade credit, as Figure 1 panel (b) shows. Similar to the previous case, when the retailer faces a fixed line of trade credit, two scenarios are possible. First, with low financing need, the retailer only uses trade credit. Second, the retailer exhausts all trade credit provided to him and then uses the bank loan. In the second scenario, the retailer’s financing constraint is:

\[ \int_0^{L_b} F(\xi) d\xi = B = wx - K_0 - \bar{L}_s, \]

which leads to \( \frac{\partial x}{\partial L_b} = \frac{\bar{F}(L_b)}{w} \); hence, the first order condition corresponding to his optimal decision is:

\[ wF(L_t) = F(x)\bar{F}(L_b), \quad (6) \]

where \( L_t = L_b + \bar{L}_s \). Combining the first order condition with the above financing constraint, the following result holds.

**Lemma 3.** \( L_b \) decreases in \( \bar{L}_s \); \( x \) increases in \( \bar{L}_s \) when \( h(L_b) > h(L_t)F(L_b) \).

To find the optimal solution, rewrite the problem as:

\[
\begin{align*}
\max & \int_0^{L_t} \bar{F}(\xi)d\xi + K - cx \\
\text{s.t.} \quad & L_t - \int_0^{L_b} F(\xi)d\xi - wx + K = 0; \\
& w\bar{F}(L_t) - F(x)\bar{F}(L_b) = 0;
\end{align*}
\]

Solving this optimization problem with \( (L_t, L_b, x) \) as decision variables, we obtain the following proposition.
Proposition 3. Compared with the scenario when trade credit is senior to the bank loan, if trade credit is junior to the bank loan:

1. for $K > \kappa_{ts}^n$, $\bar{L}_{ts}^{-} \leq L_{t} \leq \bar{L}_{ts}^{+}$, and the equality only holds when $L_{b} = 0$; \footnote{The superscript $tj$ is short for “trade credit is junior.”}
2. $\kappa_{n}^{tj} \leq \kappa_{n}^{ts}$, and the equality holds only when $\kappa_{n}^{ts} = 0$; and
3. $\pi_{tj}^{s} \geq \pi_{ts}^{s}$, and the equality holds only when $L_{tj}^{s} = 0$ or $L_{b} = 0$.

The above results are illustrated by the following numerical example. The parameters are: $c = 0.6$, $w = 0.8$, $\xi \sim \text{Uniform}[0, 1]$.

As Figure 5 shows, when trade credit is junior to the bank loan, the supplier extends credit to more retailers. However, some retailers, specifically, those with $K$ such that $K \in (\kappa_{n}^{ts}, \kappa_{n}^{tj})$, receive less trade credit and use a bank loan to complement trade credit and satisfy their financing need. For retailers in this interval, the total financing they use is less than under the case when trade credit is senior, and the chain efficiency is also lower.

At last, Figure 6 shows that when trade credit is junior, the supplier makes more profit than in the case when trade credit is senior. This result is intuitive, as all solutions under the scheme with senior trade credit are feasible in this case. On the retailer side, for those retailers that receive $\bar{L}_{s}^{s}$ when trade credit is senior, they make less profit under the case when trade credit is junior; retailers who do not receive trade credit when trade credit is senior make more profits when trade credit is junior.

Relating this result to the recent change in the bankruptcy law, that is, the introduction of Section 503(b)(9), we find that when trade credit is junior, the supplier is universally better off.
However, trade credit with high priority has its own advantages. First, for a certain interval, that is, for $K \in (\kappa_{n}^{ts}, \kappa_{u}^{ts})$, trade credit with high priority can improve chain efficiency. Second, trade credit with high priority is easy to implement as it only requires knowing whether the retailer’s cash position is above or below the threshold $\kappa_{n}^{ts}$, while the line of trade credit when trade credit is junior depends on the exact value of $K$.

3.1.5. The Bank Loan and Trade Credit Have Equal Seniority

The last scenario we study in this setting with an exogenous wholesale price is when both the bank loan and trade credit have the same priority, as Figure 1 panel (c) shows. Under this scheme, for the bank loan to break even, the retailer’s financing constraint is:

$$\frac{L_b}{L_t} \int_{0}^{L_t} \bar{F}(\xi) d\xi = B = wx - K_0 - L_s. \quad (7)$$

This constraint leads to the retailer’s first order condition satisfying:

$$w \bar{F}(L_t) = \bar{F}(x) \left[ \frac{\bar{L}_s}{(L_t)^2} \int_{0}^{L_t} \bar{F}(\xi) d\xi + \frac{L_b}{L_t} \bar{F}(L_t) \right]. \quad (8)$$

While it is difficult to obtain structural results in this scheme, Figures 5 and 6 show how trade credit usage and parties’ profits are under it. Interestingly, when trade credit and the bank loan have the same seniority, the supplier is willing to extend the least amount of trade credit among these three scenarios, and the chain efficiency and profits of different parties are the lowest as well.

3.2. The Optimal Trade Credit Contract

After focusing on the supplier’s decision on trade credit limit, we relax the constraint on the exogenous wholesale price and study the optimal trade credit when the supplier has the freedom to choose both the wholesale (trade credit) price $w$ and the line of credit $\bar{L}_s$. 
Interestingly, as the main result of this section shows, when a bank loan and trade credit are both available, it is optimal for the supplier to offer unlimited trade credit with net terms, provided that the retailer’s capital is non-negative.

**Proposition 4.** When $K \geq 0$, regardless of priority rules, the supplier offers unlimited trade credit with net terms. The retailer only uses trade credit.

While this is a strong result, it is intuitive: first, compared with other priority rules, for each fixed $w$, trade credit being a low priority claim allows the supplier to earn the highest profit. Second, when $w$ is in the supplier’s control, the maximum line of trade credit she is willing to extend is $\bar{F}^{-1}(c)$, which corresponds to charging $w = 1$ and offering trade credit that is sufficient for a retailer with $K = 0$.

4. **Trade Credit and Priority in the Presence of Multiple Risks**

The last section concludes with a strong result that when the retailer faces only demand risk and the supplier has control of the wholesale price, priority rules become irrelevant as the supplier is willing to offer unlimited trade credit under net terms; hence, she becomes the sole creditor of the retailer.

Reality, however, is more complicated. Often, the retailer faces multiple risks. For example, in addition to procuring inventory from the supplier, he may invest in other risky projects. Alternatively, he may face other uncertainties such as lawsuits and marketing campaigns. To model this situation, we introduce another risk embedded in an additional (generic) project. Intuitively, this change impacts trade credit usage from two aspects.

First, as this project requires a positive initial investment, and normally suppliers do not lend cash,\textsuperscript{15} trade credit provided by the supplier is insufficient for both inventory and the initial investment of the generic project. In this case, unlike the scenario where the supplier is the sole creditor, the retailer may have to borrow from more than one creditor, and priorities become unavoidable.

Second, in essence, our theory rests on the fact that the realized demand is correlated with the retailer’s payback power; therefore, when the realized demand is low, the retailer does not have to pay back trade credit in full. This helps lower the retailer’s marginal cost, which induces a larger order from the supplier, and in turn improves the supplier’s profit. With another risk, however, the correlation between demand and total revenue is blurred; hence, it becomes unclear how our theory applies in this case.

\textsuperscript{15} For related literature, see Burkart and Ellingsen (2004) and references therein.
With these concerns, we use several examples to show that the existence of other risks brings in many new issues, inefficiencies, for example, if the supplier still applies the optimal trade credit policy with only demand risk. Further, we examine differences in the optimal trade credit policy and the chain efficiency when different priorities apply. Following that, we show some basic theoretical results.

4.1. The Risk of Having Another Risk

Figure 7 shows the modified sequence of events. In this model, we assume the newly added generic project is a long-term project relative to inventory decision; that is, before the supply contract starts, the retailer makes a decision whether to invest in the generic project or not. If the answer is yes, he borrows a bank loan to finance this project. This project is paid off at the same time demand is realized.

Following this decision, the supplier proposes a supply contract. Further, we assume all parties may receive a noisy signal of the generic project’s payoff before the supplier proposes her contract. Two extreme cases are: first, the signal contains only noise; second, the signal is perfect; that is, the exact payoff of this project is revealed.

For simplicity, in the following illustrative scenarios, we assume both the investment payoff and demand follow a binary distribution. The joint distribution and associated payoffs are summarized in Table 1. In this table, $\sigma$ represents the state of the world, where $l$ is short for low, and $h$ is short for high.

As the table illustrates, the correlation between the generic project payoff and demand is reflected in $\epsilon \in [-1/4, 1/4]$. They are positively correlated/negatively correlated/independent if and only if $\epsilon$ is strictly greater than/strictly less than/equal to 0. When $\epsilon = 1/4$, demand and the generic project are perfectly positively correlated; when $\epsilon = -1/4$, they are perfectly negatively correlated.
Table 1  Joint Distribution of Investment Payoff and Demand

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Probability</th>
<th>Demand</th>
<th>Investment Payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$hh$</td>
<td>$1/4 + \epsilon$</td>
<td>2</td>
<td>$V$</td>
</tr>
<tr>
<td>$hl$</td>
<td>$1/4 - \epsilon$</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>$lh$</td>
<td>$1/4 - \epsilon$</td>
<td>1</td>
<td>$V$</td>
</tr>
<tr>
<td>$ll$</td>
<td>$1/4 + \epsilon$</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Note that given these payoffs, the generic project is always risky; that is, when the low state happens, the investment cannot be (fully) recovered. Further, although we show most examples assuming an exogenous wholesale price, an endogenous wholesale price does not change the picture.

4.1.1. Payoff Correlation As discussed above, one concern about having another risk is that it will blur the correlation between the realized demand and the retailer’s total payoff. To illustrate this concern, consider the following example.

**Example 1. (Payoff Correlation)** Let $K = 0$, $V = 1$, $I = 0.4$, $c = 0.4$, $w = 0.9$. Examining both investments independently, as the generic project leads to a positive NPV $(0.5(0) + 0.5V - I = 0.1)$, it is worth investing in. For the inventory decision, it is easy to show the supplier offers unlimited trade credit and the retailer orders two units. The corresponding profits are: $\pi_s = 0.5(1) + 0.5(1.8) - 2(0.4) = 0.6$, and $\pi_r = 0.1$. When not offering trade credit, the supplier’s profit is $\pi_s = w - c = 0.5$.

Now consider the case when both investment decisions are made jointly. Suppose trade credit is junior and the supplier still offers unlimited trade credit. From the retailer’s perspective, if he orders one unit, his profit $\pi_r(1) = (0.25 + \epsilon)(0.7 + 0) + (0.25 - \epsilon)(0.7 + 0) = 0.35$, and the supplier’s profit $\pi_s(1) = 0.35$. Note that in this case, both parties’ profits are independent of the correlation $\epsilon$.

If the retailer orders two units, his profit $\pi_r(2) = (0.25 + \epsilon)(0.8) = 0.2 + 0.8\epsilon$, and the supplier’s profit $\pi_s(2) = (0.25 + \epsilon)(0.8) + (0.25 - \epsilon)(0.8 + 0.8) = 0.6 - 0.8\epsilon$. Note that when $\epsilon$ increases, the supplier’s profit decreases, while the retailer’s profit increases. This is intuitive. As a creditor, the supplier prefers a smoother payoff, while the equity-holder prefers volatility.

However, for the supplier to extract more profits, the retailer’s incentive constraint needs to be satisfied; that is, $\pi_r(2) > \pi_r(1) = 0.35$. However, in this case, $\pi_s(2) < 0.8 - 0.35 = 0.45$, which is smaller than what the supplier makes when she offers no trade credit.

Interestingly, as shown in the above example, by adding another project to this picture, the retailer essentially obtains a free ride at the supplier’s expense by taking both the trade credit offered by the latter and a cheaper bank loan–making a greater profit in the process. In response, the supplier does not offer trade credit at all. It is easy to show that when trade credit has high
priority, the chain behaves exactly as when no trade credit is offered, as the retailer has no incentive
to order two units and trade credit is riskless.

4.1.2. It is Not My Fault! Another interesting case to consider is when the demand risk is
not the major risk the retailer faces.

Example 2. (It’s Not My Fault!) Let $V = 10$, $I = 4$, $\epsilon = 0$, and $K = 0$. Clearly, in this case,
when trade credit is junior, the default of either claim is completely independent of the demand
realization; that is, when the generic investment’s payoff is high, both the bank loan and trade credit
are fully paid off. Otherwise, the supplier receives nothing even though the demand realization is
high.

In this setting, consider two other actions: first, the supplier does not offer any trade credit.
If so, the retailer orders one unit and finances inventory and the project using a bank loan with
$B = 4 + 0.9 = 4.9$ and $L_b$ satisfies: $0.5L_b + 0.5(1) = 4.9$, that is, $L_b = 8.8$; the retailer’s profit $\pi_r = 
0.5(11 - 8.8) = 1.1$, and the supplier’s profit $\pi_s = 0.9 - 0.3 = 0.6$.

Second, when trade credit is assigned with high priority, the retailer still orders one unit. Trade
credit in this case is riskless; so, $\pi_r$ and $\pi_s$ are the same as when no trade credit is offered.

4.1.3. Over-Investing So far, all examples show situations that, when considered separately,
the generic project is worth investing in, that is, it has a positive net present value (NPV). However,
an interesting question arises: will anything change when the risk of the generic project and the
demand risk mingle together?

Example 3. (Over-Investing) Let $K = 0$, $\epsilon = 0$, $V = 1$ and $I = 0.7$. Apparently, when evaluated
independently, the retailer should not invest in this project, as the NPV of this project is $-0.2$.
However, suppose the supplier offers the same contract under the base case ($w = 0.9$, unlimited
credit). In this case, it is easy to show that the retailer’s optimal strategy is to invest in this project,
orders one unit from the the supplier, and makes a total profit $0.2$.\textsuperscript{16} The supplier, however, only
makes a profit of $0.3$. Therefore, under this circumstance, the supplier’s optimal strategy is not to
offer trade credit at all.

4.1.4. Debt Overhang As we mention in the model setting, as demand risk is normally
short term relative to the generic project, the payoff of the latter may be revealed (noisily) before
the supplier proposes the supply contract. Obtaining information about the project’s status, the
supplier can consequently incorporate this information into credit extension decisions. For example,

\textsuperscript{16} When the retailer does not invest in this project, he only makes $0.1$; when he invests in the project and orders two
units from the supplier, he makes $0.125$. 
when the supplier finds the company is in bad shape, she may refuse to extend trade credit if she can only hold a junior claim.

In finance, a closely related phenomenon is debt overhang, which has been a popular research topic since Myers (1977). Debt overhang means that highly levered companies may reject valuable investment opportunities because those opportunities may not benefit the equity holder. In a recent paper, He and Diamond (2010) extend the classical problem by introducing term structure. Similarly, in the following example, we show how the interaction of long-term debt (used for a long-term investment) and short-term debt (trade credit) influences trade credit usage and the retailer’s investment decisions.

**Example 4.** Consider the sequence of events as Figure 8 shows. At time 0, the retailer has the opportunity to invest in a project with a high value $V = 1$ and low value 0. The project costs him $I = 0.8$, which he can finance using a bank loan. Demand and the project payoff are perfectly positively correlated. At time 1, he receives a signal of the payoff of the project. With 90% probability, the signal is good; with 10% probability, the signal is bad. The signal is observable to every party, and it is 90% accurate; that is, when the signal is good, with 90% probability, the demand and project payoff is high, with 10% low. When the signal is bad, with 90% probability, the demand and project payoff is low, and with 10% high. For demand, $c = 0.4$, $w = 0.9$. The retailer’s capital $K = 0$.

**Figure 8  Payoffs with A Noisy Signal**

- **Generic project**
  - **Inventory decision**
  - **Realization**

First, consider the case when trade credit has low priority. Clearly, when the signal is bad, the supplier should not offer any trade credit.\(^\text{17}\) Anticipating this situation, however, the retailer should obtain a bank loan at time 0 with $B = I + w = 1.7$ and $L_b$ that satisfies: $18\%(1) + 82\%(L_b) = 1.7$, that is, $L_b = 1.85$. In addition, when the signal is good, it is optimal for the supplier to

---

\(^{17}\) When offered with unlimited trade credit, the retailer’s optimal strategy is to order two units, with a conditional profit of 0.04, leaving the supplier with a loss of $-0.26$. If extended with a credit limit of 0.9, the retailer orders one unit with a conditional profit of 0.03, leaving the supplier with a loss of $-0.04$.\)
extend trade credit despite its low priority. Conditional on the good signal, the supplier’s profit is $90\%(0.9) + 0.9 - 2(0.4) = 0.91$, and the retailer’s profit is $90\%(0.25) = 0.225$. Therefore, the unconditional profit for the supplier is: $90\%(0.91) + 10\%(0.5) = 0.869$, and the retailer’s profit is: $90\%(0.225) + 1\%(0.15) = 0.204$.

On the other hand, if trade credit has high priority, it is easy to show that the supplier’s profit is 0.5, as she can only sell one unit regardless of the signal. The retailer’s profit is $0.02 + 0.1 = 0.12$, which is dominated by the case when trade credit has low priority.

This example shows another advantage of trade credit as the low priority claim. It allows the supplier to use trade credit as a risk-sharing mechanism only when facing a good signal. In response, the retailer should obtain a larger bank loan and carry cash in the first period.

### 4.2. Priorities and Efficiency

As all of the above examples show, even in the presence of other risks, granting low priority to trade credit does not create any efficiency loss for the chain. It is true that under certain circumstances, debt overhang, for example, the supplier does not extend trade credit unless trade credit has high priority. However, in such cases, in anticipation of the supplier’s decision, the retailer could increase his cash holding by borrowing a larger amount from the bank. This result is formalized as the following proposition.

**Proposition 5.** *When the retailer has multiple creditors, compared with the case when trade credit has high priority, trade credit with low priority improves chain efficiency and the supplier’s profit.*

The intuition behind this result is that in any case that the supplier is unwilling to offer trade credit due to low priority, the retailer could also compensate for this by obtaining a larger bank loan, which would not change the retailer’s incentive and moreover would improve the supplier’s profit. Under debt overhang, this results implies that the retailer holds extra cash during the first period. As the marginal cost the retailer faces is determined by the claim with low priority, trade credit with high priority will not change the retailer’s investment decision.

### 5. Empirical Result: Aggregated Patterns

As mentioned in Section 1, one important motivation of this paper is the recently enacted BAPCPA, which strengthens the right of reclamation for trade creditors. While the intention of this act is to protect suppliers, it probably does not take into consideration of the risk-sharing effect of trade credit.
To shed some light on the extent to which practice agrees with our theory, we conduct a preliminary empirical test that looks at how relative trade credit usage (measure as trade payable divided by the sum of trade payable and debt in current liabilities) responds to the new statutes governing reclamation rights. On the one hand, high priority allows the supplier to recover more if the debtor enters bankruptcy. On the other hand, it reduces the risk-sharing incentive, hence the retailer should lower his order quantity, leaving the supplier with less profit. Quantitative analysis in this paper shows that in many circumstances, the latter incentive dominates the former one, and when trade credit has high priority, the retailer will use relatively less trade credit, but more bank debt.

We use the same data as in Yang and Birge (2010). The data is drawn from Compustat, including US retailers from 1999 – 2008. However, instead of looking at firm-level data, we use yearly aggregated data. That is,

\[ Y_t = \sum_i \left( \frac{\text{Trade Payable}}{\text{Trade Payable} + \text{Debt in Current Liability}} \right)_{it}, \]

where \( t \) represents year and \( i \) for individual firms. We test the following specification:

\[ Y_t = \alpha + \beta_1 D_1 + \beta_2 D_2 + \beta_3 X_t + \epsilon_t, \]

where \( D_1 \) and \( D_2 \) are dummies for two time periods: 1999 - 2002, and 2006 - 2008. Obviously, the second period is when BAPCPA governs reclamation rights. The first period, however, corresponds to when trade creditors were often granted high priority by the judge.\(^\text{18}\) \( X \) is time-specific control variables, including annual GDP growth and aggregated trade receivables normalized by assets \( (\sum_i (\text{Trade Receivables}/\text{Asset})_{it}) \). Intuitively, when GDP growth is high, trade creditors may have more incentive to extend trade credit, hence increasing relative usage of trade credit. Receivables, however, may influence trade accounts payable in both directions. On the one hand, if a firm wants to extend more trade credit, it may demand more trade credit from its supplier. On the other hand, if a firm has limited internal capital, it may simultaneously demand more trade credit, but extend less.

Table 2 presents our basic results. The coefficients on the dummies show that the statute change has a statistically significant impact on relative usage of trade credit. Quantitatively, when trade credit has high priority, on an aggregate level, firms reduce their relative usage of trade credit by approximately 1.4% to 4%, depending on different specifications.

\(^\text{18}\) We thank Professor Douglas Baird of the University of Chicago Law School for his help with related background.
### Table 2  Relative Trade Credit Usage in Aggregate

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999-2002</td>
<td>-0.0397*</td>
<td>-0.0215*</td>
</tr>
<tr>
<td></td>
<td>(0.0089)</td>
<td>(0.0080)</td>
</tr>
<tr>
<td>2006-2008</td>
<td>-0.0188*</td>
<td>-0.0147*</td>
</tr>
<tr>
<td></td>
<td>(0.0095)</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>GDP Growth</td>
<td>-0.0005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0023)</td>
<td></td>
</tr>
<tr>
<td>Receivable</td>
<td>-2.4319*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.7498)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.820*</td>
<td>1.026*</td>
</tr>
<tr>
<td></td>
<td>(0.0067)</td>
<td>(0.0656)</td>
</tr>
<tr>
<td>Observations</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.67</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Notes: The standard errors are shown in the parenthesis. * represent coefficients significant at 5%.

### 6. Conclusion and Future Research

This paper addresses two important questions: How should the supplier, when extending trade credit, interact with other creditors? Under different conditions, what priority rules are most efficient? Under our models, we find that assigning high priority to trade creditors would not lead to any increase in the supplier’s profit in any case. Further, it would not lead to any efficiency gain for the supply chain in most cases. Empirical results in aggregated level are consistent with the prediction based on our theory.

However, our model is not without limitation. For example, we assume no distress costs. This assumption is justified when we focus on trade credit’s role in risk-sharing. However, distress costs may be related to priority arrangements. Additionally, it is worth noting that our theory is just one of the many theories that may explain trade credit. Even though we show that granting trade credit with low priority is generally more efficient through its purpose in risk-sharing, we could not conclude that for other purposes that trade credit serves, higher priority would not benefit the supply chain. Finally, we only test the impact of BAPCPA on aggregate level data. Further empirical analysis on firm-level data would also be promising.

### References


Appendix. Proofs

A. Proofs of Section 3

**Lemma 4.** There exists a unique order quantity $x$ and the corresponding $L_s$ that satisfy $\bar{F}(x) = w\bar{F}(L_s)$ and $wx = L_s + K$.

*Proof of Lemma 4.* Substitute $L_s$ for $wx - K$. Define function $G(x) = \bar{F}(x) - w\bar{F}(wx - K)$. Obviously, as $\bar{F}(\frac{K}{w}) > w$ for $K < \kappa^*c$, we have $G(\frac{K}{w}) > 0$; on the other hand, $G(x^{max}) < 0$ where $x^{max}$ is the upper bound of the support of $\xi$. Moreover, $G'(x) = -f(x) + w^2f(wx - K) = -\bar{F}(x)(h(x) - wh(wx - K)) < 0$, where the second equality holds as $w = \frac{\bar{F}(x)}{\bar{F}(L_s)}$. Therefore, there exists a unique $x > \frac{K}{w}$ such that $G(x) = 0$.

*Proof of Lemma 1.* First, write out $L_s$ as a function of $K$ using the Implicit Function Theorem:

$$\frac{\partial L_s}{\partial K} = \frac{f(x)}{f(x) - w^2f(L_s)}; \tag{9}$$

Clearly, as $x > wx \geq L_s$, $h(x) > h(L_s)$. From the retailer’s first order condition $\bar{F}(x) = w\bar{F}(L_s)$, we have $f(x) > w\bar{F}(L_s) > w^2f(L_s)$. Therefore, $\frac{\partial L_s}{\partial K} < -1$.

Also, $\frac{\partial x}{\partial K} = w^{-1}(-\frac{\partial L_s}{\partial K} + 1) < 0$.

For the retailer’s profit $\pi_r = \int_{L_s}^{x^*} \bar{F}(\xi)d\xi - K$, write out the partial derivative of $\pi_r$ with respect to $K$:

$$\frac{\partial \pi_r}{\partial K} = \bar{F}(x)\frac{\partial x}{\partial K} - \bar{F}(L_s)\frac{\partial L_s}{\partial K} - 1;$$

$$= \bar{F}(L_s)(w - \frac{\partial L_s}{\partial K}) - 1;$$

$$= \bar{F}(L_s) - 1 < 0,$$

as desired.

*Proof of Proposition 1.* The line of trade credit $\bar{L}^*_s = \bar{F}^{-1}(\frac{c}{w})$ is shown in the previous analysis.

For the existence of $\kappa_s$, define $G(K) = \bar{F}(\frac{K + L_s}{w}) - w\bar{F}(\bar{L}_s)$. Obviously, $G(K)$ is a continuously decreasing function of $K$. As $G(\kappa_s) < 0$, a unique $\kappa_s$ in between exists if $G(0) \geq 0$; otherwise, define $\kappa_s = 0$.

When $K \geq \kappa_s$, by contradiction, suppose the retailer uses $\bar{L}^*_s$, that is, all trade credit provided to him, his marginal revenue will be $\bar{F}(\frac{K + L_s}{w})$, while his marginal cost is $w\bar{F}(\bar{L}_s)$; therefore, he is better off to order less, which means the line of credit is not binding. Therefore, all results from Lemma 1 hold.

In terms of the supplier’s profit, $\pi_s = \int_0^{L_s} \bar{F}(\xi)d\xi - cx + K_0$, its partial derivative of $\pi_s$ with respect to $K_0$ follows:

$$\frac{\partial \pi_s}{\partial K} = (1 - \frac{c}{w}) + (\bar{F}(L_s) - \frac{c}{w})\frac{\partial L_s}{\partial K}. \tag{10}$$

Obviously, when $L_s \geq \bar{F}^{-1}(\frac{c}{w})$, $\pi_s$ increases in $K_0$. On the other hand, when $L_s = 0$, $\frac{\partial \pi_s}{\partial K} < 0$, and because of the continuity of this function, the partial derivative stays negative when $L_s$ is small.

For concavity, consider the second-order derivative,

$$\frac{\partial^2 \pi_s}{\partial K^2} = -f(L_s)(\frac{\partial L_s}{\partial K})^2 + (\bar{F}(L_s) - \frac{c}{w})\frac{\partial^2 L_s}{\partial K^2}.$$ 

A sufficient condition for $\pi_s$ to be concave in $K$ is $\frac{\partial^2 \pi_s}{\partial K^2} \leq 0$. Rewriting $\frac{\partial \pi_s}{\partial K} = (1 - w\frac{h(L_s)}{\bar{F}(L_s)})^{-1}$ leads to:

$$\frac{\partial^2 L_s}{\partial K^2} = w(\frac{\partial L_s}{\partial K})^2 h(L_s)\frac{h'(L_s)}{h(x)} - \frac{h'(x)}{h(x)} \left( w + \frac{\partial L_s}{\partial K} \right);$$
Clearly, as $h'(x) \geq 0$, $\frac{\partial^2 L}{\partial K^2} \leq 0$ holds if $ln(h(x))$ is concave.

In the second scenario, for $K \leq \kappa_u$, the retailer exhausts all trade credit provided to him, orders $x = \frac{K + L_c}{w}$, which increases in $K$. The supplier’s profit $\pi_s = \int_0^{L_c} \tilde{F}(x) d\xi + K - cx$, $\frac{\partial \pi_s}{\partial K} = 1 - c\frac{\partial x}{\partial K} > 0$. For the retailer’s profit $\pi_r$, $\frac{\partial \pi_r}{\partial K} = \tilde{F}(x) - 1$. Clearly, as $x$ increases in $K$, $\pi_r$ is concave on $K$. When $K = \kappa_u$,

$$\frac{\partial \pi_r}{\partial K} = \tilde{F}(x) - 1 = \frac{w\tilde{F}(L_c)}{w} - 1 < 0.$$ 

On the other hand, when $K = 0$,

$$\frac{\partial \pi_r}{\partial K} = \tilde{F}(x) - 1 = \frac{\tilde{F}(L_c)}{w} - 1 = \frac{c}{w^2} - 1.$$ 

Obviously, when $w^2 > c$, $\frac{\partial x}{\partial K} < 0$ at $K = 0$; hence, $\pi_r$ decreases monotonically with $K$; otherwise, $\pi_r$ first increases, and then decreases.

Proof of Proposition 2. If $\exists K < \kappa_{sc}$ satisfying $\pi_s(K) = \pi_s(\kappa_{sc})$, define $\kappa_n = K$. If it does not exist, define $\kappa_n = 0$.

The shape of $\pi_s(K)$ dictates that there exists at most one $\kappa_n$ satisfying the above condition, and it can be determined by comparing the supplier’s profit when $K = 0$ and when $K = \kappa_n$. When $\pi_s(\kappa_{sc}) > \pi_s(\kappa_u)$, $\kappa_n \in (\kappa_u, \kappa_{sc})$; when $\pi_s(\kappa_{sc}) \in (\pi_s(0), \pi_s(\kappa_u))$, $\kappa_n \in (0, \kappa_u)$; when $\pi_s(\kappa_{sc}) < \pi_s(0)$, $\kappa_n = 0$.

From the supplier’s perspective, by this construction, when $K \in (0, \kappa_n)$, $\pi_s(K) < \pi_s(\kappa_u)$; hence, the supplier does not provide trade credit.

On the retailer side, as $x > \tilde{F}^{-1}(w)$ for $K \in (\kappa_n, \kappa_{sc})$; in this regime, $\frac{\partial x}{\partial K} < 0$, that is, $\pi_r > \pi_r(\kappa_{sc})$. Therefore, the retailer prefers only using trade credit.

Proof of Lemma 3. Define functions $G_1 = \int_0^{L_s} \tilde{F}(x) d\xi + L_s - wx + K$, and $G_2 = w\tilde{F}(L_c) - \tilde{F}(x)\tilde{F}(L_b)$. Partial derivatives can be derived using the Implicit Function Theorem: $\frac{\partial L_c}{\partial x} = -\frac{\Delta_1}{\Delta_2}$, and $\frac{\partial L_f}{\partial x} = -\frac{\Delta_1}{\Delta_3}$, where $\Delta_1, \Delta_2$, and $\Delta_3$ satisfies:

$$\Delta_1 = \frac{\partial G_1}{\partial x} \frac{\partial G_2}{\partial L_c} \frac{\partial G_1}{\partial L_b} \frac{\partial G_2}{\partial x} = w^2 f(L_c) - w\tilde{F}(x)f(L_b) - f(x)(\tilde{F}(L_b))^2$$ 

$$= w\tilde{F}(L_c)\tilde{F}(x)[h(L_c) - h(L_b)] - f(x)(\tilde{F}(L_b))^2$$ 

$$= -\tilde{F}(L_b)\tilde{F}(x)[\tilde{F}(L_b)h(x) - wh(L_b) + wh(L_b)] < 0,$$

where the last equation holds as $\tilde{F}(L_b) = \frac{\tilde{F}(L_b)}{\tilde{F}(x)}w > w$.

$$\Delta_2 = \frac{\partial G_1}{\partial x} \frac{\partial G_2}{\partial L_c} \frac{\partial G_1}{\partial L_b} \frac{\partial G_2}{\partial x} = w^2 f(L_c) - f(x)\tilde{F}(L_b)$$ 

$$= \tilde{F}(L_b)\tilde{F}(x)[wh(L_b) - h(x)] < 0,$$

and

$$\Delta_3 = \frac{\partial G_1}{\partial L_c} \frac{\partial G_2}{\partial L_b} \frac{\partial G_1}{\partial L_b} \frac{\partial G_2}{\partial L_c} = \frac{\partial G_1}{\partial L_c} \frac{\partial G_2}{\partial L_b} \frac{\partial G_1}{\partial L_b} \frac{\partial G_2}{\partial L_c}$$
\[ f(L_b) \bar{F}(x) - w f(L_t) F(L_b) = \bar{F}(x) F(L_b) [h(L_b) - h(L_t) F(L_b)]. \]

Therefore, \( \frac{\partial L_b}{\partial L_s} = -\frac{\Delta_1}{\Delta_2} < 0 \); and \( \frac{\partial x}{\partial L_s} = -\frac{\Delta_1}{\Delta_2} < 0 \) if \( h(L_b) - h(L_t) F(L_b) > 0 \).

**Proof of Proposition 3.** First, to show the optimal line of trade credit, consider \( L_t \) as a function of \( x \), using the Implicit Function Theorem,

\[
\frac{\partial L_t}{\partial x} = \frac{h(L_b) - F(L_b) \frac{h(x)}{F(L_b)w}}{h(L_b) - F(L_b) h(L_t)}.
\]

As the first order condition is \( \frac{\partial L_t}{\partial x} \bar{F}(L_t) = c \), the following result follows:

\[
h(L_b)[w \bar{F}(L_t) - c] = F(L_b)[h(x) - ch(L_t)].
\]

Obviously, the right hand side is strictly positive when \( L_b > 0 \); hence, \( w \bar{F}(L_t) > c \), that is; \( \bar{L}_s^t < L_t < \bar{L}_s^* \).

To show the existence of \( \kappa_s^t \), consider two scenarios. First, when \( \kappa_s^t = 0 \), it is easy to show that \( \kappa_s^t = 0 \).

Second, when \( \kappa_s^t > 0 \), we can show that for \( K = \kappa_s^t \) offering a little trade credit improves the supplier’s profit.

At last, to show that \( \pi_s^t \geq \pi_s^e \), note that the optimal trade credit policy when trade credit is senior is also a feasible solution to the supplier when trade credit is junior.

**Proof of Proposition 4.** As shown in Yang and Birge (2010), without distress costs, when the bank loan is strictly senior to trade credit, it is optimal for the supplier to offer trade credit under net terms. Further, a line of trade credit does not improve the supplier’s profit.

In terms of the optimal trade credit policy under other priorities, as Proposition 3 shows, with a fixed \( w \), the supplier makes more profit when trade credit is junior. Therefore, there is no contract under other priorities that make the supplier better off.
Inventory and Capital Structure Decisions Under Bankruptcy Risk: A One Period Model

Yasin Alan, Vishal Gaur*

May 17, 2011

Abstract

We investigate the inventory stocking and capital structure decisions of a firm in the presence of an asset-based credit limit and costly bankruptcy. The key aspect of the paper is that we model the simultaneous decisions of an equity investor, the manager of the firm, and a bank. The investor decides how much to invest in the equity of the firm. The firm, a newsvendor, takes the investor’s decision as starting equity and decides its borrowing amount and order quantity in order to maximize the total return to the investor. The bank has partial information about the firm’s demand, and sets an asset based credit limit in order to prevent the firm from over-borrowing. Our model builds on the literature in operations and corporate finance to determine the order quantity, capital structure, and probability of bankruptcy of the firm that would be realized at equilibrium in the marketplace. In particular, it shows how the probability of bankruptcy depends on the parameters of the newsvendor model and on information asymmetry between the firm and the bank. It also shows the role of the credit limit in determining the optimal capital structure regardless of the presence of taxation and bankruptcy costs.

Extended Abstract

Bankruptcy occurs when a firm is unable to meet its obligations to its creditors. In general, the risk of bankruptcy of a firm plays a major role in shaping its interactions with its creditors and investors. Classical operations management models overlook the risk of bankruptcy assuming that the firm has sufficient working capital to finance its operations. These models can mimic operational decisions of large corporations with sufficient cash and debt capacity. However, most small and medium sized businesses face liquidity constraints and bankruptcy risk.

Due to the presence of bankruptcy risk, banks commonly use asset based financing (ABF) to lend money to firms. In ABF, the borrowing amount of a firm is linked to its current assets, including inventory, cash, and accounts receivables. ABF is useful to lenders because it allows them

*Johnson Graduate School of Management, Cornell University, Sage Hall, Ithaca, NY 14853-6201, E-mail: ya47@cornell.edu, vg77@cornell.edu
to control their exposure to the firm’s bankruptcy risk by imposing an upper limit on the amount of debt on the firm’s balance sheet. ABF is also useful to borrowers with insufficient liquidity because it allows them to obtain competitive interest rates by securing their current assets.

Our paper studies the implications of ABF and costly bankruptcy on the inventory stocking and capital structure decisions of a firm. We use a game theoretic model with three players - a newsvendor firm, an investor and a bank. The investor is an expected value maximizer and allocates her funds between the newsvendor and an alternative investment opportunity. Taking the investor’s decision as starting equity, the manager of the newsvendor decides how much to borrow from the bank and how much quantity to stock in order to maximize the total return to the investor. The bank lends money to the firm with partial information about its demand. It observes the firm’s equity and sets an asset based credit limit in order to prevent the firm from over-borrowing. Through this model, we address two main questions: (i) What are the stocking quantity and capital structure outcomes of the newsvendor at equilibrium? (ii) What is her resulting probability of bankruptcy, and how does it depend on the stocking quantity, the borrowing interest rate, the demand distribution, and the price parameters of the newsvendor model? Without market frictions, stocking and capital structure decisions would be independent of each other (Modigliani and Miller 1958) and unaffected by bankruptcy risk (Stiglitz 1969). Our model includes frictions such as bankruptcy cost, corporate taxes, and information asymmetry between the firm and the bank. We use the newsvendor model as a prototypical representation of single-period capacity-constrained operational decisions of a firm.

Our paper builds on research articles on the interface of operations and corporate finance, which focus on different aspects of joint operational and financial decision-making. In particular, we extend the model of Buzacott and Zhang (2004) by making the equity investment a decision variable, rather than taking it as given, and adding practically relevant features such as taxation, information asymmetry, and bankruptcy costs. The investor’s role is important because the order quantity and borrowing amounts are functions of equity and they determine the value that the newsvendor is able to provide to its shareholders. This begets the question as to how the value of the newsvendor varies with its equity and what level of equity maximizes the value to shareholders. Thus, our paper determines the bankruptcy probability and the amounts of equity, debt, and inventory that occur at equilibrium in the marketplace.

One result of our paper is that the probability of bankruptcy takes on only two values, either zero or a positive constant that is independent of the tax rate and the order quantity. This result
implies that for firms that face a positive bankruptcy probability, the credit limit is always binding. That is, it is never optimal for the investor to give the newsvendor an amount of equity that will be small enough to entail bankruptcy risk, but large enough to allow borrowing less than the credit limit. Moreover, it is optimal for the bank to adjust the credit limit in such a way that the bankruptcy probability remains constant even though inventory and capital structure may vary. This constant probability depends on the cost parameters of the newsvendor model, borrowing interest rate, and degree of information asymmetry between the bank and the newsvendor. In practice, banks set credit limits using simple rules of thumb predicated on historical salvage values of inventory in different industries. Our result shows how banks may set credit limits in a more sophisticated way to incorporate both the demand distribution and the equity of the newsvendor.

Whether the probability of bankruptcy is zero or non-zero at equilibrium depends on the rate of return on the alternative investment available to the investor. Thus, another result of our paper shows the existence and uniqueness of a threshold rate of return below which the bankruptcy probability is zero. In such situations, there can be borrowing without bankruptcy risk, or in extreme cases, the investor may choose to create a pure equity firm. Otherwise, the investor chooses to invest a relatively small amount in the newsvendor, and the newsvendor borrows. The remaining results of the paper show the capital structure of the newsvendor at equilibrium, and the implications of taxes or the absence thereof on the equilibrium outcome. For example, we find that borrowing without risk cannot arise as an equilibrium outcome when there is no taxation. This result captures the tax shield of debt. We also show that the investor’s tendency to create a pure equity newsvendor decreases as the tax rate increases. Furthermore, we show that the debt-to-equity ratio at equilibrium is decreasing in demand uncertainty, and is non-monotone in the profit margin. These results are consistent with the empirical corporate finance literature.

References


Supply Chain Performance under Market Valuation: An Operational Approach to Restore Efficiency

Guoming Lai
McCombs School of Business, University of Texas at Austin, Austin TX 78731, guoming.lai@mccombs.utexas.edu

Wenqiang Xiao
Stern School of Business, New York University, New York 10012, wxiao@stern.nyu.edu

Jun Yang
School of Management, Huazhong University of Science and Technology, Wuhan, Hubei, China, jun.yang@mail.hust.edu.cn

Based on a supply chain framework, we study the stocking decision of a downstream buyer who receives private demand information and has the incentive to influence her capital market valuation. We first characterize a market equilibrium under a general single contract offer. We show that the buyer’s stocking decision can be distorted from the first-best level in equilibrium. Such a downstream stocking distortion is always detrimental for the buyer but might benefit (or hurt) the supplier and the supply chain for some contract terms. We further reveal scenarios where full supply chain efficiency cannot be reached under any single contract offer. Then, focusing on contract design, we show a general condition under which a menu of buyback contracts can prevent downstream stocking distortion and restore full efficiency in the supply chain in equilibrium. Our study demonstrates that in a supply chain context, a firm’s incentive to take real economic activities to influence capital market valuation can potentially be resolved through operational means.

Key words: Supply Chain, Newsvendor, Capital Market Valuation

1. Introduction

Managers of firms make their operational decisions with the considerations of their long-term goals as well as their firm’s short-term valuation in the capital market. If the capital market is able to correctly assess a firm’s real status, the interest in the capital market valuation would be aligned with the interest in the long-term firm value. Nevertheless, a firm’s real status generally is not perfectly observable to the public. External investors would have to rely on financial reports, corporate announcements and other secondary information sources (e.g., analyst studies, market news) to value a firm. In the presence of information asymmetry, managers, with short-term interest in the firm’s market value, may take real economic activities (i.e., distortions of operations), the so-called ‘real earnings management’, to influence the market perception (Graham et al. 2005, Roychowdhury 2006). Such activities however affect a firm’s long-term value.
In this paper, we study a firm’s inventory stocking decision with a short-term interest in the capital market valuation. Although the stocking decision does not directly reveal the firm’s value, it conveys information of the managers’ vision about the future business. Investors may thus react to announcements and/or market news related to specific stocking decisions. For instance, the stock price of Gome Electrical Appliances Holding Ltd., China’s second-biggest electronics retailer, rose 13.7 percent on May 26, 2010 in Hong Kong trading, the most in 10 months, when it was revealed that Gome won a distribution contract with LG Electronics Inc. that targeted $1.4 billion (9.3 billion RMB) of sales of LG products through Gome’s retail stores in 2010, a 90 percent increase from 2009, and made Gome become the biggest distributor of LG in China (Longid 2010).

In contrast, when a Wall Street Journal article revealed on December 18, 2009 that Zale, the second-largest U.S. jewelry retailer at that time, refused to accept tens of millions of dollars of inventory at the end of November 2009, its stock price plunged as much as 12.7 percent on that day; investors showed their suspicion about Zale’s sales prospect as well as its financial capability (Wahba 2009).

With the anticipation of the market response, it is interesting to analyze whether managers, when they care about their firm’s short-term market value, may potentially distort their stocking decision and, if so, to explore some mechanism that can overcome such incentives. We study this problem based on a newsvendor and supply chain framework. In our model, a downstream buyer makes a stocking decision of a product supplied by an upstream supplier for a specific selling event. The buyer receives private information (either optimistic or pessimistic) of the demand which is unknown to the other parties. Different from the classical newsvendor problem, the buyer cares about not only the firm’s true profitability in the long term but also the firm’s short-term capital market value. The investors value the firm based on the firm’s stocking decision and other

---

1 Compared to Gome’s stock price surge, the benchmark Hang Seng Index rose 1.1 percent on the same day. Ashley Cheung, an analyst at BOCI Research Ltd. in Hong Kong commented on the news, “With this contract, Gome is going to see a material positive impact on its revenue” (Longid 2010).

2 Both the Dow Jones and S&P 500 indices closed up on December 18, 2009. According to a Reuters article (Wahba 2009), cancelation of orders was generally permitted for Zale’s contracts with the suppliers; however, “the cancelation of orders at a busy time of year is an ominous sign for Zale’s sales prospects.” Milton Pedraza, Chief Executive of Luxury Institute said of the cancelation, “Anyone who thinks Christmas will be dramatically up is fooling themselves. It [cancelation] means they are in trouble, that they’re not expecting sales to be as good as expected” (Wahba 2009).

3 Our study focuses on some selling event (e.g., launch of new products or a holiday sale) that has a significant impact on a firm’s performance in a specific circumstance. Our model may not necessarily apply to general (or routinely) inventory decisions.
accessible information. In the long term, when the sales are realized, the investors can correctly assess the firm’s true value in our model; whereas, in the short term, the presence of information asymmetry about the potential demand may result in a discrepancy in the firm’s market value.

We first study a scenario where the trade between the buyer and the supplier is carried out by a single general buy-back contract. As the investors do not observe the demand information, when the demand outlook is pessimistic, the buyer with the short-term interest in the market value may have incentive to distort the stocking decision to mimic if the demand is optimistic. When the demand outlook is truly optimistic, the buyer may have to overstock in order to successfully reveal her information to the market or receive a lower return as the market cannot distinguish the real status of the firm and averages the firm’s value over the two states. We characterize a market equilibrium in which stocking distortion can arise under certain conditions. Such a downstream stocking distortion, if it occurs, is always detrimental for the buyer in equilibrium and can also influence the supplier’s profit as well as the total supply chain surplus, which, for a given contract (not necessarily optimized), may either increase or decrease in the magnitude of the buyer’s market value interest. Furthermore, we find situations where a buy-back contract that would maximize the total supply chain surplus under the classical supply chain contracting framework will not be able to do so in our model.

The second part of our study aims to prevent such a system-wise inefficient stocking distortion. Instead of addressing the issue from the perspective of corporation and/or accounting policies, we focus our study on the design of trade contracts between the supplier and the buyer. We analyze the same game under a menu of two contracts with different buy-back terms. We characterize a general condition of the buy-back contracts under which the downstream stocking distortion can be successfully prevented in equilibrium and thus full efficiency can be restored in the supply chain. A typical example would be to offer a menu in which one contract has a premium wholesale price.

---

4 In practice, firms may carry out their trades by various contracts, such as, wholesale contracts, buy-back contracts, revenue sharing contracts, trade credit, consignment, etc. We focus on buy-back contracts for their relatively simple form and also as they are commonly used in practice and well explored in the literature. Note that the result we show that the buyer may distort the stocking decision in equilibrium does not depend on the contract form we choose. The mechanism that we develop to overcome such an inefficiency does depend on the buy-back contracts. The mechanism however can be potentially generalized to other structured contracts.
while a generous buy-back term that is attractive to the buyer when the true demand outlook is pessimistic even if she takes her market valuation into account and, in contrast, the other contract has a discounted wholesale price while a stringent buy-back term that is favorable for the buyer when the true demand outlook is optimistic. The provision of alternative contracting choices serves as an operational mechanism to overcome the buyer’s incentive to influence the market value by her private information. The contract design in our study differs from those in the traditional adverse selection context since our game involves a rational capital market, and thus the design needs to be contingent on the market belief.

Hence, based on the supply chain contracting framework, our study reveals how a buyer firm with a market value interest may distort her operational decision, the stocking level, and shows the consequent impacts on the supply chain. While such a market value interest is common in reality, it is relatively new to the operations management literature. More importantly, our study shows that by implementing an appropriately designed menu of buy-back contracts, stocking distortions can be prevented in equilibrium. This finding is meaningful since, instead of amending corporation and/or accounting policies, which can sometimes be costly for a firm or even affect various industries, a firm’s incentive of taking real economic activities to influence the capital market valuation can potentially be resolved through operational means.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the model. Section 4 analyzes the game with a single contract which reveals potential distortion of the buyer’s stocking decision. Section 5 explores the design of a menu of buy-back contracts to restore supply chain efficiency. We conclude in section 6.

2. Literature

Our work relates to the inventory and supply chain literature. The newsvendor model that we apply is the cornerstone of the inventory literature. The newsvendor model is also a crucial building block of the supply chain literature, based on which various supply contracts, such as, wholesale contracts (Lariviere and Porteus 2001), buy-back contracts (Pasternack 1985, Emmons and Gilbert 1998) and
revenue sharing contracts (Cachon and Lariviere 2005), have been studied. Some research has also been conducted on contract design in settings with asymmetric demand information (Cachon and Lariviere 2001), asymmetric cost information (Ha 2001, Zhang 2010) and asymmetric forecasting information (Lariviere 2002, Taylor and Xiao 2009). However, the element that we explore, the interest in market value, and its interplay with asymmetric demand information have been little investigated. We enrich the above literature by revealing the possibility of stocking distortion in a setting like ours and providing a mitigating mechanism based on supply contract design.

The capital market interaction is however not new in the earnings management and signaling literature. Stein (1989) reveals that a firm that cares about the market value may inflate the current earnings by pulling future cash flows forward. In an operations context, Lai et al. (2011) demonstrates that firms with different levels of (true) sales may apply different magnitudes of channel stuffing (push excess inventory to the downstream channel) under inventory constraint, to inflate the current sales as well as the prospect of the future demand. These two works focus on earnings and sales; whereas, firms may also use investment to influence the capital market valuation. Bebchuk and Stole (1993) show that high productivity firms may over-invest in a project to signal their productivity to the capital market, while Kedia and Philippon (2009) find in their study that low productivity firms may hire and invest more to pool with high productivity firms. In certain environments, firms with high quality investment opportunities may also under-invest and pool with less efficient firms when the market cannot observe the true quality, as revealed in Gaur et al. (2011). The above literature however has not investigated approaches to mitigate such inefficient distortions. Several accounting works do discuss that aspect, however, they mainly focus on accounting policies. For instance, Dye and Sridhar (2004) and Liang and Wen (2007) examine the magnitude of investment distortions under different accounting regimes and discuss the advantages/disadvantages of different accounting policies. Our work extends the exploration to a supply chain setting and reveals a significant role that a supplier can play in restoring system efficiency (a similar application of supply contracting is investigated in Su and Zhang (2008) in a different context with strategic customers).
Finally, our work relates to several empirical research in operations management. Chen et al. (2005) and Hendricks and Singhal (2009) reveal that excess inventory (e.g., inventory writedowns) often negatively affects stock returns. Our results are qualitatively aligned with their hypothesis and empirical observation. In our study, when the sales are realized, for the same initial stocking level, high leftovers would directly suggest a low firm value. Furthermore, our model could be potentially extended to a multi-period setting with correlated demand; in that case, high leftovers would likely suggest a pessimistic outlook of future demand, which could also lead to a low firm value. Lai (2006) discusses that, anticipating the market response, firms may have incentive to maintain less inventory to signal their competencies measured by fill rate to inventory ratios. Lai uses aggregated inventory data from financial reports across firms and industries. Our study shares a similar motivation but has a more specific focus. We study the stocking decision of an individual firm for a specific selling event where the stocking level implies the prospect of the demand and the investors need to value the firm before any sales are realized.

3. Model

Problem Description. We consider a buyer (she) who procures a product from a supplier (he) for a selling event with uncertain demand. The buyer needs to decide the stocking level of the product prior to the selling event. The supplier produces the product at a unit cost $c$. The selling price for the buyer is fixed at $p$. After the demand is realized, there is no replenishment opportunity and thus any excess demand will be lost; in contrast, in case of overage, the leftover inventory will be salvaged at zero value by either the buyer or the supplier (if the latter buys back the excess inventory). The setting of this selling event is common knowledge.

Before deciding the stocking level, the buyer is able to observe a signal of the demand through activities such as forecasting. Ex ante, the signal is uncertain, denoted by $I$, that is high ($H$) with probability $\lambda$ and low ($L$) with probability $1 - \lambda$. We use $i$ to denote the realization of the signal. The demand conditioned on $i$ is a nonnegative random variable $X_i$ with a strictly increasing distribution $F_i(\cdot)$ (density $f_i(\cdot)$) over $\mathbb{R}^+$ for $i \in \{H, L\}$. A high-value signal implies a
stochastically larger demand with $F_H(x) \leq F_L(x)$ for all $x \geq 0$ and $F_H(x) < F_L(x)$ at least some $x > 0$ (i.e., first-order stochastic dominance). Let $F_i(\cdot) \equiv 1 - F_i(\cdot)$ for $i \in \{H, L\}$. We assume that the prior distribution of the signal and the conditional distributions of the demand are all common knowledge, but, only the buyer observes the realization of the signal who is not able to credibly communicate this information to external parties.

The trade between the buyer and the supplier is carried out through a general form of buy-back contracts. We use $(w, b, t)$ to denote a single contract in which $w$ is the per-unit wholesale price, $b$ is the per-unit buy-back price, and $t$ is the transfer price. If $b$ and $t$ are zero, the contract reduces to a wholesale contract. Given a single contract, the buyer procures $q$ units from the supplier. The trade can also be carried out through a menu of two buy-back contracts, denoted as $\{(w_H, b_H, t_H), (w_L, b_L, t_L)\}$ with the subscripts corresponding to the possible values of the signal. Given a menu of two contracts, the buyer chooses one contract $(w_i, b_i, t_i)$ and procures $q$ units from the supplier. We impose no constraint on the choice of $q$. That is, the buyer can select any $q \in \mathbb{R}^+$ that maximizes her own payoff. The contract, once taken, is legally binding and is not renegotiable. This information, including the implemented contract(s) and the stocking level, is accessible to the capital market (that we introduce below).

Deviating from the supply chain contracting framework, we include a capital market that values the buyer. The market’s valuation is the expectation of the buyer’s ending-period profit conditioned on the information the market can access. While the market would be able to precisely assess the buyer’s value when information is complete, a discrepancy of the valuation can arise in our model before the demand is fully realized, due to the private demand signal the buyer receives. We use $J$ to denote the market’s belief of the signal. Ex ante, $J$ is a random variable identical to $I$. The market will infer the realization of the signal from the stocking level (and the choice of contract if a menu is provided), based on which $J$ is updated. We use $j$ to denote the realization of $J$. The market values the buyer based on the updated belief of the signal and all other accessible information. The buyer cares not only about the true profit she will make but also about her market value. To model the buyer’s incentive scheme, we apply a simple objective function (which
has been similarly applied in the literature; see, e.g., Stein 1989, Liang and Wen 2007): the buyer places a weight $\beta \in [0,1]$ on the short-term market value and a weight $1 - \beta$ on the long-term true profit in her consideration. We consider no time discount. The buyer’s incentive scheme (captured by $\beta$) is common knowledge. Such an objective function can be motivated, for instance, if the buyer firm’s executive managers receive market-based incentives (e.g., options) or if, as discussed in Stein (1989) and Liang and Wen (2007), the buyer bears some liquidation pressure and needs to sell a fraction of her shares to the capital market before the sales revenue is fully obtained. Given this incentive scheme and the information advantage, the buyer may make the stocking decision (contract choice) purposely to influence the market valuation. We consider no accounting manipulation. That is, all the information accessible to the market is precise.

Timeline. The sequence of events is as follows. First, the supplier offers a single contract or a menu of two contracts to the buyer. Second, the buyer privately observes the signal, and chooses a contract and the corresponding stocking level or rejects the contract offer. If the buyer accepts a contract, the supplier produces to meet the buyer’s order and the buyer pays the wholesale and transfer prices. Third, the capital market observes the buyer’s contract choice and stocking level and assesses the expected profit the buyer can make which forms her market value. The buyer realizes a payoff equal to the market value multiplied by the weight $\beta$. Fourth, demand is realized, and payments between the buyer and the supplier are made according to the chosen contract. The buyer realizes another payoff equal to the true profit multiplied by the weight $1 - \beta$.

Information Structure. Similar to the settings adopted in the signaling and the supply contract design literatures (see, e.g., Stein 1989, Bebchuk and Stole 1993, Cachon and Lariviere 2001), we assume a single source of information asymmetry in our model. Such a stylized modeling approach lends tractability to the analysis while also captures the major qualitative insights. In particular, the realization of the demand signal that we assume is private to the buyer represents some information that is most difficult to obtain by external parties. Furthermore, compared to the selling price, the production cost or the details of a contract (for which invoices and other legal proofs could exist), a signal of the potential demand is also difficult to credibly communicate (as cheap talks
could arise). Finally, we assume the same information setting for the supplier and the investors to simplify the model. In practice, suppliers may have more information; however, professional investors and analysts may attempt to discover information from suppliers. In a problem like ours the supplier has no direct incentive to hold any information from (or mislead) the investors.

Benchmark. Notice that without the interaction with the capital market, the problem we have described would be the classical selling to newsvendor problem. An appropriately designed single contract would be sufficient to maximize the total supply chain surplus. That is, when \( \frac{w-b}{p-b} = \frac{c}{p} \), the buyer would stock \( q = q^o_i \equiv F_i^{-1}(\frac{c}{p}) \) which maximizes the total supply chain surplus for each signal \( i \in \{H, L\} \). This is a classical result in the supply chain literature (see, e.g., Pasternack 1985).

Nevertheless, when the buyer cares about her capital market valuation, her decision can deviate from those quantities. Hence we use the classical selling to newsvendor problem as our benchmark and call \( q^o_{i \in \{H, L\}} \) the first-best stocking levels.

To fully illustrate the impacts of the buyer’s market value interest, we start our analysis with a single contract offer in Section 4; and then, in Section 5, we design an operational mechanism with a menu of buy-back contracts to restore the efficiency in the supply chain.

4. Quantity Signaling under Single Contract

In this section, we analyze the model with a single contract offer \((w, b, t)\). We first derive the downstream market equilibrium and analyze the impact of the market value interest on the buyer’s payoff in subsection 4.1; then, we analyze the supplier’s profitability and the supply chain efficiency in subsections 4.2 and 4.3.

4.1. Downstream Market Equilibrium

Given a contract offer \((w, b, t)\), for each signal \( i \in \{H, L\} \), the expected profit of the buyer firm with a stocking level \( q \) follows

\[
\pi^B(q; i) = p\mathbb{E}[\min(q, X_i)] + b\mathbb{E}[\max(q - X_i, 0)] - wq - t
\]

\[
= (p - b) \int_0^q F_i(x)dx - (w - b)q - t.
\]
The buyer’s payoff however depends partially on the firm’s real profit and partially on the firm’s short-term market value. In the following, we formulate the buyer firm’s market value.

Since the signal is private to the buyer, the market needs to hold a rational belief to infer the realized value of the signal. In this paper, we focus only on pure-strategy equilibria. Thus we formulate a general market belief as follows:

$$J(q) = \begin{cases} H & \text{if } q \in Q_H, \\ L & \text{if } q \in Q_L, \\ I & \text{otherwise} \end{cases}$$

where $Q_H$ and $Q_L$ are disjoint subsets of $\mathbb{R}^+$. Under this belief, if the observed stocking level $q \in Q_H$, then the market believes that the realization of the signal is high; in contrast, if the stocking level $q \in Q_L$, then the market believes that the signal is low; finally, observing a stocking level falling outside $Q_H$ and $Q_L$, the market believes that a pooling scenario arises and thus the posterior belief coincides with the prior belief which is the random variable $I$. With this belief, the buyer firm’s market value follows

$$P(q) = \begin{cases} (p-b) \int_0^q \bar{F}_H(x)dx - (w-b)q - t & \text{if } q \in Q_H, \\ (p-b) \int_0^q \bar{F}_L(x)dx - (w-b)q - t & \text{if } q \in Q_L, \\ (p-b) \int_0^q (\lambda \bar{F}_H(x) + (1-\lambda) \bar{F}_L(x))dx - (w-b)q - t & \text{otherwise} \end{cases}$$

Hence, to maximize her own payoff, the buyer solves, for each signal $i \in \{H, L\}$,

$$\max_{q \in \mathbb{R}^+} \beta P(q) + (1-\beta)\pi_B^B(q; i)$$

(1)

where the weight $\beta$ represents the buyer’s interest in her market value. Before carrying out the equilibrium analysis, we first analyze the buyer’s objective function. We define

$$\bar{F}_{ij}(q) \equiv \beta \bar{F}_j(q) + (1-\beta)\bar{F}_i(q), \forall i, j \in \{H, L\},$$

where the subscript $i$ ($j$) indicates the true (market believed) signal value, and

$$G_{ij}(q) \equiv (p-b) \int_0^q \bar{F}_{ij}(x)dx - (w-b)q - t, \forall i, j \in \{H, L\},$$

which is the buyer’s expected payoff if given $q$ the market believes the value of the signal is $j$ while the true signal value is $i$. To incorporate the pooling scenario, we further define

$$\bar{F}_{i*}(q) \equiv \beta \left[ \lambda \bar{F}_H(x) + (1-\lambda) \bar{F}_L(x) \right] + (1-\beta)\bar{F}_i(q), \forall i \in \{H, L\},$$
Figure 1 Demonstration of the buyer’s all possible payoff functions (left), the intuitive criterion refinement over pooling equilibria (middle) and the thresholds \( \bar{q} \) and \( q^* \) for the characterization of separating equilibria (right).

The parameters are: \( \beta = 0.4, p = 20, c = 5, w = 8, b = 4, t = 0, \lambda = 0.5, \) and the demand follows the gamma distribution with density \( f_i(x) = \frac{1}{\Gamma(\kappa_i)\theta_i^{\kappa_i}} x^{\kappa_i-1} e^{-\frac{x}{\theta_i}} \) for \( i \in \{H, L\} \) with \((\kappa_{H}, \theta_{H}) = (1.5, 5)\) and \((\kappa_{L}, \theta_{L}) = (1.5)\) that satisfy the first-order stochastic dominance.

\[
G_{ij}(q) \equiv (p - b) \int_{0}^{q} \bar{F}_{ij}(q) \, dx - (w - b) q - t, \forall i \in \{H, L\}
\]

which is the buyer’s expected payoff if given \( q \) the market believes that a pooling scenario arises and uses the prior belief to value the firm.

**Lemma 1.** \( G_{ij}(q) \) and \( G_{i\bullet}(q) \) are concave in \( q \) for any \( i, j \in \{H, L\} \).

Lemma 1 establishes the concavity result of the buyer’s all possible payoff functions (see the left plot of Figure 1 for an illustration). Therefore, a unique maximizer of the buyer’s problem exists in any scenario. Define \( q^*_i \equiv \bar{F}_{ij}^{-1}\left(\frac{w-b}{p-b}\right) \) and \( q^*_\bullet \equiv \bar{F}_{i\bullet}^{-1}\left(\frac{w-b}{p-b}\right) \) which maximizes \( G_{ij}(q) \) and \( G_{i\bullet}(q) \), respectively. To simplify the notation, we reduce the subscript \( ij \) of \( G_{ij}(\cdot) \) and \( q^*_{ij} \) to \( i \) whenever \( i = j \). We can verify \( q^*_H > q^*_L > q^*_HL > q^*_L \) and \( q^*_H > q^*_LH > q^*_L > q^*_L \).

Now we proceed with the equilibrium analysis. Let \( q(i) \) denote the buyer’s optimal stocking level solved from (1) for each signal \( i \in \{H, L\} \). In an equilibrium, the market belief shall be consistent with the buyer’s strategy on every equilibrium path. Formally,

**Definition 1.** A downstream market equilibrium is reached if the buyer’s optimal stocking decision and the market belief satisfy: (a) \( q(H) \in \mathbb{Q}_H \) and \( q(L) \in \mathbb{Q}_L \) so that \( P(q(i)) = \pi^B(q(i); i) \)
for each signal $i \in \{H, L\}$ (i.e., a separating equilibrium), or (b) $q(H) = q(L) \in \mathbb{R}^+ \backslash \{Q_H \cup Q_L\}$ so that $P(q(i)) = \mathbb{E}[\pi^B(q(I); I)]$ for each signal $i \in \{H, L\}$ (i.e., a pooling equilibrium).

Notice that the game described in our model is a typical signaling game which can have multiple pooling and separating equilibria.\(^5\) In the following, without sacrificing the major insight, we will apply the intuitive criterion developed by Cho and Kreps (1987) to refine the equilibria. We will then focus only on those equilibria that survive the refinement. Lemma 2 holds.

**Lemma 2.** Any pooling equilibrium does not survive the intuitive criterion.\(^6\)

We prove in Appendix A why a pooling equilibrium does not survive the intuitive criterion in our model. We illustrate the main idea here in the middle plot of Figure 1. In words, for any pooling equilibrium with a stocking level $q_{PL}$, we can always find a stocking level $q'$ larger than the equilibrium stocking level such that: the buyer when observing a low signal strictly prefers the equilibrium stocking level (her payoff follows $G_L(q_{PL})$) even if a deviation to $q'$ would lead the market to think that the signal is high (her payoff would follow $G_{LH}(q')$ if she deviates); in contrast, the buyer when observing a high signal would prefer to deviate to $q'$ if such a deviation would lead the market to think that the signal is high (her payoff would follow $G_{HL}(q')$ if she deviates; without deviation her payoff follows $G_{HL}(q_{PL})$). Hence, when observing a deviation to $q'$, intuitively, the market should believe that the buyer observed a high signal, which makes the pooling equilibrium fail the intuitive criterion refinement.

With Lemma 2, we can focus only on separating equilibria onwards, for which we refine the market belief to:

$$J(q) = \begin{cases} H & \text{if } q \in Q_H, \\ L & \text{otherwise.} \end{cases}$$

We derive Lemma 3 which will be useful for characterizing separating equilibria.

---

\(^5\) We characterize some pooling and separating equilibria as examples in Appendix B1 and Appendix B2, respectively. In the pooling equilibria, since the buyer stocks the same quantity for either signal, both overstock and understock can arise. Thus the major insight we reveal that stocking distortion can occur in the presence of a short-term interest in the market value applies to the pooling equilibria as well.

\(^6\) To simplify the model, we do not consider constraints on the stocking level (i.e., the buyer can choose $q$ from $\mathbb{R}^+$). However, specific physical constraints sometimes exist in practice. With constraints, pooling equilibria that survive the intuitive criterion can exist, as discussed in Gaur et al. (2011). The intuitive criterion may also not be effective to refine the solution space if, for instance, the demand signal has more than two states.
Lemma 3. There exists a unique $q > q^*_{HL}$ that satisfies $G_{LH}(q) = G_L(q^*_L)$ and a unique $q > q^*_H$ that satisfies $G_H(q) = G_{HL}(q^*_HL)$; furthermore, $q < \bar{q}$.

We depict $q$ and $\bar{q}$ in the right plot of Figure 1 that equate $G_{LH}(q)$ to $G_L(q^*_L)$ and $G_H(q)$ to $G_{HL}(q^*_HL)$, respectively. In the following, we discuss the implications of Lemma 3.

First, it is not difficult to notice that when the signal is high, the buyer will never stock a quantity lower than $q^*_HL$ no matter how the market specifies its belief ($q^*_HL$ maximizes the buyer’s expected payoff if the market incorrectly believes the signal is low) and the buyer can secure an expected payoff, at least, equal to $G_{HL}(q^*_HL)$. Since $G_H(q)$ is concave, $\bar{q}$, as defined in Lemma 3, represents the largest quantity the buyer with a high signal is willing to stock, if needed, in order to reveal the signal to the market. In other words, if the market specifies the set $Q_H$ in a way such that any quantity in $Q_H$ is larger than $\bar{q}$, then the buyer would rather choose to stock $q^*_HL$ since to reveal the signal she would need to stock more than $\bar{q}$, which leads to an expected payoff lower than $G_{HL}(q^*_HL)$. Hence, in order for a separating equilibrium to hold, there must exist some quantity $q \in Q_H$ such that $q \leq q^*_{HL}$ and it is preferred to $\bar{q}$ by the buyer with a high signal.

Second, since the buyer will never stock less than $q^*_HL$ when the signal is high, if, as observed, the buyer’s stocking level equals $q^*_L(< q^*_HL)$ then the market should believe the signal is low. With this intuition, the buyer with a low signal can always secure an expected payoff $G_L(q^*_L)$ under a rational market belief. From Lemma 3, we can observe that $q$ is the largest quantity the buyer in the scenario with a low signal is willing to stock, to mimic the scenario with a high signal. Therefore, if the market specifies the set $Q_H$ in a way such that any quantity in $Q_H$ is larger than $q$, the buyer with a low signal would be successfully discouraged to mimic.

Finally, the result that $q < \bar{q}$ guarantees the existence of beliefs that support separating equilibria. In fact, a stream of separating equilibria exist in our model. We again use the intuitive criterion to refine the equilibria, which leads to the following proposition.

Proposition 1. Given any single contract offer $(w, b, t)$, there exists a separating equilibrium that survives the intuitive criterion, in which the stocking level follows
\[ q(i) = \begin{cases} \hat{q} & \text{if } i = H, \\ q^*_L & \text{if } i = L, \end{cases} \tag{2} \]

where \( \hat{q} = \max\{q, q^*_H\} \) and the market belief can be specified as \( J(q) = \begin{cases} H & \text{if } q = \hat{q}, \\ L & \text{o/w}. \end{cases} \tag{3} \)

Proposition 1 characterizes a separating equilibrium that survives the intuitive criterion. In fact, this is the only equilibrium in our model that can survive the refinement. In this equilibrium, the buyer stocks \( \hat{q} \) with a high signal and \( q^*_L \) with a low signal, consistent with the market belief. Notice that when the signal is low, the buyer’s stocking decision \( q^*_L \) coincides with the first-best stocking level. Thus stocking distortion does not occur. In contrast, when the signal is high, stocking distortion occurs if \( q^*_H < \hat{q} \). Such an equilibrium is sometimes called the screening equilibrium for signaling games (Cho and Kreps 1987).

Proposition 2 establishes a further result of the buyer’s equilibrium stocking decision.

**Proposition 2.** Given any single contract offer \((w, b, t)\), \( q \) increases in \( \beta \) and a unique threshold \( \hat{\beta} = \frac{G_L(q^*_L) - G_L(q^*_H)}{G_H(q^*_H) - G_L(q^*_H)} \in [0, 1] \) exists such that \( q(H) = q^*_H \), which is fixed, when \( \beta \leq \hat{\beta} \), and \( q(H) = \hat{q} \), which increases in \( \beta \), when \( \beta > \hat{\beta} \).

Proposition 2 shows that the buyer’s stocking decision when the signal is high depends on the magnitude of her interest in the market value, \( \beta \). When \( \beta \) is small (\( \beta \leq \hat{\beta} \)), the buyer stocks the first-best quantity and thus no distortion arises; when \( \beta \) is large (\( \beta > \hat{\beta} \)), the buyer stocks more than the first-best level and the overstocked amount increases in \( \beta \). The intuition of this result is as follows. When \( \beta \) is small, the buyer with a low signal has little incentive to mimic a high signal scenario. As a result, the buyer, when the signal is high, could just stock the first-best quantity and the information of the signal can be correctly revealed. When \( \beta \) is large (\( \beta > \hat{\beta} \)), the buyer’s incentive of mimicking when the signal is low becomes strong and now she would mimic by stocking \( q^*_H \) if that is still the quantity she stocks for a high signal (in particular, at \( \beta = \hat{\beta} \), \( q \) reaches \( q^*_H \)).

\(^7\)The market belief on the off-equilibrium path could be specified in other ways as long as the stocking strategies on the off-equilibrium path are dominated by the equilibrium strategy for the buyer. For instance, instead of the singleton \( Q_H = \{\hat{q}\} \), we can alternatively specify \( Q_H = \{q \in \mathbb{R}^+ : q \geq \hat{q}\} \), with which the equilibrium will not change.
Buyer's Expected Payoff \( \beta > \beta \)

Figure 2 Demonstration of the buyer's expected payoff \( \Pi_B \) as a function of \( \beta \). The parameters are: \( p = 20 \), \( c = 5 \), \( w = 8 \), \( b = 4 \), \( t = 0 \), \( \lambda = 0.5 \), and the demand follows the gamma distribution with density

\[
f_i(x) = \frac{x^{i-1} e^{-x/\theta_i}}{\kappa_i \Gamma(i)}
\]

for \( i \in \{ H, L \} \), where \( (\kappa_H, \theta_H) = (1.5, 5) \) and \( (\kappa_L, \theta_L) = (1, 5) \).

Hence, in order to convince the market when the signal is truly high, the buyer needs to stock more than \( q_H^* \); that is, she needs to stock (at least) \( q \). The stronger the buyer's mimicking incentive is when the signal is low (as \( \beta \) increases), the more units she needs to stock when the signal is high.

**Corollary 1.** The buyer’s expected payoff follows

\[
\Pi_B = \lambda G_H(q_H^*) + (1 - \lambda) G_L(q_L^*)
\]

which is fixed, when \( \beta \leq \hat{\beta} \) and follows

\[
\Pi_B = \lambda G_H(q) + (1 - \lambda) G_L(q_L^*)
\]

which decreases in \( \beta \), when \( \beta > \hat{\beta} \).

Corollary 1 concludes the impact of stocking distortion on the buyer’s payoff. Since the market correctly values the firm in equilibrium, such a stocking distortion is always costly for the buyer. This cost increases if the buyer is more interested in her market value, as depicted in Figure 2.

### 4.2. Supplier Profit

In this subsection, we investigate the impact of the downstream buyer’s market value interest on the upstream supplier’s profitability. We focus on the equilibrium characterized in Proposition 1.

We have discussed that stocking distortion could occur in equilibrium which is detrimental for the buyer. It is however not straightforward whether such a stocking distortion is detrimental or beneficial for the supplier. If the buyer overstocks, on one hand, it would seem to be beneficial for the supplier since more revenues could be collected; on the other hand, when a buy-back term is
Lai, Xiao and Yang: Supply Chain Performance under Market Valuation

Figure 3 Demonstration of the supplier’s expected profit \( \Pi^S \) as a function of \( \beta \). The common parameters are:

- \( p = 20 \), \( c = 5 \), \( w = 8 \), \( t = 0 \), \( \lambda = 0.5 \), and the demand follows the gamma distribution with density
  \[ f_i(x) = \frac{\lambda^{\kappa_i} x^{\kappa_i-1} e^{-\lambda x}}{\kappa_i^{1}(\kappa_i)} \quad \text{for } i \in \{H, L\} \]

- \( (\kappa_H, \theta_H) = (1.5, 5) \) and \( (\kappa_L, \theta_L) = (1, 5) \).

- In the left plot, \( b = 4 \); in the middle plot, \( b = 3.5 \); and in the right plot, \( b = 3 \).

\( \Pi^S \) provided, more returns could occur, which is costly for the supplier. In the following, we investigate the impact for a given contract offer \((w, b, t)\).

Based on the equilibrium characterized in Proposition 1, the supplier’s expected profit follows:

\[
\Pi^S = \begin{cases} 
\lambda \left[ (w - c - b) q^*_H + b \int_0^{q^*_H} \tilde{F}_H(x)dx \right] + (1 - \lambda) \left[ (w - c - b) q^*_L + b \int_0^{q^*_L} \tilde{F}_L(x)dx \right] + t & \text{if } \beta \leq \hat{\beta}, \\
\lambda \left[ (w - c - b) q_H + b \int_0^{q_H} \tilde{F}_H(x)dx \right] + (1 - \lambda) \left[ (w - c - b) q_L + b \int_0^{q_L} \tilde{F}_L(x)dx \right] + t & \text{if } \beta > \hat{\beta}.
\end{cases}
\]  

(4)

It is easy to notice that only \( q \) depends on \( \beta \) in (4). Therefore, the buyer’s market value interest influences the supplier’s profit only when \( \beta > \hat{\beta} \) and the signal is high. By an analytical assessment, we show Proposition 3 holds.

**Proposition 3.** When \( \beta \leq \hat{\beta} \), the supplier’s expected profit \( \Pi^S \) is independent of \( \beta \); when \( \beta > \hat{\beta} \), the following properties hold:

(i) If \( b \geq \frac{w - c}{p} p \), then \( \Pi^S \) strictly decreases in \( \beta \);

(ii) If \( b < \frac{w - c}{p} p \), then a threshold \( \beta' \in [\hat{\beta}, 1] \) exists such that \( \Pi^S \) increases in \( \beta \) when \( \beta \in [\hat{\beta}, \beta'] \) and decreases in \( \beta \) when \( \beta \in (\beta', 1] \).

First, Proposition 3 shows that \( \Pi^S \) is fixed when \( \beta \leq \hat{\beta} \) (as depicted by the straight lines in Figure 3). This result is straightforward since stocking distortion does not occur for any \( \beta \leq \hat{\beta} \).

---

8 In principle, if the supplier has the full bargaining power in the supply chain (such as a Stackelberg leader), the supplier can optimize the contract offer anticipating how the downstream market equilibrium is formed. We do not do that here since, first, a supplier may not have full bargaining power and thus the insights we reveal with a general contract would be useful, and second, such an optimization would become superfluous once we reveal in section 5 that menus of buy-back contracts exist under which full efficiency can be restored in the supply chain.
Second, Proposition 3 shows that when $\beta > \hat{\beta}$, the downstream stocking distortion could be either detrimental or beneficial for the supplier. The impact depends on the buy-back price, $b$. When $b$ is large ($b \geq \frac{w-c}{p-c} p$), the impact is detrimental (see the left plot in Figure 3). We explain the intuition as follows. Notice that the condition $b \geq \frac{w-c}{p-c} p$ can be rewritten as $\frac{p-w}{p-b} b \geq w - c$. In particular, $\frac{p-w}{p-b}$, which equals $F_H(q^*_H)$, represents the probability that a unit product will be returned to the supplier when the buyer stocks $q^*_H$ units. Thus $\frac{p-w}{p-b} b$ is the marginal refund cost for the supplier. The right hand side of the condition, $w - c$, is the marginal revenue for the supplier by selling one more unit when the buyer stocks $q^*_H$ units. Therefore, when $\frac{p-w}{p-b} b \geq w - c$, the cost is larger than the revenue for the supplier if the buyer overstocks.

In contrast, when $b < \frac{w-c}{p-c} p$, the supplier may benefit from the downstream stocking distortion. This benefit can arise because double marginalization is mitigated if the buyer overstocks. Certainly, if the buyer overstocks too much, the supplier can again be worse off. Recall from Proposition 2 that the stocking level increases in $\beta$. When $b < \frac{w-c}{p-c} p$, there is a threshold $\beta'$ such that the downstream stocking distortion is beneficial (detrimental) for the supplier when $\beta < (>) \beta'$ (see the middle plot in Figure 3). Note that $\beta'$ can possibly reach one for some given contract. Such a scenario arises if the double marginalization effect is strong (see the right plot in Figure 3). Therefore, for a given contract offer, the downstream stocking distortion is beneficial for the supplier if it mitigates double marginalization in the scenario where the buyer’s interest in her market value is intermediate (intermediate $\beta$), and it is detrimental if the stocking level is distorted to a large extent, when the buyer’s interest in her market value is strong (large $\beta$).

We further note that if the selling price $p$ becomes larger relative to the unit cost $c$, then the term $\frac{w-c}{p-c} p$ will be smaller and thus the region where the downstream stocking distortion is detrimental for the supplier will become wider. In other words, for high margin products (such as jewelries or other luxury products), the buyer’s market value interest can more likely be detrimental for the supplier. The above result suggests that selling to a downstream newsvendor-like buyer with an interest in her market value can lead to quite different outcomes for the supplier from selling to a classical newsvendor buyer.
β
Expected Total Supply Chain Surplus
β>

Lai, Xiao and Yang: Supply Chain Performance under Market Valuation

Expected Total Supply Chain Surplus
β>

4.3. Supply Chain Efficiency

The above two subsections have shown that for a given single contract offer, the buyer’s market value interest is detrimental for herself while it can be either detrimental or beneficial for the supplier. It will be interesting to see how the impact is from the perspective of the overall supply chain surplus. We analyze this impact in the following for a given contract offer \((w, b, t)\).

Based on the equilibrium in Proposition 1, the expected supply chain surplus can be derived as

\[
\Pi^{SC} = \begin{cases} 
\lambda \left[ p \int_{0}^{\hat{q}} \bar{F}_{H}(x) dx - cq_{H} \right] + (1 - \lambda) \left[ p \int_{0}^{\hat{q}} \bar{F}_{L}(x) dx - cq_{L} \right] & \text{if } \beta \leq \hat{\beta}, \\
\lambda \left[ p \int_{0}^{\hat{q}} \bar{F}_{H}(x) dx - cq_{H} \right] + (1 - \lambda) \left[ p \int_{0}^{\hat{q}} \bar{F}_{L}(x) dx - cq_{L} \right] & \text{if } \beta > \hat{\beta}.
\end{cases}
\]

(5)

It is easy to notice from (5) that the buyer’s market value interest influences the supply chain surplus only when \(\beta > \hat{\beta}\) and the signal is high. We show Proposition 4 holds (see Figure 4 for a depiction).

**Proposition 4.** When \(\beta \leq \hat{\beta}\), the expected supply chain surplus \(\Pi^{SC}\) is independent of \(\beta\); when \(\beta > \hat{\beta}\), the following properties hold:

(i) If \(b \geq \frac{w - c}{p - c} p\), then \(\Pi^{SC}\) strictly decreases in \(\beta\);

(ii) If \(b < \frac{w - c}{p - c} p\), then a threshold \(\beta'' \in [\hat{\beta}, 1]\) exists such that \(\Pi^{SC}\) increases in \(\beta\) when \(\beta \in [\hat{\beta}, \beta'']\) and decreases in \(\beta\) when \(\beta \in (\beta'', 1]\).

We see that similar to the impact for the supplier, the supply chain surplus is fixed when \(\beta \leq \hat{\beta}\) and it may decrease or increase in \(\beta\) when \(\beta > \hat{\beta}\), depending on the contract terms. When \(b\) is large
\((b \geq \frac{w-c}{p-c} p)\), the downstream stocking distortion is detrimental for the supply chain (see the left plot in Figure 4). To explain the intuition, we rewrite the condition \(b \geq \frac{w-c}{p-c} p\) to \(c \geq \frac{w-b}{p-b} p\left(= \bar{F}_H(q^*_H)p\right)\).

Hence, any overstocking beyond \(q^*_H\) will be detrimental for the supply chain since for a \(q > q^*_H\), the marginal cost \(c\) would be larger than the marginal revenue \(\bar{F}_H(q)p\). In contrast, when \(b < \frac{w-c}{p-c} p\), slight overstock could be beneficial for the supply chain. In particular, there is a threshold \(\beta''\) such that the downstream stocking distortion is beneficial (detrimental) for the supply chain when \(\beta < (>) \beta''\) (see the middle plot in Figure 4).\(^9\) \(\beta''\) may reach one for some given contract if the double marginalization effect is strong (see the right plot in Figure 4).

We note that when \(b = \frac{w-c}{p-c} p\) (or identically, \(w-b = \frac{c}{p}\)), the supply chain would be coordinated in the benchmark. Therefore, the intuition we explained in the above suggests that for a given contract offer, if the buyer’s market value interest leads the stocking level closer to the level that would maximize the supply chain surplus in the benchmark, the buyer’s market value interest improves supply chain performance; otherwise, it hurts the supply chain.

More importantly, notice that \(\frac{w-b}{p-b} = \frac{c}{p}\) is the necessary condition for supply chain coordination, while Proposition 4(i) shows that when \(b = \frac{w-c}{p-c} p\) (i.e., \(w-b = \frac{c}{p}\)), the supply chain surplus decreases in \(\beta\) when \(\beta > \hat{\beta}\). That is, supply chain coordination is not achievable if \(\beta > \hat{\beta}\). In fact, we can define a general condition for when supply chain coordination cannot be achieved. Recall from Proposition 2 that \(\hat{\beta} = \frac{G_L(q^*_L) - G_H(q^*_H)}{G_H(q^*_H) - G_L(q^*_H)}\). When \(\frac{w-b}{p-b} = \frac{c}{p}\), \(q^*_i = q^*_o\) (i.e., the first-best stocking level).

We can further notice from the definition of \(G_i(\cdot)\) that \(\hat{\beta}\) does not depend on the contract terms when \(\frac{w-b}{p-b} = \frac{c}{p}\). Hence, let \(\hat{\beta}^o = \frac{G_L(q^*_o) - G_H(q^*_H)}{G_H(q^*_H) - G_L(q^*_H)}\), which depends only on the system parameters. We obtain the following proposition.

**PROPOSITION 5.** When \(\beta > \hat{\beta}^o\), supply chain coordination cannot be achieved by any single contract offer \((w, b, t)\).

In Figure 5, we numerically demonstrate \(\hat{\beta}^o\) which decreases as the selling price \(p\) (or equivalently the margin \(p-c\)) increases, with the other parameters fixed. We can see from this figure that

\(^9\) In general, \(\beta''\) differs from \(\beta'\) in Proposition 3.
the scenario where supply chain efficiency is affected by the buyer’s interest in her market value is not restrictive. Such an inefficiency arises in this equilibrium because overstocking is the only means that the buyer can employ to reveal her demand information to the market.\footnote{The supply chain efficiency can be also affected in a pooling equilibrium where the buyer may distort the stocking level for either signal.} While such distortions have been revealed in other contexts (e.g., Bebchuk and Stole 1993), there has been little discussed in the literature about how to improve the efficiency. Naturally, one approach would be to reduce $\beta$, for instance, compensating the executive managers with less market incentives; another approach would be to amend accounting policies, such as, demanding firms to truthfully report their information. In the next section, we will investigate if other approaches exist that can be used to improve the system efficiency.

5. Restoring Efficiency with Menu of Buy-back Contracts

In this section, we aim to design an operational mechanism that can restore full supply chain efficiency. In particular, we focus on the design of a menu of two buy-back contracts. Notice that the contract design in our study differs from those in the traditional adverse selection context since our problem involves a third-party, the capital market and thus we need to establish a market...

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure5}
\caption{Demonstration of $\hat{o}$ as $p$ increases. The parameters are: $c = 5$, $\lambda = 0.5$, and the demand follows the gamma distribution with density $f_i(x) = \frac{\left(\frac{x}{\theta_i}\right)^{\kappa_i-1} \cdot e^{-\frac{x}{\theta_i}}}{\kappa_i \cdot i\left(\theta_i\right)}$ for $i \in \{H, L\}$, where $(\kappa_H, \theta_H) = (1.5, 5)$ and $(\kappa_L, \theta_L) = (1, 5)$.
\end{figure}
equilibrium. Furthermore, this market equilibrium must be separating as full efficiency would not be achieved otherwise.

Let $\tau = H$ and $L$ denote the two contracts, $(w_H, b_H, t_H)$ and $(w_L, b_L, t_L)$, respectively. With a little abuse of notation, we use $\tau$ to denote the buyer’s contract choice also. The market needs to hold a belief corresponding to each contract. Let $Q^*_H$ be the set of stocking levels corresponding to the contract $\tau$, for which the market believes that the signal is high. We use the pair of sets $Q^*_{H,L}$ to denote a market belief. Recall from section 3 that $q^*_i = \tilde{F}^{-1}_i(\xi)$ is the first-best stocking level that would maximize the total supply chain surplus for each signal $i \in \{H, L\}$. Under a menu of contracts, full efficiency can be achieved in the supply chain if and only if the first-best stocking levels can be implemented and, at the same time, the market is able to correctly infer the true value of the signal from the buyer’s contract choice and stocking decision. Formally, we define the following concept.

**Definition 2.** A market equilibrium with a menu of two buy-back contracts is system-wise efficient if the buyer’s decision follows

$$(\tau, q)(i) = \begin{cases} (H, q^*_H) & \text{if } i = H, \\ (L, q^*_L) & \text{if } i = L, \end{cases}$$

and the market belief follows

$$J(\tau, q) = \begin{cases} H & \text{if } q \in Q^*_H, \\ L & \text{otherwise}, \end{cases}$$

where the two sets $Q^*_{H,L}$ satisfy: $q^*_H \in Q^*_H$ and $q^*_L \notin Q^*_L$.

Notice from Definition 2 that in a system-wise efficient equilibrium, the set $Q^*_H$ must contain $q^*_H$ so that a high signal can be correctly inferred if the buyer with a high signal selects the $\tau = H$ contract and stocks $q^*_H$; in contrast, the other set $Q^*_L$ must not contain $q^*_L$ and then a low signal can be correctly inferred if the buyer with a low signal selects the $\tau = L$ contract and stocks $q^*_L$. Given the structure of the game, any design of the contracts needs to be conditioned on a market belief (i.e., $Q^*_{H,L}$) and we need to derive the menu of buy-back contracts together with the market belief in equilibrium. To directly search for a market belief and solve an associated mechanism
design problem could be challenging. However, the result of the following lemma will greatly reduce the complexity of the problem.

**Lemma 4.** Under any menu of two buy-back contracts, if a system-wise efficient market equilibrium is achievable with a market belief \((Q^H, Q^L_H)\), then a system-wise efficient market equilibrium is also achievable with the market belief \((\{q^o_H\}, \emptyset)\).

Lemma 4 indicates that if there exists a system-wise efficient market equilibrium of our problem then it must be achievable under a restrictive market belief in which the set \(Q^H\) is a singleton that contains just \(q^o_H\) and \(Q^L_H\) is an empty set. That is, under this market belief, if the buyer takes the \(\tau = H\) contract she must stock \(q^o_H\) in order to be recognized with a high signal; if the buyer takes the \(\tau = L\) contract she is automatically considered with a low signal. This result is powerful but also very intuitive, which we explain in the following.

Suppose there is a general belief \((Q^H, Q^L_H)\) with which a system-wise efficient market equilibrium is achieved. Then, to take the \(\tau = H(L)\) contract and stock \(q^o_H\) \((q^o_L)\) is the buyer’s best strategy with a high (low) signal, among all combinations of contracts and stocking levels.

Now, suppose we keep \(Q^L_H\) fixed but shrink the set \(Q^H\) to the singleton containing only \(q^o_H\). Such a modification will obviously not change the buyer’s payoff with a high signal from choosing the \(\tau = H\) contract and stocking \(q^o_H\); it would however reduce the payoff for the buyer to stock other quantities since any deviation from \(q^o_H\) would now be considered with a low signal. Hence the buyer’s best strategy with a high signal remains the same. It is also clear that the buyer will not mimic if the signal is low since it becomes more difficult to succeed (the buyer now would have to stock \(q^o_H\) to succeed; before, she could select an ideal quantity from the original set \(Q^H\)).

Next, we replace \(Q^L_H\) by the empty set \((\emptyset)\). This modification, if there is any impact, will only reduce the buyer’s payoff to take the \(\tau = L\) contract since now such an action will be directly considered with a low signal. Hence the buyer with a high signal will not deviate from her original best strategy. The buyer will also not deviate if the signal is low as her payoff does not change after the modification of the set.
Hence, replacing \((Q_H^H, Q_L^H)\) by \((\{q_H^H\}, \emptyset)\) will change neither the buyer’s contract choice nor the stocking level in the market equilibrium. The market belief with \((\{q_H^H\}, \emptyset)\) serves as the most conservative market belief in terms of believing in that the buyer has a high signal. Armed with Lemma 4, we can design the contracts conditional just on this market belief, which greatly reduces the complexity of the problem.\(^{11} \)

Notice that in order for the buyer to stock the first-best quantity for each signal \(i \in \{H, L\}\), the contract terms must satisfy:

\[
\frac{w_H - b_H}{p - b_H} = \frac{w_L - b_L}{p - b_L} = \frac{c}{p}.
\]

With this condition, we can determine the wholesale price \(w_i\) once the buy-back price \(b_i\) is given; vice versa. Let

\[
g_{ij}(q) \equiv \int_{q_0}^{q} F_{ij}(x) dx - \frac{c}{p} q, \forall i, j \in \{H, L\}
\]

where the subscript \(ij\) is reduced to \(i\) whenever \(i = j\), and \(q_{ij}^o \equiv F_{ij}^{-1}\left(\frac{z}{p}\right)\) for \(i \neq j\). Proposition 6 establishes the main result of this section, a general condition under which full supply chain efficiency can be re-achieved.

**Proposition 6.** With the market belief \((\{q_H^H\}, \emptyset)\), a system-wise efficient market equilibrium can be reached by an appropriately chosen pair of transfer prices \((t_H, t_L)\) if and only if the two buy-back prices in the menu of contracts satisfy

\[
\frac{p - b_H}{p - b_L} \geq K\]

where

\[
K = \max \left\{ \frac{g_{HL}(q_{HL}^o) - g_{L}(q_{L}^o)}{g_{H}(q_{H}^o) - g_{HL}(q_{HL}^o)}, \frac{g_{HL}(q_{HL}^o) - g_{L}(q_{L}^o)}{g_{H}(q_{H}^o) - g_{HL}(q_{HL}^o)} \right\}.
\]

The supplier’s profit can be any value within the range \([0, \hat{\Pi}^S]\) where

\[
\hat{\Pi}^S = p [\lambda g_H(q_{HL}^o) + (1 - \lambda) g_L(q_{L}^o)] - \lambda (p - b_L) [g_{HL}(q_{HL}^o) - g_{L}(q_{L}^o)].
\]

Proposition 6 provides a sufficient and necessary condition, satisfying which a menu of buy-back contracts can be used to restore full efficiency in the supply chain. That is, \(\frac{p - b_H}{p - b_L} \geq K\), where \(K\) depends only on the system parameters that is positive and less than one. Given this condition, a

\(^{11}\) Once the equilibrium is derived under the market belief \((\{q_H^H\}, \emptyset)\), one can specify the off-equilibrium belief in other ways so long as the equilibrium is retained.
of transfer prices always exist, based on which a system-wise efficient market equilibrium can be reached and the supplier’s profit can be accordingly allocated within the range of $[0, \bar{\Pi}^S]$.  

We have discussed in section 4 that given a single contract offer, after observing a high signal, the buyer, if she wants to reveal her signal to the market, would have to overstock to a level that she is not willing to mimic with a low signal. Such a distortion possibly would not be able to be resolved in the contexts adopted in the traditional real earnings management literature where a single (i.e., the focal) player is modeled. Within a supply chain context, however, the other parties (such as the supplier) might be able to “assist” through operational means to alleviate or overcome this distortion. In particular, in our context, the supplier can design a menu of two buy-back contracts under the condition in Proposition 6 such that: the $\tau = L$ contract is more attractive to the buyer when the signal is low and she is discouraged to overstock, and the $\tau = H$ contract is more favorable for the buyer when the signal is high and it is unneeded to overstock to reveal her signal. To put it intuitively, given a single contract offer, the buyer with a high signal would first stock $q^o_H$ units and then stock an extra amount $q - q^o_H$ (when $q > q^o_H$) to reveal to the market that her signal is high. In contrast, given a contract offer with a menu of two buy-back contracts satisfying the condition in Proposition 6, that becomes unnecessary. The buyer with a low signal will always find the $\tau = L$ contract more attractive and thus when the signal is high, she can just take the $\tau = H$ contract and stock $q^o_H$ units, by which the market can already learn her signal.

The condition for designing such a menu of buy-back contracts is not restrictive or complex. Corollary 2 shows that given any buy-back price in the $\tau = H$ ($L$) contract, as long as the buy-back price in the $\tau = L$ ($H$) contract is no higher (lower) than some threshold, a system-wise efficient market equilibrium can be reached.

**Corollary 2.** For any $b_L \in (0, p)$, there exists $\bar{b}_H(b_L) = p(1 - K) + K b_L$ such that whenever $b_H \leq \bar{b}_H(b_L)$, a system-wise efficient market equilibrium can be reached. Similarly, for any $b_H \in (0, p)$, there exists $\bar{b}_L(b_H) = b_H - p \left(\frac{1}{K} - 1\right)$ such that whenever $b_L \geq \bar{b}_L(b_H)$, a system-wise efficient market equilibrium can be reached.
A remaining issue to implement the above mechanism is how the surplus will be distributed between the parties in the supply chain. The mechanism could be difficult to implement if the resulting payoff is not satisfactory for one party, for instance, compared with the payoff s/he can obtain under an existing single contract offer. This issue is addressed in the following.

Proposition 7. With the market belief \( \{q_o^H, \emptyset\} \), when \( b_H = 0 \) and \( b_L = p - \varepsilon \), there exist
\[
\begin{align*}
\{ w_H = c & \text{ and } w_L = p - \varepsilon (1 - \frac{c}{p}), \\
t_H = pg_H(q^*_H) - \varepsilon [g_H(q^*_H) - g_L(q^*_L)] & \text{ and } t_L = \varepsilon g_L(q^*_L)
\end{align*}
\]

such that a system-wise efficient market equilibrium can be reached, in which the supplier’s profit goes to the level that is equal to the total supply chain surplus as \( \varepsilon \to 0 \).

Proposition 7 provides a special menu of buy-back contracts in which: the \( \tau = H \) contract has a discounted wholesale price while a stringent buy-back term (in fact, no returns); in contrast, the \( \tau = L \) contract has a premium wholesale price while a generous buy-back term. Given this menu of contracts, the buyer will choose the \( \tau = H \) contract when the signal is high and choose the \( \tau = L \) contract when the signal is low. Intuitively, the buyer in a status with a pessimistic demand outlook would favor generous return more, while she would favor a discounted wholesale price more when the demand outlook is optimistic. As a result, for such a menu of contracts, it might not be difficult for the market to learn the buyer’s private demand information as the market possibly only needs to detect if the buyer chooses a contract with a buy-back term or not.

Furthermore, under this menu of contracts, the supplier is able to obtain almost all of the surplus as \( \varepsilon \to 0 \). The derivation of this menu of contracts is based on an observation from (6): The upper bound of the supplier’s profit, \( \hat{\Pi}^S \), increases as \( b_L \) decreases. It is hence natural to set \( b_L = p - \varepsilon \), which would maximize the upper bound of the supplier’s profit as \( \varepsilon \) goes to zero. Combining this result with (6), we assert that the supplier’s profit in general can be allocated at any value within the range from zero to the total surplus of the supply chain. This implication is meaningful in the following sense: Suppose the trade in the supply chain is currently carried out through a single contract offer with some rule to share the surplus between the supplier and the buyer. We can
always design a menu of buy-back contracts satisfying the condition in Proposition 6 which makes each party obtain a profit at least equal to the level achieved under the single contract offer and at least one party obtain more. Therefore, pareto improvement can be achieved, which makes a potential menu of contracts attractive to both parties.

Examples do exist in practice where firms use menus of contracts to trade. For instance, IBM PC Co. ever offered menus of buy-back contracts to its resellers (Zarley 1994); menus of buy-back contracts were practiced by some book publishing and pharmaceutical firms too (Padmanabhan and Png 1995). Although the intentions in those cases may differ from ours, we show that to provide a menu of buy-back contracts, which would be unnecessary under the classical supply chain framework, could potentially be helpful to improve the system efficiency when the downstream party has private demand information and, at the same time, cares about her market value.

Certainly, to implement a menu of contracts could be more complicated for firms compared to a single contract. In particular, as the number of the underlying states increases, the complexity of the mechanism will also increase. The mechanism may become impractical when the number of the states is very large, even if analytical tractability could still be achieved.

6. Conclusion

In this paper, we explore how a downstream buyer’s short-term interest in her market value may influence the performances of the parties in the supply chain. First, we show that under a single contract offer the buyer may purposely distort the stocking level in equilibrium. Such a stocking distortion is costly for the buyer but it may either benefit or hurt the supplier depending on the contract terms. We reveal scenarios where full supply chain efficiency cannot be achieved by any single contract offer. Those findings enrich the supply chain literature. An interest in the capital market valuation is not uncommon for firms in practice, however, it has been little explored in the supply chain literature. Second, aiming to prevent stocking distortion, we investigate the role of providing a menu of buy-back contracts in such a context. We derive a general condition under which full efficiency can be restored in the supply chain. This finding contributes to the
literature that explores real earnings management. We demonstrate that in a supply chain context, operational mechanisms can possibly be designed to resolve the distortions.

We conclude by discussing several assumptions in our model. First, we have assumed that the buyer cares about her market value while the supplier does not. In practice, both firms might be interested in their market values. Whereas, as long as information is complete with respect to the supplier’s operations, our results will hold even if the supplier becomes interested in his market value. In that scenario, the market will correctly assess the supplier’s performance and thus the supplier’s objective will not differ from that in our model. It is however interesting to investigate the performance of the supply chain in a setting where the supplier cares about his market value and asymmetric information of his operations exists. Second, we have assumed that the information of the implemented contracts as well as the stocking level is accessible to the market. This assumption is important for characterizing the market equilibrium since, otherwise, in order to value the buyer, the market would need to hold a belief of the contracts and the stocking level. In practice, firms might need to disclose such information to the market; moreover, professional investors (or analysts) might also employ various approaches to discover the information. Certainly, scenarios do exist where, for instance, the information of the business environment is complete whereas the information of a firm’s decisions is not. Finally, we have assumed that there are only two states of the demand signal. Even though similar assumptions have been commonly made in the supply contract design literature, more states can arise in practice. The signaling game in such scenarios will be more complex, and to resolve the distortion, a menu of multiple contracts needs to be designed. We believe it is interesting to study the above aspects in future research.

References


Appendix A: Proofs

Proof of Lemma 1. $G_L(q)$ and $G_H(q)$ follow the classical newsvendor objective function and are thus concave. Notice that $G_{LH}(q)$, $G_{HL}(q)$, $G_{L\bullet}(q)$ and $G_{H\bullet}(q)$ are linear combinations of $G_L(q)$ and $G_H(q)$. Therefore, they are also concave. □

Proof of Lemma 2. We refer the readers to Appendix B1 for some pooling equilibria characterized in our model and Appendix B3 for a discussion of how to apply the intuitive criterion. In the following, we prove that any pooling equilibrium cannot survive the intuitive criterion in our model.

Suppose there is a pooling equilibrium, in which the buyer stocks $q_{PL}$ for each signal $i \in \{H, L\}$ and the market values the buyer firm at

$$
P(q_{PL}) = (p - b) \int_0^{q_{PL}} (\lambda \tilde{F}_H(x) + (1 - \lambda) \tilde{F}_L(x)) \, dx - (w - b) q_{PL} - t
$$

which equals the expected true profit of the firm. Given this market value, the buyer’s payoff in this pooling equilibrium follows, for each signal $i \in \{H, L\}$,

$$
G_{i\bullet}(q_{PL}) = \beta P(q_{PL}) + (1 - \beta) \pi^B(q_{PL}; i)
$$

$$
= (p - b) \int_0^{q_{PL}} (\beta (\lambda \tilde{F}_H(x) + (1 - \lambda) \tilde{F}_L(x)) + (1 - \beta) \tilde{F}_i(x)) \, dx - (w - b) q_{PL} - t.
$$

Now, we apply the intuitive criterion. In particular, we search for a stocking level $q'$ different from $q_{PL}$ such that: the buyer does not have any incentive to deviate from $q_{PL}$ to $q'$ when observing a low signal even if such a deviation would lead the market to believe that the signal is high; whereas, she does have incentive to do that with a high signal if this deviation would lead the market to believe that the signal is high. If such a $q'$ exists, then in the first step of the intuitive criterion (see Appendix B3), we will have a set $\Theta(q') = \{L\}$; that is, for the low signal, the off-equilibrium strategy $q'$ is dominated by the equilibrium strategy $q_{PL}$ for the buyer, while it is not for the high signal. The complement of $\Theta(q')$ is: $\Theta^C(q') = \{H\}$. Then, in the second step, given $\Theta^C(q') = \{H\}$, the lowest payoff that the buyer can obtain by deviating to the off-equilibrium stocking level $q'$ is derived under the market belief that the buyer observed a high signal. Hence, if the buyer does observe a high signal, she would deviate to $q'$ since in the first step we have already verified that
this off-equilibrium strategy is not dominated by the equilibrium strategy for the buyer with a high signal. Consequently, the equilibrium fails the intuitive criterion. If such a $q'$ does not exist, the equilibrium survives.

In the following, we search for such a $q'$. Assume that the market believes the signal is high when observing a stocking level $q$ different from $q_{PL}$. Then, the market value of the buyer firm follows

$$P(q) = (p - b) \int_0^q \tilde{F}_H(x) dx - (w - b) q - t.$$ 

To stock $q$, the buyer’s payoff function follows

$$G_H(q) = \beta P(q) + (1 - \beta) \pi^B(q; H)$$

$$= (p - b) \int_0^q \tilde{F}_H(x) dx - (w - b) q - t$$

with a high signal and

$$G_{LH}(q) = \beta P(q) + (1 - \beta) \pi^B(q; L)$$

$$= (p - b) \int_0^q (\beta \tilde{F}_H(x) + (1 - \beta) \tilde{F}_L(x)) dx - (w - b) q - t$$

with a low signal.

Given $F_H(x) \leq F_L(x)$ for any $x \geq 0$ and $F_H(x) < F_L(x)$ for some $x > 0$, $\tilde{F}_H(x) \geq \tilde{F}_L(x)$ for any $x \geq 0$ and $\tilde{F}_H(x) > \tilde{F}_L(x)$ for some $x > 0$. Thus $G_H(q) > G_{H*}(q_{PL})$ at $q = q_{PL}$. Since $G_H(q)$ is concave, there exists a unique $\tilde{q} > q_{PL}$ such that $G_H(\tilde{q}) = G_{H*}(q_{PL})$; furthermore, $G_H(q)$ is decreasing at $\tilde{q}$.

Now, we verify that $G_{LH}(\tilde{q}) < G_{L*}(q_{PL})$. In particular, we have

$$G_{LH}(\tilde{q}) = (p - b) \int_0^{\tilde{q}} (\beta \tilde{F}_H(x) + (1 - \beta) \tilde{F}_L(x)) dx - (w - b) \tilde{q} - t$$

$$= G_H(\tilde{q}) - (p - b) \int_0^{\tilde{q}} ((1 - \beta) \tilde{F}_H(x) - (1 - \beta) \tilde{F}_L(x)) dx$$

and

$$G_{L*}(q_{PL}) = (p - b) \int_0^{q_{PL}} (\beta (\lambda \tilde{F}_H(x) + (1 - \lambda) \tilde{F}_L(x)) + (1 - \beta) \tilde{F}_L(x)) dx - (w - b) q_{PL} - t$$
Given $\bar{G}_L(q) = G_L(q)$, we have

$$G_L(q) = G_L(q) + (1 - \beta) (p - b) \int_0^q (\tilde{F}_H(x) - \tilde{F}_L(x)) \, dx - (1 - \beta) (p - b) \int_0^{q_{PL}} (\tilde{F}_H(x) - \tilde{F}_L(x)) \, dx.$$ 

Since $\bar{q} > q_{PL}$, $G_L(q_{PL}) > G_L(q)$. Given both $G_H(q)$ and $G_LH(q)$ are continuous and $\frac{dG_H(q)}{dq} > \frac{dG_LH(q)}{dq}$ at any $q$, there must exist an $\varepsilon$ such that at $q' = \bar{q} + \varepsilon$, $G_H(q') > G_H(q)$ and $G_LH(q') < G_LH(q)$. We illustrate this result in Figure 6. Hence, for the low signal, the off-equilibrium strategy $q'$ is dominated by the equilibrium strategy $q_{PL}$ for the buyer, while it is not for the high signal. The existence of such a $q'$ asserts that any pooling equilibrium fails the intuitive criterion.

**Proof of Lemma 3.** From the definitions of $G_{LH}(q)$, $q_{LH}$ and $\bar{q}$, it is direct to see that $G_{LH}(q_{LH}^*) > G_{LH}(q_L^*) > G_L(q_L^*)$. Given $G_{LH}(q)$ is concave, there exists a unique $\bar{q} > q_{LH}^*$ that satisfies $G_{LH}(\bar{q}) = G_L(q_L^*)$. By a similar argument, we can show that there exists a unique $\bar{q} > q_H^*$ that satisfies $G_H(\bar{q}) = G_{HL}(q_H^*)$. 

**Figure 6** Demonstration of the intuitive criterion refinement over the pooling equilibria. The parameters are: $\beta = 0.4, p = 20, c = 5, w = 8, b = 4, t = 0, \lambda = 0.5$, and the demand follows the gamma distribution with density $f_i(x) = \left(\frac{\kappa_i}{\theta_i}ight)^{\kappa_i - 1} x^{\kappa_i - 1} e^{-\frac{x}{\theta_i}}$ for $i \in \{H, L\}$, where $(\kappa_H, \theta_H) = (1.5, 5)$ and $(\kappa_L, \theta_L) = (1.5)$. 

$$= G_H(q) - (p - b) \int_0^{q_{PL}} ((1 - \beta) \tilde{F}_H(x) - (1 - \beta) \tilde{F}_L(x)) \, dx.$$ 

Given $G_H(\bar{q}) = G_H(q_{PL})$, we have

$$G_L(q_{PL}) = G_L(q) + (1 - \beta) (p - b) \int_0^{\bar{q}} (\tilde{F}_H(x) - \tilde{F}_L(x)) \, dx - (1 - \beta) (p - b) \int_0^{q_{PL}} (\tilde{F}_H(x) - \tilde{F}_L(x)) \, dx.$$ 

Since $\bar{q} > q_{PL}$, $G_L(q_{PL}) > G_L(q)$. Given both $G_H(q)$ and $G_LH(q)$ are continuous and $\frac{dG_H(q)}{dq} > \frac{dG_LH(q)}{dq}$ at any $q$, there must exist an $\varepsilon$ such that at $q' = \bar{q} - \varepsilon$, $G_H(q') > G_H(q)$ and $G_LH(q') < G_LH(q)$. We illustrate this result in Figure 6. Hence, for the low signal, the off-equilibrium strategy $q'$ is dominated by the equilibrium strategy $q_{PL}$ for the buyer, while it is not for the high signal. The existence of such a $q'$ asserts that any pooling equilibrium fails the intuitive criterion.
In the following, we show \( q < \overline{q} \) by contradiction. Suppose \( q \geq \overline{q} \). Then, we have

\[
G_H(\overline{q}) - G_{LH}(q) = \left[ (p-b) \int_0^{\overline{q}} \tilde{F}_H(x)dx - (w-b)\overline{q} \right] - \left[ (p-b) \int_0^q \tilde{F}_{LH}(x)dx - (w-b)q \right]
\]

\[
= (p-b) \int_0^{\overline{q}} (1-\beta) \left[ \tilde{F}_H(x) - \tilde{F}_L(x) \right] dx + (w-b) (q-\overline{q}) - (p-b) \int_\overline{q}^q \tilde{F}_{LH}(x)dx
\]

\[
= (p-b) \int_0^{\overline{q}} (1-\beta) \left[ \tilde{F}_H(x) - \tilde{F}_L(x) \right] dx + (p-b) \int_\overline{q}^q \frac{w-b}{p-b} \tilde{F}_{LH}(x)dx
\]

\[
\geq (p-b) \int_0^{\overline{q}} (1-\beta) \left[ \tilde{F}_H(x) - \tilde{F}_L(x) \right] dx.
\]

The fourth equality holds because \( q^*_{LH} = \tilde{F}_{LH}^{-1} \left( \frac{w-b}{p-b} \right) \) and thus \( \tilde{F}_{LH}(q^*_{LH}) = \frac{w-b}{p-b} \). The last inequality holds because \( \overline{q} > q^*_{H} > q^*_{LH} \) and thus \( \tilde{F}_{LH}(q^*_L) > \tilde{F}_{LH}(x) \) for any \( x > \overline{q} \).

Further, we can obtain

\[
G_{HL}(q^*_HL) - G_L(q^*_L) = \left[ (p-b) \int_0^{q^*_{HL}} \tilde{F}_{HL}(x)dx - (w-b)q^*_{HL} \right] - \left[ (p-b) \int_0^{q^*_{L}} \tilde{F}_L(x)dx - (w-b)q^*_L \right]
\]

\[
= (p-b) \int_0^{q^*_{HL}} (1-\beta) \left[ \tilde{F}_H(x) - \tilde{F}_L(x) \right] dx + (p-b) \int_{q^*_L}^{q^*_{HL}} \tilde{F}_L(x) - \frac{w-b}{p-b} \tilde{F}_{HL}(x)dx
\]

\[
= (p-b) \int_0^{q^*_{HL}} (1-\beta) \left[ \tilde{F}_H(x) - \tilde{F}_L(x) \right] dx + (p-b) \int_{q^*_L}^{q^*_{HL}} \tilde{F}_L(x) - \tilde{F}_L(q^*_L) \right] dx
\]

\[
\leq (p-b) \int_0^{q^*_{HL}} (1-\beta) \left[ \tilde{F}_H(x) - \tilde{F}_L(x) \right] dx.
\]

Notice that \( \overline{q} > q^*_H > q^*_{HL} \). Consequently, the above two inequalities lead to \( G_H(\overline{q}) - G_{LH}(q) > G_{HL}(q^*_{HL}) - G_L(q^*_L) \) which contradicts with the definitions that \( G_H(\overline{q}) = G_{HL}(q^*_{HL}) \) and \( G_L(q^*_L) = G_L(q^*_L) \). Hence we conclude \( q < \overline{q} \), which completes the proof. 

**Proof of Proposition 1.** We refer the readers to Appendix B2 for some separating equilibria characterized in our model and Appendix B3 for a discussion of how to apply the intuitive criterion.

In the following, we prove this proposition.

From Lemma 3, it is clear that if the market holds a belief as

\[
J(q) = \begin{cases} 
H \text{ if } q = \hat{q}, \\
L \text{ o/w,}
\end{cases}
\]
where \( \hat{q} = \max\{q, q_H^*\} \), then the buyer’s strategy follows

\[
q(i) = \begin{cases} 
\hat{q} & \text{if } i = H, \\
q_L^* & \text{if } i = L.
\end{cases}
\]

That is, under this market belief, when observing a low signal, the buyer stocks \( q_L^* \) and has no incentive to stock \( \hat{q} \) to mimic the high signal strategy; the buyer also has no incentive to deviate from \( \hat{q} \) if she observes a high signal. The market belief is consistent with the buyer’s strategies. Thus, a separating equilibrium holds.

Now, we prove that this is the only separating equilibrium that can survive the intuitive criterion. To show that, suppose there is another separating equilibrium, in which the buyer’s strategy follows

\[
q(i) = \begin{cases} 
q_H^* & \text{if } i = H, \\
q_L^* & \text{if } i = L.
\end{cases}
\]

and the market belief follows

\[
J(q) = \begin{cases} 
H & \text{if } q = q_H^*, \\
L & \text{o/w},
\end{cases}
\]

where \( q_H^* \neq q_L^* \).

First, we show that if \( q_L^* \neq q_L^* \), then this equilibrium fails the intuitive criterion. Notice that by definition, \( q_L^* \) is the unique maximizer of the buyer’s payoff with a low signal if the market consistently believes that the signal is low. In case the market mistakenly believes that the signal is high, the buyer’s payoff with a low signal would be even higher by stocking \( q_L^* \). It is thus clear that if \( q_L^* \neq q_L^* \), then the off-equilibrium strategy \( q_L^* \) is not dominated by the equilibrium strategy \( q_L^* \) for the buyer when observing a low signal. Therefore, the set \( \Theta(q_L^*) \) we can derive in the first step of the intuitive criterion refinement (see Appendix B3) does not contain \( L \) while the complement \( \Theta^C(q_L^*) \) does contain \( L \). Then, in the second step of the intuitive criterion refinement, the buyer with a low signal would obviously be better off by deviating to \( q_L^* \) compared to the equilibrium strategy \( q_L^* \), even if the market believes that the signal is low for such a deviation. Hence, if \( q_L^* \neq q_L^* \), the equilibrium would fail the intuitive criterion.

Second, we fix \( q_L^* = q_L^* \) and prove that if \( q_H^* \neq \hat{q} \), then this equilibrium fails the intuitive criterion. Notice that in order for a separating equilibrium to hold, \( q_H^* \) must not be smaller than \( \hat{q} \). Now
suppose \( q^*_H \leq q \) and thus \( \hat{q} = \max\{q, q^*_H\} = q < q^*_H \). From Lemma 3, it is clear that the off-equilibrium strategy \( \hat{q} \) is dominated by the equilibrium strategy \( q^*_L \) (i.e., \( q^*_L \)) for the buyer with a low signal; in contrast, the off-equilibrium strategy \( \hat{q} \) is dominated by the equilibrium strategy \( q^*_H \) for the buyer with a high signal (since the buyer’s payoff function is always decreasing for \( q > q^*_H \)). Thus the \( \Theta(\hat{q}) \) we can derive in the first step of the intuitive criterion refinement (see Appendix B3) contains \( L \) but not \( H \), which leads to \( \Theta^C(\hat{q}) = \{H\} \). In the second step of the intuitive criterion refinement, if the buyer with a high signal deviates to \( \hat{q} \), the lowest payoff would be computed based on the market belief that the signal is high since \( \Theta^C(\hat{q}) \) contains only \( H \). As a result, the buyer with a high signal would obviously be better off by deviating to \( \hat{q} \) compared to the equilibrium strategy \( q^*_H > \hat{q} \) (since the buyer’s payoff function is decreasing). Hence the equilibrium fails the intuitive criterion.

In case \( q^*_H > q \), \( \hat{q} = \max\{q, q^*_H\} = q^*_H \) which is the maximizer of the buyer’s payoff function with a high signal if the market correctly infers the signal. Therefore, if \( q^*_H \neq \hat{q} \), it is obvious that the separating equilibrium would fail the intuitive criterion since the buyer with a high signal would have incentive to deviate to \( \hat{q} = q^*_H \) while she would never have such an incentive with a low signal given \( \hat{q} > q \).

Hence, with Lemma 2, we assert that the separating equilibrium presented in Proposition 1 is the only equilibrium that survives the intuitive criterion in our model.

**Proof of Proposition 2.** We first prove \( q \) increases in \( \beta \). As \( q \) is determined by \( G_{LH}(q) = G_L(q^*_L) \), we define

\[
H(q, \beta) \equiv G_{LH}(q) - G_L(q^*_L) = \left[ (p - b) \int_0^q F_{LH}(x) dx - (w - b)q \right] - \left[ (p - b) \int_0^{q^*_L} F_L(x) dx - (w - b)q^*_L \right] = 0.
\]

Recall \( F_{LH}(q) = \beta F_H(q) + (1 - \beta) F_L(q) \). We can easily verify that \( \frac{\partial H(q, \beta)}{\partial \beta} > 0 \) as \( G_{LH}(q) \) increases in \( \beta \) and \( G_L(q^*_L) \) is independent of \( \beta \). On the other hand, given \( q > q^*_L \) by its definition, \( G_{LH}(q) \) decreases in \( q \); thus, \( \frac{\partial H(q, \beta)}{\partial q} < 0 \). Consequently, \( \frac{\partial q}{\partial \beta} = -\frac{\partial H(q, \beta)}{\partial q} > 0 \).

In the following we show a unique threshold \( \beta = \hat{\beta} \) exists at which \( q = q^*_H \). First, when \( \beta = 0 \), from the definition of \( q \) (i.e., \( G_{LH}(q) = G_L(q^*_L) \)), we learn \( q = q^*_L < q^*_H \). Second, when \( \beta = 1 \), \( q^*_L = q^*_H \).
The definition of $q$ asserts $q > q_{LH}^* = q_H^*$. Therefore, a unique threshold $\beta = \hat{\beta}$ exists at which $q = q_H^*$.

To characterize $\hat{\beta}$, we know $q = q_H^*$ at $\hat{\beta}$ and thus the condition $G_{LH}(q) = G_L(q_L^*)$ which determines $q$ is equivalent to $G_{LH}(q_H^*) = G_L(q_L^*)$, or identically,

$$G_{LH}(q_H^*) = \hat{\beta}G_H(q_H^*) + (1 - \hat{\beta})G_L(q_H^*) = G_L(q_L^*).$$

By rearranging the terms we obtain

$$\hat{\beta} = \frac{G_L(q_L^*) - G_L(q_H^*)}{G_H(q_H^*) - G_L(q_H^*)},$$

which completes the proof. ■

**Proof of Corollary 1.** When $\beta \leq \hat{\beta}$, the buyer’s stocking decision is fixed and thus his expected payoff is also fixed; when $\beta > \hat{\beta}$, $q(H) = q$ increases in $\beta$, which implies that the buyer overstocks more units and thus his expected payoff decreases. ■

**Proof of Proposition 3.** The result for $\beta \leq \hat{\beta}$ is straightforward. To show the results of the two properties (i-ii) for $\beta > \hat{\beta}$, we take the derivative

$$\frac{d\Pi^s}{d\beta} = \lambda \left[w - c - bF_H(q)\right] \frac{dq}{d\beta}.$$

Recall from Proposition 2 that $\frac{dq}{d\beta} > 0$. Therefore, to show (i-ii), we only need to determine the sign of $w - c - bF_H(q)$.

(i) Note that $F_H(q_H^*) = \frac{p-w}{p-b}$. When $\beta > \hat{\beta}$, $q > q_H^*$ and thus $F_H(q) > F_H(q_H^*) = \frac{p-w}{p-b}$. Therefore, if $b \geq \frac{p-b}{p-w}(w-c)$, then we have $w - c - bF_H(q) < w - c - \frac{p-w}{p-b} \leq w - c - \frac{p-b}{p-w}(w-c) \frac{p-w}{p-b} = 0$, which asserts $\Pi^s$ is strictly decreasing in $\beta$. Hence, (i) holds.

(ii) Given $q$ increases in $\beta$, $q$ reaches the largest, denoted as, $q_{max}$, when $\beta = 1$. Note that when $\beta = 1$, $q_{LH} = q_H^*$. Therefore, by the definition of $q$, $q_{max}$ is the solution of $G_H(q) = G_L(q_L^*)$ that satisfies $q_{max} > q_H^*$.

Given that $F_H(q)$ reaches the largest at $q = q_{max}$, if $w - c - bF_H(q_{max}) \geq 0$, then $\Pi^s$ is increasing in $\beta$ when $\beta \in [\hat{\beta}, 1]$. That is, $\beta' = 1$. 
Now, suppose \( w - c - bF_H(q_{\text{max}}^H) < 0 \). When \( \beta > \beta', F_H(q) \) reaches the least as \( \beta \to \beta' \) (and \( q \to q_H^* \)). At \( q = q_H^* \), we have \( w - c - bF_H(q_{\text{max}}^H) = w - c - \frac{p-b}{p-b} > 0 \). Therefore, there exists a \( \beta' \in [\beta, 1] \) and a corresponding \( q \in [q_H^*, q_{\text{max}}^\prime] \) at which \( w - c - bF_H(q) = 0 \), and \( \Pi^S \) is increasing in \( \beta \) when \( \beta \in [\beta', \beta'] \) and decreasing in \( \beta \) when \( \beta \in (\beta', 1] \). This completes the proof.

**Proof of Proposition 4.** The result for \( \beta \leq \beta' \) is straightforward. To show the results of the two properties (i-ii) for \( \beta > \beta' \), we take the derivative
\[
\frac{d\Pi^SC}{d\beta} = \lambda [p \tilde{F}_H(q) - c] \frac{dq}{d\beta}.
\]
Recall from Proposition 2 that \( \frac{dq}{d\beta} > 0 \). Therefore, to show (i-ii), we only need to determine the sign of \( p \tilde{F}_H(q) - c \).

(i) Note that \( \tilde{F}_H(q_{\text{max}}^H) = \frac{w-b}{p-b} \). When \( \beta > \beta', q > q_H^* \) and thus \( F_H(q) < F_H(q_H^*) = \frac{w-b}{p-b} \). Therefore, if \( \frac{w-b}{p-b} \leq \frac{w-b}{p-b} \), then we have \( p \tilde{F}_H(q) - c < p \tilde{F}_H(q_{\text{max}}^H) - c \leq 0 \), which asserts \( \Pi^SC \) is strictly decreasing in \( \beta \). Hence, (i) holds.

(ii) When \( \beta > \beta' \), \( \tilde{F}_H(q) \) reaches the largest as \( \beta \to \beta' \) (and thus \( q \to q_H^* \)). If \( \frac{w-b}{p-b} > \frac{w-b}{p-b} \), then at \( q = q_H^* \), we have \( p \tilde{F}_H(q) - c = p \frac{w-b}{p-b} - c > 0 \), which implies that \( \Pi^SC \) is increasing at \( \beta = \beta' \).

Given \( q \) increases in \( \beta \), \( q \) reaches the largest, denoted as, \( q_{\text{max}}^\prime \), when \( \beta = 1 \), while \( \tilde{F}_H(q) \) reaches the least at \( q = q_{\text{max}}^H \). Therefore, if \( p \tilde{F}_H(q_{\text{max}}^H) - c \geq 0 \), then \( \Pi^SC \) is increasing in \( \beta \) when \( \beta \in [\beta', 1] \). That is, \( \beta'' = 1 \). In case \( p \tilde{F}_H(q_{\text{max}}^H) - c < 0 \), it is obvious that a \( \beta'' \in [\beta', 1] \) exists at which the corresponding \( q \) satisfies \( p \tilde{F}_H(q) - c = 0 \). Hence, there exists a \( \beta'' \in [\beta', 1] \) such that \( \Pi^SC \) is increasing in \( \beta \) when \( \beta \in [\beta', \beta''] \) and decreasing in \( \beta \) when \( \beta \in (\beta'', 1] \). This completes the proof.

**Proof of Proposition 5.** The proof is straightforward. Notice that in order to coordinate the supply chain, the contract must satisfy: \( \frac{w-b}{p-b} = \frac{w-b}{p-b} \). From Proposition 4(i), we notice that if \( \frac{w-b}{p-b} = \frac{w-b}{p-b} \), then \( \Pi^SC \) is strictly decreasing in \( \beta \) when \( \beta > \beta' \). Hence, the supply chain will not be coordinated by any single contract offer \( (w, b, t) \). The definition of \( \beta'' \) has been explained in the text.

**Proof of Lemma 4.** If the system can achieve the first-best with a market belief \((Q_H^H, Q_L^H)\), then these two sets must satisfy: \( q_H^* \in Q_H^H \) and \( q_L^* \notin Q_H^H \) (otherwise the buyer would need to distort the stocking levels in order for the market to correctly infer the signal). To achieve the maximum
efficiency, we need \( \frac{w_H - b_H}{p - b_L} = \frac{w_L - b_L}{p - b_L} = \frac{z}{p} \). Under such a menu of buy-back contracts, when the signal is high, the buyer’s decision follows

\[
\begin{align*}
  \Pi^B(H, H; H) &\equiv \max_{q \in \mathbb{Q}_H^H} (p - b_H) \left[ \int_0^q \bar{F}_H(x) \, dx - \frac{z}{p} q \right] - t_H \quad \text{if } \tau = H, \\
  \Pi^B(H, L; H) &\equiv \max_{q \notin \mathbb{Q}_H^H} (p - b_H) \left[ \int_0^q \bar{F}_H(x) \, dx - \frac{z}{p} q \right] - t_H \quad \text{if } \tau = H, \\
  \Pi^B(L, H; H) &\equiv \max_{q \in \mathbb{Q}_H^L} (p - b_L) \left[ \int_0^q \bar{F}_H(x) \, dx - \frac{z}{p} q \right] - t_L \quad \text{if } \tau = L, \\
  \Pi^B(L, L; H) &\equiv \max_{q \notin \mathbb{Q}_H^L} (p - b_L) \left[ \int_0^q \bar{F}_H(x) \, dx - \frac{z}{p} q \right] - t_L \quad \text{if } \tau = L.
\end{align*}
\]

when the signal is low, the buyer’s decision follows

\[
\begin{align*}
  \Pi^B(H, H; L) &\equiv \max_{q \in \mathbb{Q}_H^H} (p - b_H) \left[ \int_0^q \bar{F}_L(x) \, dx - \frac{z}{p} q \right] - t_H \quad \text{if } \tau = H, \\
  \Pi^B(H, L; L) &\equiv \max_{q \notin \mathbb{Q}_H^H} (p - b_H) \left[ \int_0^q \bar{F}_L(x) \, dx - \frac{z}{p} q \right] - t_H \quad \text{if } \tau = H, \\
  \Pi^B(L, H; L) &\equiv \max_{q \in \mathbb{Q}_H^L} (p - b_L) \left[ \int_0^q \bar{F}_L(x) \, dx - \frac{z}{p} q \right] - t_L \quad \text{if } \tau = L, \\
  \Pi^B(L, L; L) &\equiv \max_{q \notin \mathbb{Q}_H^L} (p - b_L) \left[ \int_0^q \bar{F}_L(x) \, dx - \frac{z}{p} q \right] - t_L \quad \text{if } \tau = L.
\end{align*}
\]

If a system-wise efficient market equilibrium is reached with the market belief \((\mathbb{Q}_H^H, \mathbb{Q}_H^L)\), then the buyer takes the contract \((w_H, b_H, t_H)\) and stocks \(q_H^o\) when the signal is high and she takes the contract \((w_L, b_L, t_L)\) and stocks \(q_L^o\) when the signal is low, which implies the following (IC) and (IR) constraints:

\[
\begin{align*}
  \Pi^B(L, L; L) &\geq \max \{ \Pi^B(H, H; L), \Pi^B(H, L; L), \Pi^B(L, H; L) \} \quad \text{(IC1)}, \\
  \Pi^B(L, L; L) &\geq \max \{ \Pi^B(H, H; L), \Pi^B(L, H; H), \Pi^B(L, L; H) \} \quad \text{(IC2)}, \\
  \Pi^B(H, H; H) &\geq 0 \quad \text{(IR1)}, \\
  \Pi^B(H, H; H) &\geq 0 \quad \text{(IR2)}.
\end{align*}
\]

When the thresholds in the market belief \((\mathbb{Q}_H^H, \mathbb{Q}_H^L)\) are replaced by \((\{q_H^o\}, \emptyset)\), \(\Pi^B(H, H; H)\) and \(\Pi^B(L, L; L)\) will not change and it is not difficult to find out that the right hand sides of the IC constraints will only become smaller since the market belief becomes stricter in terms of recognizing a high signal. Hence, the same separating equilibrium will be achieved. ■

**Proof of Proposition 6.** We start the proof by analyzing the (IC) and (IR) constraints. With the market belief \((\{q_H^o\}, \emptyset)\), the (IC1) constraint in (8) can be captured by the following two inequalities:

\[
(p - b_L) \left[ \int_0^{q_L^o} \bar{F}_L(x) \, dx - \frac{c}{p} q_L^o \right] - t_L \geq (p - b_H) \left[ \beta \int_0^q \bar{F}_H(x) \, dx + (1 - \beta) \int_0^{q_H^o} \bar{F}_L(x) \, dx - \frac{c}{p} q_H^o \right] - t_H
\]

and

\[
(p - b_L) \left[ \int_0^{q_L^o} \bar{F}_L(x) \, dx - \frac{c}{p} q_L^o \right] - t_L \geq (p - b_H) \left[ \int_0^{q_L^o} \bar{F}_L(x) \, dx - \frac{c}{p} q_L^o \right] - t_H.
\]
The (IC2) constraint can be rewritten as
\[
(p-b_H) \left[ \int_0^{q_H^L} F_H(x)dx - \frac{c}{p} q_H^L \right] - t_H \geq (p-b_L) \left[ \beta \int_0^{q_H^L} F_L(x)dx + (1-\beta) \int_0^{q_H^L} F_H(x)dx - \frac{c}{p} q_H^L \right] - t_L. 
\] (11)

The (IR) constraints are
\[
(p-b_H) \left[ \int_0^{q_H^L} F_H(x)dx - \frac{c}{p} q_H^L \right] - t_H \geq 0, 
\] (12)
\[
(p-b_L) \left[ \int_0^{q_L^H} F_L(x)dx - \frac{c}{p} q_L^H \right] - t_L \geq 0. 
\] (13)

From (9-11), we can obtain
\[
t_H - t_L \geq (p-b_H) g_{HL}(q_H^L) - (p-b_L) g_L(q_L^H); 
\] (14)
\[
t_H - t_L \geq (p-b_H) g_L(q_L^H) - (p-b_L) g_L(q_L^H); 
\] (15)
\[
t_H - t_L \leq (p-b_H) g_H(q_H^L) - (p-b_L) g_{HL}(q_H^L). 
\] (16)

Therefore, the right hand side of (16) must be larger than or equal to the maximum of those of (14-15) and hence we obtain the condition $\frac{p-b_H}{p-b_L} \geq K$ where
\[
K = \max \left\{ \frac{g_{HL}(q_H^L) - g_L(q_L^H)}{g_H(q_H^L) - g_{HL}(q_H^L)}, \frac{g_{HL}(q_H^L) - g_L(q_L^H)}{g_H(q_H^L) - g_L(q_L^H)} \right\}. 
\] (17)

When $\frac{p-b_H}{p-b_L} \geq K$, it is straightforward from the above reasoning that we can always find a pair of transfer payments, $t_H$ and $t_L$, which satisfy the (IC) and (IR) constraints.

In the following we argue $K$ is positive and less than one. From the definitions of the functions $g_{ij}(q)$, it is direct to observe that $K$ is always positive. Also, the second term on the right side of Equation (17) is always lower than one. Thus, to show $K$ is lower than one, it suffices to show the first term $\frac{g_{HL}(q_H^L) - g_L(q_L^H)}{g_H(q_H^L) - g_{HL}(q_H^L)} < 1$, which can be observed from
\[
g_H(q_H^L) - g_{HL}(q_H^L) = g_H(q_H^L) - \beta g_H(q_H^L) - (1-\beta) g_L(q_H^L) 
= (1-\beta) g_H(q_H^L) - (1-\beta) g_L(q_H^L) 
\]
and
\[
g_{HL}(q_H^L) - g_L(q_L^H) = \beta g_L(q_L^H) + (1-\beta) g_{HL}(q_H^L) - g_L(q_L^H) 
\]
\[
< (1 - \beta) g_H(q_H^*) - (1 - \beta) g_L(q_L^*)
\]
\[
< (1 - \beta) g_H(q_H^*) - (1 - \beta) g_L(q_L^*).
\]

Now, we analyze the range of the supplier’s profit. First, it is obvious that there always exist such \( t_H \) and \( t_L \) under which the supplier obtains zero profit (one can always reduce \( t_H \) and \( t_L \) by the same amount while keeping their difference fixed until the supplier’s profit reaches zero). Therefore, the supplier’s profit has a lower bound that is zero.

Second, the upper bound of the supplier’s profit is reached if the buyer’s profit is minimized. From (12), (13) and (14), we can find the largest \( t_H \) and \( t_L \) under which the equilibrium can hold are

\[
t_L = (p - b_L) g_L(q_L^*),
\]
\[
t_H = t_L + (p - b_H) g_H(q_H^*) - (p - b_L) g_H L(q_H^* L).
\]

Hence the largest profit that the supplier can obtain is

\[
\hat{\Pi}^S = p \left[ \lambda g_H(q_H^*) + (1 - \lambda) g_L(q_L^*) \right] - \lambda (p - b_L) [g_H L(q_H^* L) - g_L(q_L^*)].
\]

This completes the proof. \( \blacksquare \)

**Proof of Corollary 2.** This corollary directly follows from Proposition 6. \( \blacksquare \)

**Proof of Proposition 7.** This is a special scenario of Proposition 6. When \( b_H = 0 \) and \( b_L = p - \varepsilon \),

\[
\frac{p - b_H}{p - b_L} = \frac{p}{\varepsilon} \to \infty \text{ as } \varepsilon \to 0.
\]

Hence the condition for efficiency, \( \frac{p - b_H}{p - b_L} \geq K \), is satisfied. By the condition

\[
\frac{w_H - b_H}{p - b_L} = \frac{w_L - b_L}{p - b_L} = \frac{\varepsilon}{p},
\]

we can obtain \( w_H = c \) and \( w_L = p - \varepsilon (1 - \frac{c}{p}) \). We can easily verify that with

\[
t_H = pg_H(q_H^*) - \varepsilon [g_H(q_H^*) - g_L(q_L^*)] \text{ and } t_L = \varepsilon g_L(q_L^*),
\]

a system-wise efficient separating equilibrium can be reached and the supplier obtains almost all of the supply chain surplus as \( \varepsilon \) goes to zero. \( \blacksquare \)

**Appendix B: Pooling/Separating Equilibria and Intuitive Criterion**

**B1. Characterization of Pooling Equilibria**

In the following, we characterize some examples of pooling equilibria that can arise in our model. Recall from section 4 that the market belief follows
\[ J(q) = \begin{cases} H & \text{if } q \in \mathbb{Q}_H, \\ L & \text{if } q \in \mathbb{Q}_L, \\ I & \text{o/w}. \end{cases} \]

With this market belief, the buyer’s market value follows
\[
P(q) = \begin{cases} (p-b) \int_0^q F_H(x) dx - (w-b) q - t & \text{if } q \in \mathbb{Q}_H, \\ (p-b) \int_0^q F_L(x) dx - (w-b) q - t & \text{if } q \in \mathbb{Q}_L, \\ (p-b) \int_0^q (\lambda F_H(x) + (1-\lambda) F_L(x)) dx - (w-b) q - t & \text{o/w}. \end{cases}
\]

The buyer firm’s true profit follows
\[
\pi^B(q; i) = (p-b) \int_0^q F_i(x) dx - (w-b) q - t.
\]

Thus, to maximize her own payoff, the buyer solves, for each signal \( i \in \{H, L\} \),
\[
\max_{q \in \mathbb{R}^+} \beta P(q) + (1-\beta) \pi^B(q; i).
\]

We have defined
\[
\bar{F}_{ij}(q) \equiv \beta \bar{F}_j(q) + (1-\beta) \bar{F}_i(q), \forall i, j \in \{H, L\},
\]
where the subscript \( i \) (\( j \)) indicates the true (market believed) signal value, and
\[
G_{ij}(q) \equiv (p-b) \int_0^q \bar{F}_{ij}(x) dx - (w-b) q - t, \forall i, j \in \{H, L\},
\]
which is the buyer’s expected payoff if given \( q \) the market believes the value of the signal is \( j \) while the true signal value is \( i \). We have also defined
\[
\bar{F}_\bullet(q) \equiv \beta [\lambda \bar{F}_H(x) + (1-\lambda) \bar{F}_L(x)] + (1-\beta) \bar{F}_i(q), \forall i \in \{H, L\},
\]
and
\[
G_{\bullet}(q) \equiv (p-b) \int_0^q \bar{F}_\bullet(q) dx - (w-b) q - t, \forall i \in \{H, L\}
\]

Therefore, the buyer’s objective function can be organized to
\[
\max \quad \max_{q \in \mathbb{Q}_H} G_{IH}(q) = (p-b) \int_0^q \bar{F}_{IH}(q) dx - (w-b) q - t \\
\max_{q \in \mathbb{Q}_L} G_{IL}(q) = (p-b) \int_0^q \bar{F}_{IL}(q) dx - (w-b) q - t \\
\max_{q \in \mathbb{R}^+/(\mathbb{Q}_H \cup \mathbb{Q}_L)} G_{\bullet}(q) = (p-b) \int_0^q \bar{F}_{\bullet}(q) dx - (w-b) q - t
\]

A pooling equilibrium can arise in our model if and only if there exists a \( q_{PL} \in \mathbb{R}^+/(\mathbb{Q}_H \cup \mathbb{Q}_L) \) such that
\[
G_{\bullet}(q_{PL}) \geq \max_{q \in \mathbb{Q}_H} \max_{q \in \mathbb{Q}_L} \max_{q \in \mathbb{R}^+/(\mathbb{Q}_H \cup \mathbb{Q}_L)} G_{\bullet}(q), \forall i \in \{H, L\}.
\]
In fact, there can exist many pooling equilibria depending on how the two sets \( Q_H \) and \( Q_L \) are specified. In the following, we characterize some examples.

Given the functions \( G_i(q) \) and \( G_{i*}(q) \) are concave, we can solve

\[
q_L^* = \arg \max_{q \in \mathbb{R}^+} G_L(q),
q_{HL}^* = \arg \max_{q \in \mathbb{R}^+} G_{HL}(q),
q_{L*}^* = \arg \max_{q \in \mathbb{R}^+} G_{L*}(q), \quad \text{and} \quad q_{H*}^* = \arg \max_{q \in \mathbb{R}^+} G_{H*}(q).
\]

We can verify that \( q_L^* < q_{L*}^* < q_{H*}^* \) and \( q_L^* < q_{HL}^* < q_{H*}^* \).

In order for a pooling equilibrium to arise, we must not have both \( q_{L*}^* \) and \( q_{H*}^* \) in the set \( \mathbb{R}^+/(Q_H \cup Q_L) \) since, otherwise, the buyer would stock \( q_{H*}^*(q_{L*}^*) \) when observing a high (low) signal and thus the market value of the pooling scenario would not be consistent with the buyer’s strategies. Furthermore, given \( G_{L*}(q) \) is concave, \( G_{L*}(q) \) decreases when \( q \geq q_{L*}^* \). Thus, we must not have any quantity \( q \in [q_{L*}^*, q_{H*}^*] \) together with \( q_{H*}^* \) in the set \( \mathbb{R}^+/(Q_H \cup Q_L) \).

Based on the above intuition, we specify

\[
Q_H = \emptyset \quad \text{and} \quad Q_L = \{q \in \mathbb{R}^+: q < q_{PL}\}.
\]

A pooling equilibrium will arise under such a market belief for any \( q_{PL} \geq q_{H*}^* \) such that \( G_L(q_L^*) \leq G_{L*}(q_{PL}) \). In this equilibrium, the buyer stocks \( q_{PL} \) for either signal \( i \in \{H, L\} \) and the market value at \( q_{PL} \) is consistent with the buyer’s strategy. In such pooling equilibria, the buyer generally overstocks when the signal is low and she may understock when the signal is high.

We can also specify

\[
Q_H = \emptyset \quad \text{and} \quad Q_L = \{q \in \mathbb{R}^+: q \neq q_{PL}\}.
\]

A pooling equilibrium will arise under such a market belief for any \( q_{PL} \) such that \( G_L(q_L^*) \leq G_{L*}(q_{PL}) \) and \( G_{HL}(q_{HL}^*) \leq G_{H*}(q_{PL}) \). In this equilibrium, the buyer stocks \( q_{PL} \) for either signal \( i \in \{H, L\} \) and the market value at \( q_{PL} \) is consistent with the buyer’s strategy. Besides the above equilibria, many other pooling equilibria can exist, which is typical for signaling games.

**B2. Characterization of Separating Equilibria**

Recall from section 4 that for separating equilibria the market belief can be simplified to

\[
J(q) = \begin{cases} 
H & \text{if } q \in Q_H, \\
L & \text{o/w,}
\end{cases}
\]
As we discussed below Lemma 3, in order for a separating equilibrium to hold, the set $Q_H$ must not contain any quantity $q \in (q^*_L, q)$; otherwise, the buyer when observing a low signal would stock such a $q$ and then the market belief would not be consistent with the buyer’s strategy. On contrary, the set $Q_H$ shall contain at least one quantity $q \in [q, \bar{q}]$; otherwise, the buyer when observing a high signal would have no incentive to reveal her information and the market belief would not be consistent with the buyer’s strategy.

The above intuition asserts that as long as $Q_H$ contains at least one quantity in the region $[q, \bar{q}]$ and excludes the region $(q^*_L, q)$, a separating equilibrium can arise. Thus we specify $Q_H = \{q_{SP}\}$, which is a singleton. It directly follows that with this market belief, a separating equilibrium can arise for any $q_{SP} \in [q, \bar{q}]$. In such an equilibrium, the buyer stocks $q^*_H$ when observing a low signal and she stocks $q_{SP}$ when observing a high signal and the market belief is consistent with the buyer’s equilibrium strategy. Hence there exist many separating equilibria in our model.

**B3. The Intuitive Criterion**

In this subsection, we describe the intuitive criterion developed by Cho and Kreps (1987) and its application to our model. The intuitive criterion uses two steps to examine an equilibrium of a signaling game that has two players, a signal sender and a signal receiver.

(i) **The first step** of the intuitive criterion derives a set $\Theta$ of the types of the sender, with which the highest utility that the sender can obtain by taking a specific off-equilibrium strategy is lower than that by keeping the equilibrium strategy. That is, under those types, the off-equilibrium strategy is dominated by the equilibrium strategy for the sender.

Specifically, in our model, suppose we have an equilibrium in which the buyer stocks $q^*_H$ when observing a high signal and stocks $q^*_L$ when observing a low signal. If $q^*_H = q^*_L$, then the equilibrium is pooling; otherwise, the equilibrium is separating. In the first step of the intuitive criterion refinement, for any off-equilibrium stocking level $q$, we derive a set of the signals:

$$\Theta(q) = \{i \in \{H, L\} : G(q^*_i; i) > \hat{G}(q; i)\}$$

where $G(q^*_i; i)$ denotes the buyer’s equilibrium payoff while $\hat{G}(q; i)$ denotes the highest payoff that
the buyer can obtain by stocking the off-equilibrium \( q \). Note that the highest payoff for a given stocking level \( q \) is achieved if the market believes the buyer observes a high signal; that is,

\[
\hat{G}(q; i) = (p - b) \int_{0}^{q} \bar{F}_{H}(q) dx - (w - b) q - t.
\]

Therefore, \( \Theta(q) \) contains those signals under which the off-equilibrium strategy \( q \) is dominated by the equilibrium strategy \( q^c \) for the buyer.

If the set \( \Theta^C \), the complement of \( \Theta \), is an empty set, the second step becomes unnecessary since for all types the off-equilibrium strategy is always dominated by the equilibrium strategy and the sender will not deviate at all. In this case, the intuitive criterion imposes no constraint on the solution space. If \( \Theta^C \) is nonempty, then we need to carry out the second step.

(ii) **The second step** of the intuitive criterion checks if there exists a specific type in \( \Theta^C \) such that the equilibrium utility of the sender with this type is lower than the lowest utility that she can obtain by taking a specific off-equilibrium strategy given that the receiver restricts his belief to \( \Theta^C \) after observing such a deviation. If there does exist such a type, the equilibrium fails the intuitive criterion; otherwise, the intuitive criterion imposes no constraint on the solution space.

Specifically, in our model, the second step of the intuitive criterion checks, for any stocking level \( q \), if there exists a signal \( i \in \Theta^C(q) \), the complement of \( \Theta(q) \), such that with this signal the buyer’s equilibrium payoff \( G(q^c; i) \) is lower than the lowest payoff that the buyer can obtain by deviating to the stocking level \( q \) when the market belief is restricted to \( \Theta^C(q) \) for such a deviation. Let \( \hat{G}(q; i) \) denote this lowest payoff, and it follows

\[
\hat{G}(q; i) = \begin{cases} 
(p - b) \int_{0}^{q} \bar{F}_{iL}(q) dx - (w - b) q - t & \text{if } L \in \Theta^C(q), \\
(p - b) \int_{0}^{q} \bar{F}_{iH}(q) dx - (w - b) q - t & \text{o/w}.
\end{cases}
\]

That is, if the low signal \( L \) is contained in \( \Theta^C(q) \) (i.e., the strategy to deviate to \( q \) is not dominated by the equilibrium strategy for the buyer with a low signal), then the lowest payoff the buyer would obtain to deviate to \( q \) is achieved under the market belief that the buyer observed a low signal for such a deviation. If the low signal \( L \) is not contained \( \Theta^C(q) \), then \( \Theta^C(q) \) contains only the high signal and thus the lowest payoff the buyer would obtain to deviate to \( q \) is achieved under
the market belief that the buyer observed a high signal. If there exists such a signal $i \in \Theta^C(q)$ that $G(q^*_i; i) < \tilde{G}(q; i)$, then the equilibrium fails the intuitive criterion in our model; otherwise, the intuitive criterion imposes no constraint on the solution space.

We use the above procedure of the intuitive criterion to refine the equilibria in our model.
Incentive Contracts for Technology Adoption within a Supply Chain:
A Principal-Agent Perspective

Tao Yao
Harold & Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA 16802, taoyao@psu.edu

Baichun Feng
Sheldon B. Lubar School of Business, University of Wisconsin – Milwaukee, 3202 N Maryland Ave., Milwaukee, WI 53211, fengb@uwm.edu

Bin Jiang
Department of Management, College of Commerce, DePaul University, 1 E Jackson Blvd., Ste 7000, Chicago, IL 60604, bjiang@depaul.edu

This research analyzes how a buyer firm can stimulate its existing supplier to adopt a new technology via an incentive contract, which can motivate the supplier to make costly adoption efforts, truthfully reveal the information of adoption costs, and execute the adoption at the right timing in order to maximize the buyer’s benefits. Through a novel continuous-time principal-agent model with supplier’s hidden information and hidden action as well as uncertain production cost, this research obtains analytical solutions to provide a fully dynamic characterization of this incentive contract, reveals the implicational relationships among adoption timing, adoption cost and contract price. Our result gains in-depth insights into the new technology adoption in supply chain and provides important managerial implications.
Valuation and Hedging of Commodity Storage in the Presence of Term Structure Model Error

Nicola Secomandi,1 Guoming Lai,2 François Margot,1 Alan Scheller-Wolf,1 Duane J. Seppi1
1Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890, USA
2McCombs School of Business, University of Texas at Austin, 1 University Station, B6000, GSB 3.136, Austin, TX 78712-1178, USA
ns7@andrew.cmu.edu, guoming.lai@mccombs.utexas.edu, {fmargot, awolf, ds64}@andrew.cmu.edu


Valuation and hedging of financial and real option cash flows rely on models of the risk neutral dynamics of the underlying state variables. In practice, these risk neutral dynamics are not known with certainty. The result is model error. The relation between model errors and valuation and hedging performance is of critical importance for option traders. The real option we consider here is commodity storage, which allows merchants to trade a commodity at different prices over time. Our focus is specifically on contracts for natural gas storage capacity.

As described in the practice-based literature (Maragos 2002, pp. 440, 449-453, Eydeland and Wolyniec 2003, pp. 351-367, Gray and Khandelwal 2004a,b), traders use multifactor models for the term structure of futures prices to manage natural gas storage (Clewlow and Strickland 2000, Chapter 8, Seppi 2002, Eydeland and Wolyniec 2003, Chapter 5, Geman 2005, Chapter 3). We distinguish between two different approaches. The first, described by Eydeland and Wolyniec (2003, pp. 351-367) and Gray and Khandelwal (2004a,b), models the full term structure of the futures curve using as many factors as there are futures delivery dates over the term of a storage contract. This approach is related to the string and BGM models used to value fixed income options (Kennedy 1994, Brace et al. 1997, Longstaff et al. 2001). An alternative approach restricts attention to a few common factors that dominate the statistical dynamics of the futures curve. A typical number of factors is three, which have qualitative interpretations of level (shifting), slope (tilting), and curvature (bending) effects (Cortazar and Schwartz 1994, Clewlow and Strickland 2000, Chapter 8, Borovkova and Geman 2008, but see Borovkova and Geman 2006 and Bjerksund et al. 2008 for other numbers of factors).

This disagreement on the appropriate specification of the futures curve model for natural gas is the motivation for our analysis of term structure model error. The goal of our analysis is twofold: (i) To quantify the impact of different types of empirically calibrated term structure model error and (ii) to devise strategies to mitigate the impact of these model errors. We accomplish this using a flexible family of seasonal multifactor futures curve models (Blanco et al. 2002) that nests both full and lower dimensional specifications. We calibrate our price models to New York Mercantile Exchange (NYMEX) natural gas futures price data from 1997 through 2006. We find that three to four factors explain about 99% of the monthly price variance. We then use these empirically calibrated models to investigate the impact of three types of term structure model error: Model class error, in which the assumed risk neutral
dynamics of the futures curve used for valuation and hedging do not agree qualitatively with the correct risk neutral model; factor cardinality error, in which the assumed number of factors used for valuation and hedging is incorrect; and parameter error, in which the fitted factor loading coefficients are misestimated.

One complication with real options is that their value depends on an endogenous operating policy. Unfortunately, the curse of dimensionality makes computing the optimal inventory policy for natural gas storage impractical for the high-dimensional futures price models that we consider. Thus, we instead use a rolling intrinsic policy. There are two justifications for using this policy: First, although this is a heuristic policy, Lai et al. (2010) show that it yields near-optimal valuations when used with a full dimensional extension of the Black (1976) model (see also Nadarajah et al. 2011). Second, the rolling intrinsic policy is widely used in practice. Hence, the cash flows from natural gas storage that traders would, in practice, want to hedge are commonly generated by this type of operating policy.

We first study the impact of factor cardinality error on valuation. For twenty-three month storage contracts, we find that a futures price model with the most important three to six statistical factors obtains 98% of the value of storage relative to an empirically calibrated model with up to a full twenty-three factors. Hence, factor cardinality error due to omitted factors does not appear to be important for valuation purposes, and can be effectively dealt with by including the most important statistical future price factors in the valuation model. This finding is expected given our empirical calibration results.

We next study the impact of model error on delta hedging. There are several approaches to delta hedge option cash flows (Hull 2010, Chapter 6) depending on which contracts are used to do the hedging. All of these delta hedging strategies are equivalent in the absence of model error, but they may perform differently in the presence of model error. Bucket hedging trades futures contracts with delivery dates corresponding to each date on which physical spot trading has a cash flow, irrespective of the number of factors in the futures price model used for valuation. In contrast, factor hedging involves trading a number of futures contracts equal to the number of random factors in the futures price model (Cortazar and Schwartz 1994, Schwartz 1997, Clewlow and Strickland 2000, §9.5). We examine bucket hedging and two versions of factor hedging, tested using both simulated futures price paths and historical futures prices.

We find that bucket hedging is remarkably robust to empirically calibrated model error. In simulated data, it performs well in the presence of factor cardinality error and model class error (where excluded variability is idiosyncratic rather than factor-driven). In contrast, the performance of factor hedging with model error is very sensitive to which futures contracts are used to implement the hedges. For example, naïve factor hedging, which takes positions in futures contracts whose maturities are nearest to the spot trading dates, performs well without model error, but its performance is mostly disastrous in the presence of model error due to omitted factors or omitted idiosyncratic noise. In sharp contrast to our valuation results, ignoring factors that explain less than 1% of the observed price variance (or a similar amount of idiosyncratic noise) can lead to surprisingly catastrophic hedging performance. We obtain similar results for historical data, which arguably embeds model class error, factor cardinality error, and parameter error.

The difference in performance between naïve factor hedging and bucket hedging can be explained by the fact that the futures positions taken with naïve factor hedging can be extremely
large, while those of bucket hedging are not. These extreme positions tend to magnify model errors in realized prices. Bucket hedging is thus an effective approach to reduce the negative impact of model error on hedging performance. However, the parsimonious nature of the low dimensional factor hedging approach, which results in trading fewer futures contracts, is appealing. We thus develop fine-tuned factor hedging, which chooses futures contracts in a way that leads to less extreme positions relative to naïve factor hedging.

We show that the position in any individual futures contract in a factor hedging strategy can be characterized as the sum of (i) the delta of the storage option with respect to that futures price and (ii) a linear combination of additional deltas with coefficients inversely proportional to the determinant of a matrix of factor loadings corresponding to the traded futures contracts. Our fine-tuned factor hedging strategy chooses a set of traded futures contracts so as to maximize the determinant of the factor loading matrix. This yields trading positions that are much smaller than those obtained with naïve factor hedging. Both on simulated and real data, fine-tuned factor hedging dominates naïve factor hedging, achieving performance comparable to bucket hedging when the most important statistical factors are included.

Option deltas are an essential input to any delta hedging strategy. One additional contribution of this paper is that we derive pathwise delta expressions for a feasible inventory policy, under a Lipschitz continuity and derivative characterization assumption on its value function, by applying in a recursive manner the approach presented in Broadie and Glasserman (1996). We show that these expressions also hold for an optimal inventory policy, under a mild assumption on the operational parameters of our model. Use of these expressions within the rolling intrinsic policy is heuristic – as the rolling intrinsic value function can violate the Lipschitz continuity assumption – but it makes our analysis computationally tractable. Our numerical results suggest that the bias of our delta estimates, if any, is small.

Our analysis improves our understanding of the differential effect of realistic term structure model errors on the valuation and hedging of the natural gas storage real option, and provides an effective method to mitigate the negative effect of model error on factor hedging. It also may inform current practice: Due to its simplicity, practitioners are likely to favor bucket hedging over fine-tuned factor hedging. Practitioners also often prefer full dimensional over lower dimensional futures curve models. However, bucket hedging does not require a full dimensional model for valuation; our work suggests that the performance of bucket hedging is comparable whether it is based on a full or lower dimensional valuation model. Thus, current practice may entail redundancy in the number of factors used for valuation, albeit not in the number of contracts used for hedging. Although illustrated for natural gas storage, our research also has relevance for real and financial options which depend on other underlying term structures. These include storage real options for other commodities, commodity swing and put/call Bermudan options (Jaillet et al. 2004, Geman 2005, Chapter 12), and mortgages and interest rate options.

References


Optimal Operational *versus* Financial Hedging for a Risk-averse Firm

June 1, 2011
Wanshan Zhu
Lee Kong Chian School of Business, Singapore Management University, Singapore 178899, adamzhu@smu.edu.sg

Roman Kapuscinski
University of Michigan Business School, Ann Arbor, MI 48109, kapuscin@bus.umich.edu

A Multinational Risk-averse Newsvendor (MRN) produces goods at home (domestically) and sells both overseas and at home, over multiple periods. The MRN faces risks due to uncertain exchange rate, as well as uncertain demand. Intuitively, exchange rate risk should be managed by a finance department as financial risk, while uncertain demand risk should be managed by an operations department as a part of operational risk. Traditionally this is the practice of many firms. In this paper, we consider both of these risks jointly and investigate the effectiveness of two specific alternatives that allow MRN to reduce the total risk.

The first alternative is general financial hedging contracts (including futures, forwards, swaps, and options). The second one is operational hedging, which is based on optimally allocating production capacity between domestic and overseas facilities. We characterize the optimal capacity allocation decisions, financial hedging decisions, and the underlying production and transshipment decisions in a generalized model. Then, we compare the relative weaknesses and strengths of financial hedging and operational hedging. Our analysis allows for the case of either exogenous or endogenous prices.

*Key words:* operational hedging, financial hedging, operations management, risk analysis, present certainty equivalent value - PCEV
1. Introduction

Over the past two decades firms increasingly take advantage of overseas markets to both produce and sell goods. Overseas markets may, however, expose firms to the risk of an uncertain and even volatile exchange rates, which amplify the uncertainty in the earnings. Most firms do care about this uncertainty and attempt to decrease it, in order to reduce the chances of a financial distress or to directly increase the value of the firm. The company with decreased exposure to financial distress will have a lower cost of capital and a larger credit line and, thus, will provide a larger tax shield for the equity shareholder. Both the lower cost of capital and the larger tax shield are believed to increase the shareholder’s value, as discussed by Brealey and Myers (2003).

Two alternative tools are available for the MRN to hedge her exchange rate exposure. The first is financial hedging, which is often intended to explicitly target exchange rate risks. The second one is operational hedging, which assumes that total production capacity can be allocated between domestic and overseas facilities so that net foreign cash flow can be reduced, and hence, exchange rate risks are reduced. Both of these methods are frequently used in practice.

In the past two decades, as the result of breakthroughs in information technology, the variety of financial hedging contracts has grown tremendously and the cost to engage in them has become low. Consequently, many multinational enterprises have been using some of these contracts, often futures or forward contracts, to reduce their exchange rate risk exposures. For example, Cisco System Inc. (2009) states in its annual report “To reduce variability in operating expenses and service cost of sales caused by non-U.S.-dollar denominated operating expenses and costs, we hedge certain foreign currency forecasted transactions with currency options and forward contracts.” It seems intuitive that the financial hedging allows the MRN to benefit from all the advantages of the profitable overseas market, while not suffering from the consequences of a volatile exchange rate.

Operational responses experienced a similar growth: an increasing number of global firms place production facilities in multiple countries, capacity is divided between domestic and overseas locations, and transshipment is used whenever needed. While these operational responses (often labeled as operational hedging) may incur higher initial cost due to the need to establish overseas production capacity, they may also benefit the MRN in two ways. First, they allow the MRN to take advantage of the exchange rate movement by producing at a cheaper location. Second, they reduce the exchange rate risk exposure by using a part of overseas revenue for overseas operations costs.

Finance literature studies extensively the use of financial hedging to deal with currency-exchange risks, assuming risk-averse decision maker, see for example multiple papers in journals such as Risk Management, and Journal of Risk and Insurance. On the other hand, operational hedging has been studied in operations management literature under risk-neutral assumptions, although under various names. It is interesting that most of financial literature considers only financial instruments to deal with exchange-rate risk, while most of operations literature considers only operational instruments. Very few efforts have been made to assess their relative strengths and weaknesses in an integrated framework. The objective of this paper is to investigate what are the favorable conditions for the MRN to engage in either financial hedging or operational hedging, or both.

Our setting is similar to the ones described in Huchzermeier and Cohen (1996), Dasu and Li (1997), and Li et al. (2001). Over multiple periods, the MRN produces in two facilities, a domestic facility (Df) and an overseas facility (Of) to serve demands in domestic market (Dm) and overseas market (Om), see Figure 1. Demands in each period are uncertain and exchange rate changes from period to period according to Markov-modulated process. Dm’s demand can be satisfied by Df and, if needed, by Of’s production using transshipment. Similarly overseas market can be satisfied by Of and if needed by Df production. Transshipment and production costs are linear.

The MRN has a finite capacity and, in the beginning of the horizon, allocates the capacity between Df and Of. Such allocation is labeled as operational hedging. Then, the MRN operates the system for multiple periods in each period deciding how much to produce in each facility, how
much to transship, and what financial contract to sign. Due to finite capacity, some demand may be not satisfied and, therefore, lost. The papers above characterize how to run multi-location systems. While they do not state this, they effectively ask how to use operational hedging (the network capacities, production quantities) to deal with demand and exchange-rate uncertainties.

Both types of hedging that we study are special cases of this model: operational hedging means that we allocate the total capacity between both locations and consequently produce both in Df and O, but use no financial hedging; financial hedging means that all capacity is allocated to Df (consequently, we produce in Df only), while we are allowed to use any financial contracts in any period.

A critical element of our model is MRN’s risk-aversion. Even though it is argued that a firm should be risk neutral to non-systematic risks because the shareholders may be able to diversify away those risks, Smith (2004) in his review of risk attitudes points out that many firms, including large public companies, are risk-averse. Some practitioners of decision analysis (see Bickel et al. 2002) suggest deriving a firm-wide risk-averse utility function. We adopt Present Certainty Equivalent Value (PCEV) as defined in Smith (1998) to capture risk aversion. The basic model assumes that price is exogenous. As an extension, we consider price-sensitive demand, where MRN can adjust price from period to period.

Having introduced the problem, the rest of the paper seeks to analyze it. Theoretically, We find first that the firm’s optimal production decision take only finite number of values and its optimal transshipment decision features one direction – either from domestic to overseas or the other way, but not both. To maximize the PCEV, the firm also engages in optimal financial hedging. Second, we find the optimal financial contract equalizes the firm’s adjusted consumption across time and across different realizations of exchange rate. Furthermore, this optimal financial hedging implies an recursive equation for the PCEV that allows us to compute it efficiently. Third, The PCEV is concave on the domestic capacity, which implies a unique maximum PCEV with optimal operational hedging decision.

Through numerical study with reasonable normalized parameters, we find first that the operational hedging is on average better than financial hedging. Second, operational hedging becomes more effective than financial hedging with the increases of transportation cost, total production capacity, and demand coefficient of variation; but less effective with the increase of risk-aversion.
Third, both hedges are more effective with smaller profit margin and higher exchange rate coefficient of variation, but no one is preferred over the other.

As a background, Section 2 contains discussion of literature. In Section 3 we discuss the properties of PCEV. In Section 4 and Section 5 we present structural properties of optimal solution for the General Model; while in Section 6, we do so for the price-sensitive demand. In Section 7, we analyze the results of numerical analysis and discuss the relative strengths and weaknesses of the financial and operational hedging in detail. Finally, in Section 8, we present our conclusions.

2. Literature Review

Multiple streams of research are relevant to our topic. The first stream deals with multi-location production and its related issues of transshipment. The second stream studies the exchange rate uncertainty and its impact on operations under risk-neutral assumption. The third stream investigates financial and operational hedging, both empirically and theoretically, applying various risk-aversion frameworks. The fourth one describes fundamentals of modeling risk aversion and the strengths and weaknesses of different approaches.

Multi-location production
Since we allow production and sales to occur across countries, multi-location production and distribution literature is relevant. Examples from this broad literature include Krishnan and Rao (1965), who study a single-period N-location problem, or Tagaras (1989), who study the pooling effects of transshipment on service levels. Recently, Hu et al. (2007, 2008) study a general transshipment problem when capacity may be uncertain, Huang and Sosic (2010) study a game among retailers who transship goods among themselves. While these papers focus on the transshipment problem, for us it is a tactical subproblem. We concentrate on the impact of currency exchange rate uncertainty and potential hedging policies, and transshipment is an element of our operational hedging.

Currency exchange-rate uncertainty – risk-neutral approaches
A number of operations management papers address production and inventory decisions in the presence of exchange rate uncertainty. Many of them assume risk-neutral decision maker. Dasu and Li (1997) and Li et al. (2001) focus on the optimal production policies when the exchange rate uncertainty affects the production cost. They allow for fixed switchover cost and show that it is optimal to switch the production location from one to the other if the exchange rate is above one threshold, and to switch back if it is below another threshold. Li et al. assume no switchover cost and derive the optimal production as a function of initial inventory and exchange rate. Aytekin and Birge (2004) study a single capacity-allocation decision, followed by a series of production decisions, when exchange rate is uncertain, but demand is certain. Kazaz et al. (2005) consider a firm with production facilities in two foreign countries and find that the company may underserve the demand even when it has enough capacity. Dong et al. (2010) study a firm that chooses among facility networks of home only, foreign only or both, and show that the expected marginal profit determines the choices. While all these papers consider the exchange rate uncertainty, given risk-neutral framework, they do not consider using any financial instruments to hedge the risks and, consequently, do not compare the financial hedging versus operational hedging in a single framework. We next discuss the papers that study both types of hedges.

Financial hedging and operational hedging
A recent review of theoretical and empirical literature by Bandaly et al. (2010) reports that the term operational hedging is used for activities such as geographic dispersion (allocation of capacity among various locations), switching production (production decisions, given specific capacities in various countries, as in Huchzermeier and Cohen 1996), or capacity allocation postponement
(delaying the commitment of capacity to specific markets, as in Ding et al. 2007). These activities are often characterized by “mitigating risks by counterbalancing actions,” see Van Mieghem (2003). The majority of papers label operational hedging as dividing total available capacity across multiple locations (geographic dispersion). There is no ambiguity about financial hedging, which includes general financial hedges, currency derivatives, currency forwards, currency options, exotic derivatives, and foreign debt. Our paper considers allocation of capacity across countries as operational hedging (dominant interpretation in both theoretical and empirical literature, see below). Financial hedging is defined as use of any currency derivatives.

Empirical literature that studies operational versus financial hedging and it remains very active. Empirical papers define operational hedging exclusively as international and geographical dispersion. They acknowledge that companies use both financial and operational hedging, and ask which of them is used in practice and which of them is efficient or beneficial. The answers are not fully consistent.

Pantzalis et al. (2001) find support of hypothesis that the exposure of cash flows to currency risk is effectively managed by operational hedging. Allayannis et al. (2001) find that operational hedging is not an effective substitute for financial hedging and that operational hedging increases firms’ values only if used in conjunction with financial hedging. Both of these papers consider data from US companies. Kim et al. (2006), Aabo and Simkins (2005), and Faseruk and Mishra (2008) based on US, Danish, and Canadian data, respectively, find that the firms that use (or have a possibility of using) operational hedging are less likely to use financial hedging. Carter et al. (2001) suggests that firms that use both hedges face effectively smaller exchange-rate exposure. Logue (1995) notes that in a world with perfect financial markets, operational hedging would not be necessary. Since, however, the exchange rates do not perfectly follow purchase-parity and uncovered-interest-parity (theoretical) frameworks, companies should and do engage in operational hedging, labeled as natural hedge in his paper. More detailed description of the empirical literature can be found in Bandaly et al. (2010). It seems that empirical studies provide hypothetical explanations for individual interactions of financial and operational hedging, but to a smaller degree answer the question of when each of the hedging is more beneficial and what type of environment supports one versus the other. Our paper complements the empirical literature. Given lack of agreement which hedging is used and lack of answers about when they should be used, our intent is to explore the role of each of the types of hedging and to evaluate their individual and joint benefits.

The theoretical papers that attempt to compare financial and operational hedging appear both in finance and operations management literature. In theoretical finance literature dealing with currency exchange risk, a few papers consider both demand and exchange rate risks. Chowdhry and Howe (1999) in a one-period model use mean-variance objective to study financial and operational hedging. The main conclusion is that operational hedging is needed only in the presence of both demand and exchange rate uncertainty. Hommel (2003) extends the above model by including quadratic cost for foreign capacity and concludes that the operational hedging is only useful for sufficiently high exchange rate variability. This is not the case in our model.

A subset of operations management literature includes consideration of financial contracts, but without focusing on the relationship between financial and operational hedging. For example, Gaur and Seshadri (2005), Agrawal and Seshadri (2000), and Caldentey and Haugh (2006). A few operations management papers consider the relationship between financial and operational hedging.

Chod et al. (2010) bears some similarity to our work. The uncertainty is due to demand, but some portion of this uncertainty can be hedged financially. For example if demand is correlated to weather, market-traded weather derivatives can be used. Their main focus is on a comparison of two specific forms of operational flexibility (postponement and flexible capacity) and asking about value of financial hedging for each of them. Their model, however, does not compare operational versus
financial hedging. (Also, it is based on a single production facility, does not consider currency-exchange rate.)

A number of papers that study operational and financial hedging, explicitly consider the exchange rate uncertainty. Huchzermeier and Cohen (1996) consider an international network of facilities. Even though the decision maker is risk neutral, they illustrate how use of forward contracts decreases the dispersion of profits. Ding et al. (2007) study a risk-averse producer who can postpone its production decision till the exchange rate is realized. For a single-period model with mean-variance objective function they derive the optimal portfolio of call and put options and characterize the effect of postponement. They do not ask when financial hedging and when operational hedging should be used. Li and Wang (2010) extend their model to consider financial and operational hedging for more than two countries using more general financial hedging contracts. They continue to use a mean-variance objective function in a single-period model, where operational hedging is defined as postponement of production. They show numerically that financial hedging and operational hedging are substitutes. Our research complements their work by studying the broadly accepted case of geographic dispersion (dividing total capacity across multiple countries). Also, our model is a dynamic one. To the best of our knowledge, we are the first to consider the question of financial versus operational hedging in multi-period setting. Importantly, we also use PCEV framework (as explained below), which is not linked to shortcomings of many other evaluation criteria.

Risk-aversion Models

Risk aversion is central to a few sub-streams of economics and finance literature. Recently it has been increasingly studied in operations literature. Since we study decisions of a risk-averse firm over multiple periods, we start by discussing how risk and time preference are modeled.

Two major approaches to modeling the risk attitude are: (a) Value-at-Risk (VaR), and (b) various forms of utility functions. VaR is defined as the maximum loss of value that a firm can incur for a given confidence level and a given time interval. Many financial firms use Value-at-Risk to manage day-to-day operations, see Manganelli and Engle (2001). Tapiero (2005) applies VaR to a single-period inventory control problem and finds that the target cost level can be interpreted as VaR. Recently Devalkar et al. (2010) applies VaR to study the optimal decisions of a firm that procures, trades, and processes commodities.

Most theoretical models of risk-averse behavior employ a utility function, where a cash flow \( x \) in a given period is translated into concave and increasing utility function \( U(x) \), see survey by Levy (1992). Levy points out that ranking risk by non-decreasing concave utility function is consistent with the natural definition of risk premium given by Arrow Arrow (1951) and Pratt (1964) – the total utility serves as an index for risk. Gerber and Pafumi (1998) review the most often used utility functions: (i) linear, (ii) exponential, (iii) negative quadratic, and (iv) logarithmic.

Obviously, linear utility function means risk neutrality, and it is often implicitly assumed in models that do not deal with risk. Negative quadratic utility is often used in finance literature, see survey by Steinbach (2001). Its popularity is partly due to technical reasons, as it is based on only two parameters, the mean and variance of the cash flow. Since the quadratic utility function has its obvious shortcoming, penalizing superior performances (it increases only up to a certain threshold), general concave utility functions for risk-averse modeling have been widely used in economics, see survey by Fishburn (1989). In operations literature, Eeckhoudt et al. (1995) employ a concave increasing utility function to study a typical single-period news-vendor problem. Their model is extended to multiple resource capacity investment by Van Mieghem (2007) and to multi-period dynamic inventory control with set-up cost by Chen et al. (2007).

While the approaches to one-period risk modeling are relatively well agreed on, for multiple periods, two fundamental issues are intertemporal consistence and temporal risk preference. The optimal policy formulated for future periods is intertemporally consistent, if it will be carried out
as planned. To provide such consistence, economic literature typically uses discounted utility model (discounted sum of utilities of cash flows across all time periods), first proposed by Samuelson (1937) and later shown by Samuelson (1952) and Pollak (1973), to give consistent choices, when risk preference is assumed independent of time. The discounted utility model has been used for several decades, with several generalizations, for example Prakash (1977) and Fishburn (1989). Frederick et al. (2002) review extensive applications of this model and its variations in economic literature.

Early economics literature considered an alternative axiomatic approaches. Sobel (2006) provides a summary of these approaches and describes difficulty of reconciling these with the accepted evaluation criteria. For the axiomatic approach he describes (somewhat different from those used in economics literature), he shows that these axioms “carry the seeds of [imply] risk neutrality” even for single-period exponential utility with stochastic payoffs. To the best of our knowledge, the only non-trivial model compatible with these axioms, is the utility of the sum of discounted cash flows, studied in Bouakiz and Sobel (1992). This method, however, does not address the temporal risk preference (see below).

The second fundamental issue in modeling multi-period risk-aversion is the “temporal risk preference” that represents a decision maker’s preference in timing of uncertainty resolution. In general an earlier resolution of the uncertainty is preferred because it gives the decision maker more information to formulate better plans for the future consumption. The original models of discounted utility of cash flows alone, without allowing for a transfer of certain cash between periods (through bond investments), even though broadly used, do not capture this temporal risk preference. ¹

Two modified utility-based approaches correct this problem and capture temporal risk preference: recursive utility function and present certainty equivalent value (PCEV).

Kreps and Porteus (1978) first proposed recursive utility procedure to account for the timing of uncertainty resolution. Epstein and Zin (1989) generalized their models to allow independent risk-aversion specifications for a given period and across periods. Even though the model of discounted utility dominated the theoretical literature for years, the Epstein-Zin framework has now become the most used model for dynamic asset pricing applications in economic and finance literature. Many textbooks, including Singleton (2006) and Duffie (2001), discuss use of this framework in the context of specific applications. The general framework is specified by three elements: a recursive utility function, an aggregator function and a certainty equivalent operator. The freedom in choices for these elements makes the framework very powerful, yet often too general. It turns out that PCEV, as defined by Smith (1998) corresponds to a particular choice for the elements, where the recursive utility and the aggregator functions are additive exponential and the certainty equivalent operator is the PCEV operator. This particular choice makes the PCEV more conducive to any efficient evaluation, while capturing the decisions maker’s risk preference that is sensitive to both uncertainty and the time at which the uncertainty is resolved (see Smith 1998 for further discussion). Since PCEV is critical part of our model formulation, we now formally describe PCEV with risk-free bond and PCEV with financial hedging.

3. Present Certainty Equivalent Value (PCEV)

The underlying idea of Present Certainty Equivalent Value (PCEV) is to translate a stream of uncertain cash flows across multiple periods into a single certain cash value today. This is similar to the concept of Net Present Value (NPV), except in every period (concave and increasing) utility function is used to evaluate the period’s consumption and shifting money from period to period (borrowing and lending using risk-free bond) is possible. Thus, PCEV is a lump-sum amount

¹ The earlier version of this paper assumed discounted utility framework. We thank John Birge, James E. Smith, and Bryan Routlage for pointers to PCEV and Epstein-Zin frameworks.
received with certainty today that generates the same consumption utility as a stream of uncertain operational cashflows across multiple periods. Dréze and Modigliani (1972) developed the PCEV concept to study the optimal consumption decision in two periods with uncertain income. Smith (1998) extended their model to multi-period and, assuming exponential utility function, derived a recursive formula for PCEV. While utility function can be defined very generally, we follow the assumptions of Smith (1998) and consider additive exponential utilities. Exponential function has been applied widely in decision analysis practice (see, e.g., Howard 1988) and financial risk theory (see, e.g., Gerber and Pafumi 1998).

Let \( n \) be the index of the current period, with the periods counted backwards and the last period being \( 0 \). With consumption in period \( i \) denoted by \( \gamma_i \), utility function is additive across periods, i.e., \( U(\gamma_n, \gamma_{n-1}, \ldots, \gamma_0) = \sum_{i=n}^{0} u_i(\gamma_i) \) and each-period utility function has the form \( u_i(\gamma_i) = -w_i e^{\gamma_i/\rho_i} \), where \( w_i \) is period’s weight and \( \rho_i \) is risk tolerance, capturing respectively the decision maker’s time and risk preferences for consumption.

### 3.1. PCEV with Risk-free Bond

PCEV allows for borrowing and lending (investment in bonds) at risk-free interest rate in order to shift consumption across periods. Denote \( r_f \) as the risk-free interest rate. Let \( \beta_i \) be the value of bonds held at the beginning of the period, also referred to as wealth. In period \( i \), the firm realizes cash from operations \( f_i \), and then chooses its bond investment to be held in period \( i - 1 \), \( \beta_{i-1} \), for which it will pay \( \beta_{i-1}/(1 + r_f) \). Thus, the consumption in period \( i \) is \( f_i + \beta_i - \beta_{i-1}/(1 + r_f) \). Since consumption ends in period \( 0 \), we impose \( \beta_{-1} = 0 \). For given initial wealth \( \beta_n \), the maximum consumption utility of the stream of operational cash flows \( f_n, \ldots, f_0 \) is

\[
U_n(f_n, \ldots, f_0|\beta_n) = \max_{\beta_i, 0 \leq i < n} \left\{ \mathbb{E}\left[\sum_{i=0}^{n} \left( -w_i \exp\left( -\frac{f_i + \beta_i - \beta_{i-1}/(1 + r_f)}{\rho_i} \right) \right) \right] \right\}. \tag{1}
\]

The expectation is with respect to all future uncertainties.

Note that the maximum consumption utility of a deterministic cash flow in a single period is also evaluated using (1). Intuitively, the firm will spread the wealth across all periods according to each period’s utility functions. PCEV is defined as a deterministic cash flow in the initial period with the maximum utility generated by the operational cash flow and appropriate bond investments:

**Definition 1.** The present certainty equivalent value (PCEV) at period \( n \) of an uncertain operational cash flow \( f_n, \ldots, f_0 \) is defined as \( \mathcal{V}_n \) such that

\[
\mathcal{U}_n(\mathcal{V}_n, 0, \ldots, 0|\beta_n) = U_n(f_n, \ldots, f_0|\beta_n). \tag{2}
\]

Clearly, PCEV \( \mathcal{V}_n \) is a function of current wealth \( \beta_n \) and distributions of future cash flows, i.e., \( \mathcal{V}_n = \mathcal{V}_n(f_n, \ldots, f_0|\beta_n) \).

### 3.2. PCEV with Financial Hedging

The same concept of PCEV applies when a firm uses financial hedging. In our model, foreign currency exchange rate in period \( i \), \( S_i \) is a Markov-modulated process. We denote realization of \( S_i \) as \( s_i \). Financial hedge is any foreign-currency derivative, i.e., a future payoff as a function of the future realized value \( s_i \) of the foreign currency. It is easy to show that, in each period the firm needs to consider only one-period forward hedges.\(^2\)

Let \( \beta_{i-1}(S_{i-1}) \) be the payoff in period \( i - 1 \) of a derivative that the MRN buys in period \( i \). The market price of the derivatives, under risk-neutral measure, is \( \mathbb{E}[\beta_{i-1}(S_{i-1})|S_i]/(1 + r_f) \). As a result of financial hedging, the consumption in period \( i \) is \( f_i + \beta_i(s_i) - \mathbb{E}[\beta_{i-1}(S_{i-1})|S_i]/(1 + r_f) \).

\(^2\)Any longer horizon hedges would decompose into one-period hedge and a hedge decided in the next period.
Let \( \beta_n(s_n) \) be the initial wealth and let \( \beta_{-1}(\cdot) = 0 \) due to the end of consumption in period 0, then the maximum utility of a uncertain operational cash flows becomes

\[
\mathcal{U}_n(f_0, \cdots, f_n|\beta_n(s_n)) = \max_{\beta_i(s_i), s_0 \leq i < n} \left\{ E\left[ \sum_{i=0}^{n} \left( -w_i \exp\left( -\frac{f_i + \beta_i(S_i)}{\rho_i} \right) \right) \right] | S_n = s_n \} \right. (3)
\]

The definition of PCEV remains the same as in (2). This definition implies that the \( \mathcal{V}_n(\cdot) \) is additionally a function of \( s_n \).

4. Model

In this section we describe the general model that allows for both financial and operational hedging. The MRN has a finite total capacity. The first decision (operational hedging) is how to allocate the total capacity between Df and Of at the beginning of horizon. Once this decision is made, Df’s and Of’s capacities do not change in any later periods. With no operational hedging, all capacity is domestic. Financial hedging is a part of every-period decisions.

Then, in every period firm faces demands in both markets. Demands are independent from period to period, but not necessarily independent across locations. The exchange rate follows a Markov-modulated process with a finite number of possible realizations. No inventory can be held from period to period. The sales price is assumed to be exogenous. This assumption is especially appropriate in competitive market, where the MRN may be forced to be a price taker (Section 6 allows MRN to set the price). The events in each period are as follows. (1) current-period exchange rate is revealed, (2) current-period demands become known, (3) production and transshipment decisions are made, (4) operational cash flow is determined, (5) exchange rate derivatives (financial hedging) are decided, and (6) the consumption takes place.

In (3) the firm decides how much to produce at each location and how much to transship from a location to the other. In (5), if financial hedging is used, any financial derivative is allowed. Otherwise, the firm may only invest in bonds.

In each period PCEV is calculated after the realization of exchange rate, but before demand realization. The following notation is used. The first six symbols are illustrated in Figure 1.

- \( d \) and \( o \) superscripts of domestic and overseas markets or facilities, respectively,
- \( \xi^j \) market \( j \) demand, \( \xi = \{ \xi^d, \xi^o \} \),
- \( k^j \) capacity of facility \( j \), \( k = \{ k^d, k^o \} \),
- \( z^j \) quantity produced in facility \( j \) for market \( j \), \( z = \{ z^d, z^o \} \),
- \( x^j \) quantity produced in the other facility and transshipped to market \( j \), \( x = \{ x^d, x^o \} \),
- \( y^j = z^j + x^j \) total production quantity available to meet market \( j \) demand, \( y = \{ y^d, y^o \} \),
- \( K \): total capacity, \( K > 0 \),
- \( p \): unit price, \( p > 0 \),
- \( c \): unit production cost, \( c \in (0, p) \),
- \( t \): transportation cost, \( t \geq 0 \).

The exchange rate \( S_i \) is assumed to be independent of demand \( \xi \). The sale price \( p \) is assumed to be exogenous (which is relaxed in Section 6). Finally, we assume that no inventory is carried over across periods. Thus, the only interaction across periods is through the exchange rates. The

\footnote{Note that the underlying (natural) probability measures of the firm is not necessarily the same as the public exchange rate probability (risk-neutral one). It is possible to show that the effect of using natural probability measure, instead of risk-neutral one, is additive. The firm will invest additional financial hedges driven by the difference between the probability measures and these hedges will not depend on any of the uncertain cash flows. Thus, as in Smith and Nau (1995), we replace the the natural (private) probability by the risk-neutral (public) one, effectively reducing incomplete market into an equivalent complete one. Consequently, our financial derivatives are evaluated under risk-neutral measure.}
MRN’s objective is to maximize the PCEV of all future cash flows. The MRN’s problem can be simplified by two observations. First, the production and transshipment decisions can be made for each period independently from other periods. Second, in each period, maximizing profit is equivalent to maximizing PCEV.

\[
\max_{k^d \geq 0, k^o \geq 0, k^d + k^o \leq K} \mathcal{V}_n(f_n(s_n, k, \xi_n), \ldots, f_i(s_i, k, \xi_i), \ldots, f_0(s_0, k, \xi_0)),
\]

where one-period maximum cash flow is

\[
f_i(s_i, k, \xi_i) = \max_{(z, x) \in A(k, \xi_i)} \left\{ (p - c)z^d + (p - cs_i - t)x^d + s_i(p - c)z^o + (s_ip - c - t)x^o \right\},
\]

\[
A(k, \xi_i) = \{(z, x) : z^j + x^j \leq k^j, \forall j \in \{d, o\}, \forall l \in \{d, o\}, z + x \leq \xi_i, z \geq 0, x \geq 0\}.
\]

Note that financial hedging decisions are imbedded in the definition of \(\mathcal{V}_n\). We will explicitly derive these in the next section. The objective function in (5) is revenue minus costs for given production decisions \((z, x)\) in period \(i\). The constraints on \((z, x)\) in (6) ensure that the production at each facility does not exceed capacity, and sales at each market do not exceed demand. The resulting maximum operational cash is a function of only the exchange rate and demand in period \(i\). Also, the production and transshipment decisions of a period have no impact on other period’s operational cash. Maximizing the period’s operational cash flow also maximizes PCEV, since operational cash flow, at the time operational decisions are taken, is certain). This allows us to explicitly use (5).

5. Analysis

To analyze the problem, we start with initial properties of the profit function \(f\) in (5). These are building blocks for the properties of capacity decisions, described later. The period index \(i\) is suppressed for exchange rate, \(i.e., s = s_i\), unless specified otherwise.

5.1. Properties of One-period Profit Function

Function \(f\) captures the operational cash flow in a single period. Lemma 1 below formally states that the cash flow increases in capacity and demand. Also, groups of capacity and demand decisions are complementary (across locations), but within each group (demands or capacities) they are substitutes.

**Lemma 1.** (a) \(f\) is non-decreasing and concave in \((k, \xi)\).

(b) \(f\) is submodular in \((k, -\xi)\).

All omitted proofs are in the appendix.

Substituting \(y = z + x\), which denotes sales, see Figure 1,

\[
f(s, k, \xi)\) in the production problem can be expressed as follows:

\[
f(s, k, \xi) = \max_{(y, x) \in A_1(k, \xi)} (p - c)y^d + s(p - c)y^o + (s - 1)c(x^o - x^d) - t(x^o + x^d)
\]

\[
A_1(k, \xi) = \{(y, x) : \xi \geq y \geq x \geq 0, y^d \leq k^d + x^d - x^o, y^o \leq k^o + x^o - x^d\}
\]

The first constraint in the constraint set \(A_1\) states that sales must be greater than transshipped quantity but less than demand. The last two constraints state that the sales are bounded by the sum of a facility’s local capacity and net transshipment from the other facility. Since local capacity plus net transshipment define available quantity to the local market, we immediately have:
Lemma 2. For a given feasible production decision $x$, the optimal sales are $y^d = (k^d + x^d - x^o) \land \xi^d$ and $y^o = (k^o + x^o - x^d) \land \xi^o$.

Note that for any feasible $y^*$, we need $x \leq y^*$. Thus, a transshipment problem can be expressed in terms of $x$ only:

$$f(s, k, \xi) = \max_{x \in A_2(k, \xi)} (s - 1)c(x^o - x^d) - t(x^o + x^d) + (p - c)((k^d + x^d - x^o) \land \xi^d) + s(p - c)((k^o + x^o - x^d) \land \xi^o)$$

$$A_2(k, \xi) = \{x : \xi \geq x \geq 0, x^o \leq k^d, x^d \leq k^o\}$$

(8)

(9)

The above problem can be interpreted as an assignment of capacity to locations. Through the assignment, the available quantities $(k^d + x^d - x^o)$ and $(k^o + x^o - x^d)$ attempt to reach the targets, i.e., the demand in both markets.

To characterize the optimal solution of the transshipment decisions, let $\bar{k} = k - \xi$ and $\bar{K} = k^d + k^o$. A positive (or negative) $\bar{k}$ represents the capacity-shortage (or capacity-overflow) for a given demand. $\bar{K}$ is the total capacity net total demand. Let $\bar{p} = p - c$, the profit margin of sales. We also let $\bar{\bar{t}}^d = (1 - s)c - t$ and $\bar{\bar{t}}^o = (s - 1)c - t$. $\bar{\bar{t}}^d$ represents the relative cost difference of meeting domestic demand using overseas production, and $\bar{\bar{t}}^o$ vice versa.

Theorem 1. Optimal transshipment decisions are $x^d = (k^d \land \xi^d \land \bar{\bar{t}}^d)^+$ and $x^o = (k^o \land \xi^o \land \bar{\bar{t}}^o)^+$, where

$$\bar{x}^d = \begin{cases} -\infty & \text{if } \bar{\bar{t}}^d \leq -\bar{\bar{p}} \\ -k^d & \text{if } \bar{\bar{K}} \geq 0 \text{ and } 0 \geq \bar{\bar{t}}^d \leq -\bar{\bar{p}}, \text{ or } \bar{\bar{K}} = 0 \text{ and } s\bar{p} \geq \bar{\bar{t}}^d \geq (s - 1)\bar{\bar{p}} \\ \bar{\bar{k}}^d & \text{if } \bar{\bar{K}} \geq 0 \text{ and } \bar{\bar{p}} \geq \bar{\bar{t}}^d \geq 0, \text{ or } \bar{\bar{K}} \leq 0 \text{ and } (s - 1)\bar{\bar{p}} \geq \bar{\bar{t}}^d \geq -\bar{\bar{p}} \\ \infty & \text{if } \bar{\bar{t}}^d \geq \bar{\bar{p}} \end{cases}$$

(10)

$$\bar{x}^o = \begin{cases} -\infty & \text{if } \bar{\bar{t}}^o \leq -s\bar{p} \\ -k^o & \text{if } \bar{\bar{K}} \geq 0 \text{ and } 0 \geq \bar{\bar{t}}^o \leq -s\bar{p}, \text{ or } \bar{\bar{K}} = 0 \text{ and } \bar{\bar{p}} \geq \bar{\bar{t}}^o \geq (1 - s)\bar{\bar{p}} \\ \bar{\bar{k}}^d & \text{if } \bar{\bar{K}} \geq 0 \text{ and } \bar{\bar{p}} \geq \bar{\bar{t}}^o \geq 0, \text{ or } \bar{\bar{K}} \leq 0 \text{ and } (1 - s)\bar{\bar{p}} \geq \bar{\bar{t}}^o \geq -s\bar{p} \\ \infty & \text{if } \bar{\bar{t}}^o \geq \bar{\bar{p}}. \end{cases}$$

(11)

The interpretation of Theorem 1 is quite intuitive. The higher the cost differences, the higher the target transshipment level $\bar{x}^d$. Since the marginal benefit changes only at $-k^d$ and $\bar{k}^d$, i.e., the domestic capacity shortage and the overseas capacity overage, the target transshipment level is one of these values.

Note that many papers (including Chowdhry and Howe 1999) assume complete capacity pooling, where all demands are satisfied as long as the total capacity is higher than or equal to the total demand. Certainly this is not optimal if the exchange rate changes dramatically. To see this, let us consider the following example.

Example Let $k^d = 10$, $\xi^d = 2$, $k^o = 3$, $\xi^o = 5$ and $s = 0.4$, $p = 10$, $c = 5$. In this case, it is optimal to use all overseas capacity and no domestic capacity at all. Two units of overseas capacity are used for domestic demand and one unit for overseas demand. Adding any domestic production has negative marginal benefit because there are only two uses of domestic production: (1) meeting domestic demand and pushing overseas product to overseas demand, where the marginal benefit is $p - c + ps - p = -1$; and (2) meeting overseas demand; then the marginal benefit is also -1 because $ps - c = -1$. 
5.2. Properties of Bond Investments and of Financial Hedging

Having characterized the properties and optimal solutions of the production and transshipment decisions, we now proceed to study the properties of bond investments (borrowing and lending) and of financial hedging decisions. Since bond investments can be treated as a very special case of financial hedging, we describe them first.

5.2.1. Bond Only

Investment in bonds allows MRN to save cash in sunny period for a rainy day. With no discounting and equal weights, we would attempt to divide cash equally across all periods. In order to precisely capture these investments in general cases, we use two intuitive economic terms. Let the effective risk tolerance of period $i$ be defined as $R_i := \sum_{j=0}^{i}(\rho_j(1 + r_f)^{i-j})$ and the windfall $w'_i := \sum_{j=0}^{i}(\rho_j \ln((1 + r_f)^j)/w_j)/(1 + r_f)^{i-j})$.

The effective risk tolerance $R_i$ is the sum of discount risk tolerance in all periods from $n$ to 0. It reflects the decision maker’s ability to use bond to spread risk across multiple periods. It also takes into account the inter-temporal risk preference: an earlier resolution of the cash flow uncertainty is more beneficial than a later one.

The windfall $w'_i$ is a cashflow (positive or negative) in the current period that combined with weights of 1 in all periods, generates the same PCEV. It consists of discount factor from period $j$ to $i$ is $1/(1 + r_f)^{i-j}$, which is applied to the period $j$’s cash equivalent value of $w_j$.

Using these definitions, we first obtain the analytical solution of PCEV for a given single uncertain operational cash flow.

**Lemma 3.** For a given initial wealth $\beta_n$ and a given uncertain operational cash flow with $f_n$ in period $n$ and with 0 in all other periods, the optimal bond investment satisfies

$$\frac{\beta_i + w'_i}{R_i} = \frac{f_n + \beta_n + w'_n}{R_n}, \forall 0 \leq i < n;$$

(12)

the maximum utility is

$$U_n(f_n, 0, \cdots, 0|\beta_n) = E[-R_n(1 + r_f)^n \exp(-\frac{f_n + \beta_n + w'_n}{R_n})], \forall n \geq 0;$$

(13)

and the present certainty equivalent value is independent of $\beta_n$,

$$V_n(f_n, 0, \cdots, 0) = -R_n \ln \left( E[\exp(-\frac{f_n}{R_n})] \right), \forall n \geq 0.$$  

(14)

The optimal investment in (12) implies that the decision maker spreads cash consumption, adjusted by the windfalls $w'_i$, across time in proportion to the effective risk tolerance. As a result, the decision maker equalizes every period’s discounted utility and her total expected utility becomes (13). This maximum utility is the sum of utilities discounted by risk tolerance adjusted discount rate because $R_n(1 + r_f)^n = \sum_{j=0}^{n}(\rho_j(1 + r_f)^{n-j}).$ The expression of (14) shows that PCEV does not depend on the initial wealth $\beta_n$. Intuitively, PCEV is defined as an additional certain cash flow equivalent to uncertain cash flows ($f_n, \ldots, f_0$). With exponential utility function, the initial wealth is divided across periods such that it increases each period’s utility by the same multiplier (i.e., equally if all weights and risk tolerances were the same). Furthermore, (14) shows that while PCEV depends on the effective risk tolerance, it is independent of weights of any period. Different weight of future period is equivalent to a windfall, which is a certain operational cash flow. Thus, by the same logic as that for the initial wealth it does not influence PCEV. We have established these convenient simplifications (initial wealth does not matter and the weight of a period does not matter) for a single-period cash flow. We next extend these properties to any cash flows.
The optimal derivative investment satisfies

\[ V_n(f_n, \ldots, f_0) = V_n(f_n + \frac{V_{n-1}(f_{n-1}, \ldots, f_0)}{1+r_f}, 0, \ldots, 0); \quad (15) \]

the optimal bond investment satisfies

\[ \frac{\beta_{n-1} + w'_{n-1} + V_{n-1}(f_{n-1}, \ldots, f_0)}{R_{n-1}} = \frac{f_n + \beta_n + w'_n - (\beta_{n-1} + w'_{n-1})/(1+r_f)}{\rho_n}; \quad (16) \]

and the maximum utility is

\[ U_n(f_n, \ldots, f_0|\beta_n) = E[-R_n(1+r_f)^n \exp(-f_n + \beta_n + \frac{V_{n-1}(f_{n-1}, \ldots, f_0)}{1+r_f} + w'_n)/R_n)]. \quad (17) \]

Some of these results are direct extension of the results of Lemma 3: the optimal investment (16) means that the risk tolerance and time weight adjusted consumption cash are spread equally across periods. As a result, the maximum utility (17) is the sum of utilities discounted by risk tolerance adjusted discount rate. The critical result is the recursive relationship (15). It implies that PCEV can be computed by adding discounted future PCEV to the current uncertain cash flow. Using (15) and (14), we can evaluate the PCEV of any given uncertain operational cash flow. Clearly, weight of the period and initial wealth do not influence PCEV for multiple cash flows.

5.2.2. Financial Instruments

Bond investments can be viewed as most restrictive form of financial hedging: the investment and the payoff are certain. General financial derivative can be any function of currency exchange rate. Intuitively, allowing for general form of financial hedging: the investment and the payoff are certain. General financial derivative can be computed by adding discounted future PCEV to the current uncertain cash flow. Using (15) and (14), we can evaluate the PCEV of any given uncertain operational cash flow. Clearly, weight of the period and initial wealth do not influence PCEV for multiple cash flows.

Theorem 2. Consider a general cash flow with operational cash flow \( f_i \) in each periods for all \( 0 \leq i \leq n \) and initial wealth \( \beta_n(s_n) \). For all \( n \geq 1 \), the present certainty equivalent value is independent of \( \beta_n(s_n) \), and

\[ V_n(f_n, \ldots, f_0) = V_n(f_n + \frac{E[V_{n-1}(f_{n-1}, \ldots, f_0)|S_n = s_n]}{1+r_f}, 0, \ldots, 0). \quad (18) \]

The optimal derivative investment satisfies

\[ \frac{\beta_{n-1}(s_{n-1}) + w'_{n-1} + V_{n-1}(f_{n-1}, \ldots, f_0|s_{n-1})}{R_{n-1}} = \frac{f_n + \beta_n(s_n) + w'_n - E[\beta_{n-1}(S_{n-1})|S_n = s_n]}{\rho_n}; \quad (19) \]

and the maximum utility is

\[ U_n(f_n, \ldots, f_0|\beta_n(s_n)) = E[-R_n(1+r_f)^n \exp(-f_n + \beta_n + \frac{E[V_{n-1}(f_{n-1}, \ldots, f_0)|S_n = s_n]}{1+r_f} + w'_n)/R_n)]. \quad (20) \]
The results under derivative investment differ significantly, when compared bond-only investments. Based on (19), the optimal derivative is such that the adjusted consumption cash are equalized across periods, which was the case for the optimal bond investments. However, the adjusted consumption is also equalized across realizations of exchange rate in any given periods: since right-hand side of (19) is independent of exchange rate \( s_{n-1} \), the same must hold for the left-hand side. Furthermore, in PCEV recursion for bond-only investment (15), it was necessary to evaluate \( f_n + V_{n-1}/(1 + r_f) \), where \( V_{n-1} \) was a function of exchange rate \( s_{n-1} \). However, PCEV recursion for financial hedging (18) uses \( f_n + E V_{n-1}/(1 + r_f) \), where \( E V_{n-1} \) is independent of exchange rate \( s_{n-1} \). Thus, the latter equalization further improves the decision maker’s present certainty equivalent value by adding, to the current uncertain cash flow, the expected future PCEV discounted at risk free rate, as shown by (18). Using (14) and (18), we can evaluate PCEV of any operational cash flow when financial hedging is used to improve consumption utility.

Since simple financial contracts, in particular, forward contracts, are the most often used contracts in practice, it is interesting to ask about their relative effectiveness compared to optimal contracts. Clearly, their simplicity is attractive. However, one would expect that limiting oneself to forwards only might leave substantial savings unexplored. It turns out, however, that forward contracts can be optimal when exchange rate follows a certain dynamics.

**Theorem 3.** If the exchange rate follows a birth-death process (i.e., goes either one state up or one state down), then the optimal financial hedging is to use forward contract and risk-free bond.

When the exchange rate follows a birth-death process, then for its given current value, it has two realizations in the next period. Consequently, any financial contract, which is expressed as two (correctly chosen) payoff values at these two realizations will be optimal. Since forward contract is a linear payoff function, the desired optimal payoff function can be expressed as a linear combination of a forward contract and a bond. In other words, we can replicate the optimal contract by using forward contracts and risk-free bonds. This result is consistent with theory where textbooks (e.g., Luenberger 1998 or Hull 2009) show that exchange rate is often modeled as a binomial tree, which is a special case of the birth-death process. It is also consistent with practice where many firms choose to use forwards, presumably because binomial tree fairly closely reflects the behavior of currency exchange rates in time buckets considered by firms.

### 5.3. Properties of Operational Hedging

While financial hedging decisions take place in every period, operational hedging (allocation of capacity) is a major decision, taken once in the beginning of the horizon. Having characterized the properties and optimal solutions of the production and transshipment decisions as well as optimal financial hedging decisions, we now proceed to study the properties of the optimal capacity decision that maximizes MRN’s PCEV. While the maximum utility is joint concave in domestic and overseas capacity allocation, it is not obvious whether the PCEV is also concave. The following lemma is helpful to show concavity of PCEV.

**Lemma 5.** If a real valued function \( f_i(x) \) on \( x \in \mathbb{R}^n \) is concave for all \( i \), then \( \ln(\sum_i \exp[-f_i(x)]) \) is convex in \( x \).

**Proof:** By definition of convexity, we need to show that for any \( 0 < \lambda < 1 \)

\[
\lambda \ln(\sum_i \exp[-f_i(x_1)]) + (1 - \lambda) \ln(\sum_i \exp[-f_i(x_2)]) \geq \ln(\sum_i \exp[-f_i(\lambda x_1 + (1 - \lambda)x_2)])
\]

for any \( x_1 \) and \( x_2 \). Since exponential function is monotone, taking exponential on both sides does not change the inequality. Thus, it is equivalent to show

\[
\sum_i \exp[-f_i(x_1)]^\lambda (\sum_i \exp[-f_i(x_2)])^{1 - \lambda} \geq \sum_i \exp[-f_i(\lambda x_1 + (1 - \lambda)x_2)].
\]
Since \( f_i(x) \) is concave, \(-\lambda f_i(x_1) - (1 - \lambda)f_i(x_2) \geq -f_i(\lambda x_1 + (1 - \lambda)x_2)\). This implies
\[
\sum_i \exp[-\lambda f_i(x_1) - (1 - \lambda)f_i(x_2)] \geq \sum_i \exp[-f_i(\lambda x_1 + (1 - \lambda)x_2)].
\]

Hence, it is sufficient to show
\[
((\sum_i \exp[-f_i(x_1)])^\lambda(\sum_i \exp[-f_i(x_2)])^{1-\lambda}) \geq \sum_i \exp[-\lambda f_i(x_1)] \exp[-(1 - \lambda)f_i(x_2)].
\]

This inequality follows directly from the Hölder’s inequality.

**THEOREM 4.** PCEV, expressed as the objective function in (4), is concave in \( k \).

**Proof:** We now show by induction that \( V_n(\cdot) \) is concave in \( k^d \). For the initial step, by (14), \( V_0(f_0) = -R_0 \ln \left(\mathbb{E}[\exp(-f_0/R_0)]\right) \) is concave in \( k \) because Lemma 5 applies and multiplying constants (probability and effective risk tolerance) does not change the convexity. Suppose that \( V_{n-1}(\cdot) \) is concave in \( k \). Then, \( f_n + V_{n-1}(\cdot)/(1+r_f) \) is concave in \( k \) for any realization of exchange rate and demands. Thus, (14) and Lemma 5 implies that \( V_n(\cdot) \) in (15) is concave in \( k \). Since taking expectation does not change concavity, \( V_n(\cdot) \) in (18) is also concave in \( k \).

The property above completes characterization of structural properties and optimal solutions of the Model. They allow us both to understand the optimal behavior of the decision maker (MEN) as well as provide methods for fast evaluation of optimal solutions, including the financial contracts. We showed that the problem behaves in an intuitive way with concave and submodular production and transshipment decisions. Usually, derivation of financial contracts is more challenging. The properties we derived show that a number of pleasing properties can be used. The optimal financial hedging contracts (19) spread the cash consumption across time and across different realization of exchange rate. Without financial hedging, the equalization takes place only across the periods and each of the PCEV values needs to be derived for each state of exchange rate. In the special case of birth and death processes, the derivations can be further simplified due to optimality of forward contracts. Financial contracts are derived in recursive closed form for all of these situations. These results also set the stage for an efficient identification of optimal capacity allocation decisions: due to concavity of PCEV, optimal capacity allocation can be identified by a fast search.

In this section, we assumed that the price is fixed. In many situations, however, price is a decision variable used to influence demand. In the following section we study how allowing to change the price, in addition to production and transshipment, influences the optimal policy.

### 6. Price Model

Price is a powerful instrument because it allows a firm to increase profit by decreasing effective demand, and thus avoid unsatisfied demand when capacity is insufficient and, vice versa, increase demand when capacity is underutilized. We model the demand as a linear function of price \( \xi - bp \). The capacity decision and financial hedging decision are the same as those in the Basic Model (Section 4). The difference is in the third decision. Instead of production and transshipment decisions, we make joint price, production, and transshipment decisions. These decisions are made after demand is revealed. Thus, as before, the total quantity of production and transshipment does not exceed demand at a given price. Since for any case when shortage takes place, it is beneficial to increase price, optimal decisions are always market-clearing:

**Lemma 6.** Considering only combinations of price and production/transshipment decisions satisfying \( \xi - bp \geq y \geq 0 \), the optimal decisions satisfy \( p = (\xi - y)/b \).
Lemma 6 states that the optimal price is set to sell all products. It transforms a price, production, and transshipment decision problem into a production/transshipment-decision-only problem:

\[
f(s, k, \xi) = \max_{(y, x) \in A_1(k, \xi)} G(x, y, \xi) \tag{21}
\]

\[
G(x, y, \xi) = (\xi^d - bc - y^d)y^d/b + s(\xi^o - bc - y^o)y^o/b
+ (s - 1)c(x^0 - x^d) - t(x^o - x^d),
\tag{21}
\]

where \(A_1(k, \xi)\) is defined in (8). This simplified problem has structure similar to (7). The first two terms of its objective function \(G\) represent the profit if no transshipment were allowed, and the last two represent the benefit of transshipment. This formulation leads to the following property.

**Lemma 7.** \(f\) is non-decreasing and concave in \((k, \xi)\).

**Proof:** The proof follows the same logic as that of Lemma 1, where the constraint set \(A\) is replaced by \(A_1\) and objective function \(J\) is replaced by \(G\). The proof hinges on concavity of \(G\) in \((x, y, \xi)\), which is true because \(G\) is a negative quadratic function.

Lemma 7 allows us, later, to claim concavity of both financial hedging decision and capacity decision problems. Building on the latest reformulation, sales can be explicitly expressed. Let us define first \(\bar{y} \equiv (\xi - bc)/2\):

**Lemma 8.** Let \(x \leq \xi\). Considering \(x \leq y \leq \xi, y^d \leq k^d + x^d - x^o, y^o \leq k^o - (x^d - x^o), \) the optimal sales are \(y^{d*} = (x^d \vee \bar{y}) \land (k^d + x^d - x^o)\) and \(y^{o*} = (x^o \vee \bar{y}) \land (k^o + x^o - x^d)\).

**Proof:** The objective function \(G\) is concave in \(y\). Thus, the unconstrained optimal \(y = \bar{y}\). Therefore, the optimal constrained solution can be expressed as \(y^{d*} = (x^d \vee \bar{y}) \land (k^d + x^d - x^o)\) and \(y^{o*} = (x^o \vee \bar{y}) \land (k^o + x^o - x^d)\).

**Lemma 8** states that \(\bar{y}\) serve as optimal sales target levels, which are independent of the transshipment decisions \(x\), and that actual sales will be limited by the constraints in the capacity and transshipment. To further explore structural properties, optimal sales are substituted into the objective function of the problem. With definition \(\bar{y} \equiv (c - cs - t)b/2\) and \(\bar{y}^o \equiv (cs - c - t)b/(2s)\), the cost difference of cross sales due to transshipment (these definitions differ slightly from those in the previous section), the price and production/transshipment decision problem is simplified to a transshipment-decision-only problem:

\[
f(s, k, \xi) = [(\bar{y}^d)^2 + s(\bar{y}^o)^2 + \max_{x \in A_2(k, \xi)} G(x)]/b \tag{22}
\]

\[
G(x) = 2\bar{y}^d x^d - ((x^d - \bar{y}^d)^+)^2 - ((k^d - \xi^d + x^d - x^o)^-)^2
+ s(2\bar{y}^o x^o - ((x^o - \bar{y}^o)^+)^2 - ((k^o - \xi^o - x^d + x^o)^-)^2),
\tag{23}
\]

where \(A_2\) is defined in (9). Since in the optimal transshipment decisions \(x^d\) and \(x^o\) are not positive at the same time, we have:

\[
f = [(\bar{y}^d)^2 + s(\bar{y}^o)^2 + \max(f^d(s, k, \xi), sf^o(s, k, \xi))]/b \tag{23}
\]

\[
f^d(s, k, \xi) = -(s(\bar{y}^o)^+)^2 + \max_{(\xi^d, k^d) \geq x^d \geq 0} 2G^d(x^d) \tag{24}
\]

\[
G^d(x^d) = \bar{y}^d x^d - ((x^d + k^d - \bar{y}^d)^-)^2/2 - ((x^d - \bar{y}^d)^+)^2/2
- s((x^d - k^o + \bar{y}^o)^+)^2/2. \tag{24}
\]

\[
f^o(s, k, \xi) = -(\bar{y}^o)^2 + \max_{(\xi^o, k^o) \geq x^o \geq 0} 2G^o(x^o) \tag{25}
\]

\[
G^o(x^o) = \bar{y}^o x^o - ((x^o + k^o - \bar{y}^o)^-)^2/2 - ((x^o - \bar{y}^o)^+)^2/2
- ((x^o - k^d + \bar{y}^d)^+)^2/2s. \tag{25}
\]

This leads to the following properties.
Lemma 9. \( f \) is submodular in \((k, -\xi)\).

**Proof:** See Appendix. \(\Box\)

Lemma 9 formally states that the capacities and demands are complements and they are substitutes within themselves, even when price is a decision variable, mirroring Lemma 1 in the Basic Model. The submodularity leads to similar properties in the financial hedging decisions.

To characterize the optimal decision, let us define \( k = k - y \), the capacity-overage (if positive) or capacity-shortage (if negative) at two locations, and \( \bar{K} = \bar{k} + \bar{K}^o \), the total-capacity-overage (if positive) or the total-capacity-shortage (if negative). With the redefinitions above, the optimal transshipment decisions are as follows:

**Theorem 5.** The optimal solutions are \( x^d = (k^o \land \xi^d \land x^d)^+ \) and \( x^{oo} = (k^o \land \xi^o \land x^{oo})^+ \), where

\[
\bar{x}^d = \begin{cases} 
\bar{i} - \bar{k} & \text{if } \bar{i} \leq \bar{K} \leq 0, \text{ or } \bar{K} \geq 0 \text{ and } \bar{i} \leq 0 \\
\left(\bar{i} - \bar{k}^o + s\bar{K}^o\right)/(1 + s) & \text{if } \bar{K} \leq \bar{i} \leq -s\bar{K} \text{ and } \bar{K} \leq 0 \\
\left(\bar{i}^0 + \bar{y}^o + s\bar{K}^o\right)/(1 + s) & \text{if } 0 \leq \bar{K} \leq (s + 1)(\bar{y}^o \lor \bar{k}^o) - \bar{y}^o - s\bar{K}^o \text{ and } \bar{K} \geq 0 
\end{cases}
\]

(25)

\[
\bar{x}^{oo} = \begin{cases} 
\bar{i}^0 - \bar{k}^o & \text{if } \bar{i}^0 \leq \bar{K} \leq 0, \text{ or } \bar{K} \geq 0 \text{ and } \bar{i} \leq 0 \\
\left(\bar{i}^0 - \bar{k}^o + s\bar{K}^o\right)/(1 + s') & \text{if } \bar{K} \leq \bar{i} \leq -s'\bar{K} \text{ and } \bar{K} \leq 0 \\
\left(\bar{i}^0 + \bar{y}^o + s\bar{K}^o\right)/(1 + s') & \text{if } 0 \leq \bar{K} \leq (s' + 1)(\bar{y}^o \lor \bar{k}^o) - \bar{y}^o - s'\bar{k}^o \text{ and } \bar{K} \geq 0 
\end{cases}
\]

(25)

and \( s' = 1/s \).

**Proof:** See On-line Appendix. \(\Box\)

In summary, for Price Model, structural properties (concavity and submodularity), similar to those in Basic Model, are proved for all decisions. In addition, the optimal price is always market clearing. The optimal sales targets are independent of production and transshipment. Sales targets and total profits decrease in price sensitivity. Having characterized the properties of optimal price and production and transshipment decisions, concavity of Lemma 7 implies that Theorem 4 continues to hold, i.e., there is a unique optimal capacity decision.

As the structural properties and optimal solutions have been completely characterized for both the Basic Model and the Price Model, they allow us to efficiently perform numerical study.

### 7. Numerical Study

The structural results (Sections 4 and 6) enable us to perform an extensive numerical study and to answer the question: What are the favorable conditions for either financial or operational hedging (or both)? In the case we describe in biggest detail, we assume that (a) we sell both in domestic and in overseas market. We also evaluate the case (b) where we sell only overseas. In (a) and (b) price is exogenous. Finally, we look at case (c), where we sell both in domestic and overseas market and the price is endogenous. To make a fair evaluation of financial, operational, and joint (financial plus operational) hedging, their relative efficiency is compared to the base case. The base case (whether it is (a), (b), or (c)) always assumes that all capacity is located in the domestic facility and no financial hedging is used. Operational hedging allows part of the capacity to be located overseas, while financial hedging keeps all capacity in domestic location, but allows unconstrained...
use of financial instruments. The efficiency of other models is expressed as a percentage of increase over the PCEV of the base case.

We use binomial tree to model exchange rate evolution. It changes with rate \((1 + \epsilon)\) up and \(1/(1 + \epsilon)\) down in each period with mean equal to the exchange rate at the beginning of the period. Demands in each market are modelled as Erlang distribution with mean of 5. Demands and exchange rates are independent.

Since parameters of the model may significantly change the relative strength and weakness of financial hedging and operational hedging, a wide range of parameters has been chosen for the numerical study. We vary five parameters, listed below, across five values each, resulting in 3125 (55) instances. The values of parameters are as follows. Total available capacity \(K = \{9, 10, 11, 13, 16\}\) such that the utilization ranges from 110% to 66%. The sales price \(p = \{1.05, 1.1, 1.2, 1.4, 1.8\}\) such that the profit margin ranges from 5% to 80%. The medium risk-tolerance factor is set to be roughly the average profit per period, i.e., \((p - c)\times\) average total demand. To study risk-aversion effect, we use risk-tolerance value that equals a percentage of the medium risk-tolerance. The percentage value is \(\{50, 75, 100, 125, 150\}\). The exchange rate parameter \(\epsilon = \{0.05, 0.10, 0.15, 0.20, 0.253\}\), and the demand coefficient of variation ranges in \(\{10\%, 20\%, 33\%, 50\%, 100\%\}\). The production cost \(c = 1\) is kept unchanged because relative effects are revealed through the change of the price and transportation cost. In the endogenous price model, we choose average market size \(\xi\) and the price sensitivity such that at the corresponding exogenous price \(p = 1.2\) and the demand is 5. In addition we ensure that the price elasticity of demand \(\frac{\partial p}{\partial bp}\) is closes to average -1.8 (see Tellis 1988) in the central case. This implies that \(b = 7.5\) in the central case. Since we choose value around the central, \(b = \{4.5, 6, 7.5, 9, 10.5\}\). To keep \(a - bp = 5\), the corresponding \(\xi = \{10.4, 12.20, 14, 15.8, 17.6\}\). All experiments run for 12 periods because the public traded futures contract has the longest expiration period of 12 quarters (3 years). Except when we explicitly state it, all experiments assume 0 transportation costs. Transportation costs favor strongly operational hedging, as we show in one of later graphs.

### 7.1. Operational Hedging versus Financial Hedging

Our numerical study suggests that both operational and financial hedging are very effective in case (b) when we sell only in overseas market, with financial hedging being somewhat better on average (86%) than operational one (30%) and with joint hedging providing 91% benefits. The effectiveness, while still significant, decreases when we sell both in domestic and overseas markets (case (a)). In this situation, the difference between operational and financial hedging is much more significant. On average, the savings realized from operational hedging, amount to about 39% vs. 27% from financial hedging. The incremental benefit of financial hedging is around 17%. In our Price Model (case (c)), all benefits are much smaller. On average the savings from operational hedging amounted to 15% vs. 9% from financial hedging, the incremental benefit of financial hedging is about 6% - also smaller than the Basic Model. The exchange rate coefficient of variation has to be very high (as much as 0.5) for financial hedging to dominate operational hedging. Note that in case (b), all revenues are overseas while all costs are domestic. Thus, financial hedging influences a very significant portion of profit. Similarly operational hedging, can “match” most of the revenues by allocating production to overseas market. In case (a), where we sell in both domestic and overseas market, in most situations, domestic production is used anyhow to satisfy domestic demand, and financial hedging influences only part of the revenues. Operational hedging, however, has a similar increment of benefits as case (b): it can not only on average match costs with revenues, but also shift production back and forth to cheaper production site. Clearly, in Price Model, using prices reduces the overall variance of cash flow, thus making hedging less needed and less effective.

The relative efficiencies of financial and operational hedging are affected by numerous other parameters. These effects are summarized below. Unless specified otherwise, the discussion applies to selling in both markets with exogenous price (case (b)) and endogenous price (case (c)).
Transportation cost. This is the only place where we allow consider positive transportation costs. Transportation cost’s effect on operational hedging is opposite to that of financial hedging. Operational hedging savings increase with transportation costs, but financial hedging savings slowly decrease. Higher transportation costs result in less goods being transshipped and, thus, less foreign cash flow. Therefore, financial hedging is less needed. Operational hedging optimally allocates a portion of total capacity overseas, which reduces transportation costs. The higher the transportation cost, the greater the savings realized through operational hedging. This effect is more pronounced in the Basic Model (case (b)) than in the Price Model (model (c)) since being able to change prices decreases volume of transshipment. Hence, higher transportation costs favor operational hedging.

Total available capacity. The capacity effects are illustrated in Figure 2. Both financial hedging and operational hedging benefits increases when capacity increases. Note that the higher capacity implies more flexibility for operational hedging to take advantage of it, while financial hedging is mostly not influenced by it. Hence, high total capacity favors operational hedging.

Risk-tolerance. The risk-averse effects are illustrated in Figure 3. Operational hedging becomes more effective than financial hedging when risk tolerance increases because operational hedging improvement decreases at a slower rate than financial hedging, while both improvements decrease. When the risk tolerance increases, the decision maker becomes more risk-neutral and cares less about the profit variation. Since all hedge improvements come partly from the profit variation deduction, they are smaller when profit variation is less cared. Operational hedging improvement decreases at a slower rate than financial hedging because the operational hedging improvement comes in one part from profit variation deduction and in the other from expected profit improvement, while financial hedging improvement comes completely from the profit variation deduction. Since high risk tolerance reduces the value of variation deduction, but not the value of expected profit improvement, operational hedging is less affected than financial hedging. Thus, high risk-tolerance favors operational hedging, or conversely, high risk aversion favors financial hedging.
Sales price and demand sensitivity on price. The price effects do not apply to case (c) and only illustrated for case (b), see the left of Figure 4. In the exogenous price model, both financial and operational hedging improvements decreases in sales price. When price increases, the base case profit coefficient of variation decreases because its standard deviation increases proportionally, but its mean increases disproportionately faster due to higher profit margin. Since hedging is less valuable when the base profit has smaller coefficient of variation, all hedge improvements become smaller. But the sales price does not favor one hedge over the other because either hedging can decreases in price faster depending on the other system parameters. For endogenous price model, the demand sensitivity on price has an opposite effect to the price in the exogenous model.

Exchange rate coefficient of variation. The effect of exchange rate variance is illustrated in right of Figure 4. When exchange rate coefficient of variation (CV) increases, financial hedging becomes more effective because the foreign currency cash flow in the base case becomes more variable. Operational hedging also becomes much more effective because it allocates more capacity overseas. Consequently, the average of the profit under operational hedging increases much faster than the base case, but its variation increases slower. Thus operational hedging becomes more effective.

The exchange rate CV does not determine which hedge is better, but its interaction with other parameters does. For example, high exchange CV in combination with high capacity makes operational hedging much more effective than financial hedging, as shown in the left of Figure 5. On the other hand, the right of Figure 5 show that high exchange CV in combination with low capacity and other factors makes financial hedging more effective than operational hedging.

Demand coefficient of variation. In exogenous price model, financial hedging effectiveness decreases in demand coefficient of variation because demand variation increases foreign currency exposure uncertainty. Since financial hedging targets a certain level of currency exposure, it hedges too...
little when exposure is high at high demand, but too much when exposure is low at low demand. In contrast, the operational hedging decreases slower than financial hedging when demand CV increases, in fact, it may increases, as shown in the right of Figure 6. Thus, higher demand CV favors operational hedging.

Figure 6 Relative savings as a function of demand coefficient of variation.

8. Conclusion
The paper seeks to analyze, the relative strengths and weaknesses of financial hedging and operational hedging. We describe dynamic finite capacity models that allow capturing the relative benefits of both types of hedging separately and jointly. The structural properties of the models are derived. They allow describing the intuitive behavior of the model with respect to operational policy and somewhat less intuitive properties of financial hedging. The structural properties also allow us to evaluate various contributing factors numerically. The structural properties describe concavity of the present certainty equivalent value on capacity, the special structures of financial hedging contract, and analytical optimal solutions to the production and transshipment problem. We show that with exponential utility, financial hedging equalizes the cash consumption across time and various scenarios of exchange rate realizations. For the special case birth-death process of exchange rates, we show that forward contracts are optimal. In general, our results hold for both the case when price is exogenous and also when it is endogenously decided.

Some specific conclusions follow from our numerical study. For supply chains with multiple locations of demand (including the place when production takes place), operational hedging tends to dominate financial hedging, in terms of savings, except when the exchange rate variance is very high and transshipment cost is very low. Operational hedging is preferred over financial hedging when total capacity and demand variations are high. Financial hedging is preferred when risk-aversion is high. Both hedges become more effective at small profit margin and high exchange rate variations, but no one is preferred over the other. Furthermore, the marginal benefit of financial hedging is quite small once operational hedging has used. Financial hedging is especially beneficial with high variance of exchange rates, but also with low sales price, and moderate risk aversion. Maybe somewhat surprisingly, operational hedging is a significant tool for the MRN to reduce the exchange rate risk and maximizing her present certainty equivalent value. This result does not support the seemingly logical conclusion that low cost financial hedging may be sufficient. (Multiple producers increasingly use operational hedging, which is especially visible in case of automakers). We identify specific situations that are most appropriate for use of financial hedging and for use of operational hedging. Two factors are not taken into account across most of the examples. First, transportation costs very strongly favor operational hedging and we omitted them except for one example that illustrates the effect of transportation costs. Second, operational hedging requires additional initial investment, as building two plants is usually more expensive than a single one.
9. Appendix

Proof of Lemma 4: We prove the $V_n(\cdot)$'s independence on $\beta_n$ by induction, the other results will follow from the induction proof process. We first apply variable substitution $\beta_i' = \beta_i + w_i'$ for all $i \in \{0, \ldots, n\}$. When $n = 1$, we have

$$U_1(f_1, f_0 | \beta_1) = \max_{\beta_0'} \{E[-(1 + r_f) \rho_1 \exp(- \frac{f_1' + \beta_0' - \beta_0'(1 + r_f)}{\rho_1}) - \rho_0 \exp(- \frac{f_0 + \beta_0'}{\rho_0})]\}$$

where the second equation follows from (2) and (14). The first order optimality condition on $\beta_0'$ is

$$\frac{V_0(f_0) + \beta_0'}{\rho_0} = \frac{f_1 + \beta_0' - \beta_0'(1 + r_f)}{\rho_1}.$$

Substituting back the original decisions, we get optimal investment decision (16) for $n = 1$, the corresponding maximum utility is then (17). The PCEV in (15) follows from (2) and $V_1(f_1, f_0 | \beta_1)$ is independent of $\beta_1$. This concludes the proof of induction base, i.e., the truth for $n = 1$.

Now suppose $V_{n-1}(f_{n-1}, \ldots, f_0 | \beta_{n-1})$ is independent of $\beta_{n-1}$ for any $n - 1 \geq 0$, we proceed to prove the $V_n(\cdot)$'s independence on $\beta_n$ and all the other results. By (1) and variable substitution, the left side of (17) becomes

$$U'_n(f_n, \cdots, f_0 | \beta'_n) = \max_{\beta'_n | 0 \leq i < n} \{E[\sum_{i=0}^{n} (-(1 + r_f)^n \rho_1 \exp(- \frac{f_i + \beta'_i - \beta_{i-1}'/(1 + r_f)}{\rho_i})]\}$$

By (1) for $n - 1$, the objective function of the above maximization problem becomes

$$E[-(1 + r_f)^n \rho_n \exp(- \frac{f_n + \beta_n' - \beta_{n-1}'/(1 + r_f)}{\rho_n}) + U'_{n-1}(f_{n-1}, \cdots, f_0 | \beta_{n-1}')],$$

where the only decision is $\beta_{n-1}'$. By the induction assumption, (2), and (13) in Lemma 3, for $n - 1$, the function inside the expectation becomes

$$-(1 + r_f)^n \rho_n \exp(- \frac{f_n + \beta_n' - \beta_{n-1}'/(1 + r_f)}{\rho_n}) - R_{n-1}(1 + r_f)^{n-1} \exp(- \frac{\beta_{n-1}' + V_{n-1}(f_{n-1}, \cdots, f_0)}{R_{n-1}}).$$

The first order optimality condition implies

$$\beta_{n-1} = \frac{\beta_{n-1}' + V_{n-1}(f_{n-1}, \cdots, f_0)}{R_{n-1}} = \frac{f_n + \beta_n' - \beta_{n-1}'(1 + r_f)}{\rho_n}.$$

Substituting the original decision back results (16) for $n$. This equation implies that the adjusted cash consumptions are equalized between two periods. Solving this equation results the optimal bond investment

$$\beta_{n-1}' = (f_n + \beta_n' - V_{n-1}(f_{n-1}, \cdots, f_0) \rho_n/R_{n-1}) R_{n-1}/R_n.$$

This optimal decision implies that the equal adjusted cash consumption in period $n$ or $n - 1$ is $(f_n + \beta_n' - V_{n-1}(f_{n-1}, \cdots, f_0) \rho_n/R_{n-1})/R_n + V_{n-1}(f_{n-1}, \cdots, f_0)/R_{n-1}$. By simplification, it becomes

$$(f_n + \beta_n' + V_{n-1}(f_{n-1}, \cdots, f_0) \rho_n/R_{n-1})/R_n = \frac{f_n + \beta_n' - V_{n-1}(f_{n-1}, \cdots, f_0) / (1 + r_f)}{R_n},$$

where the last equation follows from $(R_n - \rho_n)/R_n - 1 = 1/(1 + r_f)$. Substituting this optimal consumption back to the objective function and then the original decision back, we have (17) for $n$. Finally, using (2) we obtain (15) for $n$, which implies that $V_n(\cdot)$ is independence of $\beta_n$.  \[\square\]
Proof of Lemma 2: We prove the PCEV’s independence on $\beta_i(s_n)$ by induction, the other results will follow from the induction proof process. We first apply variable substitution $\tilde{\beta}_i(S_i) = \beta_i(S_i) + w_i$ for all $i \in \{0, \cdots, n\}$. When $n = 1$ and by (3), the left side of (20) becomes
\[
\begin{align*}
\max_{\beta_0(S_0)} \{ - (1 + r_f) \rho_1 \exp\left( - \frac{f_1 + \beta'_1(s_1) - \frac{E[\tilde{\beta}_0(S_0)|S_1 = s_1]}{1 + r_f}}{\rho_1} \right) - E[\rho_0 \exp\left( - \frac{f_0 + \beta'_0(S_0)}{\rho_0} \right)|S_1 = s_1] \}. 
\end{align*}
\]
By (2) and (14), the objective function of the above maximization problem becomes
\[
\begin{align*}
-(1 + r_f) \rho_1 \exp\left( - \frac{f_1 + \beta'_1(s_1) - \frac{E[\tilde{\beta}_0(S_0)|S_1 = s_1]}{1 + r_f}}{\rho_1} \right) - E[\rho_0 \exp\left( - \frac{V_0(f_0|S_0) + \beta'_0(S_0)}{\rho_0} \right)|S_1 = s_1] \}.
\end{align*}
\]
By point-wise maximization, the first order optimality condition on $\beta_0(s_0)$ is
\[
\frac{f_1 + \beta'_1(s_1) - \frac{E[\tilde{\beta}_0(S_0)|S_1 = s_1]}{1 + r_f}}{\rho_1} = \frac{V_0(f_0|S_0) + \beta'_0(S_0)}{\rho_0}.
\]
This equation means that the optimal derivative investment equalizes the adjusted consumption cash not only between two periods, but also among all possible realizations of the exchange rate (the right side of the equation). Substituting the original decision back to the above equation results (19) for $n = 1$. Taking expectation of both sides conditioning on $S_1 = s_1$ results
\[
\frac{f_1 + \beta'_1(s_1) - \frac{E[\tilde{\beta}_0(S_0)|S_1 = s_1]}{1 + r_f}}{\rho_1} = \frac{E[V_0(f_0)|S_1 = s_1] + \beta'_0(S_0)|S_1 = s_1]}{\rho_0}.
\]
Simplifying this equations to obtain the price of the optimal derivative
\[
\frac{E[\tilde{\beta}_0(S_0)|S_1 = s_1]}{1 + r_f} = (f_1 + \beta'_1(s_1)) \frac{E[V_0(f_0)|S_1 = s_1]}{\rho_1} \frac{\rho_0}{(1 + r_f) R_1}.
\]
We substitute the price back in the left side of the first order optimality condition, and obtain, for both periods and all scenarios of exchange rate realization, the optimal adjusted cash consumption as follow
\[
\frac{f_1 + \beta'_1(s_1)}{\rho_1} (1 - \frac{\rho_0}{(1 + r_f) R_1}) + \frac{E[V_0(f_0)|S_1 = s_1]}{1 + r_f} R_1 = f_1 + \beta'_1(s_1) + \frac{E[V_0(f_0)|S_1 = s_1]}{1 + r_f} R_1,
\]
where the last equation follows from $1 - \frac{\rho_0}{(1 + r_f) R_1} = \rho_1 / R_1$. We substitute this consumption cash into the objective function and the original decision variable back, thus obtain (20) for $n = 1$. This and (2) imply (18) for $n = 1$, and consequently the PCEV is independent of $\beta_1(s_1)$. This concludes the proof of induction base, i.e., the truth for $n = 1$.

Now suppose $V_{n-1}(f_{n-1}, \cdots, f_0|\beta_{n-1}(s_{n-1}))$ is independent of $\beta_{n-1}(s_{n-1})$ for any $n - 1 \geq 0$, we proceed to prove that $V_n(f_n, \cdots, f_0|\beta_n(s_n))$ is independence of $\beta_n(s_n)$ and all the other results. By (3) and variable substitution, the left side of (20) becomes
\[
U_n(f_n, \cdots, f_0|\beta_n(s_n)) = \max_{\beta'_0 \leq i \leq n} \{ \frac{E\left[ \sum_{i=0}^{n} \left( -(1 + r_f)^i \rho_i \exp\left( - \frac{f_i + \beta'_i(s_i)}{\rho_i} \frac{E[\tilde{\beta}'_{n-1}(S_{n-1})|S_n = s_n]}{1 + r_f} \right) \right) | S_n = s_n \} \}.
\]
By (3) for $n - 1$, the objective function of the above maximization problem becomes
\[
-(1 + r_f)^n \rho_n \exp\left( - \frac{f_n + \beta'_n(s_n) - \frac{E[\tilde{\beta}'_{n-1}(S_{n-1})|S_n = s_n]}{1 + r_f}}{\rho_n} \right) + E[\beta'_{n-1}(f_{n-1}, \cdots, f_0|\beta'_{n-1}(S_{n-1}))|S_n = s_n],
\]
where the only decision is derivative payoff function $\beta'_{n-1}(S_{n-1})$. By the induction assumption, (2), and (13) for $n - 1$, the objective function becomes

$$-(1 + r_f)^n \rho_n \exp(-\frac{f + \beta'_n - E[\beta'_{n-1}(S_{n-1})|s_n] / (1 + r_f)}{\rho_n})$$

$$-R_{n-1}(1 + r_f)^n E[\exp(-\frac{\beta'_{n-1}(S_{n-1}) + V_{n-1}(f_{n-1}, \ldots, f_0|s_{n-1})}{R_{n-1}})|s_n].$$

The first order optimality condition on $\beta'(s_{n-1})$ is

$$\frac{f_n + \beta'_n(s_n) - E[\beta'_{n-1}(S_{n-1})|s_n] / (1 + r_f)}{\rho_n} = \frac{\beta'_{n-1}(s_{n-1}) + V_{n-1}(f_{n-1}, \ldots, f_0|s_{n-1})}{R_{n-1}}.$$ 

This equation becomes (19) for $n$ when the original decisions are substituted back. Taking expectation of both sides conditioning on $s_n$, we have

$$\frac{f_n + \beta'_n(s_n) - E[\beta'_{n-1}(S_{n-1})|s_n] / (1 + r_f)}{\rho_n} = \frac{E[\beta'_{n-1}(S_{n-1})|s_n] + E[V_{n-1}(f_{n-1}, \ldots, f_0)|s_n]}{R_{n-1}}.$$ 

This equation implies that the optimal derivative price is

$$E[\frac{\beta'_{n-1}(S_{n-1})|s_n]}{1 + r_f}] = (f_n + \beta'_n(s_n) - E[V_{n-1}(f_{n-1}, \ldots, f_0)|s_n] \rho_n) \frac{R_{n-1}}{R_n} (1 + r_f).$$

This derivative price implies that the consumption cash of period $n$ is

$$\frac{f_n + \beta'_n(s_n)}{\rho_n} \frac{1 - R_{n-1}}{R_n} + \frac{E[V_{n-1}(f_{n-1}, \ldots, f_0)|s_n]}{R_n} = \frac{f_n + \beta'_n(s_n) + \frac{E[V_{n-1}(f_{n-1}, \ldots, f_0)|s_n]}{(1 + r_f)} \rho_n}{R_n},$$

where the equation follows from $(1 - \frac{R_{n-1}}{(1 + r_f)R_n}) = \frac{\rho_n}{R_n}$. This cash consumption in period $n$ is also the one in period $n - 1$ at any realization of the exchange rate because of the optimal derivative investment. Substituting these consumption cashes into objective function yields (20) for $n$. Finally, using (20) and (2) we obtain (18) for $n$, consequently the PCEV is independent of $\beta_n(s_n)$. \(\square\)

**Proof of Theorem 3:** We prove this corollary by showing that the optimal financial contract payoff in $\beta_{n-1}(s_{n-1})$ in (19) can be obtained by bond and forward investments when the exchange rate in each state goes either up or down into the next period. To simply the notation, we let $p$ ($q$) be the probability that the exchange rate goes to $s_{n-1}^1$ up ($s_{n-1}^2$ down). The corresponding payoff at those states are $\beta^1$ and $\beta^2$. Furthermore, for simplicity, denote: $a_i = V_{n-1}(f_{n-1}, \ldots, f_0|s_{n-1}^i) + w'_{n-1}$, $i \in \{1, 2\}; a_3 = \frac{R_{n-1}}{\rho_n} (f_n + \beta_n(s_n) + w'_{n})$; $\alpha_3 = \frac{R_{n-1}}{(1 + r_f)\rho_n}$, we expressed (19) using these notation as follows

$$\beta^1 + a_1 = a_3 - \alpha_3(p\beta^1 + q\beta^2),$$

$$\beta^2 + a_2 = a_3 - \alpha_3(p\beta^1 + q\beta^2).$$

The solutions of the above equations are:

$$\beta^1 = \frac{a_3 + a_2 q - a_1 (1 + q \alpha_3)}{1 + \alpha_3},$$

$$\beta^2 = \frac{a_3 + a_1 p - a_2 (1 + p \alpha_3)}{1 + \alpha_3}.$$ 

These payoffs can be constructed by entering a forward contract with $\frac{\beta^1 - \beta^2}{s_{n-1}^1 - s_{n-1}^2} / (1 + r_f)$ risk-free bonds. \(\square\)
Proof of Lemma 1: For simplicity of the presentation, let \( \alpha_{jl} \) be the coefficient of decision variable of objective function in (5), i.e., \( \alpha_{dd} = p - c, \alpha_{od} = p - cs - t, \alpha_{do} = sp - c - t, \alpha_{oo} = s(p - c) \). To prove submodularity, we first write the dual of the production problem:

\[
    f(s, k, \xi) = \min_{\eta, \lambda: \eta \in B(\alpha)} G(\eta, \lambda, k, \xi) \\
    G(\eta, \lambda, k, \xi) = \eta k + \lambda \xi \\
    B(\alpha) = \{ \eta_j + \lambda_i \leq \alpha_{jl} \forall j \in \{d, o\}, \forall l \in \{d, o\}, \eta \geq 0, \lambda \geq 0 \}
\]

To see the relationship between \( k \) and \( -\xi \), we substitute \( \bar{\xi} = -\xi \) and \( \bar{\eta} = -\eta \). Then, the dual problem becomes:

\[
    f(s, k, \xi) = \min_{\eta, \lambda: \eta \in B_1(\alpha)} G(\bar{\eta}, \bar{\lambda}, k, \bar{\xi}) \\
    G(\bar{\eta}, \bar{\lambda}, k, \bar{\xi}) = -\bar{\eta}k - \bar{\lambda}\xi \\
    B_1(\alpha) = \{ -\bar{\eta}_j + \bar{\lambda}_i \leq \alpha_{jl} \forall j \in \{d, o\}, \forall l \in \{d, o\}, \bar{\eta} \leq 0, \bar{\lambda} \geq 0 \}
\]

The objective function \( G \) is submodular in \( (\bar{\eta}, k, \bar{\xi}) \) and the constraint set \( B_1 \) is a sublattice of \( (\eta, \lambda) \). Since minimizing submodular function over a sublattice preserves submodularity, \( f \) is submodular in \( (k, -\xi) \). \( \square \)

Proof of Theorem 1: Obviously, there exist optimal \( (x^d, x^o) \) such that \( x^d x^o = 0 \). Let us first consider scenario \( x^d \geq 0, x^o = 0 \). The problem becomes:

\[
    f(s, k, \xi) = \bar{p}x^d + s\bar{p}x^o + \max_{(x^d, x^o) \geq x^d \geq 0} \bar{t}x^d - \bar{p}(\bar{k} + x^d) - s\bar{p}(x^d - \bar{k})^+
\]

This formulation can be interpreted as follows. The total profit consists of \( \bar{p}x^d + s\bar{p}x^o \), the profit if all demand were satisfied, and the objective function under maximization, which is the impact of transshipment \( x^d \). \( \bar{t}x^d \) is the relative cost saving of cross sales. \( \bar{p}(\bar{k} + x^d) \) is the cost of not being able to meet domestic demand, if \( x^d \) is too small. \( s\bar{p}(x^d - \bar{k})^+ \) is the cost of not meeting overseas demand, if \( x^d \) is too big.

Clearly the objective function is concave in \( x^d \), as its first-order derivative decreases in \( x^d \) is \( \bar{t} + \bar{p}1_{x^d < \bar{k}} - s\bar{p}1_{x^d \geq \bar{k}} \).

Depending on the relative values of \( -\bar{k} \) and \( \bar{k} \), the above function takes one of two forms shown in Figure 7. Marginal profit depends on the sign of each of the three segments of the derivative, which results in the target level of \( \bar{x}^d \).

When \( x^d = 0 \) and \( x^o \geq 0 \), the objective function becomes \( \bar{t}x^o - \bar{p}(\bar{k} + x^o) - s\bar{p}(x^o - \bar{k})^+ \). Its derivative is illustrated in Figure 8. The same analysis can be applied to this scenario, which leads to \( \bar{x}^o \). \( \square \)

Proof of Lemma 3: We prove (12) and (13) first, then (14) will follow. By (1), the left side of (13) is

\[
    \max_{\beta_1, \beta_2, 0 \leq i < n} \{ E[ -w_n \exp(-f_n + \beta_n - \beta_{n-1}/(1+r_f)) - \sum_{i=0}^{n-1} w_i \exp(-\beta_i - \beta_{i-1}/(1+r_f)) \] 
\]
To this maximization problem, we apply variable substitutions as follow:

\[
\beta_i' = \begin{cases} 
    f_n + \beta_n + w_i' , & \text{if } i = n; \\
    \beta_i + w_i' , & \text{if } 0 \leq i \leq n - 1.
\end{cases}
\]

This variable substitution transforms the maximization problem to be

\[
\max_{\beta_i', 0 \leq i \leq n} \{ E[- \sum_{i=0}^{n} (1 + r_f)^i \rho_i \exp(-\frac{\beta_i' - \beta_{i-1}'/(1 + r_f)}{\rho_i})] \}.
\]

The first order necessary conditions are \[\beta_i' - \beta_0'/(1 + r_f)\]/\[\rho_i = \beta_0' / \rho_0\] for \(i = 0\), and for \(0 < i \leq n - 1\), \[\beta'_i - \beta'_{i-1}/(1 + r_f)\]/\[\rho_i = \beta'_i - \beta'_{i-1}/(1 + r_f)\]/\[\rho_i\]. These equations imply that the optimal decisions
satisfy $\beta'_i/R_i = \beta'_{i+1}/R_{i+1}$ for all $0 \leq i \leq n - 1$. As a result, the risk adjusted consumption cash in all periods are equalized at
\[
\frac{\beta'_n - \beta'_{n-1}/(1 + r_f)}{\rho_n} = \frac{\beta'_n - \beta'_{n-1}/[R_n(1 + r_f)]}{\rho_n} = \frac{\beta'_n}{R_n},
\]
where the last equation follows from $R_n = R_{n-1}/(1 + r_f) + \rho_n$. As a result, the maximum utility is $E[1/(1 + r_f)^n R_n \exp(-\beta'_n/R_n)]$. We obtain (12) and (13) by substituting back the original decision variables. Finally, the PCEV’s independence on $\beta_n$ and (14) follow from (2).

**Proof of Lemma 9:** $f$’s submodularity in $(k, -\xi)$ is equivalent to $f$’s being supermodular in the pairs $(k^d, -k^o)$, $(k^d, \bar{y}^d)$, $(k^o, \bar{y}^d)$, $(k^o, \bar{y}^o)$, and $(\bar{y}^d, -\bar{y}^o)$. Since supermodularity is preserved under summation and maximization, it suffices to prove that $f^d$ and $f^o$ are supermodular in those pairs. Without loss of generality, let us consider $f^d$. Recall that maximum operator on a sublattice preserves submodularity.

First, $f^d$ is supermodular in $(-k^d, k^o)$, $(k^o, \bar{y}^d)$ and $(\bar{y}^d, -\bar{y}^o)$ because $G^d$ is supermodular, and the constraint is a sublattice in $(-k^d, k^o, x^d)$, $(k^o, \bar{y}^d, x^d)$ and $(\bar{y}^d, -\bar{y}^o, x^d)$.

Second, $f^d$’s supermodularity in $(k^d, \bar{y}^d)$ and $(k^d, \bar{y}^o)$ is proved by substituting $x^{dd} = \bar{y}^d - x^d$. $G^d$ and the constraint set become:
\[
G^d(x^{dd}) = \bar{f}^d(\bar{y}^d - x^{dd}) - ((k^d - x^{dd})^2)/2 - ((\bar{y}^d - x^{dd} + k^o + \bar{y}^o)^2)/2 - s((\bar{y}^d - x^{dd} - k^o + \bar{y}^o)^2)/2
\]
\[
\bar{y}^d \geq x^{dd} \geq \max(0, \bar{y}^d - k^o).
\]

Clearly, $G^d$ is supermodular and the constraint is a sublattice in $(x^{dd}, k^d, \bar{y}^d)$ and in $(x^{dd}, k^d, \bar{y}^o)$.

Third, to show $f^d$’s supermodularity in $(k^o, \bar{y}^o)$, substitute $x^{do} = k^o - x^d$. Then, $G^d$ and the constraint become:
\[
G^d(x^{do}) = \bar{f}^d(k^o - x^{do}) - ((k^o - x^{do} - k^d + \bar{y}^d)^2)/2 - ((k^o - x^{do} - k^o + \bar{y}^o)^2)/2 - s((\bar{y}^o - x^{do} - k^o + \bar{y}^o)^2)/2
\]
\[
k^o \geq x^{do} \geq \max(0, k^o - \bar{y}^o).
\]

Clearly, $G^d$ is supermodular and the constraint is a sublattice in $(k^o, \bar{y}^o, x^{do})$. Thus, $f^d$ is supermodular. By symmetry, $f^o$’s supermodularity is proved by the same logic.

**References**


Tangaras, G. 1989. Effects of pooling on the optimization and service levels of two-location inventory system. *IIE. Trans.* **21** 250–257.


Proofs of Statements

Proof of Theorem 5: There exist optimal \((x^d, x^o)\) such that \(x^d x^o = 0\). Let us first consider scenario \(x^d \geq 0, x^o = 0\). The transshipment decision problem is represented by (24) and (24). The optimal \(x^d\) minimizes \(G^d\). \(G^d\)’s derivative

\[
G^d(x^d) = \bar{t}^d - (x^d + \bar{k}^d)^+ - (x^d - \bar{y}^d)^+ - s(x^d - \bar{k}^o)^+ \tag{EC.0}
\]

is a decreasing and piece-wise linear function, see Figure EC.1. Its shape depends on the relative positions of \(-\bar{k}^d\) and \(\bar{k}^o\), i.e., the sign of \(K\). We consider two cases:

- **Case 1:** \(K \geq 0\)
  
  \[
  G^d(x^d) = \begin{cases}
  \bar{t}^d - \bar{k}^d - x^d & \text{if } x^d \leq -\bar{k}^d \\
  \bar{t}^d - (x^d - \bar{y}^d)1_{\bar{y}^d \geq \bar{k}^o} - s(x^d - \bar{k}^o)1_{\bar{k}^o \leq \bar{y}^d} & \text{if } -\bar{k}^d \leq x^d \leq (\bar{y}^d \land \bar{k}^o) \\
  \bar{t}^d - (x^d - \bar{y}^d) - s(x^d - \bar{k}^o) & \text{if } (\bar{y}^d \lor \bar{k}^o) \leq x^d
  \end{cases}
  \tag{EC.0}
  \]

which implies:

\[
\bar{x}^d = \begin{cases}
\bar{t}^d + \bar{k}^d & \text{if } \bar{t}^d \leq 0 \\
\bar{t}^d + \bar{y}^d1_{\bar{y}^d \geq \bar{k}^o} + s\bar{k}^o1_{\bar{y}^d \geq s\bar{k}^o} & \text{if } 0 \leq \bar{t}^d \leq (s + 1)(\bar{y}^d \lor \bar{k}^o) - \bar{y}^d - s\bar{k}^o \\
(\bar{t}^d + \bar{y}^d + s\bar{k}^o)(1 + s) & \text{if } (s + 1)(\bar{y}^d \lor \bar{k}^o) - \bar{y}^d - s\bar{k}^o \leq \bar{t}^d
  \end{cases}
  \tag{EC.0}
\]
(b) $\bar{k}^{d} + \bar{k}^{o} = \bar{K} \leq 0$:

$$G^{nd} (x^{d}) = \begin{cases} 
\bar{t}^{d} - \bar{k}^{d} - x^{d} & \text{if } x^{d} \leq \bar{k}^{o} \\
\bar{t}^{d} - \bar{k}^{d} - x^{d} - s(x^{d} - \bar{k}^{o}) & \text{if } \bar{k}^{o} \leq x^{d} \leq -\bar{k}^{d} \\
\bar{t}^{d} - s(x^{d} - \bar{k}^{o}) & \text{if } -\bar{k}^{d} \leq x^{d} \leq \bar{y}^{d} \\
\bar{t}^{d} - (x^{d} - \bar{y}^{d}) - s(x^{d} - \bar{k}^{o}) & \text{if } \bar{y}^{d} \leq x^{d} 
\end{cases} \tag{EC.0}$$

The corresponding solution to $G^{nd}(x^{d}) = 0$ is:

$$\bar{x}^{d} = \begin{cases} 
\bar{t}^{d} - \bar{k}^{d} & \text{if } \bar{t}^{d} \leq \bar{K} \\
(\bar{t}^{d} - \bar{k}^{d} + s\bar{k}^{o})/(1 + s) & \text{if } \bar{K} \leq \bar{t}^{d} \leq -s\bar{K} \\
(\bar{t}^{d} + s\bar{k}^{o})/s & \text{if } -s\bar{K} \leq \bar{t}^{d} \leq s(\bar{y}^{d} - \bar{k}^{o}) \\
(\bar{t}^{d} + \bar{y}^{d} + s\bar{k}^{o})/(1 + s) & \text{if } s(\bar{y}^{d} - \bar{k}^{o}) \leq \bar{t}^{d} 
\end{cases} \tag{EC.0}$$

$\bar{x}^{d}$ in (5) is obtained by combining (9) and (9).

When $x^{d} = 0$ and $x^{o} \geq 0$, the role of $x^{d}$ and $x^{o}$ is reversed. $\square$
We study a risk management problem in ship capacity procurement where shipping companies must finance ship purchases in the face of uncertain demand. The company seeks to control capacity demand risk by selling a percentage of ship capacity to other companies. In order to facilitate this, we introduce an auction mechanism which the company can use to determine shared capacity and find partners. Acting as the auctioneer, the company announces a certain percentage of capacity to a set of buyers. The payment from the buyer is determined by auction and bidding strategy depends on two signals: the demand and the company’s own financial position. For both the risk-neutral and risk-averse utility functions, we find a unique equilibrium for buyers, and a unique percentage of sharing capacity for the company. We study the performance of second-price auctions under this framework and find that the policy not only reduces the cost of financial friction of the company, but also increases overall utilization rates of capacity. Numerical experiments illustrate that the percentage of the payments from auction is usually higher than the percentage of the capacity shared with the partner allowing the firm to improve its performance significantly. And the seller will share more capacity to mitigate its demand risk if it becomes more risk-averse.

Key words: Shipping Finance; Procurement; Capacity Sharing; Auction Applications

1. Introduction

In the shipping procurement business, the presence of uncertainty in the long-term underscores the need to integrate capacity planning with financing strategy. As external funds are used in procurement, financing costs accrue in the form of interest payments, etc., while marginal costs increase in proportionate scale due to increased risk premiums and other frictional factors. Even if the shipping company uses only internal funds, expected marginal revenues will decrease in the amount of money invested which implies increased financing cost. It is therefore imperative for the company to seek optimal strategies to balance increased financing costs with anticipated profits from capacity procurement planning decisions.

Due to very high costs, it is common for shipping companies to obtain financing to purchase ships. Besides traditional financing channels, the cost of ships and operations is often financed through the
sharing of capacity. Many of the top shippers, such as APM-Maersk, COSCO, Hanjin Shipping, use space sharing agreements or vessel sharing agreements to share capacity with other companies.\(^1\) This is beneficial as shown by the fact that agreements are even made between competitors; for example, the Korean shipper, Hanjin Shipping, shares capacity with its competitor "K" Line.\(^2\) Given this way of doing business, the shipper must effectively manage capacity and finance decisions. In this work, we focus on a key risk management problem faced by shipping companies in which capacity and cost sharing are factored into ship procurement. Here, revenue generated from the new ship is random due to randomness in future demand. However, the company must pay all the cost before the revenue realized and so profits are not guaranteed. On the other hand, high financing and purchasing costs increase financial risk. Since by sharing capacity and cost the company can achieve a balance between demand risk, financial risk and potential profits we consider a shipping company which seeks to reduce payment and control risk by selling a percentage of capacity to another company. In the approach we take, we introduce an appropriate auction mechanism to model the sale of capacity since the primary objective in this case is for the company to select a partner and determine a sharing percentage. As the auctioneer, the company (seller) announces a certain percentage \(x \in [0, 1]\) of capacity to a set of shipping companies (buyers) who compete for the sharing percentage by submitting bids which indicate their willingness to pay.

Both seller and buyers can benefit from a capacity-cost sharing scheme. From the buyer’s perspective, the scheme offers an opportunity to expand capacity to better meet demand. Here, it is possible that the buyer may not be able to procure appropriate vessels on his own. From the seller’s point of view, the scheme provides benefits. First, the seller can reduce his total cost, especially in the case of a financially constrained company for which marginal expenditure costs can increase significantly. By sharing parts of its capacity, the seller can acquire the buyer’s payment as a source of interest free financing to reduce both the purchasing and financing cost. Second, the capacity-cost sharing scheme can increase the seller’s revenue. The seller’s future demand is stochastic and a portion of its capacity may not be utilized. This unused capacity is of no use to the seller otherwise but provides an opportunity for him to increase revenue. Taking these considerations into account, a capacity-cost sharing scheme offers a good solution for the seller to manage demand and financing risk. If demand is high the seller can obtain more revenue by utilizing specific reserve capacity. On the other hand, funds obtained from selling capacity compensate for low demand.

\(^1\) The information about the sharing agreements used by shipping companies is collected from the website of Federal Maritime Commission: \url{http://ww2.fmc.gov/agreements/type_npage.aspx}

\(^2\) Cited from the following link. \url{http://www.referenceforbusiness.com/history2/12/Hanjin-Shipping-Co-Ltd.html}.
We mention that a capacity-cost sharing scheme can be applied to a broader context in business situations where capacity investment are a feature. One example is code-sharing which is extensively used in airline business. A code-sharing agreement allows participating airlines to sell their tickets on the same flight. The owner of aircraft can apply a capacity-cost sharing scheme to select an appropriate partner. Yet another example is sharing the ownership of warehouse in which a capacity-cost sharing scheme to share the capacity and cost with a partner can be used.

In a capacity sharing scheme, selecting the right partner and determining the percentages of capacity and cost to share is a problem. A company may want to share more or less capacity and cost while prospective partners may have different objectives. We therefore need an appropriate mechanism to match the company’s and its partner’s demand and determine the percentage of cost to share. Indeed, an auction mechanism is appropriate to select the partner as well as determine selling price.

In this paper we assume that the seller will announce the capacity share to sell in advance, and the buyers will then compete by submitting bids for the capacity. We will focus on the implementation of a second-price sealed auction where payment is the highest bid level except for the winning price. However our model can be extended to fit more general auction types. We will (i) study the optimal capacity decision for the seller and the bidding strategy for buyers; and (ii) attempt to understand the role of demand and financial friction in participants’ decision making as well as (iii) study the role of risk-aversion for in the problem.

In practice, there is a dearth of research available on the use of auctions to share the capacity of a ship. The main aim of this paper is to develop theoretical insights into shipping capacity procurement and management.

In this respect, three research streams in the literature are related to our work. They are (i) research on capacity reservation studied by Wu et al. (2002) and Kleindorfer and Wu (2003); (ii) research on auctions with constraints studied by Che and Gale (1998) and Malakhov and Vohra (2009); and (iii) research on auction application in procurement problem studied by Dasgupta et al (1990), Che (1993), Branco (1997) and Chen (2007).

Wu et al. (2002) studied the seller’s optimal bidding strategy and the buyers’ contracting strategies for capital-intensive goods. In their model, the seller acts as a Stackelberg leader who offered the capacity reserve cost and executive cost as the bid, and the buyer then determined the capacity reserve level according to his interest. Adopting this framework as the base case, Kleindorfer and Wu (2003) reviewed the literature related to optimal capacity reservation by contracting linked with B2B exchanges in capital-intensive industries. Wu et al. (2002) and Kleindorfer and Wu (2003)


Qing, Brian, and Kekun: *Procurement Financing in Shipping using Auctions*

**Table 1** Related literature

<table>
<thead>
<tr>
<th>Literature</th>
<th>Mechanism</th>
<th>Signal Space *</th>
<th>Risk Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wu et al. (2002)</td>
<td>Not Specified</td>
<td>-</td>
<td>Risk Neutral</td>
</tr>
<tr>
<td>Che and Gale (1998)</td>
<td>Auction</td>
<td>M C</td>
<td>Risk Neutral</td>
</tr>
<tr>
<td>Malakhov and Vohra (2009)</td>
<td>Auction</td>
<td>M D</td>
<td>Risk Neutral</td>
</tr>
<tr>
<td>Che (1993)</td>
<td>Auction</td>
<td>U C</td>
<td>Risk Neutral</td>
</tr>
<tr>
<td>Branco (1997)</td>
<td>Auction</td>
<td>U C</td>
<td>Risk Neutral</td>
</tr>
<tr>
<td>Chen (2007)</td>
<td>Auction</td>
<td>U C</td>
<td>Risk Neutral</td>
</tr>
<tr>
<td>Our Paper</td>
<td>Auction</td>
<td>M C</td>
<td>Risk Averse</td>
</tr>
</tbody>
</table>

* M - multi-dimensional, U - uni-dimensional, C - continuous, D - discrete.

They focused on the efficient integration of long-term and short-term contracting. In this work, we propose an auction as the selling mechanism for capacity-cost sharing.

Che and Gale (1998) studied the performance of first-price and second-price auctions when bidders’ financial resources are costly. In their paper, they presumed the existence of the buyer’s equilibrium for the dual-dimensional signal setting. However they did not offer a methodology to characterize the equilibrium. In this work, we derive the buyer’s equilibrium strategy for the dual-dimensional signal setting under second-price auction. Malakhov and Vohra (2009) studied the optimal auction design problem for a seller facing a group of buyers with two-dimensional private signal about the marginal valuation of the goods and its capacity constraints. They presume the signals to be finite discrete and under this critical assumption they can implement linear programming for their problem. In this work, we assume that the signals are continuous.

Dasgupta et al (1990) is one of the earliest works which studied an auction mechanism in procurement problems. They proved that under certain conditions quantity auction is the optimal mechanism for the buyer in the procurement problem. In Dasgupta et al (1990) signals are uni-dimensional whereas in the present work, signals have two dimensions. Che (1993) and Branco (1997) studied the optimal auction design problems for which bids have multi-dimensions. The auctioneer evaluates the multi-dimensional bids by a scoring function designed for its best interest. In this work, the buyer’s signals are dual-dimensional while bids only have one dimension. Indeed, one of the problems in this work is to derive a bid function for buyers that constitute an equilibrium.

In another study, Chen (2007) studied the supply contract auction which is proved to be equivalent with multi-dimensional auction studied by Che (1993) and Branco (1997) under certain...
assumptions. Vulcano et. al. (2002) studied dynamic auctions under for revenue management which is important for E-commerce. However as suggested in Talluri and van Ryzin (2004) most of the works concerning the auction applications in supply chain management do not take risk-averse into considerations. Our work will try to fill this gap.

The rest of the paper is organized as follows. Section 2 presents the framework of the model for the seller’s optimal capacity sharing decision and the buyer’s incentive compatible bidding strategy. Section 3 and Section 4 deal with the solution approach for risk-neutral and risk averse case. Numerical experiments and results are discussed in Section 5 and in Section 6, points are made towards future research.

2. Model

In this section, we construct models for buyer’s problem and seller’s problem. Some assumptions are proposed and justified in Section 2.1.

2.1. Notation and Assumptions

We consider the problem faced by seller that must determine the portion $x \in [0, 1]$ of capacity share to sell in the auction which we refer to the seller’s problem. Buyers propose appropriate bids $Y$ that best fit their private signals which is referred to the buyer’s problem. The seller acts as a Stackelberg leader who posts the capacity share to be sold which buyers then bid for. A second-price sealed auction is taken as the selling mechanism.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U(\cdot)$</td>
<td>utility function</td>
</tr>
<tr>
<td>$r$</td>
<td>revenue for the total capacity</td>
</tr>
<tr>
<td>$x$</td>
<td>capacity share to be sold in auction</td>
</tr>
<tr>
<td>$Y(x, d, \alpha)$</td>
<td>bidding function or bidding level</td>
</tr>
<tr>
<td>$C$</td>
<td>total cost for the vessel</td>
</tr>
<tr>
<td>$d_i$</td>
<td>expected demand for buyer $i$ (seller if $i = 0$)</td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>financing friction for buyer $i$ (seller if $i = 0$)</td>
</tr>
<tr>
<td>$\epsilon_i$</td>
<td>stochastic noise with $E[\epsilon_i] = 0$</td>
</tr>
</tbody>
</table>

Let $N$ be the number of buyers and let the subscript $i = 1, 2, \ldots, N$ denote different buyers. The subscript $i = 0$ will be used to denote the seller. For every buyer and seller, a dual-dimensional signal vector $(d, \alpha)^T$ is assumed. Here $\alpha$ is a financial friction factor which represents the party’s financial condition and $d$ represents expected demand. We make the following assumptions for the model.

**Assumption 1.** *Every signal is private for $(d_i, \alpha_i)^T, i = 0, 1, \ldots, N$ and $\alpha$ and $d$ are independent.*
Assumption 1 is standard in the private value auction literature. Under this assumption every buyer is privately informed with its own signal \((d, \alpha)^T\). But other buyers’ signals \((d', \alpha')^T\) are unknown random variables from the same distribution. The private information assumption is a reasonable generalization of the real world situation. Usually it is not possible for a company to know its competitors’ information exactly but a distribution can be assigned to the signals due to public apriori information. We assume that the expected demand \(d\) and financial friction \(\alpha\) are independent for computational simplicity although most results in this work do not depend on it.

**Assumption 2.** The demand of buyer \(i\) (the seller if \(i = 0\)) is \(D_i = d_i + \epsilon_i, i = 0, 1, \ldots, N\). \(\epsilon_i \sim iid\) are random noise factors with zero mean.

We can view \(D_i\) as the total demand for \(i\) in the planning horizon and it is reasonable to assume \(\epsilon\) follows normal distribution by applying the Law of Large Number. We make the iid assumption on the random noise for simplicity although in more general cases buyers’ demands can be correlated. The zero mean assumption for \(\epsilon_i\) is not necessary and for simplicity we can add the nonzero part of \(E[\epsilon_i]\) to \(d_i\).

**Assumption 3.** If the payment size is \(z\) the financing cost function is assumed to be \(c(\alpha, z) = \frac{\alpha}{2}z^2\) and the payment function is common knowledge among the seller and buyers.

There are some reasons to adopt the financing cost function in Assumption 3. First, the partial derivative on \(z\) is positive increasing which represent the increasing marginal property of the financing cost. Second, the increasing rate of the marginal financing cost is constant which equals to the financial friction \(\alpha\). Also, the marginal financing cost equals to zero while it is evaluated at \(z = 0\) which is consistent with the heuristics. The assumed financing cost function is the unique continuous function that satisfy the above three conditions. By assuming the financing cost function we can characterize the participants’ financial condition by a single parameter \(\alpha\) which facilitates the analysis and gives us useful insight into the role of financial friction.

To model the buyer’s risk preferences we propose the following assumptions concerning on the concavity of utility function.

**Assumption 4.** Both buyers and the seller are assumed to be risk-averse with the same utility function if not further explained. Thus we have \(U'(\cdot) \geq 0\) and \(U''(\cdot) \leq 0\). Further we assume \(U(\cdot) \geq 0\).

We assume the same utility function for both buyers and seller to simplify the model but do not assume a specific utility function. Only the up-to the second order analytic properties are assumed for the utility function.
2.2. Model Framework

Figure 1 illustrates the framework for the model. The seller acts as the Stackelberger leader and buyers make their decisions based on the seller’s posted capacity share $x$. For any predetermined posted capacity share $x$ every buyer needs to submits a bid $Y$ that indicates how much he is willing to pay for $x$. This bid should best fit the buyer’s interests to maximize his expected utility. The buyer’s problem (BP) is to find out the optimal bid level $Y$ according to its own signal $(d, \alpha)^T$ and the seller’s posted capacity share $x$. Bayesian-Nash equilibrium is an appropriate concept for the solution of BP as a game. A Bayesian-Nash Equilibrium is the state that everyone in the game is not incentivized to change his decision as long as other players keep their decisions unchanged. For a buyer, the Bayesian-Nash Equilibrium bid function $Y^*$ is a function which maps the buyer’s own signal $(d, \alpha)^T$ to the deterministic bid level $Y^*(x, d, \alpha)$ for any predetermined capacity share $x$. But from other buyers’ and the seller’s point of view, a given buyer’s bid level is a random function of $x$ due to the randomness of the buyer’s signal $(d, \alpha)^T$. The seller’s problem (SP) is to select an optimal capacity share $x$ for the auction according to the buyers responds, which will maximize its expected utility. Here we first model the BP and then the SP assuming the existence of the equilibrium. It will be proved in section 3 that equilibrium exists in the second-price auction for both risk-neutral and risk-averse cases.

To analyze the buyer’s problem we assume that the posted capacity share is $x$ and a second-price sealed auction is to be conducted. The buyer’s objective is to maximize its expected utility by choosing an appropriate bid level $Y$. Without loss of generality let us assume the buyer under consideration is buyer $i$, $i = 1, 2, \ldots, N$. We first assume that a unique Bayesian-Nash Equilibrium exists for buyers in the second-price auction. This assumption enables us to formulated the buyer’s problem easier. However we will verify the existence and uniqueness of the equilibrium later. Under this assumption the equilibrium bidding strategy will map the buyer’s signal $(d, \alpha)^T$ to the bid level $Y$, which is also random due to the private information assumption. Let $F(y)$ be the cumulative
distribution function of the bid level \( Y \) and \( G(y) = F_{N-1}^N(y) \) the CDF of the highest bid among any \( N-1 \) independent bids. For buyer \( i \) define \( Z_i = \max \{ Y_j : j \neq i \} \) as the highest bid level among all his competitors. If the buyer’s bid level is \( y \) there are two possible outcomes. The buyer may win the auction with probability \( G(y) \) if \( Z_i < y \). In this case the buyer’s expected utility given winning is \( E[U(\pi_i)|Z_i < y] \). With probability \( 1 - G(y) \) the buyer will lose the auction. In this case he will gain nothing but also lose nothing. Thus the conditional expected utility is \( U(0) \). The profit \( \pi_i \) equals the operational revenue \( r \min(x, d_i + \epsilon) \) minus the financing cost \( \alpha_i \frac{Z_i^2}{2} \) and his payment \( Z_i \).

For each buyer the buyer’s problem can therefore be formulated as follows.

\[
\begin{align*}
\max & \quad E[U(\pi_i) | Z < y] G(y_i) + U(0) (1 - G(y_i)) \\
\text{s.t.} & \quad \pi_i = r \min(x, d_i + \epsilon_i) - \alpha_i^2 Z_i^2 - Z_i \\
& \quad Z = \max \{ y_j : \forall j \neq i \} \\
& \quad y_i \geq 0
\end{align*}
\]

(1)

Notice that without loss of generality we can assume \( U(0) = 0 \). So the buyer’s problem can be reformulated as follows.

\[
\begin{align*}
\max & \quad E[U(\pi_i) | Z < y] G(y) \\
\text{s.t.} & \quad \pi_i = r \min(x, d_i + \epsilon_i) - \alpha_i^2 Z_i^2 - Z_i \\
& \quad Z = \max \{ y_j : \forall j \neq i \} \\
& \quad y_i \geq 0
\end{align*}
\]

(2)

The solution of buyer’s problem is an incentive compatible bidding strategy \( Y^*(x, d, \alpha) \) which is considered to be common knowledge among buyers and seller. For the seller, there’re two sources of revenue. These are revenue from its operations of the reserved capacity \( 1 - x \) and the revenue from the capacity share \( x \) that is sold in auction. The seller’s operational revenue \( OR(x, \epsilon_0) \) is determined by its reserved capacity \( x \) and its demand while the auction revenue \( AR(x, d, \alpha) \) depends on the buyer’s equilibrium and the distribution of their signals. For the cost side, there’re two terms which are financing cost and purchasing cost. The purchasing cost is the total price for the vessel \( C \). And the financing cost \( FC(x, d, \alpha) \) depends on the buyer’s payments. In a second-price auction the payment from the winning bidder equals to the second highest bid level, which is denoted by \( Y^*_{(N-1)}(x, d, \alpha) \). If the winning buyer’s payment is less than the total purchasing cost \( C \) the seller needs to pay the financing cost for the difference \( \frac{\alpha_i}{2} E \left[ \left( C - Y^*_{(N-1)}(x, d, \alpha) \right)^2 \right] \). However if the winning buyer’s payment is higher than \( C \) the seller’s financing cost is zero because he does not need to procure other financing. The seller’s problem for second-price auctions can therefore be summarized as follows.
max $E [U(\text{OR}(x, \epsilon_0) + \text{AR}(x, d, \alpha) - \text{FC}(x, d, \alpha) - C)]$

s.t. \(\text{OR}(x, \epsilon_0) = r \min(1 - x, d + \epsilon_0)\)
\(\text{AR}(x, d, \alpha) = Y^*_i(x, d, \alpha)\)
\(\text{FC}(x, d, \alpha) = \frac{\alpha}{2} E \left[ \left( C - Y^*_i(x, d, \alpha) \right)^+ \right]^2\)
\(x \in [0, 1]\)

(3)

3. Solution for Risk-Neutral Problems

In this section we will focus on the risk-neutral cases. Since the buyers are risk-neutral we set the objective function to be the buyer’s expected profit. The risk-neutral buyer’s problem for the second-price auction is presented as follows.

$$\max E \left[ r E(x, d_i) - \frac{\alpha}{2} Z^2 - Z | Z < y \right] G(y)$$

s.t. \(Z = \max \{y_j : \forall j \neq i\}\)
\(y_i \geq 0\)

(4)

An observation of the buyer’s objective function is that by increasing the bid level \(y\) the buyer’s conditional expected profit will decrease but at the same time the winning probability \(G(y)\) will increase. So the buyer needs to find a bid level that can balance these two mutual contradicting effects. Theorem 1 provides the Bayesian-Nash Equilibrium for the BP.

**Theorem 1.** The Bayesian-Nash Equilibrium bidding function \(Y^*(x, \alpha, d)\) for the buyer’s problem is the solution of the equation below.

$$\frac{\alpha}{2} y^2 + y = r E(x, d)$$

(5)

\(Y^*(x, \alpha, d)\) is concave increasing in \(x\) for any pair of signal \((d, \alpha)^T\). ■

From Theorem 1 we know that if the second-price auction is conducted the equilibrium strategy for buyers is bid to the maximum level that makes their expected profit non-negative. This theorem is consistent with the classic results on second-price auction. To understand the equilibrium we consider the following. Assume the outcome suggested by the equilibrium is \((Y^*_1, Y^*_2, \ldots, Y^*_N)\), we claim that for a bidder (say bidder 1) it is optimal to keep his bid level unchanged if other bidders have not changed their bids. Let us assume \(Y^*_1 = \max(Y^*_1, Y^*_2, \ldots, Y^*_N)\). In this case bidder 1 will win the auction and his expected profit will be \(Y^*_1 - \max(Y^*_i, i \neq 1) > 0\). It is not optimal for bidder 1 to increase the bid because no incremental expected profit is induced by doing so. Decreasing the bid level is not optimal either. Because the winning probability will decrease while the expected profit will not increase if bidder 1 wins the auction in the end. For the case that \(Y^*_1 < \max(Y^*_1, Y^*_2, \ldots, Y^*_N)\)
it is not optimal to decrease the bid level because the bidder will still lost the auction and the expected profit remains 0. Increasing the bid level is not optimal either. This is because nothing changed if the increased bid level is still lower than the winning bid, but if the increased bid level exceeded the former winning bid the bidder’s expected profit will become negative which is even worse than before.

We can characterize an iso-bid curve since we know the equilibrium bidding function. Such a curve divides the entire signal plane into two parts. Signals on the same iso-bid curve share the same bid level and signals in the upper-left area of an iso-bid curve are associated with lower bid levels, while the signals in the lower-right area of the iso-bid curve are associated with higher bid levels. The next corollary gives the formulation of the iso-bid function for the second-price auction case.

**Corollary 1.** The iso-bid function for any bidder is given by the equation below.

\[ A(d, x, y) = \frac{2r}{y^2} E(x, d) - \frac{1}{y} \]  

Figure 2 illustrates the shape of buyer’s equilibrium bid function \( Y^*(x, \alpha, d) \) and the corresponding iso-bid curves. The shape of \( Y^*(x, \alpha, d) \) is displayed in subfigure (a) and subfigure (b) while the shape of \( A(d, x, y) \) is given in subfigure (c) and subfigure (d). We performed two numerical experiments with different posted capacity share \( x \) to see the effects of \( x \) on the bid level and the iso-bid curve. For the low capacity share case \( x \) is set to be 0.3, and it is set to be 0.6 for the high capacity share case. Other parameters for this numerical experiment are \( \alpha = 0.01, \bar{\alpha} = 5, \bar{d} = 0.2, \bar{d} = 0.7, r = 1.5 \) and \( C = 1 \). All these parameters are used in the following numerical experiments if not further explained.

We observe from Figure 2 (a) (b) that the equilibrium bid function \( Y^* \) is increasing in the buyer’s expected demand while decreasing in the financial friction \( \alpha \). Buyers need to consider both demand factor and financial factor in their bidding decisions. With the same demand the buyer can bid more aggressively if his financial condition is better. While if the financial conditions are similar, the buyer will bid more if his expected demand is higher. These trends are more obvious when the capacity share \( x \) is high.

As observed in subfigure (c) and subfigure (d) the iso-bid curve is concave. Every iso-bid curve is attached to a bid level. All signals on the same iso-bid curve are assigned to the same equilibrium
Figure 2  The Equilibrium bid function

bid level. The iso-bid curves on the northwest corner are associated with lower equilibrium bid level while those on the southeast corner are associated with higher equilibrium bid level. To attain a bid level $y$ the buyer’s expected demand must increase if his financial condition becomes worse. When the bid level $y$ is low, the financial friction $\alpha$ and expected demand $d$ will move in a linear manner. However when the bid level $y$ increases a small change in financial friction will require a large change in expected demand to keep the bid level.

We can use iso-bid curves to derive the winning probability that is consistent with the equilibrium bidding function in Theorem 1. The next corollary characterizes the winning probability.

**Corollary 2.** Suppose $N$ buyers participate in the auction. If buyers’ signals are iid with probability density function $f_\alpha(a)$ and $f_d(d)$, then the winning probability is,

$$ G(y) = F^{N-1}(y) $$

(7)

Where $F(y)$ is the cumulative distribute function for bid level $Y^*$. 

$$ F(y) = \int \int \int_{\Omega_y} f_\alpha(a) f_d(d) f(\xi) d\alpha d\xi $$

(8)
\[ \Omega_y = \left\{ (a, \xi) \in [\alpha, \bar{\alpha}] \times [d, \bar{d}] : \alpha \geq A(d, x, y) \right\} \] (9)

\( \Omega_y \) is the area with signals which are associated with bid levels lower than \( y \). By taking integral on this area we obtain the cumulative distribution function (CDF) \( F_Y(y) \) for the equilibrium bid level which is consistent with Theorem 1.

Figure 3 plots the buyer’s winning probability. Subfigure (a) and subfigure (b) are the buyer’s winning probability for the low capacity share case and high capacity share case. In each case we plot the winning probability for \( N = 2, N = 4, N = 6, N = 8 \) and \( N = 10 \). When the number of buyers is small the winning bid is more evenly distributed. When the number of buyers increased the density of winning bid will converge to a higher level quickly. Hence, if there are enough buyers the seller benefits since he has a very good chance to sell the capacity share for a high price. Another observation between (a) and (b) in Figure 3 is that by increasing \( x \), the upper bound for the support of winning bid will increase, which implies benefit from the increasing competition.

Next, we examine the solution of the seller’s problem for second-price auction. For a risk-neutral seller the objective is to maximize the expected profit which consists of expected operational revenue (EOR), expected auction revenue (EAR), expected financing cost (EFC) and the purchasing cost \( C \). The decision variable for the seller is the capacity share \( x \) and the problem can be given by:
The next proposition studies the seller’s expected profit term by term.

**Proposition 1.** Let $EOR(x)$, $EAR(x)$ and $EFC(x)$ denote the expected operational revenue, expected auction revenue and expected financing cost. We have,

1. The seller’s expected operational revenue $EOR(x)$ is concave decreasing in the capacity share $x$.
2. $EAR(x)$ is concave increasing in $x$;
3. $EFC(x)$ is convex decreasing in $x$.

The marginal expected operational revenue for the seller equals to $-r \Pr(\epsilon_0 > 1 - x - d_0)$ which is the production of the revenue for the total capacity share $-r$ and the probability that the realized demand is greater than the reserved capacity $1 - x$. It is clear that the probability of the event that the realized demand is greater than $1 - x$ is increasing in the posted capacity share $x$. Hence, the marginal expected operational revenue is decreasing concave in $x$. Following the same logic we know that the expected operational revenue is increasing concave in the seller’s expected demand $d_0$. Figure 4 illustrates these observations in a more intuitive way. The surface in the figure is the seller’s operational revenue. We can observe that for any capacity share $x$, the operational revenue is increasing in the $d$ when $d$ is less than $1 - x$ but when $d$ exceeds $1 - x$ there is only very little increase in revenue obtained by increasing $d$. With this, and without considering the financial friction selecting a reserved capacity $1 - x$ to fulfill the expected demand is a good strategy for the seller. However, in the presence of financial friction this strategy is not optimal. We will illustrate this point in section 5.1.

The seller’s expected auction revenue $EAR(x)$ and expected financing cost $EFC(x)$ depend on the winning buyer’s payment which equals to the second highest bid among all buyers. We first establish the concavity of the equilibrium bid function as a function of the capacity share $x$ and then show that the concavity of the bid function is a sufficient condition for the concavity of $EAR(x)$ and $EFC(x)$.

Theorem 1 cannot guarantee the path-wise concavity of the $(N - 1)$-th order statistic $Y^*_{(N-1)}(x, \alpha, d)$. In fact it is not difficult to construct counter examples to prove that the $(N - 1)$-th
order statistic is not concave in general case. We can however prove the expectation of $Y^*_{(N-1)}$ is concave in $x$. The idea is that for any $x$ we can order $Y^*_i, i = 1, 2, \ldots, N$ strictly with probability 1 in a neighborhood of $x$. The order is disrupted with probability 0 which can be omitted while taking expectation. First, we present two lemmas which deal with the probabilistic property of the equilibrium bid function $Y^*_{(N-1)}(x, \alpha, d)$ and the continuity of the iso-bid function $A(d, x, y)$ respectively.
Figure 5 illustrates the shape of the seller’s expected auction revenue against the capacity share \( x \). When capacity is sold the seller can get more revenue from the auction. If there are more buyers participating in the auction the revenue for the seller will be higher. This phenomenon is sharper if the capacity share \( x \) is large. Figure 6 shows the buyer’s expected financing cost against his financial friction \( \alpha \) and the decision on capacity share \( x \). Generally, the expected financing cost is increasing with the financial friction. If the financial condition for the seller is bad he may need to concede high financing cost to support his plan. However, as illustrated in Figure 6 the seller can ease his financial pressure by selling more capacity. As \( x \) increases, the seller is able to collect more money from the winning buyer’s payment which is interest-free. Hence, he does not need to procure as much funds as he would have to. The expected financing cost will be reduced dramatically especially for the seller with poor financial condition.

Proposition 1 imply the existence and uniqueness of the optimal posted capacity share to be sold \( x^* \) for the seller. The following theorem gives the solution for this optimal capacity decision.

**Theorem 2.** There exists a unique optimal capacity sharing decision \( x^* \) for the seller’s problem which satisfy the following first order condition.

\[
\frac{r \Pr (\epsilon_0 > 1 - x - d_0)}{dx} = \frac{d \left( E \left[ Y_{(N-1)}^* (x) \right] \right)}{dx} - \frac{\alpha_0}{2} \frac{d \left( E \left[ \left( C - Y_{(N-1)}^* (x) \right)^+ \right]^2 \right)}{dx}
\]  

(11)
The next proposition provides the relationship between the optimal capacity share $x^*$ and the seller’s signal $(\alpha_0, d_0)$.

**Proposition 2.** The seller’s optimal capacity share to be sold $x^*$ is decreasing in the seller’s expected demand $d_0$ while increasing in the seller’s financial friction $\alpha_0$. ■

If the seller’s expected demand $d_0$ is increasing it is more likely that he will have a higher realized demand. Thus he is incented to reserve more capacity to fulfill his future demands which will lead to more revenue. This will result in a shrinkage in the optimal capacity share $x^*$. But if the seller’s financing friction $\alpha_0$ is increasing the seller’s relative advantage in financing will decrease. It is better for the seller to put more capacity share to sale as a way to release the financial pressure even his expected demand is relative high. The capacity-cost sharing scheme and the auction mechanism offer better chance for the seller to optimize its capacity and financing decisions.

4. **Solution for Risk-Averse Problems**

In this section we focus our attention on risk-averse cases. First we develop a proposition on the BP in Equation (2).

**Proposition 3.** $U(\cdot)$ is a concave increasing utility function, $U(0) = 0$. The buyer’s Bayesian-Nash equilibrium bid function $Y^*(x, \alpha, d)$ is the solution of the equation bellow.

$$E_x \left[ U \left( r \min(x, d + \epsilon) - \frac{\alpha}{2} y^2 - y \right) \right] = 0$$  \hfill (12)
The result is similar to the risk-neutral case in Section 3. The only difference is that the equilibrium bidding strategy requires the expected utility instead of expected profit equals to $U(0)$.

**Proposition 4.** Let $Y^*_{RN}(x, \alpha, d)$ and $Y^*_{RA}(x, \alpha, d)$ be the equilibrium bidding functions for risk-neutral and risk-averse buyers respectively. We have $Y^*_{RN}(x, \alpha, d) \geq Y^*_{RA}(x, \alpha, d)$, $\forall x, \alpha, d$. 

Proposition 4 states the fact that the buyer will bid less compared to the risk-neutral case due to risk-aversion. The next proposition is about the concavity of $Y^*(x, \alpha, d)$ as a function of the capacity share $x$.

**Proposition 5.** Fixing any signal $(\alpha, d)$, the equilibrium bidding function $Y^*(x, \alpha, d)$ for buyer’s problem is concave increasing in the capacity share $x$.

The concavity is reserved for the equilibrium bid function $Y^*(x, \alpha, d)$ in the case that the buyers are risk-averse. This is the result of the monotonicity of the utility function $U(\cdot)$. In the next theorem, we find that the concavity of $Y^*(x, \alpha, d)$ is a sufficient condition for the existence and the uniqueness of optimal capacity share $x^*$ for risk-averse seller.

**Theorem 3.** The optimal posted capacity share decision $x^*$ is uniquely exist for the risk-averse seller. And it solves the first order condition.

Similar results about the relationship between the optimal capacity share $x^*$ and the buyer’s signal $(\alpha_0, d_0)$ can be derived for risk-averse seller.

**Proposition 6.** The seller’s optimal capacity share to be sold $x^*$ is decreasing in the seller’s expected demand $d_0$ while increasing in the seller’s financial friction $\alpha_0$.

## 5. Numerical Examples

A numerical example will be used to illustrate the benefits of the capacity-cost sharing scheme and auction implementation. In this example there are 10 bidders participating in the auction and buyer’s expected demand $d_i$ are randomly generated from $U(0.2, 0.8)$ and financial friction $\alpha_i$ values are obtained from $U(0.01, 5.00)$. The purchasing cost $c$ is set to be 1 and the total revenue $r$ is set to be 2.25. Second-price sealed auction is implemented and all participants are risk-neutral. The benefits of the capacity-cost sharing scheme is studied in Section 5.1 and in Section 5.2, we compare the seller’s optimal strategy with an alternative strategy. The buyer’s expected cost share is compared in Section 5.3. In Section 5.4 we tests the effect of seller’s attitude on risk to its capacity decision.
5.1. The Benefits of Capacity-Cost Sharing Scheme

We study the benefits of the capacity-cost sharing scheme for the seller in this section. Second-price sealed auction is implemented to select a partner for the shipping company. As the alternative scheme the seller will choose not to share the capacity. Under this scheme the seller will pay all the purchasing cost by its own. We compare the expected profits for the seller for different signals. The result is displayed in Table 3.

<table>
<thead>
<tr>
<th>Signals</th>
<th>Expected Profits</th>
<th>Benefits for Seller</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_0$</td>
<td>$\alpha_0$</td>
<td>Share</td>
</tr>
<tr>
<td>0.20</td>
<td>0.01</td>
<td>0.3028</td>
</tr>
<tr>
<td>0.20</td>
<td>1.70</td>
<td>0.2737</td>
</tr>
<tr>
<td>0.20</td>
<td>3.30</td>
<td>0.2465</td>
</tr>
<tr>
<td>0.20</td>
<td>5.00</td>
<td>0.2180</td>
</tr>
<tr>
<td>0.40</td>
<td>0.01</td>
<td>0.6486</td>
</tr>
<tr>
<td>0.40</td>
<td>1.70</td>
<td>0.6031</td>
</tr>
<tr>
<td>0.40</td>
<td>3.30</td>
<td>0.5628</td>
</tr>
<tr>
<td>0.40</td>
<td>5.00</td>
<td>0.5219</td>
</tr>
<tr>
<td>0.60</td>
<td>0.01</td>
<td>0.9142</td>
</tr>
<tr>
<td>0.60</td>
<td>1.70</td>
<td>0.8205</td>
</tr>
<tr>
<td>0.60</td>
<td>3.30</td>
<td>0.7475</td>
</tr>
<tr>
<td>0.60</td>
<td>5.00</td>
<td>0.6801</td>
</tr>
<tr>
<td>0.80</td>
<td>0.01</td>
<td>1.1002</td>
</tr>
<tr>
<td>0.80</td>
<td>1.70</td>
<td>0.9086</td>
</tr>
<tr>
<td>0.80</td>
<td>3.30</td>
<td>0.7959</td>
</tr>
<tr>
<td>0.80</td>
<td>5.00</td>
<td>0.7085</td>
</tr>
</tbody>
</table>

As observed from Table 3 the capacity-cost sharing scheme performs much better than the alternative non-sharing scheme. Under the non-sharing scheme the shipping company’s expected profits are negative for most cases. Even the buyer’s expected demand is high (for example $d_0 = 0.8$) its profit could be negative if the financial friction is high. This is partially due to the high financing cost. The rational decision for the shipping company is to abort the purchasing plan in this case. However if the shipping company adopts the capacity-cost sharing scheme and implement the optimal strategy its expected profits will be positive for the same signals. The capacity-cost sharing scheme offers the shipping company the opportunity to make extra profits by expending its shipping capacity which it can not afford alone. The last column of Table 3 represents the benefits of the sharing scheme compared to the non-sharing scheme. The benefits are significant for the seller and buyers can also stand to benefit. Since the winning buyer only needs to pay the second highest bid, he will have a positive expected profit from the shared capacity.
5.2. The Benefits of the Optimal Strategy

Here, we will compare the performance of the seller’s optimal capacity decision (Strategy A) with an reasonable alternative strategy (Strategy B) under the capacity-cost sharing scheme.

**Strategy A:** The percentage $x^*_A$ of capacity to share is determined by Theorem 2.

**Strategy B:** The percentage $x^*_B$ of capacity to share is set to be $1 - d_0$.

<table>
<thead>
<tr>
<th>Signals</th>
<th>$d_0/\alpha_0$</th>
<th>$x^*_A$</th>
<th>$\pi^*_A$</th>
<th>$x^*_B$</th>
<th>$\pi^*_B$</th>
<th>$(\pi^<em>_A - \pi^</em>_B)/\pi^*_B$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20 0.01</td>
<td>0.6967 0.3028</td>
<td>0.80 0.2533</td>
<td>19.54%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20 1.70</td>
<td>0.7023 0.2737</td>
<td>0.80 0.2279</td>
<td>20.10%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20 3.30</td>
<td>0.7080 0.2465</td>
<td>0.80 0.2019</td>
<td>20.92%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20 5.00</td>
<td>0.7131 0.2180</td>
<td>0.80 0.1783</td>
<td>22.24%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.40 0.01</td>
<td>0.5527 0.6486</td>
<td>0.60 0.6362</td>
<td>1.94%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.40 1.70</td>
<td>0.5658 0.6031</td>
<td>0.60 0.5973</td>
<td>0.96%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.40 3.30</td>
<td>0.5830 0.5628</td>
<td>0.60 0.5606</td>
<td>0.40%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.40 5.00</td>
<td>0.5921 0.5219</td>
<td>0.60 0.5215</td>
<td>0.08%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60 0.01</td>
<td>0.4010 0.9142</td>
<td>0.40 0.9141</td>
<td>0.01%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60 1.70</td>
<td>0.4421 0.8205</td>
<td>0.40 0.8079</td>
<td>1.57%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60 3.30</td>
<td>0.4712 0.7475</td>
<td>0.40 0.7073</td>
<td>5.69%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60 5.00</td>
<td>0.4940 0.6801</td>
<td>0.40 0.6004</td>
<td>13.28%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80 0.01</td>
<td>0.2454 1.1002</td>
<td>0.20 1.0889</td>
<td>1.04%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80 1.70</td>
<td>0.3451 0.9086</td>
<td>0.20 0.7642</td>
<td>18.89%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80 3.30</td>
<td>0.4134 0.7959</td>
<td>0.20 0.4569</td>
<td>74.21%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80 5.00</td>
<td>0.4575 0.7085</td>
<td>0.20 0.1303</td>
<td>443.86%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Strategy A adopt the result with the optimal capacity share while Strategy B select to sell its expected extra capacity. Notice that expectation is the best estimation for the future demand from statistical perspective. So if the shipping company considers the capacity decision and financial friction separately it is quite reasonable to adopt strategy B.

Observed from Table 4 strategy A performs much better than strategy B. Especially when the shipping company’s expected demand is high ($d_0 = 0.80$) or low ($d_0 = 0.20$) the expected profit of strategy A is 20% higher than strategy B. We can see that when the the expected demand is low ($d_0 = 0.2$ or 0.4) the seller tends to sell less capacity than its expected extra capacity. The reason is when $x$ increases buyers will have more financial pressure on bidding higher. So the marginal benefit from increasing $x$ decreases. For the opposite case, if the expected demand is high ($d_0 = 0.60$ or 0.80) the seller tends to sell more than its expected extra capacity. By selling more the shipping company can take advantage of buyers’ financial condition and competition.
5.3. Cost Sharing Comparisons

In this section study the expected buyer’s cost as a percentage of the total purchasing cost. We compare the buyer’s cost share with its capacity share $x$. We find in Table 5 that the buyer’s share on purchasing cost is larger than its share on the capacity for all signal combinations in the numerical examples. The difference between buyer’s cost share and capacity share is significant. For most cases in our example the gap is more than 20%. This is a benefit of auction. Usually there is sufficient negotiation space between buyer and seller. The final price is determined by player’s power and information in a bilateral negotiation. But under the scenario that buyers are forced to compete with each other, the final price will move towards the winning buyer’s bottom line.

Table 5 Cost Share Comparison

<table>
<thead>
<tr>
<th>$d_0$</th>
<th>$\alpha_0$</th>
<th>Cost Share (%)</th>
<th>$x^*$ (%)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.01</td>
<td>87.06%</td>
<td>69.67%</td>
<td>17.38%</td>
</tr>
<tr>
<td>0.20</td>
<td>1.70</td>
<td>87.26%</td>
<td>70.23%</td>
<td>17.01%</td>
</tr>
<tr>
<td>0.20</td>
<td>3.30</td>
<td>87.42%</td>
<td>70.80%</td>
<td>16.62%</td>
</tr>
<tr>
<td>0.20</td>
<td>5.00</td>
<td>87.57%</td>
<td>71.31%</td>
<td>16.26%</td>
</tr>
<tr>
<td>0.40</td>
<td>0.01</td>
<td>79.53%</td>
<td>55.27%</td>
<td>24.26%</td>
</tr>
<tr>
<td>0.40</td>
<td>1.70</td>
<td>80.44%</td>
<td>56.58%</td>
<td>23.86%</td>
</tr>
<tr>
<td>0.40</td>
<td>3.30</td>
<td>81.58%</td>
<td>58.30%</td>
<td>23.28%</td>
</tr>
<tr>
<td>0.40</td>
<td>5.00</td>
<td>82.15%</td>
<td>59.21%</td>
<td>22.94%</td>
</tr>
<tr>
<td>0.60</td>
<td>0.01</td>
<td>65.57%</td>
<td>40.10%</td>
<td>25.47%</td>
</tr>
<tr>
<td>0.60</td>
<td>1.70</td>
<td>69.94%</td>
<td>44.21%</td>
<td>25.73%</td>
</tr>
<tr>
<td>0.60</td>
<td>3.30</td>
<td>72.78%</td>
<td>47.12%</td>
<td>25.66%</td>
</tr>
<tr>
<td>0.60</td>
<td>5.00</td>
<td>74.84%</td>
<td>49.40%</td>
<td>25.44%</td>
</tr>
<tr>
<td>0.80</td>
<td>0.01</td>
<td>45.17%</td>
<td>24.54%</td>
<td>20.63%</td>
</tr>
<tr>
<td>0.80</td>
<td>1.70</td>
<td>58.94%</td>
<td>34.51%</td>
<td>24.43%</td>
</tr>
<tr>
<td>0.80</td>
<td>3.30</td>
<td>66.94%</td>
<td>41.34%</td>
<td>25.60%</td>
</tr>
<tr>
<td>0.80</td>
<td>5.00</td>
<td>71.47%</td>
<td>45.75%</td>
<td>25.73%</td>
</tr>
</tbody>
</table>

5.4. The Effect of Risk-Aversion

We study the effects of the seller’s attitude to risk on the optimal capacity sharing decision. We assume the seller will evaluate its decision by a Mean-Variance objective function. Mean-Variance analysis are applied in modeling risk-averse decision making processes in operations management area. See Ding et. al. (2007) and Chen and Federgruen (2000) for examples. Under the Mean-Variance framework the seller’s objective function will become the expected profit minus $\lambda$ times of profit variance. It is known that under certain assumptions the Mean-Variance objective function is equivalent to expected utility. $\lambda$ characterize the seller’s attitude on risk. The seller is more risk-averse if its $\lambda$ is larger. In our numerical study we compare the case when $\lambda = 0, 1, 2, 4$. Figure 8 displays the results. The two subfigures in the first row are about the optimal capacity sharing
decision $x^*$ for the case that $\alpha = 0.01$ and 5.00. The subfigures in the second row of the figure display the optimal value of the Mean-Variance objective for the corresponding cases. The four curves with different colours are for different $\lambda$. The blue curve is for the case $\lambda = 0$, green for $\lambda = 1$, red for $\lambda = 2$ and aqua for $\lambda = 4$. It’s clear that when $\lambda$ become larger and larger the seller tend to share more of its capacity with its partner. The reason for this phenomenon is as bellow. If the seller is more risk-averse it will be incented to mitigate its demand risk by sharing more capacity with its partner. As the result, it is shown in Figure 8 that if its $\lambda$ is higher the seller’s Mean-Variance objective function value will be smaller. An interesting observation from Figure 8 is, compares to the case when the seller’s financial friction $\alpha_0$ is low, when $\alpha_0$ become higher (equals to 5.00) the seller will become more conservative in capacity decision if he is more risk-averse. When expected demand $d_0$ is high the seller have very good chance to utilized more capacity so the demand risk is relatively low. However if the seller decides to sell more capacity through auction the revenue from the auction is random which indicates higher risk. A risk-sensitive seller will prefer to reserve more capacity to take advantage of the relatively low demand risk while avoid high risk in revenue from the auction. However when the expected demand $d_0$ is low the demand risk is relatively high.

Figure 8  Seller’s Optimal Capacity Decision for Risk-Averse seller
so the seller tends to act more aggressive to sell more capacity through auction.

6. Conclusion and Extensions

In this paper, we studied the risk management problem in shipping capacity procurement and introduced a capacity-cost sharing scheme for the problem and a general model which employs an auction mechanism to select a partner. We derived the optimal capacity decision for the seller and the equilibrium bidding strategy for buyers for second-price sealed auction. All results are extended to the case that the buyers and seller are risk-averse.

Several assumptions are made in this paper. The private information assumption we made is commonly found in the auction literature. We modeled the auction participants’ financial situations by taking the financing cost function \( c(\alpha, z) \) to be the product of a financial friction factor \( \alpha \) and \( \frac{z^2}{2} \). This financing function is the only continuous function that satisfies the conditions which is reasonable for the financing cost and is consistent with Che and Gale (1998). The participants’ demands are assumed to be the summation of the expected demand and noise. We have assumed that noise factors are independent identical distributed and that both the seller and buyers share the same utility function.

The numerical examples illustrate the benefits of the capacity-cost sharing scheme. Compared to the non-sharing scheme, the seller will have better chance to make more profits. For many cases in the example, the buyer is unable to afford a vessel purchase without using a sharing scheme. Buyers also benefit from the sharing scheme since it offers them an opportunity to expand their shipping capacity within reasonable prices. Besides this, the utilization of the capacity is increased under the sharing scheme which benefits all. The implementation of an auction increases the competition among buyers. According to the numerical example, the expected percentage of cost shared by the winning buyer is significantly higher than the percentage of capacity it bid for. We also compared the seller’s optimal capacity decision strategy with a reasonable alternative strategy in which the seller considers the capacity decision separately from financial friction. The result demonstrates the seller’s expected profit will increase significantly if it adopts the optimal strategy suggested by this paper. We introduce the Mean–Variance objective function to test the effect of seller’s attitude to risk on its capacity decisions. The numerical results show that the seller will want to share more capacity with a partner to mitigate demand risk if he is more risk-averse.

For future research, there remains a number of areas which will be of interest. These are described briefly here:

(i) Solutions for First-Price Sealed Auction: The uniqueness of the equilibrium bid function and the concavity of the equilibrium bid level on the capacity share \( x \) are not proved. The equilibrium
bid function for multi-dimensional auctions was considered to be a hard problem in auction research for a long time. The main difficulty is that it is difficult to define an appropriate order on the multi-dimensional signals. However for a particular problem structure it is possible to take advantage of an iso-bid function. An equilibrium bid function can possibly be developed.

(ii) Comparisons of the differences between risk-neutral case and risk-averse case for both first- and second-price auctions: These are another extension and not difficult for buyer’s problem. It is possible that further research will throw light on the buyer’s decision by comparing the derivatives of their bid functions. However, the seller’s problem may be more complicated, although the approach given here will continue to work.

(iii) Comparisons of the differences between First-Price Sealed Auction and Second-Price Sealed Auction for both risk-neutral and risk-averse case: The results on these comparisons are not clear. Che and Gale (1998) developed a method to make these comparisons.

(iv) The capacity-cost sharing scheme can be researched in other industries, for example, in manufacturing and service capacity decision making.

Acknowledgments
This work is supported by research grand 10-C207-SMU-012, Singapore Management University.

Appendix A: Proof of Theorem 1.

Proof of Theorem 1. The expected profit for the buyer for bidding $y$ is

$$\Pi(y) = \left( rE(x,d) - E \left[ \frac{\alpha}{2} z^2 + z | z < y \right] \right) G(y)$$

$$= rE(x,d)G(y) - \int_0^y \left( \frac{\alpha}{2} t^2 + t \right) dG(t)$$

Set the first order derivative equals to zero. We get the equation bellow.

$$\left( rE(x,d) - \frac{\alpha}{2} y^2 - y \right) G'(y) = 0$$

$$\Rightarrow$$

$$rE(x,d) - \frac{\alpha}{2} y^2 - y = 0$$

So the bid function is

$$Y^*(x,\alpha,d) = \frac{\sqrt{1 + 2\alpha rE(x,d)} - 1}{\alpha}$$
Next we’ll prove the concavity of the equilibrium bidding strategy. Choose an arbitrary pair of signal from the sample space, keep it fixed. Without loss of generality let’s assume the signal is \((\alpha, d)\). By taking derivatives we have the following equations which proved \(E(x,d)\) is concave increasing in \(x\).

\[
\begin{align*}
E_1'(x,d) &= F_\epsilon(x - d) \\
E_{11}''(x,d) &= -f_\epsilon(x - d)
\end{align*}
\]

(18)

Here \(E_1'\) and \(E_{11}''\) are first order and second order partial derivatives on \(x\).

Due to the monotonicity of \(Y^*(x,\alpha,d)\) on \(E(x,d)\), \(Y^*(x,\alpha,d)\) is concave increasing in \(x\). Since \((\alpha, d)\) is arbitrarily selected Theorem 1 is proved. □

Appendix B: Proof of Proposition 1.


\[
\text{EOR}(x) = E_{\epsilon_0} [r \min(1 - x, d_0 + \epsilon_0)]
\]

(19)

By taking derivative of \(\text{EOR}(x)\) we have.

\[
\frac{d\text{EOR}(x)}{dx} = -r \Pr(\epsilon_0 > 1 - x - d_0) \leq 0
\]

(20)

\[
\frac{d^2\text{EOR}(x)}{dx^2} = -rf_{\epsilon_0}(1 - x - d_0) \leq 0
\]

(21)

This proved statement 1 of Proposition 1.


We need two lemmas to prove statement 2 and 3.

Lemma 1. For any given capacity share \(x\), \(\Pr(Y_i = Y_j, \forall i \neq j) = 0\).

Lemma 1 claims that fix \(x\) the probability that any two bids from different buyers equals is zero. Since the \(d\) and \(\alpha\) are continuous random variables, the even that any two bids equals is in a zero-measure set. Thus the probability is zero.

Lemma 2. Fix the expected demand \(d\), as a function of \((x,y)\) the iso-bid function \(A(d;x,y)\) is continuous.

With the continuity, there exist a neighborhood for every \(x\) in which the order of bids reserved. Together with Lemma 1 we can prove the concavity of the winning buyer’s expected payment for second price auction which is the key to prove the concavity of \(\text{EAR}(x)\) and \(-\text{EFC}(x)\).

The basical idea of is that we can prove \(\text{EAR}(x)\) (\(\text{EFC}(x)\)) is pointwise concave (convex) in the interval \(x \in [0,1]\). This implies that \(\text{EAR}(x)\) (\(\text{EFC}(x)\)) is concave (convex) in \([0,1]\). We take the advantage of the continuity of \(A(d,x,y)\) in Lemma 2. For a sample path we assume that all bids at \(x\) can be ranked strictly. Due to the continuity of \(A(d,x,y)\) this rank will not be destroyed in a neighborhood of \(x\). Since we know that the
equilibrium bid function is concave increasing in \( x \) from Proposition 1 \( \text{EAR}(x) \) (EFC(\( x \))) is concave (convex) in this neighborhood. This local concavity (convexity) is valid for any sample with positive probability because of Lemma 1. So the concavity of \( \text{EAR}(x) \) and the convexity of \( \text{EFC}(x) \) are proved.

\( \forall x_0 \in [0, 1], \) let \( D = [d, \bar{d}] \), \( A = [\alpha, \bar{\alpha}] \). Define

\[
I = \left\{ \prod_{i=1}^{N-1} (\alpha_i, d_i) \in \prod_{i=1}^{N-1} (A \times D) : \exists i \neq j, Y_i(\alpha_i, d_i; x) = Y_j(\alpha_j, d_j; x) \right\}
\]

(22)

\( \forall \prod_{i=1}^{N-1} (\alpha_i, d_i) \in I^c \)

Without lose of generality assume that

\[
Y_1 < Y_2 < \cdots < Y_{N-1} < Y_N
\]

(23)

\( \forall Y_i < Y_{i+k}, \exists Y \) that \( Y_i < Y < Y_{i+k} \). And

\[
\left\{ \begin{array}{l}
A(d_i; x, Y) < \alpha_i \\
A(d_{i+k}; x, Y) > \alpha_{i+k}
\end{array} \right.
\]

(24)

Denote \( r_1 = \alpha_i - A(d_i; x, Y) \) and \( r_2 = A(d_{i+k}; x, Y) - \alpha_{i+k} \). Due to the continuous of iso-bid function (refers to Lemma 2). There exist \( \delta_1 \) and \( \delta_2 \) that

\[
\forall (\hat{x}_1, \hat{y}_1) \in \{ (\hat{x}, \hat{y}) : \| (\hat{x}, \hat{y}) - (x, Y) \|_2 < \delta_1 \} \text{ we have } A(d_i; \hat{x}_1, \hat{y}_1) < \alpha_i;
\]

\[
\forall (\hat{x}_2, \hat{y}_2) \in \{ (\hat{x}, \hat{y}) : \| (\hat{x}, \hat{y}) - (x, Y) \|_2 < \delta_2 \} \text{ we have } A(d_{i+k}; \hat{x}_2, \hat{y}_2) < \alpha_i;
\]

Let \( \delta = \min(\delta_1, \delta_2), \exists \Delta > 0 \) that

\[
(x + \Delta, y) \in \{ (\hat{x}, \hat{y}) : \| (\hat{x}, \hat{y}) - (x, Y) \|_2 < \delta_2 \}, \text{ which satisfy the following equations.}
\]

\[
\left\{ \begin{array}{l}
A(d_i; x + \Delta, y) < \alpha_i \\
A(d_{i+k}; x + \Delta, y) > \alpha_{i+k}
\end{array} \right.
\]

(25)

This means for all \( x \in (x_0, x_0 + \Delta), Y_i(x) < Y_{i+k}(x) \) if \( Y_i(x_0) < Y_{i+k}(x_0) \). So \( Y_{(N-1)(x)} = Y_{N-1}(x), \forall x \in (x_0, x_0 + \Delta) \).

Since \( Y_{N-1}(x) \) is concave increasing in \( x \) for all \( \prod_{i=1}^{N-1} (\alpha_i, d_i) \in I^c \). We have

\[
\text{EAR}(x) = E[Y_{(N-1)}] = \left[ Y_{(N-1)} \right] \mathbb{E} \left[ \prod_{i=1}^{N-1} (\alpha_i, d_i) \in I^c \right] \mathbb{P} \left( \prod_{i=1}^{N-1} (\alpha_i, d_i) \in I^c \right)
\]

\[
+ \mathbb{E} \left[ Y_{(N-1)} \right] \mathbb{P} \left( \prod_{i=1}^{N-1} (\alpha_i, d_i) \in I \right) \mathbb{P} \left( \prod_{i=1}^{N-1} (\alpha_i, d_i) \in I \right)
\]

\[
= \left[ Y_{(N-1)} \right] \mathbb{E} \left[ \prod_{i=1}^{N-1} (\alpha_i, d_i) \in I^c \right]
\]

(26)

Which is concave increasing in \( x \), \( \forall x \in [x_0, x_0 + \Delta] \).
Since \( [(C - Y_i(x))^+]^2 \) is convex decreasing in \( x \), \( \forall x \in [0, 1] \), for the same reason presented in the prove of (2) we can prove (3).

4. Proof of Lemma 1.

\[
\Pr(Y_i = Y_j | Y_j = t) = \int\int_{(\alpha, d) \in \{(\alpha, d): \alpha = A(d, x, t)\}} f_\alpha f_d d\alpha dd = 0
\]

\[
\Rightarrow \Pr(Y_i = Y_j) = \int_{L(x)}^{U(x)} \Pr(Y_i = Y_j | Y_j = t) f_{Y_j}(t) dt = 0
\]

Here \( L(x) = \sup\{y : F_{Y}(y) = 0\} \) and \( U(x) = \inf\{y : F_{Y}(y) = 1\} \). This proved Lemma 1.

\[\square\]

**Appendix C: Proof of Theorem 2.**

**Proof of Theorem 2.** Since \( \text{EOR}(x) \) and \( \text{EAR}(x) \) are concave increasing in \( x \) and \( \text{EFC}(x) \) is convex decreasing in \( x \), the seller’s objective function is concave increasing in \( x \). So the seller’s problem has unique optimal solution. The solution is determined by the first order condition which is presented as below.

\[
rPr(\epsilon_0 > 1 - x - d_0) = \frac{d\left(\frac{E[Y_{(N-1)}(x)]}{dx}\right)}{\alpha_0} - \frac{\alpha_0}{2} \frac{d\left(\frac{E\left[(C - Y_{(N-1)}(x))^+\right]^2}{dx}\right)}{dx}
\]

This proved Theorem 2. \[\square\]

**Appendix D: Proof of Proposition 2.**

**Proof of Proposition 2.** Define \( H(x, \alpha, d) \) as followed.

\[
H(x, \alpha, d) = \frac{r \partial E[1 - x, d_0]}{\partial x} + \frac{\partial (\text{EAR}(x))}{\partial x} - \frac{\alpha_0}{2} \frac{\partial \left(\frac{E\left[(C - Y_{(N-1)}(x))^+\right]^2}{dx}\right)}{\partial x}
\]

\[
\frac{\partial H}{\partial x^*} = \frac{d^2\text{EOR}(x)}{dx^2} + \frac{d^2\text{EAR}(x)}{dx^2} - \frac{d^2\text{EFC}(x)}{dx^2} \leq 0
\]

\[
\frac{\partial H}{\partial \alpha_0} = -\frac{1}{\alpha_0} \frac{d\text{EFC}(x)}{dx} \geq 0
\]

\[
\frac{\partial H}{\partial d_0} = -rf_{\alpha_0}(1 - x - d_0) \leq 0
\]
So the partial derivatives of \( x^* \) are as followed.

\[
\frac{\partial x^*}{\partial \alpha_0} = \frac{\partial H}{\partial \alpha_0} \geq 0 \tag{34}
\]

\[
\frac{\partial x^*}{\partial d_0} = -\frac{\partial H}{\partial d_0} \frac{\partial H}{\partial x^*} \leq 0 \tag{35}
\]

These proved Proposition 2. \( \square \)

**Appendix E: Proof of Proposition 3.**

*Proof of Proposition 3.* Since \( \epsilon \) is independent with \( Y_i, \forall i \).

\[
E [U (\pi(x, \alpha, d, Z, \epsilon))] | Z < y] = E_Z [E, [U (\pi(x, \alpha, d, Z, \epsilon))] | Z < y] \tag{36}
\]

We have

\[
E [U (\pi(x, \alpha, d, Z, \epsilon))] | Z < y] G(y) = \int_0^y E, [U (\pi(x, \alpha, d, t, \epsilon))] G'(t) dt \tag{37}
\]

Taking the derivative of the buyer’s expected profit refers to the bid level \( y \) results the following equation.

\[
E, [U (\pi(x, \alpha, d, y, \epsilon))] G'(y) - U(0) G'(y) = 0 \tag{38}
\]

So we have

\[
E, [U (\pi(x, \alpha, d, y, \epsilon))] = U(0) \tag{39}
\]

\[\Rightarrow\]

\[
E, \left[ U \left( r \min(x, d + \epsilon) - \frac{\alpha}{2} y^2 - y \right) \right] = U(0) \tag{40}
\]

Which proved the proposition. \( \square \)

**Appendix F: Proof of Proposition 5.**

*Proof of Proposition 5.*

\[
F(x, y) \overset{\text{def}}{=} E, \left[ U \left( r \min(x, d + \epsilon) - \frac{\alpha}{2} y^2 - y \right) \right] - U(0) \tag{41}
\]

We know that by definition

\[
F(x, y) \equiv 0 \tag{42}
\]
So we have

\[
\begin{align*}
\begin{cases}
\frac{dF(x,y)}{dx} = 0 \\
\frac{d^2F(x,y)}{dx^2} = 0
\end{cases}
\end{align*}
\]

Thus,

\[
\begin{align*}
\begin{cases}
y'(x) = -\frac{F'_1}{F'_2} \\
y''(x) = -\frac{F''_{11}+F''_{12}y'+(F''_{21}+F''_{22})y'}{F'_2}
\end{cases}
\end{align*}
\]

It is obvious that there's no cross term in \( F \) so we know that

\[
F''_{12} = F''_{21} = 0
\]

Thus,

\[
y''(x) = \frac{F''_{11} + F''_{22} \left( \frac{F'}{F'_2} \right)^2}{F'_2}
\]

Since

\[
F'_1 = \frac{\partial}{\partial x} \left( E \left[ U \left( r \min(x, d + \epsilon) - \frac{\alpha}{2} y^2 - y \right) \right] \right)
= \int_\xi^\zeta U' \cdot \left( r \frac{d(\min(x, d + \epsilon))}{dx} \right) f_\epsilon(t) \, dt 
= \int_\xi^\zeta U' \cdot (rI_{1 : x < d + \epsilon}) f_\epsilon(t) \, dt \geq 0
\]

\[
F''_{11} = \frac{\partial^2 F'_1}{\partial x^2}
= \int_\xi^\zeta U'' \cdot (rI_{1 : x < d + \epsilon})^2 f_\epsilon(t) \, dt \leq 0
\]

\[
F'_2 = \frac{\partial}{\partial y} \left( E \left[ U \left( r \min(x, d + \epsilon) - \frac{\alpha}{2} y^2 - y \right) \right] \right)
= \int_\xi^\zeta U' \cdot \left( d \left( -\frac{\alpha}{2} y^2 - y \right) \right) f_\epsilon(t) \, dt 
= - \int_\xi^\zeta U' \cdot (\alpha y + 1) f_\epsilon(t) \, dt \leq 0
\]

\[
F''_{22} = \frac{\partial^2 F'_2}{\partial y^2}
= - \int_\xi^\zeta \left( -U''(\alpha y + 1)^2 + \alpha U' \right) f_\epsilon(t) \, dt \leq 0
\]

Therefore,

\[
y''(x) = -\frac{F''_{11} + F''_{22} \left( \frac{F'}{F'_2} \right)^2}{F'_2} \leq 0
\]

This proved the proposition. \( \square \)
Appendix G: Proof of Theorem 3.

Proof of Theorem 3. Since the concavity (convexity) of the operational revenue term and auction revenue term (financing cost term) are reserved, and the utility function is monotonically increasing the objective function is concave in $x$. So the first order condition gives the optimal solution. □

References


The Effect of Payment Timing on Inventory Decisions in a Newsvendor Experiment

Li Chen • A. Gürhan Kök • Jordan Tong

The Fuqua School of Business, Duke University, Durham, North Carolina 27708, USA
li.chen@duke.edu • gurhan.kok@duke.edu • jordan.tong@duke.edu

September 19, 2010

In the newsvendor problem, a decision maker chooses an inventory order quantity prior to the realization of a random demand. The decision maker faces a trade-off between ordering too many and having leftover inventory versus ordering too few and missing out on potential sales. Keeping the net profit structure constant, we study how the timing of the payments affects the inventory decisions. Specifically, we examine three payment schemes which can be interpreted as the inventory order being financed 1) by the newsvendor herself, 2) by the supplier, and 3) by the customer. Models that are neutral with respect to the payment timings, such as risk aversion, risk seeking, and other static utility preferences, would predict the same ordering behavior across the three payment schemes. On the contrary, we find in laboratory experiments that ordering decisions are significantly different across the three payment schemes. Specifically, the order quantity under newsvendor own financing is greater than that under supplier financing, which is, in turn, greater than the order quantity under customer financing. These findings are also in disagreement with what a regular or hyperbolic time-discounted utility model would predict. Instead, we propose a model of underweighting order-time payments based on the principles of mental accounting and show that the experimental results are consistent with this new model. We further validate the robustness of our model under different profit-margin conditions. Our findings contribute to the understanding of the psychological processes involved in newsvendor decisions and have implications for supply chain financing practices such as trade credit and supply chain contract design.

Key words: Behavioral operations, newsvendor, decision under uncertainty, mental accounting, coupling, payment depreciation, trade credit, supply chain contracts.
1 Introduction

In the newsvendor problem, a decision maker chooses an inventory order quantity to meet a random future demand. The objective is to choose an order quantity that optimally balances the expected cost of overordering (i.e., having leftover inventory) with the expected cost of underordering (i.e., missing out on potential sales). The newsvendor framework is commonly used in practice for managing products with a short selling season and limited replenishment opportunities, such as fashion apparels and hi-tech products. Although the optimal solution to the newsvendor problem can be fairly easily determined, human subjects are often observed deviating significantly from the optimal solution in experiments (e.g., Schweitzer and Cachon 2000). In this paper, we study how a seemingly innocuous change in payment timing can lead to significantly different ordering decisions in a repeated newsvendor experiment.

In the newsvendor problem, payments can occur at different points in time. Consider the three different payment timing schemes below. First, consider the most natural payment scheme for the newsvendor problem. The newsvendor pays for the purchase order at a cost of \( c \) per unit. After demand is realized, she receives revenue for units sold at a price of \( p \) per unit. We assume that any leftover inventory has a salvage value of zero. In this payment scheme, the newsvendor finances her own inventory. We refer to it as own-financing or “payment scheme O.” Second, consider the case when the supplier finances the newsvendor’s inventory order. Under supplier financing, which we refer to as “payment scheme S,” the supplier permits a delayed payment for inventory. Thus, the newsvendor does not need to make a payment to the supplier upon ordering. After demand is realized, the newsvendor receives a net profit of \( p - c \) per unit sold and pays the supplier for any leftover inventory at a unit cost of \( c \). Third, consider the case when the customer finances the newsvendor’s inventory order. Under customer financing, which we refer to as “payment scheme C,” the newsvendor receives advanced payment from the customer for each unit that she orders and pays the cost to the supplier. Thus, she receives a net profit of \( p - c \) per unit ordered at the time of ordering. However, after demand is realized, she refunds the advanced payment \( p \) per unit for the leftover inventory that is not demanded by the customer. The net payments and the timing of these payments under payment schemes O, S, and C are summarized in Table 1.
Payments at time of order
Payments after demand realization
per unit ordered per unit sold per unit leftover

<table>
<thead>
<tr>
<th>Payment Scheme</th>
<th>Payments at time of order</th>
<th>Payments after demand realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own Financing (O)</td>
<td>$-c$</td>
<td>$+p$</td>
</tr>
<tr>
<td>Supplier Financing (S)</td>
<td>$0$</td>
<td>$+(p-c)$</td>
</tr>
<tr>
<td>Customer Financing (C)</td>
<td>$+(p-c)$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

Table 1: Net payments and transaction timing under different payment schemes.

The net profit structure is the same for all three payment schemes because there is no time value of money in the model set-up (i.e., no interest can be earned on current wealth). The unit cost of overordering (also called the overage cost) is $c$, as the newsvendor loses an amount equal to purchase cost minus the salvage value per unit of leftover inventory. The unit cost of underordering (also called the underage cost) is $p - c$, as the newsvendor loses the opportunity to earn the profit margin if demand exceeds the order quantity. Because the overage and underage costs remain unchanged across payment schemes, intuition predicts that ordering behavior should be the same under each scheme. In fact, any model based on a static evaluation of utility of the actual net profit, that does not explicitly consider the timing of the payments, predicts the same ordering behavior under each payment scheme. For example, static utility models of risk aversion, risk seeking, loss aversion, and other utility preferences described in Schweitzer and Cachon (2000) all predict the same ordering behavior under each payment scheme. On the contrary, we find that ordering decisions are significantly different across the three payment schemes in our experimental studies.

We present two experimental studies in this paper. In the first study, we test the ordering behavior when overage cost is equal to underage cost. In this setting, theory predicts that the profit-maximizing solution is to order the median demand under all three payment schemes. Previous newsvendor behavioral studies suggest that individuals are in general biased toward ordering the mean demand\(^1\) — the pull-to-center effect (Schweitzer and Cachon 2000; Bostian et al. 2008; Bolton and Katok 2008; and Ho et al. 2010). Because our focus is on the payment timing effect, we want to control for the pull-to-center effect. Thus, we set the optimal solution at the center (median).

\(^1\)These studies consider only symmetric demand distributions, which implies that the mean and the median are the same.
of the distribution, allowing us to isolate the effect of payment timing on inventory decisions. Under this experimental setting, we find that ordering behavior is significantly different under the schemes O, S, and C. Individuals order significantly more under scheme O than under scheme S, and significantly more under scheme S than under scheme C; they order approximately the optimal quantity (median demand) under scheme S.

The differences in ordering behavior under each payment scheme suggest that the timing of the payments plays an important role. Consider the commonly used time-discounted utility formulation of the newsvendor problem, in which future payments are discounted relative to payments that occur closer to the present. This utility formulation can indeed predict differences in ordering decisions between payment schemes; it actually predicts that orders will be highest under C and lowest under O because the deferred outgoing payments are largest under the former and smallest under the latter. However, this is exactly the opposite of the ordering behavior we observe in our experiments. Similarly, our results cannot be explained by a hyperbolic time-discounted model, in which a relatively high discount rate is assigned over short time horizons and a relatively low discount rate over long time horizons (Laibson 1997). In our repeated newsvendor experiment, the decision horizon is effectively a single period because the current-period inventory decision has no effect on the rewards in future periods. Thus, in our setting, there is no distinction between a hyperbolic time-discounted model and a regular time-discounted model.

To explain our results, we propose a model based on the principles of mental accounting (see Thaler 1999 for a review). Briefly, mental accounting provides a framework for explaining how individuals track the costs and benefits of a transaction. Thaler (1980, 1985) proposes that a consumer opens a mental account upon entering a transaction and then closes it when the transaction is completed, thereby spanning the length of time between the costs and benefits of the transaction. Importantly, he suggests that when payment precedes consumption, the individual does not feel loss at the moment of payment, but rather holds it in a mental account until the transaction is completed. The “coupling” (Prelec and Loewenstein 1998) or mental linking between the costs and benefits of a transaction (e.g., the extent to which the individual mentally “holds” the upstream loss) is subject to many factors, but generally is stronger the more transparent the relationship between payment and consumption. Below, we describe how newsvendor payment schemes affect
individuals’ ordering decisions through this coupling process.

In the newsvendor framework, the three payment schemes considered in this paper lead to different mental accounting schemes. Under payment scheme O, the order payment precedes and is temporally separated from the revenue received at the time of demand realization (see Table 1). Because of this separation, individuals tend to treat the order as an investment that will lead to future benefits. Thus, the cost of the order is perceived to be less than $c$ because the experience of the actual cost is “buffered” by the thought of future benefits (Prelec and Loewenstein 1998). We call this effect “underweighting order-time payments,” or simply “payment underweighting” when the meaning is clear in the context. It explains why individuals tend to order more aggressively under scheme O. Conversely, under payment scheme C, the profit received at the order time is temporally separated from the potential refund payment at the time of demand realization (see Table 1). Thus, individuals tend to treat the profit received at the order time as borrowed benefits, which may require a repayment in the future. As a result, the profit is perceived to be less than $p - c$ because it is “attenuated” by the thought of future repayments (Prelec and Loewenstein 1998). This explains why individuals order more conservatively under scheme C. Here, the underweighted profit is the payment received at the order time, so it is the same order-time payment underweighting effect described above. In this case, the underweighting resembles the “debt aversion” effect in consumer behavior described in Prelec and Loewenstein (1998). Finally, under payment scheme S, all payments occur at the same time of demand realization, so there is no temporal separation (see Table 1). Thus, profit $p - c$ and cost $c$ are weighted equally, which yields the result that individuals order approximately the optimal quantity (median demand) under scheme S.

Furthermore, in our repeated newsvendor experiment, individuals are likely to learn from past outcomes to inform future decisions. When evaluating the outcome of a previous round, the individual also underweights the payments that occurred at the order time compared to the payments that occurred after the demand realization. This is because the order time payments occurred earlier before the demand realization, while the payments after the demand realization occurred more recently. Such an evaluation process is similar to what Gourville and Soman (1998) call “payment depreciation,” or the gradual reduction of the relevance of sunk costs. It further reinforces the payment underweighting effect and makes the order deviation robust over more rounds of play.
This is indeed what we observe in our experiment (see Section 4).

In the second study, we test the robustness of the payment timing effect for products with high and low profit margins, hereafter denoted as high-profit and low-profit products. For both types of products, we find that the effect of payment timing is still significant and the order quantities under the schemes O, S, and C exhibit the same relative order found in the first study—highest under scheme O and lowest under scheme C. However, the relative differences between order sizes under each payment scheme are not the same under high- and low-profit conditions. Specifically, we find that, for a high-profit product, orders under schemes O and S are relatively similar, but both are significantly higher than orders under scheme C. On the other hand, for a low-profit product, orders under scheme O are significantly higher than orders under schemes S and C, and orders under schemes S and C are relatively similar. This distortion can also be explained by the underweighting of order-time payments effect observed in the first study. Under the high-profit condition, profit $p - c$ is greater than cost $c$. Thus, from Table 1, the magnitude of the order-time payment is greater under scheme C than under scheme O. Hence, payment underweighting has a greater impact under scheme C than under scheme O. Symmetrically, under the low-profit condition, cost $c$ is greater than profit $p - c$, so payment underweighting has a greater impact under scheme O than under scheme C and, thus, the observation is reversed. Furthermore, as expected, we also observe the pull-to-center effect under both high- and low-profit conditions, i.e., order quantities are shifted toward the mean demand in both settings.

Our experimental findings have the following implications. First, our results contribute to the understanding of the psychological processes involved in newsvendor ordering decisions. We find that payment timing affects inventory decisions through the perceived costs and benefits resulting from individuals’ mental accounting. Second, the payment schemes studied in this paper have natural interpretations in the context of trade credit and supply chain financing. For reasons such as capital constraints or credit availability, suppliers sometimes choose to finance their downstream retailer’s inventory in order to induce higher orders (e.g., Smith 1987, and Xu and Zhang 2010). Our results suggest that there is a behavioral effect of payment timing that works against this intended objective: when capital constraint is not tight and the time value of money is negligible, individuals order less when inventory is financed by the supplier than when it is own-financed. Thus,
managers should carefully take this behavioral effect into account when facing similar circumstances. Finally, our results also have some implications for supply chain contract design. For example, if a supplier offers a wholesale-price contract, it is well-known that a rational retailer will underorder due to double marginalization (Lariviere and Porteus 2001). The payment timings under the wholesale-price contract resemble our payment scheme O. Thus, due to the order-time payment underweighting effect, the retailer is likely to order more than the rational order quantity. The resulting inventory decision, though suboptimal for the retailer, may inadvertently achieve better supply chain performance than theory would predict. Similar conclusions can also be drawn for other common supply chain contracts, such as buy-back (Pasternack 1985) and revenue-sharing contracts (Cachon and Lariviere 2005).

The rest of the paper is organized as follows. We provide a literature review in Section 2. We present three decision models and develop our hypotheses in Section 3. We present our experimental findings and discuss the results in Section 4. Section 5 contains our concluding remarks.

2 Literature Review

There is a growing literature on behavioral operations, in which researchers study how human subjects make inventory decisions in experimental settings. For example, Schweitzer and Cachon (2000) find the pull-to-center effect in newsvendor experiments. Bostian et al. (2008) further examine various decision heuristics that may lead to this effect, and find support for an adaptive learning model, i.e., the weighted attraction model proposed by Camerer and Ho (1999). Bolton and Katok (2008) find that ordering decisions do not improve significantly over time — they find only a slight improvement over 100 rounds of play. Lurie and Swaminathan (2009) find that aggregating feedback across multiple periods and limiting the frequency of choices can partially improve newsvendor performance. Feiler et al. (2010) further show that making demand distribution unknown and lost sales unobserved introduces bias in subjects’ ordering decisions. In a serial supply chain setting, researchers also find that human subjects do not sufficiently account for the leadtime delay and subsequently overreact to their inventory levels (Sterman 1989; and Croson and Donohue 2005, 2006). To account for some of these experimental findings, Su (2008) proposes a descriptive model for decision making under bounded rationality. Our paper differs from the previous research in
that we find that different payment timings can significantly affect individuals’ inventory decisions in the newsvendor experiment. Furthermore, we propose a descriptive model to explain the results based on the principles of mental accounting.

Mental accounting theory has long been used to help understand the psychology behind choice behavior (Kahneman and Tversky 1979; Tversky and Kahneman 1981; Thaler 1980, 1985). It provides an explanation for many phenomena in human behavior that seem irrational—most notably in consumer behavior (e.g., Thaler 1985; Heath and Soll 1996), but also in other functional areas, such as finance (Shefrin and Statman 1985) and accounting (Burgstahler and Dichev 1997). Our experimental findings provide another example of mental accounting in operations management. Inventory managers are equally as likely to succumb to the mental accounting process as are consumers. Thus, there are many analogies we can draw from the consumer behavior mental accounting literature (e.g., Prelec and Loewenstein 1998, Gourville and Soman 1998). For instance, Shafir and Thaler (2006) find that the typical wine connoisseur thinks of her initial purchase of the wine as an investment (so the cost is felt less) and later thinks of the wine as free when she drinks it. This phenomenon is very much like the ordering behavior we observe under the own financing payment scheme.

Our paper is also related to the literature on trade credit and supply chain contract design. There are many theories regarding what drives trade credit terms and why trade credit exists (Peterson and Rajan 1997; and Ng et al. 1999). Trade credit terms can certainly affect a firm’s ordering policy. Gupta and Wang (2009) characterize the optimal inventory policy for a setting with stochastic demand and trade credit. Other financial considerations can also affect inventory decisions, such as asset-based financing (Buzacott and Zhang 2004) and capital constraints (Xu and Birge 2004; Babich and Sobel 2004; and Xu and Zhang 2010). Our experimental results suggest that there is also a potential behavioral factor that influences inventory decisions through the altered financial transaction timings. Furthermore, payment timing is an important aspect of supply chain contracts between firms. Loch and Wu (2008) study how social preferences, such as fairness, status, and reciprocity, affect contracting behavior in a supply chain. Katok and Wu (2009) test buy-back and revenue-sharing contracts in an experimental setting and compare their effectiveness at coordinating supply chains with predications from theory. Our results suggest that
the payment timing effect could also play a role in determining the effectiveness of these contracts.

3 Models of Newsvendor Decision Making

In the newsvendor problem, a decision-maker chooses an order quantity \( q \) of a product to meet a future random demand \( D \). Let \( F(\cdot) \) denote the cumulative distribution function for the random demand. We assume that backlogs are not allowed (i.e., unmet customer demand is lost) and leftover inventory cannot be carried over to the subsequent period and has zero salvage value. The unit cost of the product is \( c \) and it is sold at price \( p \) (with \( p > c \)).

In this paper, we consider three payment schemes that are summarized in Table 1. (1) Own financing (payment scheme O): the newsvendor pays the cost \( c \) per unit at the time of ordering and receives a revenue \( p \) per unit sold when demand is realized. (2) Supplier financing (payment scheme S): the supplier permits the newsvendor to delay her payment for units ordered until the time of demand realization. Thus, under this scheme, the newsvendor pays nothing at the time of ordering; at the time of demand realization, she receives \( p - c \) per unit sold and pays the supplier \( c \) per unit leftover. (3) Customer financing (payment scheme C): the customer advances payment \( p \) per unit that the newsvendor orders, but requires that the newsvendor refund the payment for the units that are not demanded. Thus, under this scheme, the newsvendor receives \( p - c \) per unit ordered at the time of ordering, but must refund \( p \) per unit leftover to the customer at the time of demand realization. No interest rate discounting is considered in the above three payment schemes. Next, we describe three models that differ in how the decision maker takes payment timings into account. In order to distinguish between different uses of the term “utility,” we use the term “reward function” to denote the “perceived” monetary gains (possibly dependent on payment timing) given a demand outcome scenario and use the term “utility function” to denote the mapping from the monetary gains to the decision-maker’s final utility based on her utility preferences.

3.1 Static Rewards: Payment-Timing Neutral Preferences

Let \( R^i(q, D) \) denote the reward function given the quantity \( q \) and demand realization \( D \) under the payment scheme \( i \in \{O, S, C\} \). Because there is no interest rate discounting in the problem, if the
decision-maker has a neutral preference for payment timing, \( R^i(q, D) \) can be expressed as follows.

\[
R^i(q, D) = \begin{cases} 
-cq + p \min(q, D) & \text{if } i = O, \\
(p - c) \min(q, D) - c \max(q - D, 0) & \text{if } i = S, \\
(p - c)q - p \max(q - D, 0) & \text{if } i = C.
\end{cases}
\]

This formulation essentially is based on a static model as it does not take the time dimension into account. Let \( U(\cdot) \) be a utility function reflecting utility preferences (e.g., risk aversion, risk seeking, and loss aversion). The optimal order quantity that maximizes the expected utility is

\[
q^* = \arg \max_q E_D[U(R^i(q, D))].
\]

It is easy to verify that \( R^O(q, D) = R^S(q, D) = R^C(q, D) \) for any \( q \) and \( D \). Thus, any utility preference model based on these reward functions will yield the same decision under the three payment schemes. In particular, it is well known that the expected-profit-maximizing quantity (so called “newsvendor quantity”) is given by

\[
q^O = q^S = q^C = F^{-1}\left(\frac{p - c}{p}\right).
\]

The term \( (p - c)/p \) is known as the critical fractile. This solution is for a decision maker with risk-neutral preferences. The following result formally states this equivalence for any utility function.

**Proposition 1** If the decision maker is payment-timing neutral, then the utility-maximizing quantities under the three payment schemes are identical, i.e., \( q^O = q^S = q^C \).

**Proof:** For any given utility function \( U(\cdot) \) based on the reward function, we have \( U(R^O(q, D)) = U(R^S(q, D)) = U(R^C(q, D)) \). Thus, the result follows. \( \square \)

### 3.2 Time-Discounted Rewards: Preference for Payment Deferment

The decision-maker may prefer to receive benefits earlier and delay costs until later, i.e., future payments are discounted (Samuelson 1937). This is also known as the time-discounted utility model. Under this model, the reward derived from the payment at the time of demand realization is discounted, even though there is no explicit monetary interest rate. Let \( \delta \) (\( 0 < \delta < 1 \)) denote this reward discount factor. The time-discounted reward function \( R^i(q, D) \) can then be expressed
as follows.

\[
R^i(q, D) = \begin{cases} 
-cq + \delta p \min(q, D) & \text{if } i = O, \\
\delta (p - c) \min(q, D) - \delta c \max(q - D, 0) & \text{if } i = S, \\
(p - c)q - \delta p \max(q - D, 0) & \text{if } i = C.
\end{cases}
\]

Evidently, this preference model yields different utility-maximizing quantities under the three payment schemes. In particular, we have the following result:

**Proposition 2** Let \( U(\cdot) \) be a continuous and increasing utility function based on the rewards. With time-discounted rewards, the expected-utility-maximizing quantities under the three payment schemes have the following order: \( q^O < q^S < q^C \).

**Proof:** The critical fractiles for payment schemes O, S, and C are \((\delta p - c)/\delta p, (p - c)/p, \) and \((p - c)/\delta p, \) respectively. Because \((\delta p - c)/\delta p < (p - c)/p < (p - c)/\delta p, \) we have \( q^O < q^S < q^C \) for the expected-reward maximizing quantities (i.e., for a linear utility function of the reward).

Next, we generalize to any monotone increasing utility function. Define \( 1(q < D) \) as the binary indicator function. By rewriting \( R^S(q, D) = \delta[-cq + p \min(q, D)], \) we see that \( \partial R^S(q, D)/\partial q = -\delta c + \delta p 1(q < D) > -c - \delta p 1(q < D) = \partial R^O(q, D)/\partial q \) for any \( D. \) We can also rewrite \( R^S(q, D) = \delta[(p - c)q - p \max(q - D, 0)], \) so that \( \partial R^S(q, D)/\partial q = \delta (p - c) - \delta p 1(q \geq D) < (p - c) - \delta p 1(q \geq D) = \partial R^C(q)/\partial q. \) Thus, the derivatives follow the order of \( \partial R^O(q, D)/\partial q < \partial R^S(q, D)/\partial q < \partial R^C(q)/\partial q \) for any value of \( D. \) To compute the optimal solution of \( q^i = \arg \max_q \mathbb{E}_D[U(R^i(q, D))], \) we need to derive first-order conditions. Because the reward functions have bounded derivatives, they satisfy the Lipschitz condition of order one. This means that we can exchange expectation and derivatives to get the first order conditions (see Glasserman 1994). As a result, one can show that the first derivatives of the expected utility functions under the three schemes follow the increasing order O, S, and C. Thus, we have \( q^O < q^S < q^C \) for the expected-utility-maximizing quantities. \( \square \)

### 3.3 A Model of Mental Accounting: Underweighting Order-Time Payments

According to the mental accounting process described in the introduction, the decision-maker is likely to underweight payments made at the time of order (Prelec and Loewenstein 1998). Here, we assign an underweighting factor \( \beta (0 < \beta < 1) \) to the order-time payment. The reward function
$R^i(q, D)$ under the payment underweighting model is given below.

$$R^i(q, D) = \begin{cases} 
-\beta c q + p \min(q, D) & \text{if } i = O, \\
(p - c) \min(q, D) - c \max(q - D, 0) & \text{if } i = S, \\
\beta(p - c)q - p \max(q - D, 0) & \text{if } i = C.
\end{cases}$$

Again, this model yields different utility-maximizing quantities under the three payment schemes. In particular, we have the following result:

**Proposition 3** Let $U(\cdot)$ be a continuous and increasing utility function based on the rewards. With order-time payment underweighting, the expected-utility-maximizing quantities under the three payment schemes have the following order: $q^O > q^S > q^C$.

**Proof:** The proof is similar to that of Proposition 2. □

Thus, the order of optimal decisions under the payment underweighting model is in disagreement with the order of optimal decisions under models with static or time-discounted rewards.

4 Newsvendor Experiments

In this section, we present two repeated newsvendor experiments to examine the behavioral effect of payment timing on inventory decisions. To isolate the behavioral effect of timing, we eliminate factors such as capital constraints and interest rates in our experimental designs. In the first study, we test whether ordering behavior can be described by the models presented in the previous section. We find that ordering behavior is consistent with the model of underweighting order-time payments, and is thus inconsistent with the static and time-discounted rewards models. In the second study, we test the robustness of the model predictions under high- and low-profit conditions.

4.1 Study 1: A Simple Payment Timing Experiment

4.1.1 Experimental Design and Hypotheses

In Study 1, we test the three payment schemes O, S, and C under very simple parameters $c = 1$, $p = 2$ in a repeated newsvendor setting. The resulting payment schemes under each condition are determined according to Table 1.
In each round, subjects roll three fair six-sided dice, the sum of which determines the demand for that round. Thus, demand is independent, identically distributed, and symmetric with a minimum value of 3, maximum value of 18, and mean value of 10.5. Recall that all payment schemes are equivalent in the sense that they produce identical total net profits or losses for any given ordering decision and demand realization. Furthermore, the actual overage cost and underage costs are equal at $1 each and the risk neutral critical fractile is given by \((p - c) / p = 50\%\) for all three payment schemes. The expected-profit-maximizing solution under all payment schemes is to order either 10 or 11 units every period. The newsvendor pull-to-center effect suggests that participants are biased towards the mean of the demand distribution, or 10.5.

Recall that according to Proposition 1, a static rewards model predicts orders will be the same under all three payment schemes. Nevertheless, the payment schemes O, S and C differ in the timing of payments, which we believe creates significant differences in ordering behavior. Therefore, we predict that ordering behavior will be inconsistent with a static rewards model. This is our first hypothesis.

**Hypothesis 1:** Payment scheme has an effect on order quantities.

Specifically, in this repeated newsvendor experiment, we propose that individuals tend to underweight the order-time payments, which yields ordering behavior consistent with the model prediction of Proposition 3 and thereby inconsistent with the time-discounted rewards model prediction of Proposition 2. This is our second hypothesis.

**Hypothesis 2:** Average order size in O is greater than the average order size in S, which , in turn, is greater than the average order size in C.

In the subsequent sections, we describe our experimental methods, present our results, and provide a discussion of the results for Study 1.

### 4.1.2 Methods

We recruited 99 undergraduate and graduate students from Duke University using the Fuqua Behavioral Lab recruiting system. The experimental conditions were assigned sequentially to the participants.\(^2\) In exchange for their participation, participants received a minimum of $5 plus a

\(^2\)We conducted this study in two parts. In the first part, we ran conditions O and S (57 subjects). In the second part, we ran conditions S and C (42 subjects). We found no significant differences between the two repetitions of
$1 bonus for every 50 play dollars they had at the end of the game (each participant began with 100 play dollars).

Participants were given an instruction sheet explaining the details of the game for the payment scheme to which they were assigned. Instructions were also read out loud by a research assistant before beginning play. Participants were told that they would be selling “widgets” (represented by poker chips) and that customer demand for the widgets in a given time period was represented by the sum of the rolling of three standard dice. Each participant interacted one-on-one with a research assistant, who facilitated payment transfers and recorded ordering decisions and dice rolls. A participant decided an order quantity vocally, placed that many poker chips into the “store” (represented by a square drawn on an index card), and made appropriate payment transfers. Then, the participant rolled the three dice, determined how many units were sold and/or leftover, and again made appropriate payment transfers. Finally, the participant removed all chips from the store to begin the next round.

Payment transfers were conducted in the form of play paper currency in denominations of 1, 5, and 10. All payments to the participant were conducted by the research assistant, while all payments from the participant were conducted by the participant. Appropriate payment transactions occurred immediately following the ordering decision and immediately following demand realization. The participant also moved the poker chips and rolled the dice themselves, which facilitated their understanding of the process. Game play was for 25 rounds, after which a follow-up question was administered: “If you could play the game again choosing only one order quantity, what number would you choose?” Also two written comprehension questions were administered at this time: “What is the minimum demand possible you can roll with three dice?” and “What is the maximum demand possible you can roll with three dice?”

4.1.3 Results

All 99 participants completed the study. One participant in the S condition incorrectly answered both comprehension questions and also made multiple orders of more than 18, and was therefore removed from the analysis (though all results hold when included). The resulting average ordering

condition S, and therefore aggregated the data for analysis, yielding 29 subjects for condition O, 49 for condition S, and 21 for condition C.
decisions in each round are shown in Figure 1, and a summary of our main results can be found in Table 2.

We conducted a repeated measures generalized linear model to analyze the 25 inventory order decisions under each payment scheme. We found that payment scheme significantly affected ordering behavior \( F(2, 95) = 18.88, p < .0001 \). Specifically, we found that orders were highest under payment scheme O and lowest under payment scheme C. In order to test these differences, we conducted planned contrast tests. These tests showed that all three differences were significant: orders under O were significantly greater than orders under S \( F(1, 95) = 18.10, p < .0001 \), orders under S were significantly greater than orders under C \( F(1, 95) = 7.46, p = .0075 \), and orders under O were significantly greater than orders under C \( F(1, 95) = 35.83, p < .0001 \). There was no significant difference in the overall ordering levels over time (Wilks’ Lambda = .701, \( F(24, 72) = 1.28, p = .212 \)). In other words, there was no main effect for round. We also found no significant interaction between payment scheme and experience gained as more rounds were played (Wilks’ Lambda = .614, \( F(48, 144) = .83, p = .774 \)).

<table>
<thead>
<tr>
<th>Payment Scheme</th>
<th>Mean order quantity (standard deviation in parentheses)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own Financing (O)</td>
<td>11.728 (1.392)</td>
<td>29</td>
</tr>
<tr>
<td>Supplier Financing (S)</td>
<td>10.573 (1.031)</td>
<td>48</td>
</tr>
<tr>
<td>Customer Financing (C)</td>
<td>9.749 (1.058)</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Average over 25 rounds</th>
<th>Round 1</th>
<th>Follow-up question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own Financing (O)</td>
<td>11.069 (3.390)</td>
<td>11.759 (1.766)</td>
<td></td>
</tr>
<tr>
<td>Supplier Financing (S)</td>
<td>10.271 (2.210)</td>
<td>10.448 (1.234)</td>
<td></td>
</tr>
<tr>
<td>Customer Financing (C)</td>
<td>9.238 (2.406)</td>
<td>9.571 (1.207)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrast tests</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( q^O - q^S )</td>
<td>1.155***</td>
<td></td>
</tr>
<tr>
<td>( q^O - q^C )</td>
<td>.824**</td>
<td>1.033</td>
</tr>
<tr>
<td>( q^S - q^C )</td>
<td>1.979***</td>
<td>1.831*</td>
</tr>
</tbody>
</table>

Table 2: Mean and standard deviations of ordering quantities, and significance tests for differences between payment schemes in Study 1. * \( p < .05 \), ** \( p < .01 \), *** \( p < .001 \).

As expected, average orders across the 25 rounds for each participant were highest under payment scheme O, and lowest under payment scheme C \( q^O = 11.728, q^S = 10.573, q^C = 9.749 \). We compared these average orders with the mean of the demand distribution, 10.5, because both the expected profit-maximizing criterion and the pull-to-center effect predicted orders near mean demand. We found that average orders under O were significantly greater than mean demand
Figure 1: Average order quantities of subjects in each round of play in Study 1 under own financing (O), supplier financing (S), and customer financing (C).

\( t(28) = 4.751, p < .001 \), average orders under C were significantly less than mean demand \( t(20) = -3.256, p = .004 \), while average orders under S were not significantly different from mean demand \( t(47) = .493, p = .624 \).

Next, we investigated subjects’ first ordering decision, which was a decision not confounded by experience or feedback. Again, we found that average orders in the first round were highest under payment scheme O, and lowest under payment scheme C \( (q^O = 11.069, q^S = 10.271, q^C = 9.238) \). However, follow-up planned contrasts showed that only one difference was significant. The difference between orders in O and S was not significant \( (F(1, 95) = 1.64, p = .2035) \), and the difference between orders under S and C was not significant \( (F(1, 95) = 2.22, p = .1397) \). Only the difference between orders under O and C was significant \( (F(1, 95) = 5.81, p = .0178) \).

Additionally, we analyzed subjects’ answers to the follow-up question, “If you could play the game again choosing only one order quantity, what number would you choose?” We found that payment scheme significantly affected subjects’ responses to this question such that their responses were highest under payment scheme O, and lowest under payment scheme C \( (q^O = 11.759, q^S = \)
This time, follow-up planned contrasts showed that all three differences were significant: the difference between responses under O and S was significant \( F(1, 95) = 15.69, p < .0001 \), the difference between responses in S and C was significant \( F(1, 95) = 5.67, p = .0193 \), and the difference between responses in O and C was significant \( F(1, 95) = 29.43, p < .0001 \).

The actual demands generated by rolling the three dice were relatively consistent with the theoretical predictions. The means were 10.739, 10.557, and 10.764, under O, S, and C, respectively. Also, all participants (except the one eliminated participant in condition S) correctly answered 3 and 18 for the minimum and maximum possible demand that could be generated by rolling 3 dice.

### 4.1.4 Discussion

Study 1 establishes that payment timing has a significant effect on ordering behavior in the newsvendor problem. In support of Hypothesis 1, we found that orders were significantly different under payment schemes O, S, and C. A static rewards model cannot explain the observed differences in behavior, as it predicts the same order quantities across payment timing schemes. For example, none of the static-reward-maximizing decision models proposed by Schweitzer and Cachon (2000) (i.e., risk-neutral, risk-averse, risk-seeking, loss-averse, prospect-theory, stockout-averse, waste-averse preferences, or preferences for minimizing ex-post inventory error) can explain differences in orders under payment schemes O, S, and C, because they do not take into account the effects of payment timing. In support of Hypothesis 2, we found that orders were generally largest under payment scheme O and smallest under payment scheme C. This result is consistent with the order-time payment-underweighting model. However, it is inconsistent with the discounted-rewards model, which would predict orders to be largest under C and smallest under O.

In order to obtain a rough estimate for the underweighting factor \( \beta \), we calculated the upper and lower bounds on the value of \( \beta \) that lead to each ordering decision for each subject-round under payment schemes O and C. The average of these lower bounds is \( \beta = 0.6785 \), while the average of these upper bounds is \( \beta = 0.8935 \). Taking the midpoint of these two bounds yields an estimate of \( \beta = 0.786 \). In other words, in this experiment we found that on average, individuals order in such a way that they only take into account 78.6% of payments that occur at the order time. For example, this suggests that an individual, who orders 10 units at $1 each, perceives the $10
Study 1 also provides us with some insights into the effect of feedback and the robustness of the payment timing effect. From Figure 1, it appears that the effect of payment timing on average orders remains quite consistent over the 25 rounds. Also, the answers to the follow-up question suggest that the effect of payment timing on ordering behavior extends beyond the rounds of play. The standard deviation of order quantities under each scheme becomes much smaller in the responses to the final question than the standard deviation in the first ordering decision (see Table 2, which results in a much sharper distinction in order decisions among the three schemes. This suggests that the payment timing effect becomes more robust over the 25 rounds, which is consistent with the learning from previous outcome feedback described in the introduction. Specifically, in evaluating the outcome of the previous round, the payment at the time of ordering is subject to the payment depreciation effect (Gourville and Soman 1998), and thus is weighted less than the more recent payment after the demand realization. This evaluation process reinforces the payment underweighting effect and makes the order deviation robust over time.

In the next study, we test the robustness of our results by conducting a similar study for products that have different profit margins.

### 4.2 Study 2: Payment Timing Experiments with High and Low-Profit Products

#### 4.2.1 Experimental Design and Hypotheses

In Study 2, we implement two repeated newsvendor experiments to test the effect of payment timing for products with two different profit margins. The high-profit condition is conducted for a product with parameters $c = 1$, $p = 4$, which implies an actual overage cost of 3 and an actual underage cost of 1. Under the expected-profit-maximizing model, this yields a critical fractile of 75%. The low-profit condition is conducted for a product with parameters $c = 3$, $p = 4$, which implies an actual overage cost of 1 and actual underage cost of 3. Under the expected-profit-maximizing model, this yields a critical fractile of 25%. Within each high- and low-profit condition, we again test payment schemes O, S and C. One can substitute the appropriate values of $c$ and $p$ into Table 1 to obtain a description of the payment schemes for the high and low-profit conditions, shown in Table 3. As in Study 1, demand is determined by the sum of three standard dice rolled by the
subject in each round.

<table>
<thead>
<tr>
<th>Payment Scheme</th>
<th>Payments at time of order</th>
<th>Payments after demand realization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>per unit ordered</td>
<td>per unit sold</td>
</tr>
<tr>
<td>High-profit product</td>
<td>Own Financing (O)</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>Supplier Financing (S)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Customer Financing (C)</td>
<td>+3</td>
</tr>
<tr>
<td>Low-profit product</td>
<td>Own Financing (O)</td>
<td>-3</td>
</tr>
<tr>
<td></td>
<td>Supplier Financing (S)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Customer Financing (C)</td>
<td>+1</td>
</tr>
</tbody>
</table>

Table 3: Net payments and transaction timing under different payment schemes for the high-profit product (c=1, p=4) and the low-profit product (c=3, p=4) in Study 2.

For each payment scheme O, S, and C, the expected profit-maximizing solution is 13 for the high-profit condition and 8 for the low-profit condition. The pull-to-center effect predicts that individuals are biased towards the center of the distribution, 10.5, causing actual orders to be somewhere between 13 and 10.5 for the high-profit condition, and somewhere between 8 and 10.5 for the low-profit condition. Nevertheless, the pull-to-center effect still predicts no difference between the payment schemes. In fact, as in Study 1, any static rewards model would predict no difference between order quantities due to the payment scheme. Therefore, in Study 2, we test the following hypothesis, which is analogous to Hypothesis 1.

**Hypothesis 3:** Payment scheme has an effect on order quantities for high- and low-profit products.

Specifically, we expect ordering decisions to be consistent with the model of underweighting order-time payments. However, given the complex nature of the interaction between profit margin, critical fractile, and payment underweighting, the effect of payment scheme may not be the same under high- and low-profit products. For this reason, we formulate three separate sub-hypotheses in each of Hypotheses 4 and 5 (which parallel Hypothesis 2 in Study 1).

**Hypothesis 4:** For the high-profit product,

(OS) orders under O are greater than orders under S,

(SC) orders under S are greater than orders under C,

(OC) orders under O are greater than orders under C.
Hypothesis 5: For the low-profit product,

(OS) orders under O are greater than orders under S,
(SC) orders under S are greater than orders under C,
(OC) orders under O are greater than orders under C.

In the subsequent sections, we describe our experimental methods, present our results, and provide a discussion of the results for Study 2.

4.2.2 Methods

We recruited 130 undergraduate and graduate students from Duke University using the Fuqua Behavioral Lab recruiting system—70 for the high-profit condition and 60 for the low-profit condition. The payment scheme conditions were assigned sequentially to the participants within each experiment. In exchange for their participation, participants received a minimum of $5, with a bonus based on how much play cash they earned in the game. In the high-profit condition, participants earned a $1 bonus for every 100 play dollars they had at the end of the game (each participant began with 100 play dollars). In the low-profit condition, participants earned a $1 bonus for every $50 play dollars they had at the end of the game (each participant began with 150 play dollars). Participants appeared to be motivated to perform well on the task. For Study 2, each participant played the game for 20 rounds. In all respects except for the payment scheme parameter changes and the reduced number of rounds, the experimental design and methods were the same as in Study 1. The same written final and follow-up questions were administered.

4.2.3 Results

All 130 participants completed the study. The resulting average ordering decisions for each round are shown in Figures 2 and 3, and a summary of our findings can be found in Table 4. As in Study 1, we first analyzed the 20 inventory decisions for both the high- and low-profit conditions using a repeated measures generalized linear model. We then examined the effect of payment scheme over the 20 rounds, as well as for the first round, and for the responses to the follow-up questions.
High-Profit Condition

For the high-profit condition, the repeated measures generalized linear model showed that payment scheme significantly affected ordering behavior ($F(2, 67) = 18.61, p < .0001$). We found that average orders were highest under payment scheme O, and lowest under payment scheme C. By observing the average order sizes over the 20 rounds ($q^O = 11.821, q^S = 11.233, q^C = 9.900$), we found that the difference between S and C was larger than the difference between orders under O and S. Follow-up planned contrasts showed that some of these differences were significant, while others were not. Orders under O were not significantly greater than orders under S ($F(1, 67) = 3.34, p = .0719$). Orders under S were significantly greater than orders under C ($F(1, 67) = 16.80, p = .0001$), and orders under O were significantly greater than orders under C ($F(1, 67) = 35.75, p < .0001$).

We also investigated whether experience affected ordering decisions over time. As one can observe from Figure 2, orders appear to be increasing over time in the high-profit condition, however, the effect was not significant (Wilks’ Lambda = .641, $F(19, 49) = 1.44, p = .1502$). We also found no significant interaction between payment scheme and round (Wilks’ Lambda = .719, $F(38, 98) = .46, p = .9957$).

Subjects’ first ordering decisions followed the same order, highest under O and lowest under C ($q^O = 11.250, q^S = 10.523, q^C = 9.130$). The difference between orders under O and S was not significant ($F(1, 67) = 1.69, p = .1980$), the difference between orders under S and C was significant ($F(1, 67) = 6.04, p = .0166$), and the difference between orders under O and C was significant ($F(1, 67) = 14.32, p = .0003$).

Regarding subjects’ responses to the final question (“If you could play the game again choosing only one order quantity, what number would you choose?”), the average quantity ordered by subjects in scheme O was greater than that ordered by subjects in group S, which, in turn, was greater than that ordered by subjects in group C ($q^O = 12.250, q^S = 11.608, q^C = 10.261$). The difference between orders under O and S was not significant ($F(1, 67) = 2.69, p = .1058$), while the difference between orders under S and C and between orders under O and C were significant ($F(1, 67) = 11.62, p = .0011$ and $F(1, 67) = 25.86, p < .0001$, respectively).

The actual demands generated by rolling the three dice were relatively consistent with the
Figure 2: Average order quantities of subjects in each round of play in the high-profit condition of Study 2 under own financing (O), supplier financing (S), and customer financing (C).

Theoretical predictions. Mean demands were 10.787, 10.679, and 10.670, under O, S, and C, respectively. Also, all participants correctly answered 3 and 18 for the minimum and maximum possible demand that could be generated by rolling 3 dice.

Low-Profit Condition

For the low-profit condition, the repeated measures generalized linear model showed that payment scheme significantly affected ordering behavior \( F(2, 57) = 7.15, p = .0017 \). We again found that average orders were highest under payment scheme O, and lowest under payment scheme C. However, the differences that were significant were not the same as in the high-profit condition. The difference between average order size under each payment scheme was larger between O and S than between S and C \( (q^O = 10.478, q^S = 9.305, q^C = 9.137) \). Order quantities under O were significantly greater than orders under S \( (F(1, 57) = 9.22, p = .0036) \) but orders under S were not significantly greater than orders under C \( (F(1, 57) = .19, p = .666) \). Orders under O were significantly greater than orders under C \( (F(1, 57) = 12.04, p = .001) \). Though from Figure
Table 4: Mean and standard deviations of ordering quantities, and significance tests for differences between payment schemes in Study 2. * $p < .05$, ** $p < .01$, *** $p < .001$.

<table>
<thead>
<tr>
<th>Payment Scheme</th>
<th>Mean order quantity (standard deviation in parentheses)</th>
<th>Contrast tests</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average over 20 rounds Round 1 Follow-up question</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High-Pro fit Condition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own Financing (O)</td>
<td>11.821 (1.336) 11.250 (2.691) 12.250 (1.595) 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplier Financing (S)</td>
<td>11.233 (1.020) 10.522 (1.344) 11.610 (1.373) 23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Financing (C)</td>
<td>9.900 (.892) 9.130 (1.359) 10.261 (.964) 23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$q^O - q^S$ .588 .728 .640</td>
<td>$q^S - q^C$ 1.333*** 1.392* 1.349**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$q^O - q^C$ 1.921*** 2.120*** 1.989***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Low-Pro fit Condition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own Financing (O)</td>
<td>10.478 (1.270) 11.100 (2.900) 10.350 (.988) 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplier Financing (S)</td>
<td>9.305 (.974) 9.600 (2.210) 8.850 (1.089) 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Financing (C)</td>
<td>9.137 (1.382) 9.050 (2.012) 8.700 (1.719) 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$q^O - q^S$ 1.173** 1.500 1.500***</td>
<td>$q^S - q^C$ .168 .550 .150</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$q^O - q^C$ 1.340*** 2.050** 1.650***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 it appears that orders were decreasing over time, the effect was not significant (Wilks’ Lambda $= .728, F(19, 39) = .77, p = .73$). We also found no significant interaction between payment scheme and round (Wilks’ Lambda $= .440, F(38, 78) = 1.04, p = .428$).

We examined subjects’ first ordering decisions and found that orders were highest under O and lowest under C ($q^O = 11.100, q^S = 9.600, q^C = 9.050$). However, contrast tests showed that not all of the differences were significant: the difference between orders under O and S was marginally significant ($F(1, 57) = 3.89, p = .0534$), the difference between orders under S and C was not significant ($F(1, 57) = .52, p = .472$) and the difference between orders under O and C was significant ($F(1, 57) = 7.27, p = .0092$).

Subjects’ responses to the final question were also highest under O and lowest under C ($q^O = 10.350, q^S = 8.850, q^C = 8.700$) with two contrast tests being significant: the difference between orders under O and under S was significant ($F(1, 57) = 13.18, p = .0006$), the difference between
orders under S and under C was not significant \((F(1,57) = .13, p = .7179)\), and the difference between orders under O and under C was significant \((F(1,57) = 15.95, p = .0002)\). These were consistent with the observed quantity choices of the subjects.

The actual demands generated by rolling the three dice were relatively consistent with the theoretical predictions. Mean demands were 10.523, 10.570, and 10.690, under O, S, and C, respectively. Also, all participants correctly answered 3 and 18 for the minimum and maximum possible demand that could be generated by rolling 3 dice.

### 4.2.4 Discussion

Study 2 examines the effect of payment timing for high and low-profit products. We find that for both types of products, payment timing significantly affects ordering decisions. Therefore, as in Study 1, we find support for Hypothesis 3, which again rejects a static rewards model.

A second objective of Study 2 is to test the robustness of the model of underweighting order-time payments. First, we find that under both high and low-profit conditions, orders trend from highest to lowest \(q^O > q^S > q^C\). Nevertheless, not all of these differences are significant. That
is, we do not find significant results for all three parts of Hypotheses 4 and 5. Specifically, for the high-profit condition, we find significant support for $q^S > q^C$ and $q^O > q^C$ but not for $q^O > q^S$. That is, Hypothesis 4 parts (SC) and (OC) are supported, but not (OS). On the other hand, for the low-profit condition, we find significant support for $q^O > q^S$ and $q^O > q^C$ but not for $q^S > q^C$. That is, Hypothesis 5 parts (OS) and (OC) are supported, but not (SC).

We offer the following explanation for this distortion. Because of the high/low profit margin parameters, the amount of order-time payment subject to underweighting is different under schemes O, S, and C. Refer to Table 3. Under the high-profit condition, the cost is $1 per unit and the profit is $3 per unit. The magnitudes of the order-time payments per unit under schemes O, S, and C are $1, $0, and $3, respectively. Thus, the payment underweighting effect has a much greater impact on payment scheme C compared to schemes O and S. This is consistent with our observations that differences between O and C and between S and C are significant, but the difference between O and S is not. Similarly, for the low-profit case, the order-time payments per unit are $3, $0, and $1, under O, S, and C, respectively. Thus, the payment underweighting effect has a much greater impact on payment scheme O compared to schemes S and C, leading to significant differences between O and C and between S and C, but not between O and S.

To further understand this phenomenon, we calculate the expected-reward-maximizing order quantities based on the model of underweighting order-time payments using the estimated underweighting factor $\beta = 0.785$ obtained from Study 1. The results are shown in Table 5, along with the observed average order quantities in Study 2. From this table, we make two observations. First, we see that the payment underweighting model is consistent with the findings of Study 2 in terms of the relative differences between order sizes under O, S, and C. Second, we also see a pull-to-center effect across all payment schemes. For the high-profit condition, the observed values are lower than the model’s prediction, while for the low-profit condition, the observed values are higher than the model’s predictions. As a result of these two observations, we conclude that while the pull-to-center effect shifts order quantities towards the center, the order and the relative distances between the payment schemes are preserved. Furthermore, we also note that the “downward pull” in the high-profit case appears to be stronger than the “upward pull” in the low-profit case. This is consistent with what Ho et al. (2010) observe in their experiments.
Theoretical order quantity under payment underweighting with $\beta=0.786$

<table>
<thead>
<tr>
<th>Payment Scheme</th>
<th>Theoretical order quantity under payment underweighting with $\beta=0.786$</th>
<th>Average order quantity in Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Profit</td>
<td>Low Profit</td>
</tr>
<tr>
<td>Own Financing (O)</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Supplier Financing (S)</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Customer Financing (C)</td>
<td>11</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 5: Average observed order quantities in Study 2 and the theoretical order quantities according to the payment depreciation model.

5 Concluding Remarks

In this paper, we find that payment timing can significantly affect ordering behavior in the newsvendor problem. These differences can be explained by the effect of underweighting order-time payments, which results from individuals’ mental accounting. In most newsvendor experimental studies in the literature, including, for example, Schweitzer and Cachon (2000) and Bolton and Katok (2008), individuals were provided with feedback concerning their net profit at the end of each round (or set of rounds). We show that significantly different ordering behavior can be generated by providing more specific feedback, such as making explicit payments to and from the participant at different time points during a round. In a sense, we demonstrate that the inventory ordering behavior is subject to the framing of the problem in terms of different payment schemes. Prelec and Loewenstein (1998) argue that one can frame a choice problem to facilitate “decisional efficiency” in consumer behavior by promoting tight linkage between spending and consumption. Here we can do the same: the supplier-financing payment scheme facilitates the tightest linkage between costs and benefits, and it has the highest decisional efficiency – individuals are observed ordering the optimal quantity when other effects are controlled for (Study 1). Furthermore, understanding this decisional effect can help firms gain advantage in designing financial transactions and contract terms with their counterparts.

For instance, suppliers often offer retailers trade credit, allowing retailers to delay payment for goods until they make the sale, hoping that this will encourage higher orders. This intended effect of trade credit is captured by the time-discounted rewards model in Section 3.2. When capital constraint is not an issue and the interest rate is negligible, the practice of trade credit
(corresponding to the supplier financing scheme) might inadvertently lower the retailer’s order quantity relative to that without trade credit (corresponding to the own-financing scheme) as shown by our experiments. When the interest rate is significant, however, both the time-value of money and the payment underweighting effects could be significant and either can dominate depending on the system parameters. Therefore, suppliers should carefully evaluate these effects before proposing trade credit terms to their downstream partners.

Another application is in supply chain contract design and coordination. A wholesale-price contract typically has payment transactions resembling the own-financed scheme in this paper. Our results suggest that the retailer may place larger-than-optimal orders due to the payment underweighting effect, reducing some of the supply chain inefficiency due to wholesale-price contracts. If the supplier can estimate the retailer’s underweighting factor $\beta$ as we did in Section 4, then she may coordinate the supply chain by setting the wholesale price equal to the unit production cost divided by the underweighting factor. Under a buy-back contract, the retailer receives a refund for leftover inventory upon demand realization. To the retailer, the refund payment is likely to be weighted more than the purchase cost incurred earlier at the order time. Thus, the supplier may exploit this effect to achieve supply chain coordination by offering a smaller buy-back price for leftover inventory. For a similar reason, under a revenue-sharing contract, with payment underweighting, the supplier may be able to charge a higher wholesale price to the retailer and still achieve supply chain coordination. It would be interesting to empirically investigate these potential implications. However, we acknowledge that there could be many other factors at work simultaneously in real-world contract settings, so isolating the effect of payment timing may present a non-trivial challenge.

Acknowledgements

The authors thank Rick Larrick, Jack Soll, Jeannette Song, and Daniel Feiler for helpful discussions on the paper and research assistants Nathan Glencer and Lucy Zheng for their help with running the experiments.
References


