MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 6, No. 1, Winter 2004, pp. 53–72
ISSN 1523-4614 | EISSN 1526-5498 | 04 | 0601 | 0053

INFORMS
DOI 10.1287/msom.1030.0029

On Measuring Supplier Performance Under Vendor-Managed-Inventory Programs in Capacitated Supply Chains

Ki-Seok Choi
Samsung SDS Co., Ltd., 707-19 Yoksam-2Dong, Kangnam-Gu, Seoul, 135-918, Korea, kiseok2003.choi@samsung.com

J. G. Dai
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, dai@isye.gatech.edu

Jing-Sheng Song
The Fuqua School of Business, Duke University, Durham, North Carolina 27708, jssong@duke.edu

As widely accepted performance measures in supply chain management practice, frequency-based service levels such as fill rate and stockout rate are often considered in supply contracts under vendor-managed-inventory (VMI) programs. Using a decentralized two-party capacitated supply chain model consisting of one manufacturer and one supplier in a VMI environment, we demonstrate that supplier's service level is in general insufficient for the manufacturer to warrant the desired service level at the customer end. The method by which the supplier achieves her service level to the manufacturer also affects customer service level.

By developing bounds on the customer service level, we show that the expected backorders at the supplier should also be taken into account. We suggest a supply contract that offers a menu of different combinations of supplier's service level and expected backorders according to a linear function. Under this contract, the manufacturer can control the end customer service regardless of how the supplier manages her inventory. The supplier has complete flexibility on which combination of the two quantities on the menu to choose according to her own cost functions. Because it does not require any detailed information on supplier's operational characteristics nor her costs, this kind of contract is expected to be easily implementable. In addition, we derive an estimate of the customer service level in terms of the new measures.

Our findings have direct implications to supply chain metrics in general: The local service levels are insufficient measures to guarantee the system wide performance. Alternative local measures and/or coordination mechanisms should be employed to achieve desired system performance. Our analysis illustrates a possible way to explore such alternative measures.

Key words: service-level guarantees; supply contracts; supplier performance measures; vendor-managed inventory
History: Received: April 11, 2003; Senior Editor: Kaj Rosling.

1. Introduction

Frequency-based customer service measures, commonly known as service levels, such as fill rate and stockout rate, have always been important performance indicators that all companies care about. This is reflected by their repeated appearance in business press, company Web sites, and business advertisements across industries. (See various examples cited in Sobel 2002. See also Kleijnen and Smits 2002.) While the issue of how to manage a centralized supply chain with a target customer service level has received much attention, especially in recent years, to our knowledge how to coordinate different parties in a decentralized supply chain to achieve a target end customer service level has not been sufficiently addressed in the literature. For example, when different parties are involved in a supply chain, a common belief is that as long as the immediate upstream party has a higher service level than the desired downstream party's target service level, the downstream party would be able to manage his operations to meet his target. However, this rule of thumb does not appear to have any theoretical justification. Also, there exists no guideline as to how much higher the upstream party's service level should be. This paper explores these issues using a simple two-party capacitated supply chain consisting of one manufacturer and one supplier.

We find that the above-mentioned rule of thumb is not valid. We provide an example in which the supplier meets her pre-agreed service level such as a low stockout rate, yet the manufacturer is not able to meet his customer service level. The method by which the supplier achieves her service level to the manufacturer—such as increasing inventory, improving production reliability, or increasing production capacity—also affects the customer service level. This example also indicates that any contract that is primarily based on supplier's service level is flawed. Our analysis suggests an alternative contract form that contains a menu of two easily observable supplier delivery performance measures: the supplier's stockout rate *and* expected backorders. When the supplier meets these performance measures, the manufacturer is guaranteed to meet his customer service level. This type of contract leaves the supplier complete flexibility on which combination of the two quantities on the menu to choose according to her own cost functions. Because it does not require any detailed information on supplier's operational characteristics nor her costs, this kind of contract is expected to be easily implementable.

Our research was motivated by the questions faced by the managers of electronic manufacturing service providers, such as Solectron and Flextronics, through our industry experiences with them. In the last decade or so, in response to increasingly shorter product life cycles, higher customer expectations, and fierce global competition, the electronics industry has experienced extensive growth and restructuring, and its supply chains have become highly decentralized. To improve supply chain efficiency, the industry has adopted two successful innovations. One is the make-to-order (MTO), also known as build-to-order, manufacturing strategy, led by Dell Computer in the computer industry (see Simchi-Levi et al. 2000). Under MTO, the manufacturer does not keep finished-product inventory and releases production orders only after receiving customer orders. This approach eliminates the risk of wasted investment in unwanted finished-product inventory due to technological obsolescence or unmatched demands. Another innovation is the vendor-managed-inventory (VMI) programs, led by Wal-Mart and Procter & Gamble in the retail industry and by Campbell Soup in the grocery industry (see Buzzell and Ortmeyer 1995 and Fisher 1997) and followed by other industries (see Thompson and Strickland 2001 and Barnes et al. 2000). Under VMI in the supplier-manufacturer setting, the supplier is responsible for all decisions regarding the component inventory at the manufacturer.

Despite the success and potential benefits of MTO and VMI, companies that are adopting these business models often face complex execution challenges. For example, to ensure satisfactory order fulfillment performance, MTO calls for more critical component inventory management than ever, which in turn requires closer and improved relationships with suppliers. A common issue is how to monitor the supplier's performance so as to guarantee a satisfactory service level at the customer end. Without exception, the manufacturer would like to have components available whenever he needs them, but this would cost the supplier too much to be feasible.

A solution to this dilemma is often a contract between the manufacturer and the supplier, in which both parties agree on certain requirements regarding the supplier's performance in component delivery. The requirements can be expressed in various ways. To simplify the monitoring process, which is often done by third-party companies, measurable quantities are preferred. As widely adopted performance measures in supply chain practice, service levels such as fill rate and stockout rate, which quantify how much and/or how soon downstream orders are delivered, easily come to managers' minds as desirable candidates. Such measures are indeed being used under VMI contracts in various companies; see, e.g., Fry et al. (2001).

However if the manufacturer tries to control his manufacturing process to achieve certain customer service level, how can he accurately quantify the service level to specify in a supply contract, and how does that specified service level affect the end customer service level? These are the questions asked by the managers.

We address these questions by considering a decentralized two-party supply chain consisting of one manufacturer and one supplier. The manufacturer has a finite capacity and makes a product to serve market demand following an MTO policy. The supplier provides components for the product and manages the component inventory at a warehouse near or at the manufacturer's site under a VMI program. In the general model, we do not impose any assumptions about the supplier's operational characteristics, such as her capacity and inventory policy. The two parties are connected through component requirement and fulfillment. The manufacturer can influence the supplier only through requirements on the supplier's component delivery performance.

We examine several commonly used service level measures, and demonstrate the irrelevance of supplier's service level to customer service level. For instance, in §4 we consider a detailed supplier model in which the supplier can manipulate three parameters in order to achieve her service level, say a stockout rate no more than 5%, specified by the manufacturer. The three parameters are excess capacity, production reliability (yield), and base-stock level. We show that many different choices of these parameters can result in the desired 5% service level. However, the resulting stockout rate at the end customer level demonstrates a wide range and sometimes can be as high as 34.46%. In other words, the customer service level depends not only on the quantitative measure of the service level from the supplier, but also on *how* the supplier achieves that service level. A major contributor to the fluctuating customer service is the manufacturer's capacity. These findings imply the necessity of other supplier performance measures that can secure the manufacturer's customer service.

By developing bounds on the customer service level, we identify plausible supplier performance measures that allow the manufacturer to control the end customer service regardless of how the supplier manages the inventory. The results suggest that in addition to the service level, we also need to measure the average component shortage or the component backlogs. We propose a new contract form that consists of a menu of combinations of these two quantities. The design of the menu is through a linear function of the two quantities that guarantee the manufacturer's customer service level. The parameters of the linear function depend on the manufacturer's capacity and demand distribution only. Therefore, the contract design is independent of the supplier's operating characteristics and cost information. The supplier can choose any combination of the two quantities listed on the menu, which leaves the supplier complete flexibility in optimizing her own cost. Thus, this type of contract is "detail free" (Wilson 1987), so it is expected to be easily implementable. Moreover, the development of the bound (the linear function) assumes no information on the supplier's operating characteristics, and hence the result is robust.

Our results have direct implications to supply chain performance metrics in general. That is, the local (or internal) service levels in a supply chain are insufficient measures to guarantee the entire supply chain's service level to its end customer. (It is worth mentioning that this observation is true even for uncapacitated supply chains but with positive transportation leadtimes between stages.) Thus, alternative local measures and/or coordination mechanisms should be employed in order to achieve desired system performance. Although individual researchers may have noticed or suspected the ineffectiveness of local service levels in supply chain coordination, to our knowledge, our paper provides the first documented study to bring the awareness of this issue. Our study also goes a step further by demonstrating that how the supplier achieves the prespecified service level also affects the performance of the downstream party, and thus provides deeper insights.

Developing performance measures solely from what can be observed from supplier's output process, without knowing the supplier's operating characteristics, presents a tremendous technical challenge. As such, the analysis should be, in general, system specific. We hope that the approach we take here—using bounds on the downstream service level—provides some encouragement and inspiration for future research endeavor along this line.

Finally, we note that while VMI has been discussed in numerous papers, our paper appears to be the first to examine the environment in which the downstream party is a capacitated manufacturer. The majority of the VMI literature to be reviewed in the next section focuses on a supplier-retailer setting, so there is no downstream capacity issue. In such a setting, the supplier's performance can be easily aligned with the customer service level: because customers are served directly from the supplier managed inventory, the supplier's service level is precisely the customer service level. Because this is no longer true in a capacitated supply chain as demonstrated in our paper, it is reasonable to conjecture that the results obtained in the supplier-retailer setting need to be reexamined in the supplier-manufacturer setting.

The remainder of the paper is organized as follows. Section 2 reviews the literature. Section 3 describes a general model and provides some preliminary discussions on the relationship between the supplier's and the manufacturer's service levels. Section 4 analyzes several special cases of the general model and demonstrates the ineffectiveness of using service levels to measure the supplier's performance. Section 5, focusing on the general model again, derives alternative supplier performance measures. Section 6 provides an example on how to design a supply contract using the alternative measures. We end the paper with a few concluding remarks, including some discussions on other possible supplier coordination mechanisms.

2. Literature Review

We first review the literature on VMI. Plambeck and Zenios (2003) study a make-to-stock model, in which the manufacturer bears the inventory-holding and backorder costs of the finished good but delegates the production of the finished good to a supplier. The supplier dynamically controls the production rate and incurs a convex production cost. The manufacturer cannot monitor the production rate, but can draw inference from increases in the inventory level. By making payments contingent on the inventory level, the manufacturer motivates the supplier to control the production rate in a manner that will minimize the manufacturer's total expected discounted cost. They show that the optimal incentive payment scheme consists of piece rates and inventory penalties that vary dynamically with the inventory level. This scheme coordinates the system if the supplier is risk neutral. Otherwise operational performance is degraded by the conflict in incentives between manufacturer and supplier. Their model setting is different from ours in several ways. First, in our model the manufacturer controls the finished-good inventory, while the supplier controls the component inventory, so there are two stocking positions in the supply chain instead of 1. Second, because the manufacturer, in their model, delegates the production to the supplier, there is no capacity issue at the manufacturer; whereas in our model, the manufacturer has a finite production capacity. Third, they assume the manufacturer tries to minimize total discounted costs, which includes backorder costs, while we do not consider backorder cost explicitly but assume that the manufacturer tries to achieve certain service level.

Fry et al. (2001) consider the (z, Z)-type VMI contract in a one supplier, one retailer supply chain: The retailer sets a minimum inventory level z and a maximum inventory level Z, and the supplier is agreed to pay a penalty to the retailer for every unit of retailer's inventory that is outside this band after customer demand. Both parties know the retailer's demand distribution. The supplier produces every T periods with no capacity limit. It also has the option of outsourcing in order to maintain the desired retailer's inventory level. The supplier's decisions are thus how much to produce in each production cycle, how much to outsource, and how much to send to the retailer in each period. With the outsourcing option (so that the supplier can always supply what is needed at the retailer), the retailer's problem becomes a single-location inventory problem, whose backorder costs influence the supplier's costs. Recall that our model treats a two-location inventory problem. In their paper, there is also no capacity issue.

As indicated by these authors, in all the VMI agreements they observed in practice, "the penalties are not incurred immediately (i.e., on a daily basis), but are based on long-term (approximately yearly) performance, often as part of 'balanced scorecard' evaluation." The service level considered in our paper measures long-term performance.

Cachon (2001) studies how to achieve channel coordination in a one-supplier, multi-retailer competitive supply chain using VMI. Both the supplier and the retailers incur inventory and backorder costs. Cachon shows that VMI is not guaranteed to coordinate the chain unless all members are willing to accept or pay fixed transfer payments. A numerical study shows that VMI provides no improvement in supply chain costs when fixed transfer payments are forbidden. Narayanan and Raman (2002) examine a retailer and a supplier under a newsvendor setting. The retailer carries a private label product that is a substitute to the product he carries from the supplier. Thus, the cost associated with a stockout is different for the supplier and the retailer, and consequently their target fill rates are different. They derive conditions under which stocking decisions should be transferred from retailer to supplier (VMI). Clark and Hammond (1997) and Cachon and Fisher (1997) study the issue of whether VMI coupled with information sharing provides greater benefits than information sharing alone. Bernstein and Federgruen (2003) analyze the constant-demand-rate case and consider a model of VMI where the replenishment decision is transferred to the supplier, but the retailer is able to make his own pricing decisions. Other papers on VMI study logistics issues; Fry et al. (2001) provide an excellent review. Our study here has a different focus from these works.

We now review the literature on supply chains with particular concerns on service levels. Much of this literature has dealt with the problem of achieving a target service level at the most downstream stage in a centralized supply chain (see van Houtum et al. 1996 and Diks et al. 1996 for literature reviews). Most authors focus on how to coordinate the elements in the supply chain to achieve a system-wide target. Operational decisions on production, distribution, and/or inventory control at each stage are coordinated by a central planner. Under a VMI program, however, the manufacturer and the supplier are separate organizations, so each party in the supply chain makes its own operational decisions. Usually, the supply chain parties can affect each other's operations only by specifying requirements on observable measures in a contract; 1 cannot dictate how the other accomplishes those requirements. The manufacturer, for instance, may include in a supply contract service performance requirement regarding component supply but not detailed inventory policies.

Several authors adopt a decomposition approach (Bollapragada et al. 2000, Cohen and Lee 1988, Lee and Billington 1993, Paschalidis and Liu 2003). Instead of centralizing the entire operation in the supply chain to achieve a target system service level, a *local* service level target is set for each stage of the supply chain. These local targets work as links between stages so that the entire system accomplishes the target service level.

The decomposition method is relevant to the VMI setting we consider, because defining local service targets can be viewed as a way of specifying service requirements in the supply contract. When it comes to how to define local targets, however, the existing literature provides no notable results. Cohen and Lee (1988) simply assume that they are given while Lee and Billington (1993) use a simple search heuristic to find the best local targets, assuming each stage follows a base-stock policy. In Bollapragada et al. (2000) and Paschalidis and Liu (2003), the service level of the supplier is set to be greater than or equal to the target customer service level. This rule-of-thumb is based on a common belief that the downstream service level is guaranteed regardless of how the upstream service level is achieved, as long as the upstream service level is high enough. In this paper, we show that this belief is not necessarily valid.

Several researchers have studied fill rates in centralized capacitated serial systems under modified base-stock policies; see, for example, Glasserman (1997) and Sobel (2002). However, these works focus on the system fill rate and do not discuss local fill rates and their relationship with the system fill rate.

3. Model and Preliminaries

We consider the following model: There is a single manufacturer who makes a product to order (we relax this assumption later), and there is a single supplier who provides components for the product. The demands for the product in different periods are independent and identically distributed random variables. One unit of component is used to produce one unit of the product. The supplier manages the component inventory at a warehouse near or at the manufacturer's site under a VMI program. The manufacturer has no knowledge about the supplier's capacity or inventory policy. Periodic-review systems are used to control inventory and production at both the manufacturer and the supplier.

At the beginning of each period, after observing customer demand for the period, the manufacturer decides production quantity considering demand and production capacity. Then it retrieves components from the component inventory. If there are not enough components in inventory, the manufacturer reduces production quantity to the number of available components. At the end of the period, the product is delivered to the customer and the supplier restocks the component inventory according to its inventory policy. When the manufacturer is not able to deliver the entire customer demand due to capacity limit or component shortage, the unsatisfied portion of the demand is backordered.

The following notation is used throughout the paper.

D_t: demand for period t,

B_t: backorders of the product at the beginning of period t, $B_1 = 0$,

I_t: component inventory level at the beginning of t before retrieval by the manufacturer,

c: production capacity of the manufacturer in a period,

P_t: number of the product manufactured in period t,

R_t: number of the component requested in period t,

Q_t: number of the component requested but not available in period t.

For any stochastic process $\{X_t,\ t = 1, 2, \ldots\}$ with a stationary distribution, denote X to be a random variable that follows the stationary distribution.

There are several definitions of service levels, common in both industry practice and the academic literature; see, for example, Schneider (1981). One of the most popular measures is the α-type service level, which measures the likelihood of stockout. In particular, let α be the long-run fraction of periods that has stockout. Then $1 - \alpha$ is called the α-type service level. The α-type service level of the manufacturer in our model is

$$\begin{aligned} 1 - \alpha_m &= 1 - \text{long-run fraction of periods with demand backorders} \\ &= 1 - \overline{\lim_{T\to\infty}} \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{B_t > 0\}}. \end{aligned}$$

Similarly, the α-type service level of the supplier is

$$\begin{aligned} 1 - \alpha_s &= 1 - \text{long-run fraction of periods with component stockout} \\ &= 1 - \overline{\lim_{T\to\infty}} \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{Q_t > 0\}}. \end{aligned}$$

Defined using upper limits, α_m and α_s take into account even the case where $(1/T)\sum_{t=1}^{T} \mathbf{1}_{\{B_t>0\}}$ or $(1/T)\sum_{t=1}^{T} \mathbf{1}_{\{Q_t>0\}}$ does not have a limit. (When $\{B_t\}$ and $\{Q_t\}$ have a stationary distribution and satisfy a strong law of large numbers, $\alpha_m = \mathsf{P}\{B > 0\}$ and $\alpha_s = \mathsf{P}\{Q > 0\}$ with probability 1, where B and Q are the corresponding stationary random variables. See §4 for an example.)

Let β be the *unfill rate*, which is the long-run proportion of demands that cannot be fulfilled immediately. It can be shown that β equals the ratio of the expected backorders to the expected demand. Then $1 - \beta$ is the β-type service level, and its corresponding expression for the manufacturer is

$$\begin{aligned} 1 - \beta_m & \\ &= 1 - \frac{\text{expected unsatisfied demand per unit time}}{\text{expected demand per unit time}} \\ &= 1 - \frac{\mathsf{E}[U]}{\mathsf{E}[D]}, \end{aligned}$$

where $U_t = \min\{B_{t+1}, D_t\}$ is the unsatisfied demand out of D_t.

While β_m uses the new backorders incurred in the current period, γ_m uses the cumulative backorders up to a period.

$$\begin{aligned} 1 - \gamma_m &= 1 - \frac{\text{expected cumulative unsatisfied demand per unit time}}{\text{expected demand per unit time}} \\ &= 1 - \frac{\mathsf{E}[B]}{\mathsf{E}[D]}. \end{aligned}$$

The β- and γ-type service levels of the supplier in our model are identical because unsatisfied component requirement is not backlogged.

$$1 - \beta_s = 1 - \gamma_s = 1 - \frac{\mathsf{E}[Q]}{\mathsf{E}[R]}.$$

For simplicity in exposition, most of this paper focuses on the α-type service level, but we remark that the major findings hold for the other two types of service levels as well and provide more detailed discussions whenever possible.

We assume the manufacturer follows a modified base-stock policy to control its production. That is, in each period the manufacturer produces as much as possible, within the manufacturing capacity, so as to keep the inventory level as close as possible to a target base-stock level. This type of policy has been shown to be optimal for the manufacturer to minimize the long-run average inventory-backorder cost, provided the demand is stationary and there is infinite component supply (i.e., the only restriction of the production level is the manufacturer's own capacity limit). See Federgruen and Zipkin (1986). More recently, Parker and Kapuscinski (2001) show that this kind of policy is also optimal for both finite-horizon and infinite-horizon problems if the manufacturer's capacity does not exceed the supplier's capacity. Because we consider an MTO manufacturer here, there is no finished-goods inventory; the base-stock level is set at zero.

Let R_t be the planned production quantity for period t, then this policy implies

$$R_t = \min\{B_t + D_t, c\}. \tag{1}$$

Recall that P_t is the actual production quantity in period t. So, Q_t, the difference between the required and available component quantities, can be expressed as

$$Q_t = R_t - P_t.$$

The number of product backorders at the end of period t is then

$$\begin{aligned} B_{t+1} &= B_t + D_t - P_t \\ &= B_t + D_t - R_t + Q_t \\ &= Q_t + (B_t + D_t - c)^+, \end{aligned} \tag{2}$$

where $(x)^+ = \max\{x, 0\}$.

Equation (2) provides some preliminary insights into how the supplier's performance is related to the customer service level $1-\alpha_m = \mathsf{P}\{B=0\}$. There are two conditions for B_{t+1} to be 0. First, Q_t should be 0. Component stockout immediately causes backorders of customer demand and thus α_m is at least as much as α_s. The second condition for $B_{t+1}=0$ from (2) is $B_t + D_t \le c$. A recursive expansion of B_t changes the condition to $Q_{t-1} + (B_{t-1} + D_{t-1} - c)^+ + D_t \le c$. From this inequality, it is clear that not only the frequency (the first condition, whether $Q_t > 0$ or not) but also the amount of component shortage (Q_{t-1} in the second condition) has influence on α_m. If this quantity is large, due to the capacity limit c, it can take many periods for the manufacturer to clear a large amount of backorders, which would result in a high value of α_m.

Of course, Q_t depends on the supplier's operational characteristics. To see this, we assume for the moment that the supplier produces components in a manufacturing facility whose capacity is $\{V_t\}$. There is ample upstream supply, so the supplier's production quantity is solely constrained by her own production capacity. Similar to the manufacturer, the supplier follows a revised base-stock policy to control the component inventory. Let s be the target component base-stock level. Then the actual component inventory level $\{I_t\}$ is updated as follows:

$$I_t = \min\{s, I_{t-1} - P_{t-1} + V_{t-1}\}. \tag{3}$$

Because

$$P_t = \min\{R_t, I_t\}, \tag{4}$$

Q_t can be expressed as

$$\begin{aligned} Q_t &= (R_t - I_t)^+ \\ &= \max\{0, R_t - s, R_t - I_{t-1} + P_{t-1} - V_{t-1}\}. \end{aligned}$$

Thus, the supplier's operational characteristics, such as $\{V_t\}$ and s, directly influence the size of Q_t, which in turn influences α_m. In other words, the supplier's production/inventory operations eventually have an effect on the manufacturer's customer service level.

This raises the question of whether the supplier's service level α_s provides sufficient information about the supplier's performance for the manufacturer to predict the customer service level. In the next section, we study a more detailed model to shed light on this issue. Questions under investigation include: Are there definitive relationships between the supplier's service level and the manufacturer's service level? Does it matter how the supplier achieves the target service level specified by the manufacturer? What is the role of the supplier's operating characteristics such as inventory level and capacity?

4. Relevance of Service Levels

4.1. The α-Type Service Level

For simplicity and to highlight the impact of the supplier's operational characteristics, in this section we assume that both the customer demand and the manufacturer's capacity are constant. Let demand $D_t = d > 0$ for all periods t and the manufacturer's capacity $c = d + b$. The supplier has a finite and variable production capacity. Its capacity in period t is

$$V_t = \begin{cases} d+e & \text{with probability } p, \\ 0 & \text{with probability } 1-p, \end{cases} \tag{5}$$

where $0 < p < 1$.

We assume $c > d$ or $b > 0$ to prevent backorders from exploding. For a similar reason, the supplier's maximum capacity is assumed to be greater than d, in other words, $e > 0$. The manufacturer's capacity is always b units greater than demand. On the other hand, the supplier's capacity depends on two parameters: e and p. Parameter e is interpreted as the maximum extra capacity of the supplier, while parameter p represents the reliability of the supplier's resource. A higher value of p means less variance in the supplier's capacity. The expected capacity, $p(d+e)$ is assumed to be greater than d. Otherwise, the supplier could not deliver enough components to satisfy customer demand.

We set the initial component inventory level I_1 equal to the base-stock level s. If $s \le d$, then there will always be backorders after the first backorder occurs, because the base-stock level is the maximum number of components available. (If $B_t > 0$, then $R_t > d$, which implies $Q_t = R_t - P_t > d - s \ge 0$ and $B_{t+1} > 0$.) For this reason, we assume s to be greater than d.

We will show shortly that the process $\{(I_t, B_t)\colon t = 1, 2, \ldots\}$ is a discrete-time Markov chain. Using its stationary distribution, we can compute α_m and α_s. This allows us to make several key observations as illustrated in Table 1, which lists several combinations of e, p, and s generating α_s close to 5% when $d = 10$.

As we can see, although α_s is kept roughly at the same level (5%), α_m varies significantly (7% to 34%) with different combinations of the parameters. Thus, the impact of the supplier on the customer service level cannot be predicted simply by the supplier's service level. How the supplier achieves its service level matters.

The results also help us to see which of the supplier's characteristics has the greatest impact on the customer service level. We observe that the gap between α_m and α_s becomes smaller when the manufacturer has more extra capacity b. But even with the same b, the customer service level still depends heavily on the supplier's characteristic parameters such as extra capacity e and reliability factor p. It appears that having reliable resources, e.g., $p = 0.95$, has the greatest effect on improving the customer service level $(1-\alpha_m)$. This is perhaps the most expensive method, however, for the supplier to improve its operations.

Now we show how α-type service levels have been computed for Table 1. It follows from (1) and (4) that

$$P_t = \min\{B_t + d, d + b, I_t\}, \tag{6}$$

and

$$\begin{aligned} I_{t+1} &= \begin{cases} \min\{s, I_t - P_t + d + e\} & \text{with probability } p, \\ I_t - P_t & \text{with probability } 1-p, \end{cases} \\ &= \begin{cases} \min\{s, \max\{I_t + e - B_t, I_t + e - b, d + e\}\} \\ \quad \text{with probability } p, \\ \max\{I_t - B_t - d, I_t - d - b, 0\} \\ \quad \text{with probability } 1-p. \end{cases} \end{aligned} \tag{7}$$

Backorders are updated by

$$\begin{aligned} B_{t+1} &= B_t + d - P_t \\ &= \max\{0, B_t - b, B_t + d - I_t\}. \end{aligned} \tag{8}$$

Table 1 **Supplier's Characteristic Parameter Sets Yielding $\alpha_s \approx 5\%$**

b	e	p	s	α_s (%)	α_m (%)
1	2	0.95	19	5.00	13.26
	3	0.80	55	4.98	32.39
	9	0.70	37	5.01	30.74
	15	0.70	31	5.02	34.46
2	3	0.80	63	5.09	21.30
	3	0.95	19	5.00	7.80
	5	0.79	30	4.97	20.04
	10	0.84	20	5.00	18.14
5	6	0.95	15	5.00	5.82
	9	0.80	29	5.02	7.26
	28	0.64	38	4.96	8.62
	34	0.55	44	5.01	11.17

From (7) and (8), it is clear that $\{(I_t, B_t)\colon t = 1, 2, \ldots\}$ is a discrete-time Markov chain. Its state space $\mathcal{S}$ is

$$\mathcal{S} = \{(i, j)\colon (i, j) \sim (s, 0),\ 0 \le i \le s,\ j \ge 0\},$$

where $(i, j) \sim (s, 0)$ means (i, j) communicates with $(s, 0)$. Because $(I_1, B_1) = (s, 0)$, only the states that communicate with $(s, 0)$ are included in $\mathcal{S}$. Not all inventory positions can be reached from the initial state. For example, if $d = b = e = 2$ and $s = 2k$ for a positive integer k, any state in $\{(2(k-i)-1, 2j+1)\colon i = 0, 1, \ldots, k-1, j = 0, 1, \ldots\}$ does not communicate with $(s, 0)$.

It can be easily shown that $\{(I_t, B_t)\colon t = 1, 2, \ldots\}$ is positive recurrent if $p(d+e) > d$. Because the Markov chain is irreducible and positive recurrent, there is a unique stationary distribution $\{\pi(i, j)\colon (i, j) \in \mathcal{S}\}$. Further, the strong law of large numbers holds for $\{(I_t, B_t)\}$. In particular, with probability 1,

$$\alpha_m = \overline{\lim_{T\to\infty}} \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{B_t > 0\}} = \mathsf{P}\{B > 0\}.$$

Thus, using the stationary probability, α_m can be expressed as

$$\alpha_m = \mathsf{P}\{B > 0\} = \sum_{i=0}^{s} \sum_{j=1}^{\infty} \pi(i, j). \tag{9}$$

Similarly, the expression for α_s is

$$\begin{aligned} \alpha_s &= \mathsf{P}\{Q > 0\} \\ &= \sum_{j=0}^{b} \sum_{i=0}^{j+d-1} \pi(i, j) + \sum_{j=b+1}^{\infty} \sum_{i=0}^{b+d-1} \pi(i, j). \end{aligned} \tag{10}$$

The following results reveal that under certain conditions there is a definitive relationship between α_m and α_s. The proofs can be found in the appendix.

Proposition 4.1. *If $b \ge e$, $\{\pi(i, j)\}$ is independent of b.*

Proposition 4.1 implies that adding capacity does not help the manufacturer improve the customer service level when its capacity is greater than or equal to the supplier's peak capacity.

Proposition 4.2. *If $b > e$, $\alpha_m = \alpha_s$.*

Proposition 4.2 indicates that α_m is equal to α_s if the manufacturer always has excessive capacity ($b > e$). In this case, the manufacturer knows exactly what kind of service level it can offer to the customer once it makes an agreement with the supplier on the expected stockout rate α_s. Similarly, once the manufacturer sets a target customer service level, it should require a supplier's service level not less than the target level when signing a supply contract. Proposition 4.2 illustrates a special case where the customer service level is predictable from the supplier's service level.

Unfortunately, such a direct relationship between the customer and the supplier's service levels does not always exist. As demonstrated in Table 1, if $b \le e$, there is no general result the manufacturer can use to estimate the customer service level from the supplier's service level.

4.2. Other Types of Service Levels

We now show that similar conclusions can be reached for β- and γ-type service levels as well. We use the same example as in §4.1 for which stationary distribution of $\{B_t\}$ and $\{Q_t\}$ exist. Recall that demand and the manufacturer's capacity are assumed to be constant. The supplier can change its performance by three parameters: e, p and s. Table 2 lists several combinations of the supplier's characteristic parameters which yield $\beta_s (=\gamma_s) \approx 0.5\%$ when $d = 10$ and $b = 5$.

While $\beta_s (=\gamma_s)$ remains close to 5%, α_m ranges from 12% to 20%. It turns out that β_m and γ_m vary significantly as well (7% to 11% and 9% to 19%, respectively). Thus, no matter what type of definition is used, the customer service level is not predictable just by β_s (or γ_s). Note that β_s involves the average amount of component stockout, which is one of the measures recommended in §5. This example demonstrates again that a single quantity is not adequate to serve as the supplier performance measure. The manufacturer must use two supplier performance measures: the frequency and the average amount of component stockout, jointly.

Table 2 Supplier's Characteristic Parameter Sets Yielding $\beta_s (=\gamma_s) \approx 5\%$

e	p	s	$\beta_s (=\gamma_s)$ (%)	α_m (%)	β_m (%)	γ_m (%)
6	0.87	18	4.95	18.51	7.89	9.74
10	0.71	29	5.02	12.35	10.75	18.40
15	0.74	25	4.95	19.60	9.17	13.44
20	0.63	33	5.00	14.57	10.56	18.89

The examples in this section show that the customer service level can vary even when the supplier's service level is constant. The degree of customer service improvement depends not only on the supplier's service level, but also on *how* the supplier has achieved that level. Thus, in general, the manufacturer needs more information on the supplier's operation in order to guarantee the customer service level.

5. New Supplier Performance Measures

To overcome the shortcomings of the traditional service level as a measure of supplier performance, in this section we develop alternative supplier performance measures that can allow the manufacturer to secure its customer service level regardless of how the supplier delivers the performance.

Note that an upper bound on α_m gives a lower bound on $1-\alpha_m$, which in turn can serve as the minimum customer service level the manufacturer can guarantee. Our goal is thus to construct an upper bound on α_m, say u_m, in terms of the supplier's delivery statistics of $\{Q_t\colon t=1,2,\ldots\}$. This way, if we require the supplier to deliver a performance measured by a specified value u_m, then the manufacturer is assured to deliver a customer service level $1-u_m$.

To obtain a robust performance measure, it is important that both the bound and the derivation of the bound depend only on the observable quantities $\{Q_t\}$. Thus, in our analysis here we do not make any specific assumption about the supplier's operational characteristics. Equation (2) plays a key role in this analysis.

Our first main result, an upper bound on α_m, is stated in the theorem below. All the proofs can be found in the appendix.

THEOREM 5.1. *Assume demands are i.i.d. and have the same distribution as D. Let c be the manufacturer's production capacity in each period. Let*

$$u_m = \frac{\bar{Q} + c(\alpha_s+\nu) + \mathsf{E}[(D-c)^+]}{c(1+\nu) - \mathsf{E}[D] + \mathsf{E}[(D-c)^+]},$$

where $\bar{Q} = \overline{\lim}_{T\to\infty}(1/T)\sum_{t=1}^{T} Q_t$ *and* $\nu = \mathsf{P}\{D > c\}$. *Then, with probability* 1,

$$\alpha_m \le u_m. \tag{11}$$

The result of Theorem 5.1 and its derivation do not make use of any information on the supplier's side except $\{Q_t\colon t=1,2,\ldots\}$. In particular, the bound is valid whichever policy the supplier adopts to manage the component inventory. The following theorem provides a lower bound on α_m.

THEOREM 5.2. *Let D, c, and $\bar{Q}$ defined as in Theorem* 5.1, *and let*

$$l_m = \frac{\bar{Q} + \mathsf{E}[(D-c)^+]}{c - \mathsf{E}[D] + \mathsf{E}[(D-c)^+]}.$$

If $\lim_{T\to\infty} B_T/T = 0$, *then, with probability* 1,

$$\alpha_m \ge l_m. \tag{12}$$

The smaller the gap between u_m and α_m is, the better the upper bound u_m is. The following lemma gives a bound on the gap.

LEMMA 5.1. *If* $\lim_{T\to\infty} B_T/T = 0$,

$$u_m - \alpha_m \le \frac{c(\alpha_s+\nu)}{c(1+\nu) - \mathsf{E}[D] + \mathsf{E}[(D-c)^+]}. \tag{13}$$

The next theorem shows the convergence of the upper bound u_m. Its proof makes use of Lemma 5.1 and can be found in the appendix. In the theorem, superscript r is used as an index of sequences.

THEOREM 5.3. *Assume that there exist constants C and $\bar{r}$ such that*

$$\frac{c^r}{c^r - \mathsf{E}[D^r]} \le C \quad \text{for } r > \bar{r}.$$

If $\alpha_m^r \to 0$ as $r\to\infty$, then u_m^r converges to 0 as well.

Tighter upper bounds are available when demand is deterministic, as in the previous section.

COROLLARY 5.1. *If demand is a constant, i.e., $D_t = d < c$ for all t,*

$$\alpha_m \le \frac{\bar{Q}}{c-d} + \alpha_s. \tag{14}$$

Furthermore, if d and c are integers and $\{Q_t\colon t=1,2,\ldots\}$ takes only integral values,

$$\alpha_m \le \frac{\bar{Q}}{c-d} + \left(1 - \frac{1}{c-d}\right)\alpha_s. \tag{15}$$

Theorem 5.1 and its corollary give a hint as to how the supplier's performance should be measured to guarantee the customer service level. The manufacturer can control the supplier's influence on the customer service level through service level (α_s) and average stockout ($\overline{Q}$) together by including an agreement on these two measures when signing the supply contact. We can also learn from the expression of the bounds that the requirements on the supplier must be adjusted depending on the manufacturer's capacity.

To demonstrate the performance of the bounds, we first consider the deterministic demand case, as in §4. The examples in Table 1 satisfy the assumptions for (15) of Corollary 5.1. Table 3 shows the corresponding lower bound (l_m) and upper bound (in the right-hand side of (15), denoted by u_m^d) with the service levels. Recall that b is the difference of capacity and demand ($b = c - d$).

As mentioned in the previous section, the customer service level ($1 - \alpha_m$) varies significantly from 65% to 94% depending on the supplier's operating parameters, even though the supplier's service level ($1 - \alpha_s$) remains almost the same as 95% in all instances. In Table 3, we can see that u_m^d, which takes into account both α_s and $\overline{Q}$, reflects the variation of the customer service level effectively. The gap between the upper bound and the customer service level is less than 3%. Especially when $b = 1$, $\alpha_m = u_m^d = \overline{Q}$ ($\alpha_m \leq \overline{Q}$ from (15) and $\alpha_m \geq \overline{Q}$ from (12)).

The gap between l_m and α_m is also not significant—less than 3% in most cases. The lower bound as well as the upper bound might be used to estimate the customer service level. However, for the purpose of controlling the supplier's performance to guarantee the customer service level ($1 - \alpha_m$), upper bounds on α_m have a better use than lower bounds. Thus, we focus on the upper bound.

Table 3 **Upper Bound on α_m (Deterministic Demand)**

b	e	p	s	α_s (%)	l_m (%)	α_m (%)	u_m^d (%)
1	2	0.95	19	5.00	13.26	13.26	13.26
	3	0.80	55	4.98	32.39	32.39	32.39
	9	0.70	37	5.01	30.77	30.77	30.77
	15	0.70	31	5.02	34.46	34.46	34.46
2	3	0.80	63	5.09	21.27	21.30	23.82
	3	0.95	19	5.00	5.59	7.80	8.09
	5	0.79	30	4.97	19.41	20.04	21.89
	10	0.84	20	5.00	18.14	18.14	20.64
5	6	0.95	15	5.00	5.82	5.82	9.82
	9	0.80	29	5.02	4.24	7.26	8.26
	28	0.64	38	4.96	6.31	8.62	10.28
	34	0.55	44	5.01	9.43	11.17	13.44

To illustrate the performance of the bounds for the general i.i.d. demand case (Theorem 5.1), we consider three settings of the supplier's operational characteristics. The first setting is similar to the example in §4. The supplier follows a revised base-stock policy with base-stock level s to control production of components. Instead of taking one of two possible values (see (5)), the supplier's capacity $\{V_t\}$ is now assumed to have a more general distribution. Another difference is that the demand is an i.i.d. process instead of a constant. In the second setting, the supplier uses an (s, S) policy to control the component inventory. In each period, if the inventory position is below or equal to s after the manufacturer retrieves components, the supplier starts to produce or places an order from an upstream source to bring the inventory position to S. The replenishment takes place after a fixed lead time, LT. The last setting is similar to the second one except that the inventory policy is an (R, S) policy. The parameter R represents a review period and the other parameter S has the same meaning as in the (s, S) policy. After checking the component inventory position every R periods, the supplier initiates a replenishment to bring the inventory position back to S.

Tables 4, 5, and 6 show the simulation results when $\{D_t: t = 1, 2, \ldots\}$ and $\{V_t: t = 1, 2, \ldots\}$ follow normal, Poisson, and gamma distribution, respectively. $\mathsf{E}[D]$ is set to 20 in all examples. σ_X^2 and c_X^2 denote the variance and squared coefficient of variation of a random variable X, respectively. All simulation results are averaged values from 10 replications. Each replication runs for $T = 10^5$ periods. In simulations with normal distributions, negative random numbers have been reset to 0.

For each inventory policy, the simulation is done with two different levels of the variations of demand and/or the supplier's capacity (three with gamma distribution). For each set of demand and supplier parameters, we illustrate the effect of manufacturer's capacity c on the performance of both the supplier and the manufacturer. While increasing c improves manufacturer's service level α_m, it tends to have the

Table 4 Supplier Performance Measures (with Normal Distributions)

Supplier's Inventory Policy		c	α_s (%)	$\bar{Q}$	ν (%)	l_m (%)	α_m (%)	u_m (%)	θ_m (%)
Revised	$\sigma_D^2=50$	26	5.45	0.265	19.82	15.39	31.22	63.83	31.31
Base-Stock	$\sigma_V^2=50$	28	6.70	0.378	12.87	9.85	22.40	53.05	24.78
Policy		30	8.10	0.481	7.83	7.14	16.68	43.64	20.46
$\mathsf{E}[V]=25$		32	8.73	0.570	4.45	5.77	13.41	36.32	17.67
$s=40$		34	9.12	0.642	2.37	5.01	11.64	31.03	15.87
		40	9.50	0.784	0.24	3.95	10.12	23.31	13.63
	$\sigma_D^2=20$	26	0.29	0.008	8.96	3.13	10.67	30.52	11.11
	$\sigma_V^2=20$	28	0.39	0.012	3.66	0.96	4.30	13.34	4.80
		30	0.44	0.015	1.26	0.35	1.73	5.25	2.02
		32	0.46	0.017	0.37	0.18	0.83	2.37	1.01
		34	0.47	0.018	0.09	0.13	0.56	1.49	0.69
		40	0.48	0.019	0.00	0.09	0.49	1.05	0.57
(s, S) Policy	$s=80$	26	4.46	0.365	19.77	16.73	33.02	62.19	31.84
	$S=200$	28	4.48	0.387	12.84	9.89	22.30	47.04	22.95
	$LT=3$	30	4.42	0.388	7.82	6.19	15.25	34.07	16.53
	$\sigma_D^2=50$	32	4.36	0.386	4.47	4.23	10.89	24.53	12.23
		34	4.32	0.382	2.37	3.15	8.31	18.23	9.49
		40	4.31	0.385	0.24	1.94	5.77	10.96	6.47
	$s=70$	26	0.32	0.009	8.96	3.14	10.88	30.62	11.15
	$S=150$	28	0.34	0.012	3.66	0.95	4.38	13.26	4.77
	$LT=2$	30	0.39	0.013	1.26	0.33	1.74	5.07	1.94
	$\sigma_D^2=20$	32	0.39	0.013	0.37	0.15	0.80	2.16	0.91
		34	0.40	0.014	0.09	0.11	0.51	1.28	0.59
		40	0.40	0.014	0.00	0.07	0.42	0.87	0.47
(R, S) Policy	$R=9$	26	4.98	0.636	19.77	20.69	36.47	65.58	35.53
	$S=250$	28	5.01	0.665	12.84	13.16	25.37	50.55	26.25
	$LT=3$	30	5.00	0.683	7.82	9.06	18.03	37.79	19.68
	$\sigma_D^2=50$	32	5.00	0.697	4.47	6.79	13.44	28.33	15.26
		34	5.00	0.708	2.37	5.46	10.68	21.97	12.39
		40	5.01	0.731	0.24	3.68	7.45	14.07	8.88
	$R=5$	26	0.85	0.040	8.96	3.65	11.49	32.60	12.05
	$S=160$	28	0.89	0.044	3.66	1.35	5.06	15.21	5.63
	$LT=2$	30	0.90	0.045	1.26	0.65	2.40	6.86	2.76
	$\sigma_D^2=20$	32	0.91	0.046	0.37	0.42	1.42	3.77	1.68
		34	0.91	0.046	0.09	0.34	1.11	2.74	1.33
		40	0.91	0.046	0.00	0.23	0.98	2.05	1.14

opposite effect on supplier's performance measured by α_s and $\bar{Q}$. We also observe that α_m improves (gets smaller) with smaller variations in demand and supplier capacity. The upper bound u_m also decreases as α_m decreases to 0, which is expected from Theorem 5.3. When α_m is less than 1%, u_m is not more than 3%.

To decide what kind of performance to ask of the supplier, the manufacturer needs an estimate of α_m that is expressed in terms of the supplier performance measures. An ideal candidate would be a tight upper bound on α_m so that it can be used to derive supplier's performance requirements that guarantee a target customer service level. In the case of deterministic demand, the upper bounds in (14) and (15) serve exactly this purpose. In general, Theorem 5.3 shows that u_m gets smaller when α_m decreases. But, unlike the deterministic demand case, the numerical examples in Tables 4–6 reveal that in most cases the difference between u_m and α_m is not small enough for u_m itself to be used to approximate the customer service level. To resolve this discrepancy, we develop the following closer estimate of α_m in terms of α_s and $\bar{Q}$, which is another main result in this section:

$$\theta_m = \frac{\bar{Q} + (c - \mathsf{E}[D])((1-\nu)\alpha_s + \nu) + \mathsf{E}[(D-c)^+]}{(c - \mathsf{E}[D])(1+\nu) + \mathsf{E}[(D-c)^+]}.$$

Table 5 Supplier Performance Measures (with Poisson Distributions)

Supplier's Inventory Policy		c	α_s (%)	$\bar{Q}$	ν (%)	l_m (%)	α_m (%)	u_m (%)	θ_m (%)
Revised	$E[V]=22$	26	8.90	0.446	7.27	10.18	17.49	60.04	23.63
Base-Stock		28	10.04	0.575	2.86	7.93	13.12	47.94	19.88
Policy		30	10.53	0.681	0.94	7.00	11.51	40.21	18.18
$s=40$		32	10.72	0.769	0.26	6.45	10.99	35.48	17.35
		34	10.79	0.841	0.06	6.01	10.85	32.31	16.85
		40	10.82	0.987	0.00	4.94	10.82	26.57	15.76
	$E[V]=25$	26	0.46	0.016	7.27	3.24	8.90	27.38	10.01
		28	0.59	0.023	2.86	1.09	3.63	11.88	4.37
		30	0.66	0.027	0.94	0.47	1.60	5.10	2.04
		32	0.68	0.030	0.26	0.29	0.94	2.79	1.23
		34	0.69	0.032	0.06	0.23	0.75	2.06	0.99
		40	0.70	0.034	0.00	0.17	0.70	1.57	0.87
(s, S) Policy	$s=80$	26	2.41	0.140	7.27	5.24	11.90	35.18	13.51
	$S=200$	28	2.39	0.142	2.86	2.56	6.22	18.91	7.49
	$LT=3$	30	2.38	0.141	0.94	1.61	3.88	11.21	4.85
		32	2.36	0.141	0.26	1.21	2.99	8.15	3.82
		34	2.37	0.141	0.06	1.02	2.66	6.90	3.44
		40	2.37	0.142	0.00	0.71	2.43	5.45	3.08
	$s=70$	26	0.24	0.007	7.27	3.09	8.70	26.53	9.68
	$S=150$	28	0.28	0.010	2.86	0.92	3.35	10.74	3.92
	$LT=2$	30	0.29	0.011	0.94	0.30	1.27	3.89	1.52
		32	0.30	0.012	0.26	0.14	0.57	1.63	0.70
		34	0.30	0.012	0.06	0.09	0.37	0.97	0.46
		40	0.31	0.012	0.00	0.06	0.31	0.67	0.37
(R, S) Policy	$R=9$	26	2.99	0.277	7.27	7.45	14.23	38.72	16.06
	$S=250$	28	2.99	0.281	2.86	4.29	8.17	22.36	9.73
	$LT=3$	30	2.99	0.284	0.94	3.03	5.54	14.39	6.86
		32	2.99	0.286	0.26	2.42	4.42	11.02	5.65
		34	2.99	0.287	0.06	2.05	3.92	9.45	5.10
		40	2.99	0.288	0.00	1.44	3.35	7.43	4.43
	$R=5$	26	0.78	0.040	7.27	3.63	9.49	28.70	10.63
	$S=160$	28	0.82	0.044	2.86	1.35	4.09	12.83	4.83
	$LT=2$	30	0.83	0.046	0.94	0.65	1.95	5.78	2.38
		32	0.83	0.046	0.26	0.43	1.20	3.32	1.51
		34	0.84	0.046	0.06	0.34	0.96	2.51	1.23
		40	0.83	0.047	0.00	0.23	0.85	1.90	1.07

Tables 4–6 list θ_m in the last column. In all cases, θ_m is closer to α_m than u_m.

In general, θ_m is not an upper bound on α_m. It would be an upper bound if the following inequalities held (see (26) and (28)). All the new notations have been used for the proof of Theorem 5.1 and their definitions can be found in the appendix.

$$\overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_{\{t\in K\cup G\}}(c-R_t) \le (c-\mathsf{E}[D])\cdot \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_{\{t\in K\cup G\}}$$

and

$$\overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_{\{t\in N\cap L\}} \le \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_{\{t\in N\}}\cdot \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_{\{t\in L\}}.$$

Due to the dependency of B_{t+1} and Q_t on D_t, it cannot be guaranteed that the two inequalities hold in general. Nonetheless, as shown in Tables 4–6, θ_m is greater than α_m except for 2 out of 108 instances.

Figure 1 plots the numerical results for the revised base-stock policy systems reported in Tables 4–6. As expected, the variation of demand has substantial influence on the customer service level because

Table 6 Supplier Performance Measures (with Gamma Distributions)

Supplier's Inventory Policy		c	α_s (%)	$\bar{Q}$	ν (%)	l_m (%)	α_m (%)	u_m (%)	θ_m (%)
Revised	$c_D^2 = 2$	55	13.12	3.596	8.98	17.91	27.64	43.99	34.34
Base-Stock Policy	$c_V^2 = 2$	65	14.28	4.652	6.60	15.03	24.10	39.98	31.93
$\mathrm{E}[V] = 50$		80	15.30	6.416	3.74	11.88	20.62	35.86	28.88
$s = 100$		100	15.84	7.476	2.49	10.49	19.42	32.12	27.52
	$c_D^2 = 1$	55	3.53	0.722	6.24	5.38	12.39	18.43	13.78
	$c_V^2 = 1$	65	3.98	0.930	3.79	3.65	9.02	13.91	10.74
		80	4.29	1.155	1.77	2.48	6.67	10.26	8.28
		100	4.45	1.343	0.63	1.83	5.59	8.18	6.90
	$c_D^2 = 0.5$	55	0.30	0.041	2.62	0.97	3.60	5.29	3.75
	$c_V^2 = 0.5$	65	0.36	0.054	1.10	0.40	1.70	2.47	1.83
		80	0.40	0.065	0.29	0.16	0.77	1.08	0.85
		100	0.42	0.071	0.05	0.10	0.52	0.68	0.56
(s, S) Policy	$s = 80$	55	16.43	5.496	8.98	22.87	34.34	52.59	41.37
	$S = 200$	65	16.42	6.104	6.60	18.08	29.29	45.47	36.59
	$LT = 3$	80	16.37	6.766	4.29	13.70	24.79	38.35	31.79
	$c_D^2 = 2$	100	15.98	7.383	2.49	10.37	21.66	32.17	27.54
	$s = 70$	55	10.28	2.369	6.24	9.92	20.46	31.91	23.82
	$S = 150$	65	10.37	2.595	3.79	7.28	16.96	25.95	20.07
	$LT = 2$	80	10.31	2.781	1.77	5.17	14.37	20.67	16.70
	$c_D^2 = 1$	100	10.45	2.931	0.63	3.81	13.13	17.48	14.71
	$s = 70$	55	1.03	0.112	2.62	1.17	4.24	6.59	4.64
	$S = 150$	65	1.15	0.140	1.10	0.59	2.51	3.77	2.79
	$LT = 1$	80	1.23	0.163	0.29	0.33	1.74	2.35	1.84
	$c_D^2 = 0.5$	100	1.26	0.174	0.05	0.23	1.55	1.86	1.53
(R, S) Policy	$R = 9$	55	19.05	7.702	8.98	28.63	39.10	61.02	48.71
	$S = 250$	65	20.50	9.317	6.60	24.84	35.10	56.80	46.35
	$LT = 3$	80	20.81	11.387	4.29	21.18	31.55	50.89	42.94
	$c_D^2 = 2$	100	21.95	13.678	2.49	18.14	29.02	46.85	40.73
	$R = 5$	55	13.37	3.840	6.24	13.97	25.08	39.88	30.28
	$S = 160$	65	13.69	4.408	3.79	11.24	21.82	34.17	26.91
	$LT = 2$	80	13.99	4.955	1.77	8.77	19.31	28.94	23.77
	$c_D^2 = 1$	100	14.57	5.458	0.63	6.96	18.04	25.70	21.90
	$R = 4$	55	1.25	0.205	2.62	1.43	4.58	7.16	5.10
	$S = 160$	65	1.32	0.226	1.10	0.78	2.78	4.19	3.14
	$LT = 1$	80	1.36	0.242	0.29	0.46	1.89	2.65	2.10
	$c_D^2 = 0.5$	100	1.38	0.250	0.05	0.32	1.65	2.11	1.75

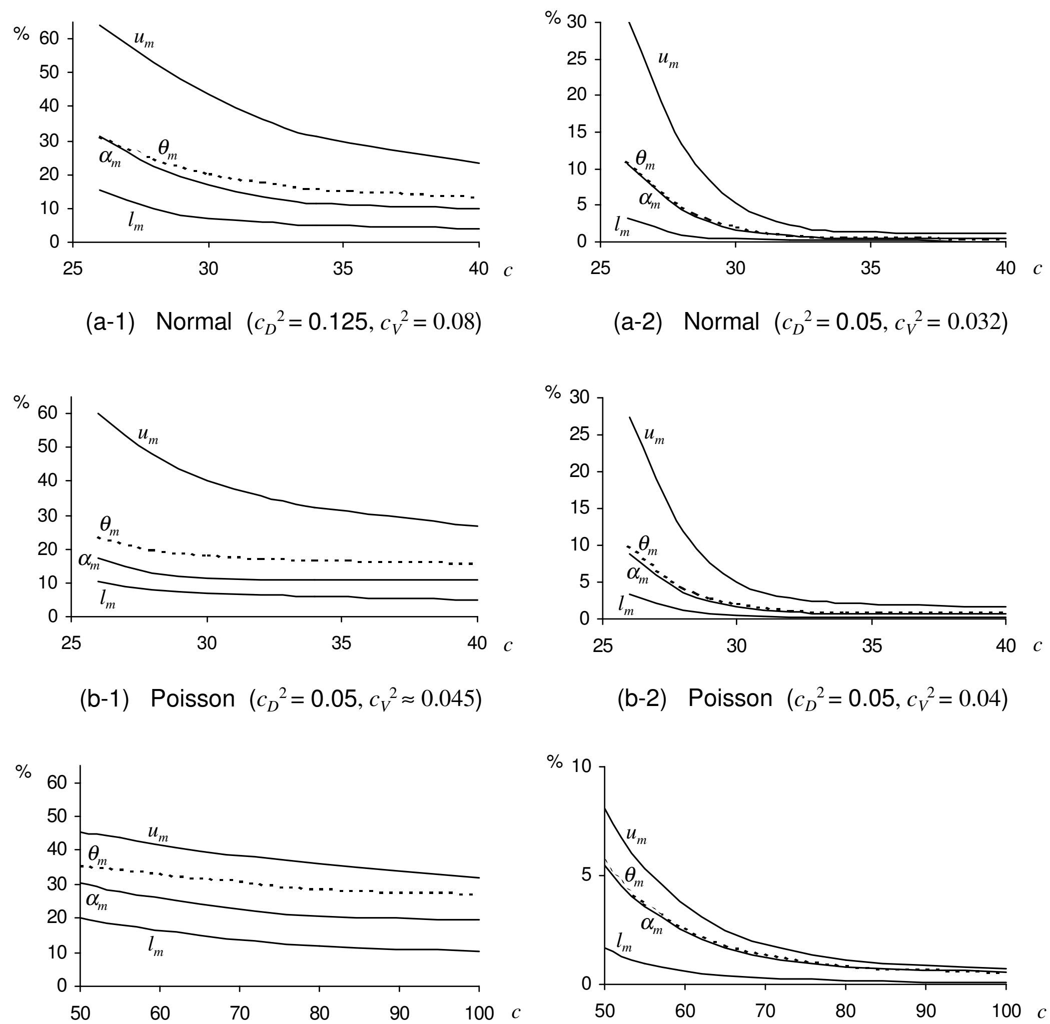
demand exceeding manufacturer's capacity results in backorders. It is also obvious that the variation of the supplier's performance affects α_m. For example, in the case with a revised base-stock policy, fluctuations of the supplier's capacity cause more frequent component stockout. The customer service level $(1 - \alpha_m)$ gets lower with higher variations (larger values of c_D^2 and/or c_V^2) than with lower variations (smaller values c_D^2 and/or c_V^2).

Figure 1 shows that θ_m is an accurate estimate of α_m when α_m is small (<10%). We have obtained the same results with the other inventory policies. Thus, when the target customer service level is high, i.e., greater than 90%, the manufacturer can derive from θ_m the supplier's performance requirements, which will result in a customer service level close to the target.

6. Proposed Contract Form

As shown in the previous section, we can use θ_m to approximate the manufacturer's customer service level α_m with confidence. We now show how to use this result to design a supply contract.

Figure 1 Effects of the Variation of Demand and Supplier's Capacity



(a-1) Normal ($c_D{}^2 = 0.125$, $c_V{}^2 = 0.08$)
(a-2) Normal ($c_D{}^2 = 0.05$, $c_V{}^2 = 0.032$)
(b-1) Poisson ($c_D{}^2 = 0.05$, $c_V{}^2 \approx 0.045$)
(b-2) Poisson ($c_D{}^2 = 0.05$, $c_V{}^2 = 0.04$)
(c-1) Gamma ($c_D{}^2 = 2$, $c_V{}^2 = 2$)
(c-2) Gamma ($c_D{}^2 = 0.5$, $c_V{}^2 = 0.5$)

Note that θ_m is a linear function of α_s and $\bar{Q}$;

$$\theta_m = \xi_1 \alpha_s + \xi_2 \bar{Q} + \eta,$$

where $\xi_1 = (c - \mathsf{E}[D])(1-\nu)/((c - \mathsf{E}[D])(1+\nu) + \mathsf{E}[(D-c)^+])$, $\xi_2 = 1/((c - \mathsf{E}[D])(1+\nu) + \mathsf{E}[(D-c)^+])$, $\eta = ((c - \mathsf{E}[D])\nu + \mathsf{E}[(D-c)^+])/((c - \mathsf{E}[D])(1+\nu) + \mathsf{E}[(D-c)^+])$, and $\nu = \mathsf{P}\{D > c\}$.

If we set θ_m to a target level α_m^*, then we get a linear relationship between α_s and $\bar{Q}$:

$$\xi_1 \alpha_s + \xi_2 \bar{Q} = \alpha_m^* - \eta. \tag{16}$$

Frequency and average of shortage are positive quantities unless both are zero. Thus the range of α_s satisfying (16) is

$$0 < \alpha_s < \frac{\alpha_m^* - \eta}{\xi_1}. \tag{17}$$

Now for any fixed α_s in this range, we can solve (16) to obtain the corresponding $\bar{Q}$. If the supplier can achieve the service level no greater than α_s and at the same time can control the average backorders not to exceed $\bar{Q}$, then the manufacturer's customer service level (stockout rate) is bounded above by α_m^*, and therefore the target customer service level will be guaranteed. This pair of values $(\alpha_s, \bar{Q})$ can then be specified in a supply contract.

In fact, the supply contract can be composed of a list of such pairs. Once a base pair is specified, we can use

$$\Delta\bar{Q} = -\left(\frac{\xi_1}{\xi_2}\right)\Delta\alpha_s. \qquad (18)$$

to choose the other pairs to be included in the menu. Note that (18), derived from (16), provides the (approximate) trade-off between α_s and $\bar{Q}$ that yields the same customer service level.

For example, if the customer demand has a Poisson distribution with mean $E[D] = 20$ and the manufacturer's capacity is $c = 30$, we have

$$\xi_1 \approx 0.970, \qquad \xi_2 \approx 0.0984, \qquad \eta \approx 0.0164.$$

From (17) the range of α_s adequate for 95% target service level (i.e., $\alpha_m^* = 0.05$) is

$$0 < \alpha_s < \frac{0.05 - 0.0164}{0.970} \approx 3.5\%,$$

and the slope of the linear trade-off in (18) is

$$-\frac{\xi_1}{\xi_2} \approx 9.86.$$

These lead to the following sample menu for the 95% target customer service level:

α_s (%)	$\bar{Q}$
1	0.242
2	0.144
3	0.045

7. Concluding Remarks

In this paper we examine the role of the conventional service levels such as fill rate and stockout rate in decentralized supply chains. We specifically study the relevance of local service levels in a supply chain to its end customer service level. The research was motivated by our interactions with managers who are interested to know how to specify supplier's service level in a VMI contract so to allow the manufacturer to control his manufacturing process to achieve certain desired customer service level. Using a two-party capacitated supply chain model, we demonstrate that even if the supplier provides steady performance in terms of these service measures, the eventual customer service level at the end of the supply chain can vary considerably. Thus, the rule-of-thumb of specifying a higher upstream service level indicated in the literature proves to be invalid. Although our model assumes an MTO manufacturer, the results can be easily extended to a make-to-stock (MTS) manufacturer, with a positive base-stock level.

This finding also indicates that the commonly used VMI contract that requires a minimum inventory level that the supplier must maintain, while working for a supplier-retailer setting, fails to work for a supplier-manufacturer setting when the manufacturer has a finite production capacity. Note that the probability that the inventory level is below a lower limit is similar in nature to the stockout rate (in the latter, the lower limit is zero). So, maintaining a minimum inventory level with high probability is equivalent to a α-type service level. In a supplier-retailer setting, customers withdraw directly from the inventory supplier maintains, so supplier's service level is precisely the customer service level. Under the supplier-manufacturer setting, even if the component supply is available, the manufacturer's capacity may limit his ability to fulfill the customer demand in time. Indeed, the irrelevance of supplier's service level holds even if we ask the supplier to maintain at least c units (the manufacturer's capacity) component inventory.

By establishing a bound on the manufacturer's customer service level, we show that, in addition to the supplier's stockout rate, the average component backorders is another important measure that should be specified in a supply contract for the manufacturer to guarantee his desired customer service level. The supply contract can be designed as a menu of different combinations of stockout rate and average component backorders along a linear function. The virtue of this kind of contract is that it requires minimum information sharing and is easy to monitor. Also, its implementation is independent of the supplier's cost.

It is worth mentioning that our findings on the irrelevance of supplier's service level hold even if demand information is shared with the supplier. Assume the manufacturer retrieves a component whenever a demand occurs, even though it might not be required for production in the current period. This way the actual demand information is passed to the component supplier, which should help prevent the manufacturer's capacity from being wasted due to component stockout. However, using D_t instead of R_t to request components in the example in §4, with unavailable components backlogged at the supplier, we have observed the same lack of connection between customer and supplier's service levels (for all α-, β-, and γ-type service levels; the data are not reported here).

Instead of local performance measures, transfer payments from the manufacturer to the supplier based on the supply chain performance is another way for the manufacturer to control the supplier's impact on the final customer service level. Examples include the linear transfer payment agreement discussed in Cachon (1999), Cachon and Zipkin (1999), Caldentey and Wein (2003), or the fixed payment agreement proposed in Cachon (2001). See a review by Cachon (2003). To carry out these types of contracts, besides demand information sharing, several additional requirements are in need. First, to figure out the payment schemes, the manufacturer needs to assess the penalty associated with customer backorders. So the supply chain performance is not only measured by the service level but also by the backorder level. Second, because the supplier is measured by the end customer service, it is necessary for the manufacturer to share his demand, capacity, and holding cost information with the supplier, so that the supplier can be in effect a central planner. Or, both the supplier and the manufacturer share their capacity and cost information with a central planner, who can then determine the payment schemes. These requirements may render these types of contracts difficult to implement.

Acknowledgments
The authors thank Gerard Cachon for valuable comments on an earlier version of this paper. The paper has also benefited from stimulating discussions with Derek Atkins, Martin Puterman, Paul Zipkin, and other seminar participants at Duke University, the University of British Columbia, the University of Minnesota at Twin Cities, and the University of Texas at Dallas. The research was supported in part by National Science Foundation Grants DMI-9457336, DMI-9813345, DMI-0300599, and DMI-0084922, and by The Logistics Institute—Asia Pacific, a partnership between The National University of Singapore and The Georgia Institute of Technology.

Appendix

Proof of Proposition 4.1. We show that b has no influence on $\{P_t\colon t=1,2,\ldots\}$. It implies that $\{B_t\colon t=1,2,\ldots\}$ and $\{I_t\colon t=1,2,\ldots\}$ are updated independently of b.

From (6), if

$$B_t \le b \text{ or } I_t \le d+b, \qquad t=1,2,\ldots \tag{19}$$

P_t is not dependent on b.

For $t=1$, it is obvious that (19) holds. Assuming that (19) holds for $t=k$, we consider three exclusive and exhaustive conditions on B_k and I_k:

(i) if $B_k \le b$ and $I_k \ge d$, $P_k \ge I_k \ge d$ implies $B_{k+1} = B_k - P_k + d \le B_k \le b$.

(ii) if $B_k \le b$ and $I_k < d$, $P_k = I_k$ implies $I_{k+1} \le I_k - P_k + d + e \le d+b$.

(iii) if $B_k > b$ and $I_k \le d+b$, $P_k = I_k$ as in (ii). This leads to $I_{k+1} \le d+b$.

Therefore, (19) also holds for $t=k+1$. By induction, it holds for $t=1,2,\ldots$. □

Proof of Proposition 4.2. We will show that if $b > e$, $B_{t+1} > 0$ is equivalent to $I_t < R_t$. This implies $\alpha_m = \alpha_s$.

(i) if $B_t \le b$, $R_t = B_t + d$. Therefore, $B_{t+1} > 0$ iff $I_t < R_t$.

(ii) if $B_t > b$, $R_t = d+b$ and

$$B_{t+1} \ge B_t + d - R_t > 0.$$

We need to show that $I_t < R_t$. We can modify (19) in the proof of Proposition 4.1 as follows. If $b < e$,

$$B_t \le b \quad \text{or} \quad I_t < d+b, \qquad t=1,2,\ldots.$$

It implies $I_t < d+b$ if $B_t > b$. Therefore, $I_t < d+b = R_t$. □

Proof of Theorem 5.1. First, it is convenient for the proof to define the following subsets of time horizon $\{1,2,\ldots\}$.

$$L=\{t\colon D_t > c\}, \qquad M=\{t\colon B_t>0\}, \qquad N=\{t\colon Q_t>0\},$$
$$F=\{t\in M^c\colon B_{t+1}>0\}, \qquad G=\{t\in M\colon B_{t+1}=0\},$$

where A^c denotes the complement set of A.

L is the set of the periods when manufacturer's capacity is less than customer demand. M and N are the set of the periods when there is stockout of finished product and component, respectively. The α_m and α_s defined in §3 can be expressed as

$$\alpha_m = \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{\{B_t>0\}} = \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{\{t\in M\}},$$
$$\alpha_s = \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{\{Q_t>0\}} = \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{\{t\in N\}}.$$

F is the set of the periods when the manufacturer does not have backorders at the beginning but does at the end. G is for the opposite case. Note that cardinality of G is either equal to or less than that of F by 1.

From the second equality in (2) that, for any T,

$$B_{T+1} = B_1 + \sum_{t=1}^{T}(D_t - P_t). \tag{20}$$

Because $B_1 = 0$, $B_{T+1} \ge 0$ implies

$$\sum_{t=1}^{T}(R_t - D_t) \le \sum_{t=1}^{T} Q_t. \tag{21}$$

By definition, $R_t = \min\{D_t, c\}$ for $t \in M^c$. Thus, $R_t = D_t$ for $t \in M^c \cap L^c$. Using this, the right-hand side in (21) is reduced to

$$\begin{aligned}\sum_{t=1}^{T}(R_t - D_t) &= \sum_{t\in M^c\cap L^c}(R_t - D_t) + \sum_{t\in(M^c\cap L^c)^c}(R_t - D_t)\\ &= \sum_{t\in M\cup L}(R_t - D_t)\\ &= \sum_{t\in M}(R_t - D_t) + \sum_{t\in M^c\cap L}(R_t - D_t).\end{aligned} \tag{22}$$

From (21) and (22),

$$\sum_{t\in M}(R_t - D_t) \le \sum_{t=1}^{T} Q_t + \sum_{t\in M^c\cap L}(D_t - R_t).$$

After adding $\sum_{t\in M}(c - R_t)$ on both sides, we get

$$\sum_{t\in M}(c - D_t) \le \sum_{t=1}^{T} Q_t + \sum_{t\in M}(c - R_t) + \sum_{t\in M^c\cap L}(D_t - R_t). \tag{23}$$

For $t \in L\cup(M\backslash G\cap N^c)$, $R_t = c$. It is because

(i) L is the set of the periods when customer demand exceeds manufacturer's capacity. Because the manufacturer needs components not less than demand and not more than capacity, $R_t = c$ for $t\in L$.

(ii) If $t \in M\backslash G\cap N^c$, the manufacturer has retrieved all the components it requires ($t \in N^c$ and it means $P_t = R_t$). But, there are backorders at the *end* of t ($t \notin G$ which means $B_{t+1} > 0$). This implies that $R_t = \min\{D_t + B_t, c\} = c$. Otherwise ($R_t = D_t + B_t > c$), there would be no backorder at the end ($B_{t+1} = B_t + P_t - D_t = B_t + R_t - D_t = 0$), which contradicts $t \in M\backslash G$. Thus, (23) becomes

$$\sum_{t\in M}(c-D_t) \le \sum_{t=1}^{T} Q_t + \sum_{t\in M\cap(L\cup(M\backslash G\cap N^c))^c}(c-R_t) + \sum_{t\in M^c\cap L}(D_t - c). \tag{24}$$

To simplify the index set of the second summation on the right-hand side, we define $K = M\cap L^c\cap N$. Then,

$$\begin{aligned}M\cap(L\cup(M\backslash G\cap N^c))^c &= (M\cap L^c)\cap(M^c\cup G\cup N)\\ &= (M\cap L^c\cap G)\cup(M\cap L^c\cap N)\\ &= G\cup(M\cap L^c\cap N) = G\cup K.\end{aligned}$$

The third equality is from the fact that G is a subset of M by definition and also of L^c because $t\in L$ ($D_t > c$) implies $B_{t+1} > 0$. Because

$$\sum_{t\in M^c\cap L}(D_t - c) = \sum_{t\in M^c}(D_t - c)^+ = \sum_{t=1}^{T}(D_t - c)^+ - \sum_{t\in M}(D_t - c)^+,$$

we get the following from (24)

$$\sum_{t\in M}[(c - D_t) + (D_t - c)^+] \le \sum_{t=1}^{T} Q_t + \sum_{t\in K\cup G}(c - R_t) + \sum_{t=1}^{T}(D_t - c)^+,$$

which is equal to

$$\sum_{t\in M}(D_t - c)^- \le \sum_{t=1}^{T} Q_t + \sum_{t\in K\cup G}(c - R_t) + \sum_{t=1}^{T}(D_t - c)^+, \tag{25}$$

where $(x)^- = (-x)^+$.

We can rewrite (25) using indicator functions. The range of summations is suppressed for simple exposition.

$$\sum_t \mathbf{1}_{\{t\in M\}}(D_t - c)^- \le \sum_t Q_t + \sum_t \mathbf{1}_{\{t\in K\cup G\}}(c - R_t) + \sum_t (D_t - c)^+.$$

Using nonnegativity of R_t, we get

$$\sum_t \mathbf{1}_{\{t\in M\}}(D_t - c)^- \le \sum_t Q_t + c\sum_t \mathbf{1}_{\{t\in K\cup G\}} + \sum_t (D_t - c)^+. \tag{26}$$

Because $K\cap G = \varnothing$ ($K\subset N^c$, $G\subset N$) and $\sum_{t=1}^{T}\mathbf{1}_{\{t\in G\}} \le \sum_{t=1}^{T}\mathbf{1}_{\{t\in F\}}$, the second summation on the right-hand side of (26) is changed to

$$\begin{aligned}\sum_t \mathbf{1}_{\{t\in K\cup G\}} &= \sum_t \mathbf{1}_{\{t\in K\}} + \sum_t \mathbf{1}_{\{t\in G\}}\\ &\le \sum_t \mathbf{1}_{\{t\in K\}} + \sum_t \mathbf{1}_{\{t\in F\}}.\end{aligned} \tag{27}$$

Now we show that $F = M^c\cap(N\cup L)$. For $t\in M^c$,

$$B_{t+1} = D_t - P_t = D_t - R_t + Q_t = (D_t - c)^+ + Q_t.$$

This implies that $B_{t+1} > 0$ iff $D_t > c$ ($t\in L$) or $Q_t > 0$ ($t \in N$).

$$M^c\cap\{t\colon B_{t+1} > 0\} = M^c\cap(N\cup L).$$

The left-lend side is F by definition.

With this equality about F, (27) is changed to

$$\begin{aligned}\sum_t \mathbf{1}_{\{t\in K\cup G\}} &\le \sum_t \mathbf{1}_{\{t\in(M\cap L^c\cap N)\cup(M^c\cap(N\cup L))\}}\\ &= \sum_t \mathbf{1}_{\{t\in(N\cup L)\backslash(M\cap L)\}} = \sum_t \mathbf{1}_{\{t\in N\cup L\}} - \sum_t \mathbf{1}_{\{t\in M\cap L\}}\\ &= \sum_t \mathbf{1}_{\{t\in N\}} + \sum_t \mathbf{1}_{\{t\in L\}} - \sum_t \mathbf{1}_{\{t\in N\cap L\}} - \sum_t \mathbf{1}_{\{t\in M\cap L\}}\\ &\le \sum_t \mathbf{1}_{\{t\in N\}} + \sum_t \mathbf{1}_{\{t\in L\}} - \sum_t \mathbf{1}_{\{t\in M\cap L\}}.\end{aligned} \tag{28}$$

Using (28), (26) is reduced to

$$\sum_t \mathbf{1}_{\{t\in M\}}(D_t-c)^- + c\sum_t \mathbf{1}_{\{t\in M\cap L\}} \le \sum_t Q_t + c\left[\sum_t \mathbf{1}_{\{t\in N\}} + \sum_t \mathbf{1}_{\{t\in L\}}\right] + \sum_t (D_t-c)^+.$$

Dividing by T and taking upper limits as $T\to\infty$ lead to the inequality below.

$$\overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}(D_t-c)^- + c\cdot \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^T \mathbf{1}_{\{t\in M\cap L\}} \le \bar{Q} + c\left(\alpha_s + \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^T \mathbf{1}_{\{t\in L\}}\right) + \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^T (D_t-c)^+. \quad (29)$$

Because $\{D_t: t=1,2,\ldots\}$ is i.i.d., by the strong law of large numbers, the two summations on the right-hand side converge to ν and $\mathsf{E}[(D-c)^+]$ almost surely as $T\to\infty$, respectively. Once the two upper limits on the left-hand side are shown to be equivalent to $\alpha_m \mathsf{E}[(D-c)^-]$ and $\alpha_m\nu$ respectively, the proof is finished because (29) gets reduced to the following, which equals (11).

$$\alpha_m \mathsf{E}[(D-c)^-] + c\alpha_m\nu \le \bar{Q} + c(\alpha_s+\nu) + \mathsf{E}[(D-c)^+].$$

We first prove that $\overline{\lim}_{T\to\infty} 1/T\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}(D_t-c)^-$ is equal to $\alpha_m \mathsf{E}[(D-c)^-]$ with probability 1.

$$\overline{\lim_{T\to\infty}} \frac{\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}(D_t-c)^-}{T} = \overline{\lim_{T\to\infty}} \frac{\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}}{T} \cdot \frac{\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}(D_t-c)^-}{\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}}.$$

The first part of the right-hand side is the definition of α_m. If $\{D_t: t\in M\}$, in turn $\{(D_t-c)^-: t\in M\}$, is i.i.d, then the strong law of large numbers can apply to the second part and its limit is $\mathsf{E}[(D-c)^-]$.

Let $\tau_n = \inf\{t>\tau_{n-1}: B_t>0\}$ for $n=1,2,\ldots$ and $\tau_0=0$. Then, $\{D_t: t\in M\}$ can be represented by $\{D_{\tau_n}: n=1,2,\ldots\}$. We use f_X for the p.d.f. (or p.m.f.) of random variable X. Since D_t is independent of B_t and has the same distributions as D, $f_{D_{\tau_n}} = f_D$ for an arbitrary n. Thus, $\{D_{\tau_n}: n=1,2,\ldots\}$ is identically distributed. It is also independent because

$$\begin{aligned}
f_{D_{\tau_1,\tau_2,\ldots,\tau_n}}(x_1,x_2,\ldots,x_n) &= \mathsf{E}_\tau[f_{D_{\tau_1,\tau_2,\ldots,\tau_n}|\tau}(x_1,x_2,\ldots,x_n)]\\
&= \mathsf{E}_\tau[f_{D_{\tau_1}|\tau}(x_1) f_{D_{\tau_2}|\tau}(x_2)\cdots f_{D_{\tau_n}|\tau}(x_n)]\\
&= \mathsf{E}_\tau[f_{D_{\tau_1}}(x_1) f_{D_{\tau_2}}(x_2)\cdots f_{D_{\tau_n}}(x_n)]\\
&= \mathsf{E}_\tau[f_D(x_1) f_D(x_2)\cdots f_D(x_n)]\\
&= f_D(x_1) f_D(x_2)\cdots f_D(x_n)\\
&= f_{D_{\tau_1}}(x_1) f_{D_{\tau_2}}(x_2)\cdots f_{D_{\tau_n}}(x_n),
\end{aligned}$$

where $\tau = (\tau_1,\tau_2,\ldots,\tau_n)$.

Because $\{D_t: t\in M\}$ is i.i.d., so is $\{(D_t-c)^-: t\in M\}$. If $\alpha_m>0$ (this is assumed here because (11) holds obviously when $\alpha_m=0$), $\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}$ goes to infinity as T goes to infinity. Now we can apply the strong law of large numbers and get

$$\frac{\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}(D_t-c)^-}{\sum_{t=1}^T \mathbf{1}_{\{t\in M\}}} \xrightarrow{\text{a.s.}} \mathsf{E}[(D-c)^-] \quad \text{as } T\to\infty.$$

It can be proven similarly that the second upper limit in (29) is equal to $\alpha_m\nu$ with probability 1. □

Proof of Theorem 5.2. From (20)–(25), we can see the following equation holds.

$$B_{T+1} + \sum_{t\in M}(D_t-c)^- = \sum_{t=1}^T Q_t + \sum_{t\in K\cup G}(c-R_t) + \sum_{t=1}^T (D_t-c)^+.$$

Because $R_t\le c$ for all t,

$$B_{T+1} + \sum_{t\in M}(D_t-c)^- \ge \sum_{t=1}^T Q_t + \sum_{t=1}^T (D_t-c)^+.$$

Using the similar arguments as in the proof of Theorem 5.1, we get

$$\overline{\lim_{T\to\infty}} \frac{B_{T+1}}{T} + \alpha_m \mathsf{E}[(D_t-c)^-] \ge \bar{Q} + \mathsf{E}[(D_t-c)^+], \quad (30)$$

which equals (12) if $\overline{\lim}_{T\to\infty} B_T/T = 0$. Because $B_t\ge 0$ for all t, the condition $\overline{\lim}_{T\to\infty} B_T/T=0$ is equivalent to $\lim_{T\to\infty} B_T/T = 0$. □

Proof of Lemma 5.1. Because $u_m-\alpha_m \le u_m - l_m$,

$$\begin{aligned}
u_m-\alpha_m &\le \frac{\bar{Q}+c(\alpha_s+\nu)+\mathsf{E}[(D-c)^+]}{c(1+\nu)-\mathsf{E}[D]+\mathsf{E}[(D-c)^+]} - \frac{\bar{Q}+\mathsf{E}[(D-c)^+]}{c-\mathsf{E}[D]+\mathsf{E}[(D-c)^+]}\\
&\le \frac{c(\alpha_s+\nu)}{c(1+\nu)-\mathsf{E}[D]+\mathsf{E}[(D-c)^+]}. \quad \square
\end{aligned}$$

Proof of Theorem 5.3. Let $\delta = \overline{\lim}_{T\to\infty} B_T/T$. From (30) and a similar argument in the proof of Lemma 5.1,

$$\begin{aligned}
u_m^r &\le \alpha_m^r + \frac{c^r(\alpha_s^r+\nu^r)+\delta^r}{c^r(1+\nu^r)-\mathsf{E}[D^r]+\mathsf{E}[(D^r-c^r)^+]}\\
&\le \alpha_m^r + \frac{c^r(\alpha_s^r+\nu^r+\delta^r/c^r)}{c^r-\mathsf{E}[D^r]}.
\end{aligned}$$

By the assumption, for $r>\bar{r}$,

$$u_m^r \le \alpha_m^r + C\left(\frac{\alpha_s^r+\nu^r+\delta^r}{c^r}\right).$$

We show that α_s^r, ν^r, and δ^r/c^r converge to 0 as $\alpha_m^r\to 0$.

Demand exceeding production capacity and component shortage give rise to backorders of finished product, which implies

$$\alpha_m^r \ge \nu^r \quad \text{and} \quad \alpha_m^r \ge \alpha_s^r.$$

Thus, as α_m^r converges to 0, ν^r and α_s^r also converge to 0.

Now we show the convergence of δ^r/c^r to 0. Assume that $\delta^r = \overline{\lim}_{T\to\infty} B_T^r/T > 0$. Then, for an arbitrary $0 < \varepsilon < \delta^r$, B_T^r is greater than $(\delta^r - \varepsilon)T$ infinitely often. Because the backorders cannot be reduced more than c^r in one period,

$$\alpha_m^r \geq \lim_{T\to\infty} \frac{\lfloor(\delta^r - \varepsilon)T/c^r\rfloor}{T + \lfloor(\delta^r - \varepsilon)T/c^r\rfloor} = \frac{\delta^r - \varepsilon}{c^r + \delta^r - \varepsilon}.$$

Thus, when $\alpha_m^r < 1$,

$$\delta^r - \frac{\alpha_m^r c^r}{1 - \alpha_m^r} \leq \varepsilon.$$

Because the inequality holds for an arbitrary $\varepsilon > 0$,

$$\delta^r - \frac{\alpha_m^r c^r}{1 - \alpha_m^r} \leq 0,$$

which equals

$$\frac{\delta^r}{c^r} \leq \frac{\alpha_m^r}{1 - \alpha_m^r}.$$

Hence, δ^r/c^r converges to 0 as $\alpha_m^r \to 0$. □

Proof of Corollary 5.1. If $D_t = d$, (25) is reduced to the following inequality and we use it instead of (26) to get an upper bound:

$$(c-d)\sum_t \mathbf{1}_{\{t\in M\}} \leq \sum_t Q_t + (c-d)\sum_t \mathbf{1}_{\{t\in(M\cap N)\cup G\}}.$$

Also, if d, c, and $\{Q_t\colon t = 1, 2, \ldots\}$ are all integers, so is $\{R_t\colon t = 1, 2, \ldots\}$. Then $R_t \geq d+1$ for $t \in M$, and the above inequality can be replaced with a tighter 1,

$$(c-d)\sum_t \mathbf{1}_{\{t\in M\}} \leq \sum_t Q_t + (c-d-1)\sum_t \mathbf{1}_{\{t\in(M\cap N)\cup G\}}.$$

The remaining procedure is similar, as in Theorem 5.1. □

References

Barnes, E., J. G. Dai, S. Deng, D. Down, M. Goh, H. C. Lau, M. Sharafali. 2000. Electronics manufacturing service industry. TLI-AP Report, Georgia Institute of Technology.

Bernstein, F., A. Federgruen. 2003. Pricing and replenishment strategies in a distribution system with competing retailers. *Oper. Res.* **51**(3) 409–426.

Bollapragada, R., U. S. Rao, J. Zhang. 2000. A decomposition approach for managing assembly systems with demand and supply uncertainty. Working paper, Carnegie Mellon University, Pittsburgh, PA.

Buzzell, R. D., G. Ortmeyer. 1995. Channel partnership streamline distribution. *Sloan Management Rev.* **36**(3) 85–96.

Cachon, G. 1999. Competitive and cooperative inventory management in a two-echelon supply chain with lost sales. Working paper, Duke University, Chapel Hill, NC.

Cachon, G. 2001. Stock wars: Inventory competition in a two echelon supply chain. *Oper. Res.* **49**(5) 658–674.

Cachon, G. 2003. Supply chain coordination with contracts. A. de Kok, S. Graves, eds. *Handbooks in Operations Research and Management Science: Supply Chain Management*. North-Holland, Amsterdam, The Netherlands.

Cachon, G., M. Fisher. 1997. Campbell Soup's continuous product replenishment program: Evaluation and enhanced decision rules. *Production Oper. Management* **6**(3) 266–276.

Cachon, G., P. Zipkin. 1999. Competitive and cooperative inventory policies in a two-stage supply chain. *Oper. Res.* **45**(7) 936–953.

Caldentey, R., L. Wein. 2003. Analysis of a decentralized production-inventory system. *Manufacturing Service Oper. Management* **5** 1–17.

Cohen, M. A., H. L. Lee. 1988. Strategic analysis of integrated production-distribution systems: Models and methods. *Oper. Res.* **36**(2) 216–228.

Diks, E. B., A. G. de Kok, A. G. Lagodimos. 1996. Multi-echelon systems: A service measure perspective. *Eur. J. Oper. Res.* **95**(2) 241–263.

Federgruen, A., P. Zipkin. 1986. An inventory model with limited production capacity and uncertain demands I. The average-cost criterion. *Math. Oper. Res.* **11**(2) 193–207.

Fisher, M. L. 1997. What is the right supply chain for your product? *Harvard Business Rev.* (March–April) 105–116.

Fry, M., R. Kapuscinski, T. Olsen. 2001. Coordinating production and delivery under a (z, Z)-type vendor-managed inventory contract. *Manufacturing Service Oper. Management* **3**(2) 151–173.

Glasserman, P. 1997. Bounds and asymptotics for planning critical safety stocks. *Oper. Res.* **45** 244–257.

Hammond, J., T. Clark. 1997. Reengineering channel reordering processes to improve total supply chain performance. *Production Oper. Management* **6**(3) 248–265.

Kleijnen, J., M. Smits. 2002. Performance metrics in supply chain management. *Production Oper. Management* **6**(3) 248–265.

Lee, H. L., C. Billington. 1993. Material management in decentralized supply chains. *Oper. Res.* **41**(5) 835–847.

Narayanan, V., A. Raman, S. Kraiselburd. 2002. Contracting for inventory in a supply chain with stochastic demand and substitute products. *Production Oper. Management*. Forthcoming.

Parker, R., R. Kapuscinski. 2001. Optimal policies for a capacitated two-echelon inventory system. *Oper. Res.* Forthcoming.

Paschalidis, I. C., Y. Liu. 2003. Large deviations-based asymptotics for inventory control in supply chains. *Oper. Res.* **51**(3) 437–460.

Plambeck, E., S. Zenios. 2003. Incentive efficient control of a make-to-stock production system. *Oper. Res.* **51** 371–386.

Schneider, H. 1981. Effect of service-levels on order-points or order-levels in inventory models. *Int. J. Prod. Res.* **19**(6) 615–631.

Simchi-Levi, D., P. Kaminsky, E. Simchi-Levi. 2000. *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies*. McGraw-Hill, New York.

Sobel, M. 2002. Fill rates of single-stage and multi-stage supply systems. *Manufacturing Service Oper. Management*. Forthcoming.

Thompson, A. A., A. J. Strickland. 2001. *Strategic Management: Concepts and Cases*, 12th ed. McGraw-Hill, New York.

van Houtum, G. J., K. Inderfurth, W. H. M. Zijm. 1996. Materials coordination in stochastic multi-echelon systems. *Eur. J. Oper. Res.* **95**(1) 1–23.

Wilson, R. 1987. Game theoretic analyses of trading processes. T. F. Bewley, ed. *Advances in Economic Theory: Fifth World Congress*. Cambridge University Press, Cambridge, U.K., Chapter 2, 33–70.