Technical Memorandum Number 817

Subcontracting Strategies with Stochastic Third-Party Capacity and Tardiness Penalty Contracts

by

Tolga Aydinliyim
George L. Vairaktarakis

September 2007

Department of Operations
Weatherhead School of Management
Case Western Reserve University
330 Peter B Lewis Building
Cleveland, Ohio 44106

Subcontracting Strategies with Stochastic Third-Party Capacity and Tardiness Penalty Contracts

Tolga Aydinliyim [*] George L. Vairaktarakis [†]

September 17, 2007

Abstract

We study a subcontracting model where a manufacturer cannot process all his workload before the customer due date. Therefore, he subcontracts part of his workload to a third-party who has prior customer commitments. Prior committed orders require processing that follows a general distribution and hence the available capacity at the third-party is uncertain. The manufacturer wants to maximize his expected profits by partitioning his workload among his in-house production capacity and the third-party. We study two different contracts; (i) the manufacturer pays a unit processing fee for workload processed by the third-party, (ii) in addition to the unit processing fee, the third-party shares the tardiness penalties incurred by the manufacturer. We show that this production chain cannot be coordinated under the first contract unless the third-party accepts to merely break even. Under the second contract, coordination is possible if the unit processing fee and the tardiness sharing fraction are negotiated jointly. In this case, the coordinating contract Pareto-dominates non-coordinating contracts. We show that with the coordinating contract, the tardiness sharing fraction coincides with the third-party's share of the additional profits generated by the coordinated chain. We also studied the tardiness penalty sharing contract when one of the parameters is chosen exogenously and compared its merit to the unit processing fee contract. Despite being more difficult to administer and verify in this case, the tardiness penalty sharing contract proved to be more *efficient* than the unit processing fee contract.

[*]Decision Sciences Department, Lundquist College of Business, 1208 University of Oregon, Eugene, OR 97403-1208, tolga@uoregon.edu

[†]Department of Operations, Weatherhead School of Management, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7235, gxv5@case.edu

1 Introduction

Subcontracting to third-parties has recently become prevalent business practice across many industries. According to Day (1956), subcontracting refers to "the procurement of an item or service that a firm is normally capable of economic production in its own facilities and which requires a prime contractor to make specification available to the supplier". Outsourcing, on the other hand is a special case where the firm has no means to produce on its own. Firms strategically outsource their non-core operations and focus on their core competencies to enhance their effectiveness in the long-term (Greaver (1999)). Whereas firms use subcontracting as a short-term solution to increase their flexibility, reduce their exposure to risk, improve their response to unexpected increases in demand, and reduce costs. However, it is not easy to make decisions on when to subcontract and how much to subcontract.

Capacity and subcontracting decisions arise in many activities ranging from the production of semiconductor chips to staffing of call centers, from transportation contracts to textile-apparel retail channels. In semiconductor manufacturing most *Fabless Original Equipment Manufacturers* completely outsource the manufacturing of the customized Sytem-on-Chip (SoC) solutions developed by foundries such as United Microchips Corporation (UMC). However, a number of *Integrated Device Manufacturers* which handle semiconductor manufacturing in-house. These companies often times use subcontractors to compete for contracts that require additional production capacity and time beyond their normal capabilities. Subcontracting provides short-term flexibility, is mostly cheaper than overtime and less capital-intensive than capital investments which may not be easily reversible. In operating call centers the major cost factor is the size of the workforce. Hiring new employees entails training costs. Moreover, once the number of operators to employ is decided, their salary has to be paid even if the anticipated demand does not materialize. Also, there may be a limit to the amount of overtime (Gans and Zhou (1999)). In logistics, the carrier guarantees a constant capacity and charges the deliveries in excess of the agreed upon capacity at a higher rate. One can view this excess capacity as the subcontracted demand. In the textile-apparel industry, one can see agreements between producers and subcontractors, where the producers who want to prevent themselves from demand peaks use subcontractors instead of allocating larger capacity than the amount needed to meet the average demand; see Tan (2004) and Tan and Gershwin (2004). Tan (2006) reports that in a textile-apparel retail channel, a primary contractor bids for contracts from large buyers such as Wal-Mart, Federated, GAP and then uses a chain of small to medium size enterprises (SME's) as subcontractors. Sometimes, temporary entities called *network organizations* are formed by these SME's so that they achieve the capacity to win the contract.

Over the last twenty years, numerous subcontractors have evolved as specialized manufacturers who provide third-party services for certain manufacturing operations. For example, Electronic Contract Manufacturing (ECM) is a term used for companies that offer contracts for electronic assembly for another company and it is a sector that is forecasted to expand at 20-25% annually. These contract manufacturers have great economies of scale as they can utilize the most appropriate technology and offer reasonable prices and excellent operational efficiency that is reflected in improved lead times. The ECM revolution started with IBM in 1981. Subsequently, it became a trend for the Original Equipment Manufacturers to entirely outsource non-core operations and sell off their production units to ECM's. Erickson for example sold seven Swedish plants to Flextronics and Solectron in 1997. Key ECM's today include Foxconn, Solectron, Flextronics, Sanmina-SCI and others.[1]

Coordination among the customers of the third-party contractors may be achieved more easily when the third-party has the internet-based technologies to provide capacity and production schedule information to his customers (e.g. MyUMC of UMC[2]), or when the manufacturers are members of the extended supply chain of the parent company, such as Cisco (Grosvenor and Austin (2001)). However these opportunities are not available all the time. Having recognized this fact and the increasing trend towards subcontracting and outsourcing, the United Nations Industrial Development Organization (UNIDO) has formed Subcontracting and Partnership Exchanges (SPX).[3] SPX's are technical information, promotion and match making centers for industrial subcontracting and partnership between main contractors, suppliers and subcontractors, aiming at the optimal utilization (the most complete, rational and productive) of the manufacturing capacities of the affiliated industries. Today more than 44 SPX's in over 30 countries facilitate production linkages between small, medium and large manufacturing firms and link up with global markets and supply chain networks.

Subcontracting of industrial production by SPX's members is generally based on the short-term need for additional production capacity. When the available production capacity with a main contractor is not sufficient to cope with the total volume of production necessary to execute an

[1] http://en.wikipedia.org/wiki/Electronic_Contract_Manufacturing

[2] MyUMC is UMC's full-service customer information portal that gives customers easy access to UMC's foundry services by providing a total online supply chain solution. This application offers 24-hour access to detailed account information such as manufacturing, engineering, design and financial data through each user's own personalized MyUMC start page.

[3] http://www.unido.org/doc/4547

order and when further creation of an in-house capacity is either infeasible or undesirable, the main contractor has to depend on a subcontractor to manufacture the balance of the order quantity. In such a setting, the customers served by the third-party are sometimes not equally important to the third-party. For example, consider the significance of Toyota for a bumper manufacturer, or Intel for a chip manufacturer. Therefore, one can see various subcontracting settings where the third-party acts as a rational profit maximizer who prioritizes one customer over the others (for a related model where the manufacturers served by the third-party are of similar importance to the third-party see Vairaktarakis and Aydinliyim (2007)).

In this article, we study a subcontracting model where a manufacturer cannot process all his workload before a given due date, and hence subcontracts part of his workload to a third-party. The relationship between the manufacturer and the third-party is a one time business relationship, which provides increased capacity and flexibility to the manufacturer required to take a large contract from a buyer. The third-party is also committed to another customer. Order quantities due to this customer follow a general distribution and hence the available capacity at the third-party is uncertain. Such situations may arise when the third-party is a committed supplier to a large customer, who sole-sources from the third-party (see Elmaghraby (2000) for a review of sourcing strategies). In most manufacturing industries, a powerful buyer such as Toyota favors rationalized supplier chains and maintain long-term business relationships with a single or few third-parties. The manufacturer wants to maximize his expected profits by partitioning his workload among his in-house production capacity and the third-party. We study two different contracts:

(i) *Unit Processing Charge Contracts:* The manufacturer is charged a unit processing fee for his workload processed at the third-party.

(ii) *Tardiness Penalty Sharing Contracts:* In addition to the terms of the first contract, the third-party agrees to share the tardiness penalties incurred due to the delays on his schedule.

The third-party acts as the Stackelberg leader, decides on the contract type and the contract parameters and then offers the contract to the manufacturer as a take-it-or-leave-it deal. If the manufacturer takes the offer, he responds by subcontracting part of his workload to the manufacturer.

The rest of this article is organized as follows. In Section 2 we survey the related literature. In Section 3 we define our model formally. In Section 4, we consider the centralized problem. In Section 5 we first consider the unit processing charge contract (see Section 5.1). Then we analyze the tardiness penalty sharing contract (see Section 5.2) with emphasis on the allocation of the additional chain profits when both parties coordinate. In Section 6 we consider the case where one of the parameters in the tardiness penalty sharing contract is exogenously determined and compare it with the unit processing charge contract. We discuss our findings and make concluding remarks in Section 7.

2 Literature Review

The literature in supply chain contracts that investigates the single buyer single supplier relationship is quite rich. In the context of a newsvendor model, researchers typically study three important questions. First, which contracts coordinate the supply chain? A contract is said to coordinate the supply chain if the set of supply chain optimal actions is a Nash Equilibrium. Second, which contracts have sufficient flexibility (by adjusting parameters) to allow for any division of the supply chain profits? If a contract can allocate rents arbitrarily, then there exists a contract that Pareto-dominates a non-coordinating contract. Third, which contracts are worth adopting? Although coordination and flexible rent allocation are desirable features, a contract that satisfies these conditions might be difficult to administer, or it might be difficult for the coordinating parties to verify if the terms of the contract are granted. In such a case, simpler contracts with tolerable supply chain performance are utilized. Lariviere (1998) and Cachon (2003) surveyed related results. For detailed discussions of such contracts, see Lariviere and Porteus (2001) for wholesale price contracts, Pasternack (1985) for buy-back contracts, Cachon and Lariviere (2005) for revenue-sharing contracts, Tsay (1999) for quantity-flexibility contracts, Taylor (2002) for sales-rebate contracts, Bernstein and Federgruen (2005) for price-discount contracts, etc. Our study differs from this huge body of literature in the following regards. In the single supplier single buyer models, the retailer completely sources from the supplier, the retailer faces a stochastic demand, and there are backlog penalties incurred by both parties. In our model the manufacturer allocates his workload between his production capacity and the third-party capacity, the demand for the manufacturer is deterministic but the third-party capacity is stochastic, and the third-party partially compensates for the unmet demand. Also, we consider tardiness penalties which reflect the delivery time of the production activities.

The problem of subcontracting has been studied in the economics literature in the context of vertical integration without taking capacity constraints into account. Kamien et al. (1989), Kamien and Li (1990) first model capacity constraints, either implicitly or explicitly, in the context of subcontracting production. A surney of the literature in supply contract design and analysis, where capacity is explicitly taken into account as a decision variable, is presented for example in Tsay *et al.* (1998) and Lariviere (1998).

In Van Mieghem (1999) the option of subcontracting to improve financial performance is considered by analyzing a sequential stochastic investment game. The manufacturer and the subcontractor first decide independently on their capacity investment levels and then the manufacturer has the option to subcontract part of his production to the subcontractor. Our work is related in that the manufacturer has the option to produce in-house or subcontract. We consider deterministic workload for the manufacturer whose short-term capacity represented by the due date is limited, and hence the timing of the production activities is important.

Atamturk and Hochbaum (2001) provided a multi-period treatment of subcontracting that focuses on the production aspect. They consider the trade-off between acquiring capacity, subcontracting, production, and holding inventory by providing analytical models, structural results on the optimal solutions, and algorithms that simultaneously optimize these interrelated decisions. Following a Markov Decision Process approach, Tan (2004) considered a model with a subcontractor who guarantees long-term availability. Similarly, using a stochastic optimal control problem formulation, Tan and Gerscwin (2004) analyzed the production and subcontracting strategies for a manufacturer with limited capacity and volatile demand where there are a number of subcontractors available. Our model differs from these studies in the sense that we consider a single period model, where demand uncertainty is captured by the committed workloads at the third-party and the penalties for the unmet demand are handled in an alternative way.

Literature in subcontracting and outsourcing mostly focuses on pricing issues. It only considers production at the aggregate level disregarding the *timing of the production activities* and the *ability to meet delivery schedules* by the subcontractors. According to a 2003 survey in Wall Street Journal, the Original Equipment Manufacturers (OEM's) rated the ability to meet delivery schedules as the most significant factor in choosing contractors, whereas price was ranked only fifth (Ansberry and Aeppel (2003)). As noted in Anupindi and Bassok (1999) delivery commitments are of crucial importance. As noted in Li (1992), the delivery delays for chips provided by subcontractors cause costs to increase significantly as the production lines need to be shut down when chips are not available. Bernstein and Véricourt (2006) considered procurement contracts with service guarantees. Specifically they considered a market with two suppliers and a set of buyers in search of procurement contracts with one of the suppliers. Cachon and Zhang (2006, 2007) considered the importance of delivery commitments in single and multi-sourcing settings. In our model, we consider the timing issues in subcontracting and investigate the impacts of the contract choice from the manufacturer's, the third-party's and the chain's point of view.

Research at the shop floor level is conspicuously scarce even though supply coordination necessitates the coordination of production activities. For a review of cooperative scheduling games introduced by Curiel *et al.* (1989), see Curiel *et al.* (2002). In a practical setting motivated by the coordination of manufacturing operations at Cisco's supply chain Aydinliyim and Vairaktarakis (2006) studied a setting where a group of manufacturers outsource operations to the same third-party whose limited capacity is represented by manufacturing windows. Cai and Vairaktarakis (2006) considered a similar model where tardiness penalties are considered and optional overtime capacity is available for windows booked by the manufacturers. In a subcontracting setting, Vairaktarakis (2006) considered a model where multiple manufacturers compete for earlier positions at the third-party schedule. Vairaktarakis and Aydinliyim (2007) considered the centralized strategies for the same problem, provided an allocation scheme for the additional savings due to centralization, and numerically compared the value of centralization and the efficiency of the incentive rules suggested by Vairaktarakis (2006). In the later article, each manufacturer is of the same importance to the third-party and the capacity at the third-party is deterministic. Alternatively, our model captures the stochastic availability of the third-party capacity and the effect of prior commitments to a more important customer for the third-party.

3 The Model

Let m be a manufacturer who owns a contract to process P hours of workload for a customer. He makes a profit of r_m for each unit of processing and accepts a penalty g_m for every unit of workload that is completed past a given due date d which is smaller than P. Otherwise, the manufacturer would be able to meet the due date by processing the entire workload in-house. For this reason, m subcontracts part x of his workload to a third-party ($3P$) and processes the rest $P - x$ in-house. Suppose that sending a unit of workload to $3P$ costs c_m more to manufacturer m than the in-house production. Note that c_m includes all additional variable costs that m incurs when he subcontracts instead of producing in-house, for example unit transportation costs.

The $3P$ has prior commitments to other customers. The total workload he receives from these customers is a random variable y which follows a general distribution with CDF F which is twice-differentiable, has positive density f on $[0, \infty)$, and expected value μ_y. We assume that the $3P$ has prior agreements with these customers which force him to prioritize the processing of workload y in his production schedule. Such contract commitments are commonly found between Just-In-Time manufacturers and their suppliers. The $3P$ acts as a Stackelberg leader and offers a contract to manufacturer m. If the parties reach agreement the $3P$ receives a transfer payment, T. In its simplest form, the transfer payment is a function of the subcontracted amount, x, and can take on additional parameters depending on the contract. Then, $3P$ processes y hours of workload followed by x hours of workload from manufacturer m. A typical schedule for m and $3P$ is presented in Figure 1.

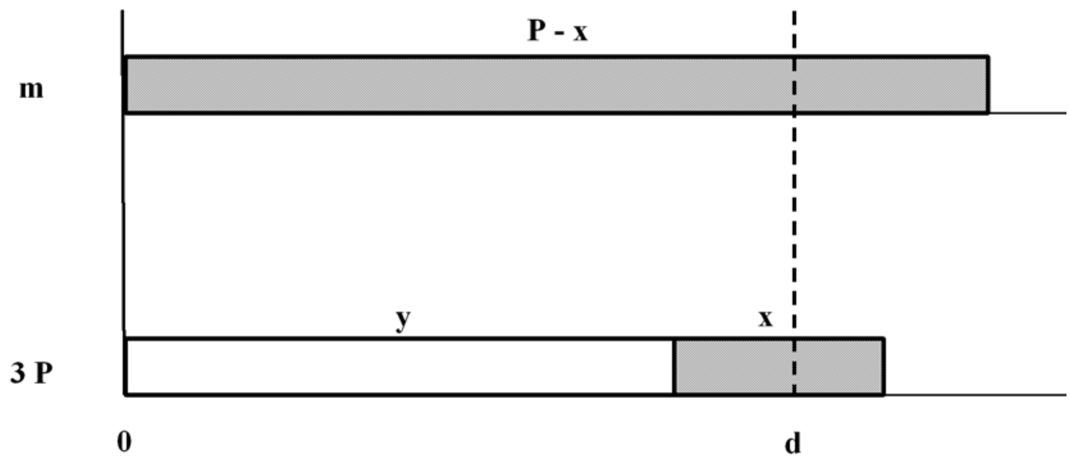

Figure 1: A typical schedule for m and $3P$

Let $\pi_m(x)$ be the expected profit for manufacturer m when he chooses to subcontract x hours of workload to $3P$. The expected profit for $\pi_{3P}(x)$ is defined similarly. Let $L_{3P}(x)$ and $L_m(x)$ be the tardiness incurred by $3P$, m respectively. Then, we can write

$$\pi_m(x) = r_m P - c_m x - g_m \cdot (E_y[L_{3P}(x)] + L_m(x)) - T, \tag{1}$$

where $E_y[L_{3P}(x)]$ is the expected tardiness at the $3P$ schedule. The first term in (1) is manufacturer m's profit due to processing; the second term is the additional subcontracting cost for x hours of workload; the third term is the expected total tardiness; and the last term is the transfer payment between m and $3P$. Manufacturer m incurs tardiness due to the $3P$ schedule only when $y + x$ exceeds d; the customer due date. Therefore,

$$\begin{aligned} E_y[L_{3P}(x)] &= E_y(y + x - d)^+ \\ &= \int_{d-x}^{\infty} (y + x - d) f(y) dy \end{aligned}$$

$$
\begin{aligned}
&= \int_{d-x}^{\infty} y f(y)dy + \int_{d-x}^{\infty} (x-d)f(y)dy \\
&= \mu_y - \int_{0}^{d-x} y f(y)dy + (x-d)[1 - F(d-x)] \\
&= \mu_y - [yF(y) \mid_0^{d-x} - \int_{0}^{d-x} F(y)dy] + (x-d)[1 - F(d-x)] \\
&= \mu_y - [(d-x)F(d-x) - \int_{0}^{d-x} F(y)dy] + (x-d)[1 - F(d-x)] \\
&= (\mu_y + x - d) + \int_{0}^{d-x} F(y)dy.
\end{aligned}
$$

Similarly, the tardiness incurred by manufacturer m due to his own schedule is

$$
L_m(x) = \begin{cases} P - x - d & , \text{ if } \; x < P - d \\ 0 & , \text{ if } \; x \geq P - d. \end{cases} \tag{2}
$$

Note that $L_m(x)$ is piecewise linear convex function. Let c_{3P} be the hourly processing cost for $3P$, and r_{3P} be the profit for processing each unit of the committed workload y. Then, the expected profit for $3P$ is

$$
\pi_{3P}(x) = r_{3P}\mu_y - c_{3P}x + T. \tag{3}
$$

The first term in (3) is the expected net profit from the processing of the committed workload; the second term captures the production cost of processing x units for manufacturer m; and the last term is the transfer payment received from manufacturer m.

4 The Centralized Schedule

In this section, we study the problem of finding the centralized optimal workload x^* which maximizes the expected profit for the entire chain, i.e.,

$$
\begin{aligned}
\Pi(x) &= \pi_m(x) + \pi_{3P}(x) \\
&= r_{3P}\mu_y + r_m P - (c_m + c_{3P}) \cdot x - g_m \cdot (E_y[L_{3P}(x)] + L_m(x)).
\end{aligned} \tag{4}
$$

Note that $\Pi(x)$ is not differentiable at $x = P - d$. Define the function $\Sigma(x)$ as

$$
\Sigma(x) = r_{3P}\mu_y + r_m P - c\,x - g_m \cdot (E_y[L_{3P}(x)] + P - x - d)\,, \tag{5}
$$

where $c = c_m + c_{3P}$. Recall that the tardiness function $L_m(x)$ is a piecewise linear convex function (see (2)), and hence we can re-write (4) as

$$\Pi(x) \quad = \begin{cases} \Sigma(x) & , \text{if} \quad x < P - d \\ \Sigma(x) + g_m \cdot (P - x - d) & , \text{if} \quad x \geq P - d. \end{cases} \tag{6}$$

Note that $g_m \cdot (P - x - d) \leq 0$ when $x \geq P - d$, which implies that it is never optimal to subcontract more than $P - d$ units of workload. Intuitively, when $x > P - d$, manufacturer m has idle time before d in his own schedule. Therefore, we restrict our attention to the interval where $x \leq P - d$, or $\Pi(x) = \Sigma(x)$. Observe that

$$\begin{aligned} \frac{dE_y[L_{3P}(x)]}{dx} &= \frac{d\{(\mu_y + x - d)\}}{dx} + \frac{d\{\int_0^{d-x} F(y)dy\}}{dx} \\ &= 1 + (-1) \cdot [F(d - x) - F(0)] \\ &= 1 - F(d - x), \end{aligned}$$

and hence, the first order order derivative of (5) equals

$$\frac{d\Sigma(x)}{dx} = -c + g_m F(d - x). \tag{7}$$

Differentiating (7) we obtain

$$\frac{d^2\Sigma(x)}{dx^2} = -g_m f(d - x) < 0. \tag{8}$$

Therefore, $\Sigma(x)$ is strictly concave and attains its unique maximum when the first order conditions (FOC) are satisfied, i.e., there exists $x_c^* \leq P - d$ such that

$$F(d - x_c^*) = \frac{c}{g_m}. \tag{9}$$

Note that, for (9) to hold true, we need $c \leq g_m$. Otherwise, the chain optimum is obtained when m incurs all penalty costs and does not subcontract at all. In (9), it is possible to have $x_c^* < 0$ when $\Sigma(x)(= \Pi(x))$ is decreasing at $x = 0$. But then, $F(d) < \frac{c}{g_m}$ and m's optimal action is not to subcontract. Therefore, the centralized optimal workload to be subcontracted to $3P$ is

$$x^* = \max(x_c^*, 0). \tag{10}$$

A clarification is in order. It should be noted that $\Sigma(x) \geq \Pi(x)$ for all possible x values with strict inequality for $x > P - d$ (see (6)). We already noted that $x^* \leq P - d$ which is on the interval where $\Sigma(x) = \Pi(x)$, and hence $\Sigma(x^*) = \Pi(x^*)$. However, $\Pi(x)$ is not differentiable at $x = P - d$, so optimizing $\Sigma(x)$ over all possible values of x captures the unique optimum for $\Pi(x)$ even when $x^* = P - d$.

The optimal workload x^* is always less than d when the the total workload for the manufacturer is less than twice his due date, i.e. $P \leq 2d$. However, when $P > 2d$, tardiness cannot be avoided unless the manufacturer subcontracts to more than one third-party, or when the third-party's technology allows processing the subcontracted amount faster (or more efficiently) than the manufacturer. In our model, we assume that the manufacturer and the third-party have the same processing speed. Even though the manufacturer is able to process the workload himself, he subcontracts as a result of the need for additional capacity, i.e., $P > d$. In this study we do not consider other modes of production for various reasons. For example, overtime might be costly or simply unavailable. Investing in new resources is capital intensive and can only be justified on the basis of long-term planning. In our model, we focus on the value of subcontracting as a one-time business transaction employed as a short-term solution. Evaluating the technological advantage of the third-party is a fruitful research direction due to the increasing popularity of contract manufacturing as discussed in the introduction.

5 Decentralized Schedules

We now consider the problem of manufacturer m who tries to find the optimal allocation of his workload between his own and $3P$'s capacity with the objective of maximizing his own expected profit. Manufacturer m's expected profit depends on the terms of the contract between him and $3P$. We study two such contracts.

5.1 Unit Processing Charge Contracts

The unit processing charge contract is of the form $T(x, u) = u \cdot x$, i.e., $3P$ charges u for each unit of m's subcontracted workload. Then, the manufacturer's expected net profit can be re-written as

$$\pi_m(x) = r_m P - (c_m + u) \cdot x - g_m \cdot (E_y[L_{3P}(x)] + L_m(x)). \tag{11}$$

As in Section 4, the optimal workload, say x^*_m, for m to subcontract is at most $P - d$. For notational simplicity we do not define a function similar to $\Sigma(x)$ of Section 4. Instead, we present results by assuming

$$\pi_m(x) = r_m P - (c_m + u) \cdot x - g_m \cdot (E_y[L_{3P}(x)] + P - x - d) \tag{12}$$

for all possible values of x, knowing that the profit maximizing subcontracted workload is less than $P - d$. Hence, similar to the analysis in Section 4, $\pi_m(x)$ is concave for $x \leq P - d$ because

$\frac{d^2\pi_m(x)}{dx^2} = -g_m f(d-x) < 0$. Therefore, there exists a unique $x_u^* \leq P - d$ for which the FOC for m's expected profit (12) are satisfied, i.e.,

$$F(d - x_u^*) = \frac{c_m + u}{g_m}. \tag{13}$$

Note that $g_m > c_m + u$, otherwise manufacturer m is always better-off by not subcontracting. All other parameters being constant x_u^* decreases with increasing u. The manufacturer may choose not to subcontract at all if the hourly fee u is too high. This is the case when $F(d) > \frac{c_m+u}{g_m}$ which implies that the manufacturer's expected profit in (11) is decreasing and hence in this case x_u^* is negative. Therefore, we have

$$x_m^* = \max(x_u^*, 0). \tag{14}$$

Similar to the centralized schedule, $x_m^* \leq d$ holds trivially when $P \leq 2d$. In what follows we discuss coordination possibilities. Suppose that $x^* = x_c^*$ and $x_m^* = x_u^*$. Coordination is achieved if the supplier's best action is to propose a contract which makes $x^* = x_m^*$.

Lemma 1 *If F is strictly increasing on $[0, \infty)$, coordination is possible only if $u = c_{3P}$.*

Proof of Lemma 1: Since $F(y)$ is strictly increasing in $y \in [0, \infty)$, there is a unique x_m^* for which $F(d - x_m^*) = \frac{c_m+u}{g_m}$. So, $x^* = x_m^*$ only when $F(d - x^*) = \frac{c_m+u}{g_m}$, or $\frac{c_m+u}{g_m} = \frac{c}{g_m}$ (due to (9)), which holds when $u = c_{3P}$. $\diamond$

Lemma 1 implies that $3P$ cannot coordinate the entire chain unless he agrees to break even for his business with manufacturer m. This is because of *double marginalization*, which points to coordination failure as there are two separate margins for the manufacturer and the third-party and neither party considers the margin for the entire chain. Similar to our finding, the wholesale-price contract does not coordinate the supplier and the retailed in a newsvendor setting. A two-echelon supply chain with a single supplier and a single retailer cannot be coordinated with a wholesale-price contract unless the supplier sets a whole-sale price less than his marginal cost (see Lariviere and Porteus (2001), Cachon (2003)). On the other hand, when the marginal cost is not constant, marginal cost pricing does not necessarily lead to zero profit (see Cho and Gerchak (2001), Bernstein, Chen, and Federgruen (2002)). In our case, the third-party already makes a positive profit on the committed workload y, and can accept zero margin on his business with the manufacturer in order to offset fixed costs when he does not utilize his excess capacity.

In what follows we show that the third-party in fact prefers to set the fee u higher than his marginal cost c_{3P}. The amount of workload that the manufacturer subcontracts depends on u (see

(13)). For strictly increasing F which is also differentiable everywhere on $[0, \infty)$, we can find a one-to-one mapping between the subcontracted workload u and x. Given

$$u(x) = -c_m + g_m \cdot F(d - x),$$

the $3P$'s objective function becomes

$$\begin{aligned} \pi_{3P}(u(x), x) &= r_{3P}\mu_y - c_{3P}x + u(x)x \\ &= r_{3P}\mu_y - c_{3P}x - [c_m + g_m F(d - x)] \cdot x \\ &= r_{3P}\mu_y - c \cdot x + g_m \cdot x \cdot F(d - x). \end{aligned} \tag{15}$$

Note that $\pi_{3P}(u(x), x)$ is twice-differentiable everywhere on $x \in [0, d)$ because F is differentiable everywhere on $[0, \infty)$. We can write

$$\begin{aligned} \frac{d\pi_{3P}(u(x), x)}{dx} &= -c + g_m \cdot [F(d - x) - f(d - x)x] \\ &= -c + g_m F(d - x) \cdot [1 - \tfrac{f(d-x)x}{F(d-x)}]. \end{aligned}$$

Note that $F(d - x)$ is decreasing in x, so the above expression is decreasing in x if $\frac{f(d-x)x}{F(d-x)}$ is increasing. But $\frac{f(d-x)}{F(d-x)} = -\frac{d \ln F(d-x)}{dx}$. As $F(d-x)$ is decreasing in x, $\ln F(d - x)$ is also decreasing. So, $\frac{d \ln F(d-x)}{dx} < 0$ and hence $\frac{f(d-x)x}{F(d-x)}$ is increasing in x. This result implies that $\pi_{3P}(u(x), x)$ strictly concave for $x \in [0, d)$. Then,

Lemma 2 *If F is strictly increasing in $[0, \infty)$, then $x^*_{3P} < x^*$.*

Proof of Lemma 2: As $\pi_{3P}(u(x), x)$ is continuous and concave, there must be a unique x^*_{3P} for which $\frac{d\pi_{3P}(u(x),x)}{dx} = 0$. At the supply chain optimum, $x = x^*$, we have $F(d - x) = \frac{c}{g_m}$. Replace $F(d - x)$ with $\frac{c}{g_m}$ in (15) to get

$$\begin{aligned} \frac{d\pi_{3P}(u(x), x)}{dx}\Big|_{x=x^*} &= -c + g_m \cdot [F(d - x^*) - f(d - x^*)x^*] \\ &= -c + g_m \cdot [\tfrac{c}{g_m} - f(d - x^*)x^*] \\ &= -g_m f(d - x^*)x^* < 0, \end{aligned}$$

which implies that $\pi_{3P}(u(x), x)$ is decreasing at $x = x^*$. The uniqueness of x^*_{3P} and the concavity of $\pi_{3P}(u(x), x)$ imply $x^*_{3P} < x^*$. $\diamond$

From Lemma 2, we can conclude that setting $u = c_{3P}$ is not the optimal strategy for the third-party. In summary:

Proposition 1 *Unit processing charge contract $T(x, u) = u \cdot x$ does not coordinate the manufacturer and the third-party.*

Contracts similar to the unit processing charge contract are common in practice as they are simple to administer and easy to verify. In the supply chain coordination literature it has been shown that the contracts of this type can actually coordinate the chain in the event that there are competing players (Cachon (2003)). For instance, Wang and Gerchak (2001) consider a model where the total demand (deterministic in nature) is allocated to the competing retailers proportional to their order quantities. Parlar (1988), Karjalainen (1992), Anupindi and Bassok (1999), and Anupindi, Bassok and Zemel (1999) assign independent demands to each retailer and then redistribute the excess demand. Lippman and McMardle (1997) allocate the demand using a splitting rule that uses the realization of the total demand. Another part of the literature considers demand allocated dynamically; see for example Gans (2002).

5.2 Tardiness Penalty Sharing Contracts

In an effort to coordinate the third-party and the manufacturer, in this subsection, we consider a contract that holds $3P$ responsible for the tardiness incurred on his resource, i.e. $T(u, x, \lambda) = u \cdot x - \lambda g_m \cdot (E_y[L_{3P}(x)])$ for $\lambda \in (0, 1)$. Under this contract, the manufacturer's expected net profit is

$$\pi_m(x) = r_m P - (c_m + u) \cdot x - (1 - \lambda) g_m \cdot (E_y[L_{3P}(x)]) - g_m L_m(x). \tag{16}$$

Although $3P$ agrees to share the tardiness penalties, m still does not want to subcontract more than $P - d$ units of his workload because (16) is strictly decreasing for $x > P - d$ even when $\lambda = 1$ (Note that we assume $L_m(x) = P - x - d$ for all x values knowing that the optimal subcontracted workload will be no more than $P - d$). For $x \leq P - d$, the expected profit function for m is strictly concave as $\frac{d^2 \pi_m(x)}{dx^2} = -(1 - \lambda) g_m f(d - x) < 0$, and hence there exists a unique x_t^* for which (16) is maximized. The FOC imply

$$F(d - x_t^*) = \frac{c_m + u - \lambda g_m}{(1 - \lambda) g_m}. \tag{17}$$

Note that $0 < \frac{c_m + u - \lambda g_m}{(1-\lambda) g_m} < 1$ implies $c_m + u < g_m$ and $\lambda < \frac{c_m + u}{g_m}$. Then, we have

$$x_m^* = \max(x_t^*, 0). \tag{18}$$

In what follows we discuss coordination issues for this penalty sharing contract. Suppose $x^* = x_c^*$ and $x_m^* = x_t^*$. Coordination is achieved if the supplier's best action is to propose a contract which makes $x^* = x_m^*$.

Lemma 3 *If F is strictly increasing on $[0, \infty)$, coordination is possible only if*

$$u = c_{3P} + \lambda(g_m - c). \tag{19}$$

Proof of Lemma 3: As $F(y)$ is strictly increasing in $y \in [0, \infty)$, there is a unique x_m^* for which $F(d - x_m^*) = \frac{c_m + u - \lambda g_m}{(1-\lambda) g_m}$. So, $x^* = x_m^*$ only when $F(d - x^*) = \frac{c_m + u - \lambda g_m}{(1-\lambda) g_m}$, or $\frac{c_m + u - \lambda g_m}{(1-\lambda) g_m} = \frac{c}{g_m}$ which is possible only when $u = c_{3P} + \lambda(g_m - c)$. $\diamond$

Note that, in Lemma 3, $g_m > c$ and hence $3P$ can coordinate the entire chain by offering u which is strictly bigger than his marginal cost c_{3P} by a fraction λ of the marginal lateness cost $(g_m - c)$ incurred by m when he does not subcontract.

5.2.1 The Third-Party's Profit

When the third-party offers a tardiness sharing contract, he decides: (i) the fraction λ of the manufacturer's expected tardiness shared by $3P$, and (ii) the unit processing fee u. Due to Lemma 3, chain coordination is possible if the third-party chooses (λ, u) pair so that $u(\lambda) = c_{3P} + \lambda(g_m - c)$. Then, the third-party's profit is

$$\begin{aligned}
\pi_{3P}(x) &= r_{3P}\mu_y - c_{3P}x + u(\lambda)x - \lambda g_m(E_y[L_{3P}(x)]) \\
&= r_{3P}\mu_y - c_{3P}x + [c_{3P} + \lambda(g_m - c)]x - \lambda g_m(E_y[L_{3P}(x)]) \\
&= r_{3P}\mu_y + \lambda(g_m - c)x - \lambda g_m(E_y[L_{3P}(x)]) \\
&= r_{3P}\mu_y + \lambda g_m(P - d) - \lambda \cdot [cx + g_m(P - x - d) + g_m(E_y[L_{3P}(x)])] \\
&= r_{3P}\mu_y + \lambda g_m(P - d) - \lambda \cdot [cx + g_m L_m(x)) + g_m(E_y[L_{3P}(x)])] \\
&= r_{3P}\mu_y + \lambda g_m(P - d) - \lambda \cdot [r_{3P}\mu_y + r_m P - \Pi(x)] \\
&= (1 - \lambda) r_{3P}\mu_y - \lambda \cdot [r_m P + g_m(P - d)] + \lambda \Pi(x) \\
&= \text{constant} + \lambda\, \Pi(x).
\end{aligned} \tag{20}$$

Some ambiguity arises when λ equals 1 or 0 and $u(\lambda) = c_{3P} + \lambda(g_m - c)$. For $\lambda = 1$, the manufacturer's profit becomes

$$\begin{aligned}
\pi_m(x) &= r_m P - (c_m + u)x - (1 - \lambda) g_m(E_y[L_{3P}(x)]) - g_m L_m(x) \\
&= r_m P - g_m x - g_m L_m(x) \\
&= r_m P - g_m(P - d),
\end{aligned}$$

i.e., constant. Therefore, x^* or any other quantity (including no subcontracting at all) is optimal for m. In order to avoid the administrative costs of subcontracting (not included in our model), the manufacturer might prefer not to subcontract. Similarly, for $\lambda = 0$, the third-party's profit function becomes

$$\begin{aligned}
\pi_{3P}(x) &= r_{3P}\mu_y - c_{3P}x + u(\lambda)x - \lambda g_m(E_y[L_{3P}(x)]) \\
&= r_{3P}\mu_y - c_{3P}x + c_{3P}x \\
&= r_{3P}\mu_y,
\end{aligned}$$

i.e., constant. Hence, any amount is optimal for $3P$ and his profit is $r_{3P}\mu_y$ which is his profit without subcontracting. In the previous section, we showed that in such a situation the $3P$ does not follow the coordinating contract, $u = c_{3P}$. Instead he offers $u > c_{3P}$ and the manufacturer subcontracts $x^*_{3P} < x^*$. Using the simplified form of $\pi_{3P}(x)$ in (20), we have the following proposition.

Proposition 2 *For a given $\lambda \in (0,1)$, contract $T(u, x, \lambda(u)) = u(\lambda) \cdot x - \lambda g_m \cdot (E_y[L_{3P}(x)])$ coordinates m and $3P$, i.e. $x^* = x^*_{3P} = x^*_m$, if $3P$ chooses unit price $u(\lambda) = c_{3P} + \lambda(g_m - c)$.*

Proof of Proposition 2: From the concavity of $\Pi(x)$, $\pi_{3P}(x)$ is also concave and hence there exists a unique maximizing x, say x^*_{3P}. Everything except $\Pi(x)$ in $\pi_{3P}(x) = (1-\lambda)r_{3P}\mu_y - \lambda \cdot [r_m P + g_m(P-d)] + \lambda\Pi(x)$ is constant, so the maximizer x^* of $\Pi(x)$ also maximizes $\pi_{3P}(x)$. $\diamond$

5.2.2 Allocation of the Chain Profits

In this subsection we investigate the profit functions of $3P$ and m, and the allocation of the chain profits when the third-party charges at the coordinating unit price. Let $\pi^0_m = r_m P - g_m(P-d)$ and $\pi^0_{3P} = r_{3P}\mu_y$ be the profits of the manufacturer and the third-party when the manufacturer does not subcontract to the third-party at all. Also let $\Delta_m(x) = \pi_m(x) - \pi^0_m$ and $\Delta_{3P}(x) = \pi_{3P}(x) - \pi^0_{3P}$ be the additional profits of m and $3P$ due to subcontracting. Then, $\Delta(x) = \Delta_{3P}(x) + \Delta_m(x)$ is the total additional profits due to centralization. We have the following proposition:

Proposition 3 *Given $\lambda \in (0,1)$ and coordinating contract $T(u, x, \lambda(u)) = u(\lambda) \cdot x - \lambda g_m \cdot (E_y[L_{3P}(x)])$, then both m and $3P$ are better-off when subcontracting occurs, i.e., $\Delta(x) > 0$ implies $\Delta_m(x) > 0$ and $\Delta_{3P}(x) > 0$.*

Proof of Proposition 3: Re-write (20) as follows:

$$\begin{aligned}
\pi_{3P}(x) &= (1-\lambda)r_{3P}\mu_y - \lambda \cdot [r_m P + g_m(P-d)] + \lambda\Pi(x) \\
&= (1-\lambda)\pi^0_{3P} - \lambda\pi^0_m + \lambda\Pi(x) \\
&= (1-\lambda)\pi^0_{3P} + \lambda \cdot [\Pi(x) - \pi^0_m] \\
&= (1-\lambda)\pi^0_{3P} + \lambda \cdot [\pi_{3P}(x) - \pi^0_{3P} + \pi^0_{3P} - \pi_M(x)\pi^0_m] \\
&= (1-\lambda)\pi^0_{3P} + \lambda\pi^0_{3P} + \lambda \cdot [\Delta_{3P}(x) + \Delta_m(x)] \\
&= \pi^0_{3P} + \lambda\Delta(x).
\end{aligned} \tag{21}$$

From (21), it holds that if $\Delta(x) > 0$, then

$$\pi_{3P}(x) - \pi^0_{3P} = \Delta_{3P}(x) = \lambda\Delta(x) > 0.$$

Consequently,

$$\pi_m(x) \quad = \pi_m^0 + (1-\lambda)\Delta(x). \tag{22}$$

If $\Delta(x) > 0$, then $\pi_m(x) - \pi_m^0 = \Delta_m(x) = (1-\lambda)\Delta(x) > 0$. This completes the proof of the proposition. $\diamond$

Proposition 3 means that, if coordination results in larger chain profits, the coordinating contract with parameters $(u(\lambda), \lambda)$ and $u(\lambda) = c_{3P} + \lambda(g_m - c)$ Pareto-dominates any non-coordinating contract. Note that for $\lambda < 1$, the manufacturer is better-off by outsourcing.

The third-party can determine his share of the additional chain profits while choosing the parameters of the contract and the quantities subcontracted by the manufacturer and $3P$ are equal to the social optimum, i.e., $x^* = x^*_{3P} = x^*_m$ (see Proposition 2). Therefore, when the manufacturer responds to $(u, \lambda(u))$ optimally, $\Pi(x) = \Pi(x^*)$, and $\Delta(x) = \Delta(x^*)$ regardless of the choice of $\lambda \in (0,1)$. Note that, $3P$ can keep all additional chain profits $\pi_{3P}(x^*) = \Pi(x^*) - \pi_m^0 = \pi_{3P}^0 + \Delta(x^*)$ as λ approaches 1. However, as λ approaches 1, m may not be willing to subcontract at all, which means $\Delta(x^*)$ becomes 0. This discussion suggests that coordination is possible with a tardiness sharing contract, if m and $3P$ jointly agree on the contract parameters $(u, \lambda(u))$ and $3P$ does not attempt to keep all additional profits. The extent to which the third-party can increase fraction λ and the extent to which the manufacturer can decrease λ depend on the negotiating power of the two parties. For a discussion of the contract parameter choices when they are negotiated separately (i.e $u \neq c_{3P} + \lambda(g_m - c)$), see Section 6.

Suppose that both parties agree on the terms depending on their individual capability of generating profits without subcontracting. Let $\hat{\lambda} = \frac{\pi_{3P}^0}{\pi_m^0 + \pi_{3P}^0}$ denote the fraction of the chain profits attained by the third-party when he does not subcontract. If $3P$ chooses contract parameters $(\hat{\lambda}, u(\hat{\lambda}))$, his profit after coordination becomes

$$\begin{aligned}
\pi_{3P}(x) \quad &= (1-\hat{\lambda})\pi_{3P}^0 - \hat{\lambda}\pi_m^0 + \hat{\lambda}\Pi(x) \\
&= (1 - \tfrac{\pi_{3P}^0}{\pi_m^0 + \pi_{3P}^0})\pi_{3P}^0 - \tfrac{\pi_{3P}^0}{\pi_m^0 + \pi_{3P}^0}\pi_m^0 + \hat{\lambda}\Pi(x) \\
&= \hat{\lambda}\Pi(x),
\end{aligned}$$

and $\pi_m(x) = (1-\hat{\lambda})\Pi(x)$. Therefore, when $\lambda = \hat{\lambda}$, the chain profits are allocated to m, $3P$ in proportion to their profits without subcontracting. A tardiness sharing contract with parameters $(\hat{\lambda}, u(\hat{\lambda}))$ is *fair* in the sense that it reflects the individual profitability of the coordinating parties.

5.2.3 Additional Chain Profits

Depending on the cost parameters of the third-party and the manufacturer, subcontracting may not be profitable. Also the availability of the third-party may be an issue depending on the distribution of the committed workload at the third-party. In what follows, we state conditions under which $\Delta(x) > 0$ is guaranteed to hold, and hence the third-party and the manufacturer can create savings by using a tardiness sharing contract with $u = c_{3P} + \lambda(g_m - c)$.

Proposition 4 *If* $\frac{d-\mu_y}{d} > \frac{c}{g_m}$ *and* m *acts optimally, then* $\Delta(x) > 0$.

Proof of Proposition 4: Note that if $u = c_{3P} + \lambda(g_m - c)$ then $x_m^* = x^*$. Also note that, $F(d - x^*) = \frac{c}{g_m}$, so we can re-write (4) as

$$
\begin{aligned}
\Pi(x^*) &= \pi_m(x^*) + \pi_{3P}(x^*) \\
&= r_{3P}\mu_y + r_m P - cx^* - g_m E_y[L_{3P}(x^*)] + g_m L_m(x^*)) \\
&= r_{3P}\mu_y + r_m P - cx^* - g_m(\mu_y + x^* - d + \textstyle\int_0^{d-x^*} F(y)dy) - g_m(P - x^* - d) \\
&= r_{3P}\mu_y + r_m P - cx^* - g_m(\mu_y - d) - g_m(P - d) - g_m \textstyle\int_0^{d-x^*} F(y)dy \\
&\geq r_{3P}\mu_y + r_m P - cx^* - g_m(\mu_y - d) - g_m(P - d) - g_m(d - x^*)F(d - x^*) \\
&= r_{3P}\mu_y + r_m P - cx^* - g_m(\mu_y - d) - g_m(P - d) - g_m(d - x^*)\tfrac{c}{g_m} \\
&= r_{3P}\mu_y + r_m P - g_m(P - d) + g_m(d - \mu_y) - g_m(P - d) - cd \\
&= \pi_{3P}^0 + \pi_m^0 + g_m(d - \mu_y) - cd.
\end{aligned}
$$

If $\frac{d-\mu_y}{d} > \frac{c}{g_m}$, then $g_m(d - \mu_y) - cd > 0$. Hence,

$$\Delta(x^*) = \Pi(x^*) - (\pi_{3P}^0 + \pi_m^0) = g_m(d - \mu_y) - cd > 0,$$

and the proof is completed. $\diamond$

The condition in Proposition 4 is sufficient but not necessary for positive expected additional profits to be generated. Then, $\frac{d-\mu_y}{d} > \frac{c}{g_m}$ has the following economic interpretation. If the expected third-party availability as a percentage of his total capacity (including the committed workload) is more than the unit cost of subcontracting relative to the unit tardiness penalty, then subcontracting results to additional savings. Obviously the manufacturer cannot observe the expected committed workload of the third-party, μ_y in advance. Even though $3P$ may have past data to forecast the expected committed workload and has knowledge of his cost structure c_{3P}, he may not observe the cost structure, c_m and g_m, of the manufacturer. Therefore, sharing information about each other's demand and cost structure is beneficial to both parties. Benefits of sharing demand forecasts in supply chains have recently been studied in the literature (see for example Cachon and Lariviere (2001) where optimal supply chain performance requires the manufacturer to share her initial forecast truthfully).

6 Tardiness Sharing Contracts with Exogenous Penalty Sharing Fraction λ

Recall that the manufacturer chooses the optimum subcontracted workload x_m^* as a response to (λ, u) offered by the third-party, that solves

$$F(d - x_m^*) = \frac{c_m + u - \lambda g_m}{(1-\lambda)g_m}.$$

Suppose that a tardiness sharing fraction λ is exogenously determined. Recall that $\frac{c_m+u-\lambda g_m}{(1-\lambda)g_m} < 0$ if $c_m + u < \lambda g_m$. Therefore, for a given λ, $\bar{u} = \lambda g_m - c_m$ is the highest unit processing fee that the third-party can offer. To satisfy $\frac{c_m+u-\lambda g_m}{(1-\lambda)g_m} < 1$, we need $c_m + u < g_m$ which is always satisfied for $u < \bar{u}$.

Recall that under a unit processing charge contract, $T(u, x) = ux$, the third-party optimally sets a unit fee, say $u_1^* > c_{3P}$, to which the manufacturer responds by subcontracting less than the chain optimum, x_1^* (see Lemma 2). Let $\pi_1(u_1(x_1^*), x_1^*)$ be the third-party's optimal profit. Generally, let $\pi_1(u_1(x), x)$ be the third-party's profit function under the unit-processing charge contract, and $\pi_2(u_2(x), x)$, u_2^*, x_2^* be defined similarly for the tardiness sharing contract $T(u_2, x, \lambda) = u_2 x - \lambda g_m E_y[L_{3P}(x)]$. Note that $u_1(x)$ and $u_2(x)$ are the unit processing fees for which m subcontracts x units of workload, i.e.

$$u_1(x) = -c_m + g_m F(d - x) \tag{23}$$

and

$$u_2(x) = -c_m + \lambda g_m + (1-\lambda) g_m F(d - x). \tag{24}$$

Then, $\pi_1(x) = r_{3P}\mu_y - cx + g_m F(d-x)x$ and $\frac{d\pi_1(u_1(x),x)}{dx} = -c + g_m \cdot [F(d-x) - f(d-x)x]$ (see Section 5.1). Similarly,

$$\begin{aligned}
\pi_2(u_2(x), x) &= r_{3P}\mu_y - c_{3P}x + u_2(x)x - \lambda g_m E_y[L_{3P}(x)] \\
&= r_{3P}\mu_y - c_{3P}x + [-c_m + \lambda g_m + (1-\lambda) g_m F(d-x)] \cdot x - \lambda g_m E_y[L_{3P}(x)] \qquad (25) \\
&= r_{3P}\mu_y + (\lambda g_m - c)x + (1-\lambda) g_m F(d-x)x - \lambda g_m E_y[L_{3P}(x)].
\end{aligned}$$

Note that $\pi_2(u_2(x), x)$ is twice-differentiable on $[0, \infty)$ if F is differentiable on $[0, \infty)$. Then,

$$\begin{aligned} \frac{d\pi_2(u_2(x),x)}{dx} &= -c + \lambda g_m + (1-\lambda)g_m \cdot [F(d-x) - f(d-x)x] + \lambda g_m \cdot [1 - F(d-x)] \\ &= -c + (1-\lambda)g_m \cdot [F(d-x) - f(d-x)x] + \lambda g_m F(d-x) \\ &= -c + (1-\lambda)g_m F(d-x) \cdot [1 - \tfrac{f(d-x)x}{F(d-x)}] + \lambda g_m F(d-x). \end{aligned} \tag{26}$$

We have already shown that $\frac{f(d-x)x}{F(d-x)}$ is increasing in x and obviously $F(d-x)$ is decreasing in x. Therefore $\frac{d\pi_2(u_2(x),x)}{dx}$ is decreasing, which means $\pi_2(u_2(x), x)$ is concave for $x \in [0, d)$. Hence,

Lemma 4 *If F is strictly increasing on $[0, \infty)$, then $x_2^* < x^*$.*

Proof of Lemma 4: As $\pi_2(u_2(x), x)$ is continuous and concave, there must be a unique x_2^* for which $\frac{d\pi_2(u_2(x),x)}{dx} = 0$. At the supply chain optimum, $x = x^*$, we have $F(d-x) = \frac{c}{g_m}$. Replace $F(d-x)$ with $\frac{c}{g_m}$ in $\frac{d\pi_2(u_2(x),x)}{dx}$ to get

$$\begin{aligned} \frac{d\pi_2(u_2(x),x)}{dx}\Big|_{x=x^*} &= -c + (1-\lambda)g_m \cdot [F(d-x^*) - f(d-x^*)x^*] + \lambda g_m F(d-x^*) \\ &= -c + (1-\lambda)g_m \cdot [\tfrac{c}{g_m} - f(d-x^*)x^*] + \lambda g_m \tfrac{c}{g_m} \\ &= -c - (1-\lambda)g_m \cdot f(d-x^*)x^* < 0 \end{aligned}$$

which implies that $\pi_2(u_2(x), x)$ is decreasing at $x = x^*$. From the uniqueness of x_2^* and the concavity of $\pi_2(u_2(x), x)$, we have $x_2^* < x^*$. $\diamond$

Consider contract $T(u_1, x) = u_1 x$ for which $3P$ charges fee u_1^* and m responds with workload x_1^*. Also, consider tardiness sharing contract $T(u_2, x, \lambda)$ where $\lambda < \bar{\lambda}$ is exogenously determined, for which $3P$ charges u_2^* and m responds with x_2^*. Then,

Lemma 5 *We have $x_1^* < x_2^* < x^*$.*

Proof of Lemma 5: We have already shown that $x_1^* < x^*$ (Lemma 2) and that $x_2^* < x^*$ (Lemma 4). Thus, we only need to show that $x_1^* < x_2^*$. Note that

$$\begin{aligned} \frac{d\pi_1(u_1(x),x)}{dx} &= -c + g_m \cdot [F(d-x) - f(d-x)x] \\ &= -c + g_m F(d-x) - g_m f(d-x)x \end{aligned}$$

and

$$\begin{aligned} \frac{d\pi_2(u_2(x),x)}{dx} &= -c + (1-\lambda)g_m \cdot [F(d-x) - f(d-x)x] + \lambda g_m F(d-x) \\ &= -c + g_m F(d-x) - (1-\lambda)g_m f(d-x)x. \end{aligned}$$

It immediately follows that

$$\frac{d\pi_1(u_1(x),x)}{dx} < \frac{d\pi_2(u_2(x),x)}{dx}, \text{ for } x \in [0, d).$$

Therefore, at $x = x_1^*$ where $\frac{d\pi_1(u_1(x_1^*),x_1^*)}{dx} = 0$, $\frac{d\pi_2(u_2(x_1^*),x_1^*)}{dx}$ is positive which implies that $\pi_2(u_2(x), x)$ is still increasing at $x = x_1^*$. From the concavity of $\pi_2(u_2(x), x)$, we have $x_1^* < x_2^*$. ⋄

The above results are used in the next subsection to evaluate the efficiency of contracts $T(u_1, x) = u_1 x$ and $T(u_2, x, \lambda) = u_2 x - \lambda g_m E_y[L_{3P}(x)]$.

6.1 Performance Measures of Contracts

When a contract does not coordinate the chain, various measures are proposed in the literature to evaluate its performance. One such measure (see Lariviere and Porteus (2001)) is the profit of the offering party (in our case the third-party) as a fraction of the total chain profit of the non-coordinating outcome, i.e.

$$\frac{\pi_{3P}(u(x_{3P}^*), x_{3P}^*)}{\Pi(x_{3P}^*)}, \tag{27}$$

where x_{3P}^* is the optimal subcontracted workload when the third-party offers a contract with parameter $u(x_{3P}^*)$. If this ratio is high, then the contract is generally an attractive option for the offering party. However, a situation where the total chain profit differs appreciably from the centralized optimum is not desirable. For instance, suppose that at the chain optimum the parties make \$1000 collectively, while at the third-party optimum $x = x_{3P}^*$ the chain makes \$500, 90% of which (\$450) is captured by the third-party. The third-party certainly prefers another contract where the chain makes \$800, 70% of which (\$560) is captured by him. Therefore, another measure referred to as the *efficiency* of the contract is also considered, which is the ratio of the chain profit at the third-party optimum to the optimal total chain profit, i.e.,

$$\frac{\Pi(x_{3P}^*)}{\Pi(x^*)}. \tag{28}$$

The *efficiency* of the contract measures the extent to which the chain comes close to the centralized chain performance. Using a high efficiency contract is not only preferred by the third-party, but also by the entire chain as it reduces the *cost of decentraliztion* (see Cachon (2003)). We have the following result.

Proposition 5 *Consider contracts $T(u_1, x) = u_1 x$ and $T(u_2, x, \lambda) = u_2 x - \lambda g_m E_y[L_{3P}(x)]$ where $\lambda < \bar{\lambda}$ is exogenously determined, with response strategies u_1^*, x_1^* and u_2^*, x_2^* for $3P$ and m for each contract, respectively. The efficiency of the tardiness sharing contract is greater than the efficiency of the unit processing charge contract, i.e.*

$$\frac{\Pi(x_2^*)}{\Pi(x^*)} > \frac{\Pi(x_1^*)}{\Pi(x^*)}.$$

Proof of Proposition 5: We know that $x_1^* < x^*$ (Lemma 2), $x_2^* < x^*$ (Lemma 4), and $x_1^* < x_2^*$, (Lemma 5). From the concavity of $\Pi(x)$, we conclude that $\Pi(x)$ is increasing for $x < x^*$ (see Figure 2). Therefore,

$$\Pi(x_1^*) < \Pi(x_2^*),$$

and hence

$$\frac{\Pi(x_2^*)}{\Pi(x^*)} > \frac{\Pi(x_1^*)}{\Pi(x^*)}.$$

This completes the proof. $\diamond$

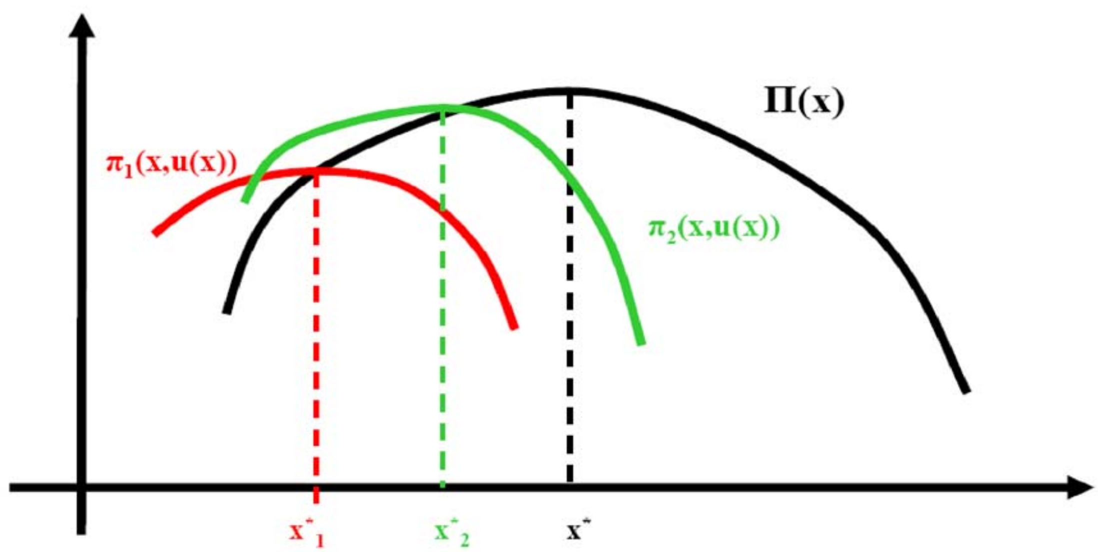

Figure 2: The profit functions $\Pi(x)$, $\pi_1(x, u(x))$, and $\pi_2(x, u(x))$

7 Conclusions

In this article, we considered a subcontracting model where a manufacturer who cannot process all his workload before a given due date subcontracts part of his workload to a third-party who is also committed to another customer. Order quantities due to this customer follow a general distribution and hence the available capacity at the third-party is uncertain. The manufacturer wants to maximize his expected profits by partitioning his workload among his in-house production capacity and the third-party. We studied two different contracts; (i) a unit processing charge contract, and (ii) a tardiness penalty sharing contract. We showed that coordination is possible under the second contract if the parameters are negotiated simultaneously. Then, the resulting contract Pareto-dominates all non-coordinating contracts.

In supply chain contracting literature, similar coordination results have been obtained for *whole-sale price contracts* (Lariviere and Porteus (2001)), *buy-back contracts* (Pasternack (1985)) and *revenue-sharing contracts* (Cachon and Lariviere (2003)). Our findings point to significant differences from the above literature when the manufacturer has adequate production capability and he uses the option of subcontracting only when it is profitable. In the whole-sale price contract, the supplier can coordinate the chain only when he sets the price less than his marginal cost. In our model, he can coordinate the chain by setting a unit processing fee equal to marginal cost. In case the third-party has high fixed costs (not considered in our model), chances of agreeing on a non-coordinating simple to administer contract are higher. Also, similar to the buy-back or the revenue-sharing contracts, our tardiness penalty sharing contract can coordinate the chain and allocate the additional rents arbitrarily. In addition, the contract parameter λ reflects the fraction of the additional profits the third-party (the offering party) captures under a coordinating contract. This is a significant administrative advantage for the third-party. On the other hand, it makes it difficult for the third-party to resist manipulating the parameters so as to reap all the additional chain profits (which is possible with $\lambda = 1$). But then, the manufacturer will not agree upon a contract where the third-party explicitly states that he will have the entire additional chain profits.

Our analyses include the case where the contract parameters are negotiated separately. Despite its administrative complexity, the tardiness penalty sharing contract proved superior to the unit processing charge contract because when λ is chosen exogenously, the tardiness penalty sharing contract has a higher *efficiency* than that of the unit processing charge contract.

To the best of our knowledge, this article is the first effort to study contracts in the context of manufacturing scheduling operations. We think that this is a fruitful research direction and that the analysis of the multiple manufacturers case will further verify the value of subcontracting. In the multiple manufacturers case, there will be competition for earlier capacity usage at the third-party. Conflicting interests of the manufacturer and the third-party will create a capacity allocation subproblem. Similarly, the case with multiple third-parties is also worth considering as it leads to more subcontracting choices for the manufacturer.

References

[1] Ansberry, C. and Aeppel, T. Battling Imports: Surviving the Onslaught; U.S. Companies Custimize, Rethink Strategies to Compete With Products from Abroad, *The Wall Street Journal*, New York, NY, October 6, 2003.

[2] Anupindi, R. and Bassok, Y. *in S. Tayur, R. Ganeshan, and M. Magazine eds. 'Quantitative Models for Supply Chain Management'*, Kluwer, Dordrecht, Netherlands, 1998.

[3] Aydinliyim, T. and Vairaktarakis, G. L. Coordination of Outsourcing Operations, *Manufacturing & Service Operations Management*, under review, 2006.

[4] Bernstein, F. and Federgruen, A. Decentralized Supply Chains with Competing Retailers Under Demand Uncertainty, *Management Science*, 51 (1), 18-29, 2005.

[5] Cachon, G. P. Supply Chain Coordination with Contracts. *in S. Graves, T. de Kok, eds. 'Handbooks in Operations Research and Management Science: Supply Chain Management'*, North-Holland, New York, NY, Chapter 6, 2003.

[6] Cachon, G. P. and Lariviere, M. Contracting to Assure Supply: How to Share Demand Forecasts in a Supply Chain, *Management Science*, 47 (5), 629-646, 2001.

[7] Cachon, G. P. and Harker, P. T. Competition and Outsourcing with Scale Economies, *Management Science*, 48 (10), 1314-1333, 2002.

[8] Cai, X.Q. and Vairaktarakis, G. L. Cooperative Strategies for Manufacturing Planning with Negotiable Third-Party Capacity, *Management Science*, under revision, 2007.

[9] Calleja, P., Estévez-Fernández, M. A., Borm, P., Hamers, H. Job Scheduling, Cooperation and Control, *CentER Discussion Papers*, Tilburg University, Tilburg, The Netherlands, 2004-65.

[10] Curiel, I., Pederzoli, G. and Tijs, S. Sequencing Games, *European Journal of Operational Research*, 40, 344-351, 1989.

[11] Curiel, I., Hamers, H. and Klijn, F. Sequencing Games: A Survey, *in P. Borm and H. Peters, eds. 'Chapters in Game Theory: In Honor of Stef Tijs'* Kluwer Academic Publishers, Boston, pp. 27-50, 2002.

[12] Elmaghraby, W.J. Supply Contract Competition and Sourcing Policies, *Manufacturing & Service Operations Management*, 2, 350-371, 2000.

[13] Greaver, M. F. Strategic Outsourcing, AMA Publications, New York, NY, 1999.

[14] Hain, R. and Mitra, M. Simple Sequencing Problems with Interdependent Costs, *Games and Economic Behavior*, 48, 271-291, 2004.

[15] Kamien, M. I. and Li, L. Subcontracting, Coordination, Flexibility, and Production Smoothing in Aggreagate Planning, *Management Science*, 36 (11), 1352-1363, 1990.

[16] Lariviere, M. Supply Chain Contracting and Co-ordination with Stochastic Demand. *in S. Tayur, R. Ganeshan, and M. Magazine eds. 'Quantitative Models for Supply Chain Management'*, Kluwer, Dordrecht, Netherlands, 1998.

[17] Lariviere, M. and Porteus, E. Selling to the Newsvendor: An Analysis of Price-Only Contracts, *Manufacturing & Service Operations Management*, 3(4), 293-305, 2001.

[18] Li, L. The Role of Inventory in Delivery-Time Competition, *Management Science*, 38 (2), 182-197.

[19] McMillan, J. Managing Suppliers: Incentive Systems in Japanese and U.S. Industry, *California Management Review*, Summer, 1990.

[20] Nash, J.F. Non-Cooperative Games. *Annals of Mathematics*, 54, 286-295, 1951.

[21] Pasternack, B. Optimal Pricing and Returns Policies for Perishable Commodities, *Marketing Science*, 4 (2), 166-176, 1985.

[22] Pinedo, M. *Scheduling: Theory, Algorithms and Systems*, Prentice Hall, Englewood Cliffs, NJ, 2002.

[23] Shapley, L. Cores of Convex Games, *International Journal of Game Theory*, 1, 11-26, 1971.

[24] Taylor, T. Coordination Under Channel Rebates with Sales Effort, *Management Science*, 48 (8), 992-1007, 2002.

[25] Tsay, A. Quantity-Flexibility Contract and Supplier-Customer Incentives, *Management Science*, 45 (10), 1339-1358, 1999.

[26] Vairaktarakis, G. L. Non-Cooperative Outsourcing Games, *Manufacturing & Service Operations Management*, under review, 2007.

[27] Vairaktarakis, G. L. and Aydinliyim, T. Competition vs. Centralziation in Subcontracting Operations, *Management Science*, under review, 2007.

[28] Van Mieghem, J. A. Coordinating Investment, Production and Subcontracting, *Management Science*, 45 (7), 954-971, 1999.