Managing a Two-Stage Serial Inventory System under Demand and Supply Uncertainty and Customer Service Level Requirements

Ramesh Bollapragada‡ Uday S. Rao† Jun Zhang*

‡Bell Laboratories Room 4G-423A, Lucent Technologies, Holmdel, NJ 07733

† CBA, University of Cincinnati, Cincinnati, OH 45221

*A.B. Freeman School of Business, Tulane University, New Orleans, LA 70118

March 2001, Revised April, July 2002, February 2003

Abstract

We consider a two-echelon serial inventory system with demand and supply uncertainty, non-zero lead times for component procurement and end-product assembly, and a minimum customer service level requirement. We present two supply models which incorporate both quantity and timing uncertainty; these models correspond to current and proposed supply environments. Assuming that installation base-stock ordering policies are followed and that the demand distribution is quasi-concave, we show that the chance-constrained problem of determining optimal base-stock levels which minimize the total inventory investment (cost-weighted stock levels) subject to a service constraint is a convex programming problem. We characterize the relation between the optimal base-stock levels of the component and the end-product. We also illustrate how an optimal internal (component) service level can be computed, which permits decomposition of the two-stage serial system into two coordinated single-echelon systems. Computational experiments illustrate insights on the effects of supply uncertainty and other problem parameters on stock-positioning in a two-echelon serial system. In particular, we evaluate the benefits of switching from one supply environment to another.

Keywords: Two-echelon Supply Chain; Installation Base-stock Policy; Customer Service Level; Supply Modeling.

1 Introduction

In an uncertain world, operations managers are faced with the important problem of distributing the right amount of inventories at different levels (components and finished goods) in a supply chain. In our experience, the goal of this stock-positioning is to meet management-set customer service targets, without violating budget constraints on inventory investment (cost-weighted stock levels). The work reported here is motivated by issues faced by a large telecommunications company in one of its advanced manufacturing facilities and is representative of supply coordination problems in many high-tech equipment manufacturing industries.

The basic setting is a two-echelon supply chain, consisting of components and finished goods. Components are ordered from outside suppliers and are assembled into the end-product. Supply lead times are comparable to manufacturing lead times since the assembly process is complex. There is uncertainty in both supply and demand. Consequently, the manufacturer holds safety stock inventories of both components and end-products. Currently, the firm uses a discrete time, decision support system (similar to MRP or ERP) that manages inventory at both the component and end-product levels, independently of each other. That is, the information system uses "node"-specific input parameters such as the component supply lead time or the product assembly lead time to determine safety stock levels of components and finished goods, respectively. Although the deployed system is quite robust and considers many important parameters that play a role in this setting, there is a strong belief among the planners that *the component and end-product levels should be linked through an over-arching model that takes a more integrated view of inventory management.* Furthermore, one of the critical component suppliers has recently suggested that he would be willing to switch from his traditional approach of "capacitated" supply (with no restrictions on backlog quantity and duration) to a new approach where he guarantees delivery either within his quoted lead time, L, or at most one period late. Hence, management needs methods to *evaluate the benefits of the proposed guaranteed lead time for supply.*

With this motivation, we study the total inventory investment (measured in base-stock dollars) and its positioning in a two-echelon supply chain. Due to the high-tech, innovative nature of the products there is currently negligible component commonality and different end-products are treated independently. Hence, we consider the simple case of a single end-product produced from one critical component (or module) delivered by an external, unreliable supplier. The case with multiple components / echelons forms the basis for future research (Section 6.3). Two supply models, based on the industrial environment, are developed. The main issues addressed in this paper include the following:

A. How should the optimal inventory investment and its distribution between components and end-products be determined under supply and demand uncertainty and with a desired target customer service level specified (instead of a shortage penalty cost)?

B. How can the integrated inventory planning model be replaced by individual models for components and end-products?

C. How does the proposed guaranteed lead time $(L,\ L+1)$ supply model compare with the current capacitated supply model (with unlimited backlog quantity and duration)?

D. How will inventory positioning be affected in the proposed supply model by changes in problem parameters such as customer service targets, demand and supply uncertainty, lead times, and unit costs?

The main contributions of this work are: (i) We develop two models with supply uncertainty and present solution approaches to determine optimal base-stock levels for components and finished products, such that the total inventory investment (measured in total base-stock dollars, see expression (1)) is minimized and the desired customer service level is achieved. (ii) Under both demand and supply uncertainty, we show that the optimal component stock level is a convex-decreasing function of the stock level of finished products. (iii) We investigate a simple internal service level based approach to decompose and coordinate component and end-product replenishment in a two-echelon service-constrained supply network. (iv) Our computational experiments illustrate managerial insights, obtained from what-if sensitivity analyses, on how stock positioning in two-echelon serial systems changes with problem parameters and supply model. In particular, we identify the circumstances under which the benefits of guaranteed supply lead time are likely to be large or small.

After a brief literature review, we present our supply models and a mathematical problem formulation in Section 3. In Section 4, we develop properties which are used within a search algorithm for computing optimal base-stock levels. Results from computational experiments are detailed in Section 5. We conclude in Section 6 with a summary and some model extensions.

2 Literature Review

Multi-echelon inventory problems with demand uncertainty have been widely studied since the seminal work by Clark and Scarf (1960). See, for instance, Axsäter (1993) (continuous review policies), Federgruen (1993) (periodic review policies), and Diks, de Kok, and Lagodimos (1996) (service levels instead of shortage penalty cost).

We use an installation-based, periodic review, order-up-to (base-stock) policy, under a service level requirement, since this corresponds to the industrial environment that motivated this research. Our work is similar to Gallego and Zipkin (1999) and Boyaci and Gallego (2001). The main difference is that we explicitly incorporate both supply and demand uncertainty, in a two-echelon framework, to ensure that the desired customer service level will be provided and total inventory investment (cost-weighted base-stock levels) will be minimized. Further, we compare and contrast two different supply models corresponding to the current and proposed supply environments in the application that motivated this work. Supply uncertainty in single-echelon systems has been studied by (a) Henig and Gerchak (1991)− random yield, (b) Ciarallo, Akella, and Morton (1994)− random supply capacity, (c) Anupindi and Akella (1993)− order placement with dual unreliable suppliers, and (d) Moinzadeh and Lee (1989)− order delivery in two partial shipments. In this paper, our supply uncertainty is similar to the model in Moinzadeh and Lee (1989) and to Model III in Anupindi and Akella (1993). However, we model a two-echelon environment and *we are interested in understanding how supply and demand uncertainty together affect the positioning of inventory across the supply chain.* This has not yet been adequately addressed in previous research with the possible exception of the random supply lead time, contract assembly environment in Song, Yano, and Lerssrisuriya (2000).

3 Model: Minimizing Cost with Service Constraints

Without loss of generality, one unit of the component goes into making one unit of the end-product. Customer demand for the product in each time period t is an independent, identically distributed, nonnegative, continuous random variable, ξ_t, with cumulative distribution function $F(\cdot)$ and density $f(\cdot)$. The replenishment of components and the assembly of products are governed by discrete-time, periodic-review policies. Product assembly has a non-zero lead time, l, when the critical component is available. The component has a unique supplier who quotes a delivery lead time L. However, this supplier is not 100% reliable since he may not always deliver orders within the quoted lead time.

Typically, supply uncertainty results from several factors such as capacity constraints on the supplier's resources, random yield in the supplier's manufacturing processes, variability in delivery lead times, poor planning and product rationing by suppliers. Supply performance is also affected by the supplier's relation with the manufacturer and the structure of the supply contract signed between supplier and manufacturer. Currently, it appears that the supplier satisfies orders based on its production capacity (which is random). Consequently, orders may be delivered in several portions, and delivery of some portion of an order may be delayed by several time periods (beyond the quoted lead time). Although we will briefly present the current supply model, the main focus will be on the supplier's proposed lead time guarantee of at most a single period delay beyond quoted lead time in delivery of components to the manufacturing plant.

3.1 The Supply Models

In our problem environment, the end-product manufacturer has significant leverage with component suppliers. Consequently, the supplier strives to provide the manufacturer a high service level at low cost. When a supplier cannot deliver a manufacturer's order on time, the unsatisfied demand is backlogged. The supply environment may be modeled by assuming that the component supplier has random production capacity, η, with CDF $G(\cdot)$ and density $g(\cdot)$. η represents the maximum amount of the component that the supplier can provide in one time period (without expediting, overtime or subcontracting). Whenever the manufacturer places an order in period t for y units of the component, a random capacity (also denoted by η_t) is realized. Note that this capacity is only an artifact used to model the supplier's delivery behavior (capability) as observed (perceived) by the manufacturer.

Current Supply Model - Unlimited backlogging: Let "backlog" denote the accumulated component orders not satisfied by the supplier. Then, in period t, the supplier ships $\min(y + \text{backlog}, \eta_t)$; any unsatisfied demand is backlogged. To ensure system stability (based on standard results from queuing theory), we assume that η_t is a stationary process with $E[\eta_t] > E[\xi_t]$.

Proposed Supply Model - Backlog limited to one period: In this case, if $\eta_t \geq y$, then the manufacturer will receive y units of the product after L periods. Otherwise, the manufacturer will receive η_t units after L periods, and the remainder $y - \eta_t$ after $L + 1$ periods[1]. Note that the delivery of $y - \eta_t$ delayed units that were ordered $L + 1$ periods earlier does not affect the supplier's ability to deliver η_{t+1} units in the same period. Suppliers usually know if a shortage will occur for a specific manufacturer's order before this order is due. Typically, due to the importance of the manufacturer, suppliers can take corrective actions (such as overtime) to satisfy the manufacturer's order on time or at most one period late. Such reactive actions in one period often do not affect the normal capacity available in the next period. This model of supply uncertainty is motivated by the authors' experience with suppliers in the telecommunications industry. The use of η_t to model supply uncertainty is similar to Ciarallo et al. (1994); delivery lead times of L and $L + 1$ have been considered in Moinzadeh and Lee (1989). Our model of supply uncertainty is different from the random yield models considered in the literature (Yano and Lee 1995), since in the random yield case, the supplier may never deliver the total amount ordered by the manufacturer. Although random yield adequately models uncertainty in manufacturing processes, it does not accurately capture the supply delivery and the backordering process when external suppliers are unreliable, but their supply behavior can be influenced by purchasing managers. In the next section, we present details of our integrated planning model.

3.2 Problem Formulation

All events occur at the beginning of each period and in the following sequence: (1) Receive component deliveries from the supplier (based on orders placed at least L time periods earlier) and add completed assemblies (started l time periods ago) to finished goods inventory. (2) Demand for end-product assemblies materializes and is communicated to both echelons. Use available inventory to satisfy as much of this demand as possible; any excess customer demand is backordered. (3) Initiate new end-product assembly (based on available components) and place purchase orders for components from the external supplier.

[1]The extension to the case where the shipment is delayed to $L + k$ is straightforward, provided excess capacity in future periods can not be used to expedite past delayed deliveries. An application of two supply lead times, corresponding to dual supply modes (regular and expedited), is presented in Rao, Scheller-Wolf, and Tayur (2000).

In instances where the desired amount of end-product assembly exceeds the limit imposed by component availability, the amount of this excess is backordered and will be called the "delayed-assembly." This delayed-assembly is caused by a "component-backlog." In this paper, "work-in-process" denotes end-products whose assembly has been initiated but not completed. Similarly, "component pipeline inventory" denotes the stock in transit from the supplier to the assembly system. The installation stock (inventory position) of end-products is equal to the on-hand finished goods inventory + work-in-process + delayed-assembly − customer backorders. The installation stock of components is equal to the on-hand inventory + component pipeline inventory − component-backlog. In each period, an installation base-stock policy is used to decide the number of components to procure and the number of end-products to assemble. That is, when the installation stock falls below the base-stock level, an order is placed to raise it back to that level. Our objective is to determine the base-stock levels for the end-product, S^p, and for the component, S^c, that minimizes the total inventory investment subject to achieving a specified customer service level. For simplicity, we will restrict attention to the Type-1 service level measure corresponding to the probability of satisfying customer demand directly from available product inventory. In Section 6.1, we extend our results to the case when a Type-2 service level (fill rate) is used.

Let C^c and C^p denote, respectively, the per unit cost of the component and the end-product. Let $IL^p(S^c, S^p)$ denote the steady state inventory level of the end-product immediately after demand materializes in a period. Let α be the desired customer service level for the end-product. Then the problem can be formulated as,

$$\min \quad U(S^c, S^p) \equiv C^c S^c + C^p S^p \quad s.t. \quad P\{IL^p(S^c, S^p) \geq 0\} \geq \alpha. \tag{1}$$

We use the linear objective in (1) to simplify presentation and because such a model is used in the company that motivated this research[2]. We use the term "inventory investment" to refer to the objective in (1), since the base-stock level in the objective may be viewed as a surrogate for inventory. However, the true inventory will likely be different from the base-stock level. Hence, even if the inventory holding costs are proportional to the unit costs, a model that minimizes inventory holding costs will be different from the model in (1). In Section 6.2 we consider a model with an inventory holding cost objective. By using the general functional form $U(S^c, S^p)$ as the objective, we show that our analysis applies

[2]Implementation and use of a similar model has been reported in the press, see, for instance, OR/MS Today, Feb 1996 issue, "Engine for Inventory Management", pages 14-17.

to both objective functions (since, the requirement that $U(\cdot)$ is non-decreasing is satisfied in both cases). Our computational experiments suggest that the inventory holding cost corresponding to stock levels obtained using an inventory investment objective are typically within one percent of the minimum holding cost objective value. Similar results have also been recently reported for assembly systems in Bollapragada, Rao, and Zhang (2001).

In the following sections, we show that, for our demand and supply model, the above formulation corresponds to a convex programming problem. Consequently, a subgradient-based search may be used to find the optimal stock levels. In each step of this search, we need to estimate the gradient and to ensure feasibility (service level requirement). Checking feasibility is relatively easy using simulation, however computation of the gradient is non-trivial (Wets 1989). Although simulation-based optimization approaches such as Infinitesimal Perturbation Analysis may be used to estimate the gradient, these approaches are complicated by "noise" (Bashyam and Fu 1998) in the service constraint and can be computationally intensive when the number of scenarios is large. Hence, our approaches will only require feasibility checks, without any gradient computation.

4 Problem Analysis

In this section, we present analytical properties and bounds which we use to compute optimal base-stock levels. In Section 4.1, we characterize the inventory level of the end-product in steady state. The analysis is similar to the case with constant (deterministic) supply lead-time, except that a random number of components (denoted by ζ) may not be delivered due to the supply uncertainty. To avoid confusion, our analysis will focus on the proposed model of supply uncertainty with guaranteed lead time $(L,\ L+1)$. For the current supply model, we can obtain similar results (see discussion at the end of this section).

4.1 System Properties

Let $\xi_{t,t+L} \equiv \xi_t + \cdots + \xi_{t+L-1}$ and let $\xi^{(L)}$ denote the cumulative demand over L periods. Let $I_t^c(\cdot)$ denote the component inventory level at the beginning of period t; the absence of subscript t will denote steady state quantities. We first characterize the delayed-assembly (Lemma 1), then we evaluate the inventory level of the end-product in steady state (Lemma 2). As is common practice, we use the notation $x^- = \max(0, -x)$ and $\wedge$ for the minimum operator.

Lemma 1 *The delayed-assembly of the end-product in steady state is* $D(S^c) = (I^c(S^c) - \xi)^- = [S^c - \zeta - \xi^{(L)}]^-$, *where* $\zeta \equiv \xi - \xi \wedge \eta$.

Proof. Since base-stock replenishment policies are followed, the demand observed by the component echelon is identical to end-product demand. Hence, the component inventory level at the beginning of period $t+L$, before demand materializes, is $I^c_{t+L}(S^c) = S^c - \xi_t + \xi_t \wedge \eta_t - \xi_{t+1,t+L}$. Since the demand in period $t+L$ is ξ_{t+L}, exactly ξ_{t+L} units of the component are needed to raise the inventory position of the end-product back to S^p. If $I^c_{t+L}(S^c) \geq \xi_{t+L}$, there is no "delayed-assembly" of end-products; otherwise the delayed-assembly due to component shortage is $\xi_{t+L} - I^c_{t+L}(S^c)$. Hence, the delayed-assembly is $(I^c_{t+L}(S^c) - \xi_{t+L})^-$. In steady state, $I^c(S^c) = S^c - \zeta - \xi^{(L-1)}$ and, thus, $D(S^c) = [S^c - \zeta - \xi^{(L)}]^-$. ■

Lemma 2 *In steady state, the inventory level of the end-product at the beginning of a period, before demand materializes, is* $I^p(S^c, S^p) = S^p - D(S^c) - \xi^{(l-1)}$,*where* $D(S^c)$ *is defined in Lemma 1. Thus the inventory level after demand materializes is* $IL^p(S^c, S^p) = I^p(S^c, S^p) - \xi = S^p - D(S^c) - \xi^{(l)}$.

Proof. Recall that, for a base-stock policy, the end-product installation stock just after the order epoch in period $t+L$ will be S^p. Since the assembly lead-time is l, the inventory level of the end-product at the beginning of period $t+L+l$ will be $S^p - D_{t+L}(S^c) - \xi_{t+L+1,t+L+l}$. Hence, in steady state, $I^p(S^c, S^p) = S^p - D(S^c) - \xi^{(l-1)}$. ■

By the above lemmas, we see that base-stock levels may be directly related to component and end-product inventory levels. Furthermore, each base-stock level consists of a cycle-stock portion (based on duration of the period), a pipeline stock portion (based on mean lead time), and a safety stock portion (due to uncontrollable factors such as supply uncertainty). From Lemma 2, the two-echelon serial system problem (SSP) may be formulated as follows:

$$\min \quad U(S^c, S^p) \qquad s.t. \quad P\{S^p - [S^c - \zeta - \xi^{(L)}]^- - \xi^{(l)} \geq 0\} \geq \alpha.$$

Let $A(S^c, S^p) = \{(S^c, S^p) | P\{IL^p(S^c, S^p) \geq 0\} \geq \alpha\}$ denote the feasible region for SSP. The following results will be useful in characterizing and computing optimal base-stock levels. Note that a probability measure is called quasi-concave if its probability density function is of the form $e^{-Q(\cdot)}$ with $Q(\cdot)$ being quasi-convex. Many common distributions, such as the Normal and Gamma, are quasi-concave.

Theorem 1 *(i) If the joint distribution of* ζ, $\xi^{(l)}$, *and* $\xi^{(L)}$ *is quasi-concave, then* $A(S^c, S^p)$ *is a closed convex set. (ii) The optimal base-stock vector* (S^{c*}, S^{p*}) *satisfies* $P\{IL^p(S^c, S^p) \geq 0\} = \alpha$.

Proof. *(i)* Since $S^c - \zeta - \xi^{(L)}$ is a concave function of S^c, $D(S^c) = [S^c - \zeta - \xi^{(L)}]^-$ is convex. Consequently, $IL^p(S^c, S^p)$ is concave in (S^c, S^p). Applying Proposition 3.1 of Wets (1989), $A(S^c, S^p)$ is a closed convex set. *(ii)* Follows from the fact that the objective function of SSP and $IL^p(S^c, S^p)$ are both continuous and increasing in (S^c, S^p), implying that any solution with $P\{IL^p(S^c, S^p) \geq 0\} > \alpha$ is suboptimal since it can be improved by reducing either S^c or S^p (or both). ■

4.2 Relationship Between S^c and S^p

In this section we explore how changes in the component stock level, S^c, affect the corresponding end-product stock level, S^p, and vice-versa. Proofs are included in Appendix A.

Theorem 2 *Let* $S^c(S^p) = \min\{S^c | P\{IL^p(S^c, S^p) \geq 0\} = \alpha\}$, *then* $S^c(S^p)$ *is decreasing and convex in* S^p. *Similarly,* $S^p(S^c) = \min\{S^p | P\{IL^p(S^c, S^p) \geq 0\} = \alpha\}$ *is decreasing and convex in* S^c.

The complementary nature of S^C and S^p in meeting customer service levels has been demonstrated in Boyaci and Gallego (2001). However, the convexity of $S^p(S^c)$ and $S^c(S^p)$ which results from the quasi-concavity assumption has not been seen. Theorem 2 also yields lower and upper bounds on S^{p*}, corresponding respectively to $S^p(\infty)$ and $S^p(0)$. These bounds on S^{p*} in turn define bounds on S^{c*}. Furthermore, we see that the optimal inventory investment objective is of the form $C^c S^c(S^p) + C^p S^p = C^c S^c + C^p S^p(S^c)$. Hence, the optimal values of S^c and S^p may be found using a one-dimensional search over, say, S^c.

Theorem 3 *(i)* $\frac{\partial S^c(S^p)}{\partial S^p} \leq -1$, *where* $S^c(S^p)$ *is defined as in Theorem 2. (ii) For* $(S^c, S^p) = (S^{c*}, S^{p*})$ *and a linear objective* $U(S^c, S^p)$, $\frac{\partial S^c}{\partial S^p} = -\frac{C^p}{C^c}$ *when* $C^p \geq C^c$.

This theorem shows that a decrease in the component's base-stock level, S^c, entails an increase in the product's base-stock level, S^p, so as to ensure that the same service level is obtained. Further, the decrease in S^c will be greater than the increase in S^p, and the total inventory will decrease. When $C^p \leq C^c$, the total inventory cost will also decrease, which yields the following result:

Corollary 1 *When $C^p \leq C^c$, $S^{c*} = 0, S^{p*} = S^{*(L+l)}$, where $S^{*(L+l)}$ is the optimal base-stock level for a single echelon system with supply uncertainty and lead time of $L + l$.*

The above proposition implies that we only hold stocks of the component when it is cheaper than the end-product. This fact has been observed by Gallego and Zipkin (1999) for serial systems with inventory holding and penalty costs and no customer service level requirement or supply uncertainty.

We conclude this section by considering the current model of supply uncertainty (capacitated supply with no lead time guarantee, described in Section 3.1). Using an approach similar to the one in Section 4.1, it is possible to show that the formulation defined in (1) with the current capacitated supply model remains a convex programming problem. Further, the steady state inventory levels (and shortfall) may be characterized using simple recursive equations; a convex decreasing relationship between optimal base stock levels remains valid. Finally, the problem may be solved using a subgradient-based search. In the next section, we implement our solution method and compute the optimal (S^{c*}, S^{p*}) for several instances of SSP. We also compare stock levels under the current and proposed supply environments, and present insights from a sensitivity analysis.

5 Computational Results

In this section, we first provide a computational comparison between the current and proposed supply models. Next, we study the effect of parameter changes on inventory positioning in our two-echelon system with the proposed L, $L + 1$ supply model. Computational studies on the effects of problem parameters in serial systems have been described in Glasserman and Tayur (1995) and Gallego and Zipkin (1999). In the following sections, we only report on issues which have not been effectively addressed before.

In all our experiments demand is modeled using a Gamma distribution with scale parameter $\lambda > 0$, shape parameter $w > 0$, and density

$$\phi(x) = \begin{cases} \frac{1}{\Gamma(w)} \lambda^w x^{w-1} e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\Gamma(w)$ is defined to make $\int_0^\infty \phi(x) = 1.0$. This demand distribution is quasi-concave and allows us to isolate the effects of the scale and shape parameters, which correspond respectively to the demand magnitude and its coefficient of variation. Note that the Gamma random variable has mean $\mu = w/\lambda$, variance $\sigma^2 = w/\lambda^2$, and coefficient of variation $cv \equiv \sigma/\mu = 1/\sqrt{w}$. By proper choice of w, we can obtain a wide range of cv values for demand (this should be contrasted with the commonly used Normal distribution for which the probability of negative demand becomes unacceptably high when $cv > 0.4$).

5.1 Effects of Supply Parameters & Supply Model

In this computational experiment, we varied several supply parameters for both the proposed L, $L+1$ supply model and the current capacitated supply model. Besides studying the effects of supply parameters, this experiment permits a direct comparison between the two supply environments. The main supply parameters include the quoted lead time L and the supply service level, $\gamma = P\{\xi \leq \eta\}$. In addition, we also consider the effects of uncertainty in available capacity η by modeling it using a gamma distribution. For each supply model, we ran a full-factorial experiment with L and l each taking on values 2, 4, or 8, unit costs $C^c = 1$ and $C^p = 2$, 4, or 8, gamma demand with mean 10 and $cv = 0.1$, 0.4, or 1.0, customer service levels $\alpha = 0.75$, 0.85, or 0.95, supply service level values of 0.7 or 0.9, and supply $cv = 0.1$, 0.4 or 1.0. We use a relatively high supply service level since the buyer has greater leverage than the supplier in our problem environment. Lower supply service levels translate into lower supply capacity η. Consequently, the inventory investment will be higher. When C^p is small (resp. large), this typically implies that inventories will shift towards holding more end-product (resp. component) inventories. Extremely low supply service levels correspond to very low supply capacity η, and consequently, the system without guaranteed lead time can be unstable (incapable of satisfying mean demand). In short, the data chosen above were considered representative, while ensuring system stability. For each supply model, a total of $2(3^6) = 1458$ problem instances were solved and analyzed, from which we draw the following conclusions regarding the behavior of total inventory investment and stock positioning.

1. *Inventory Investment*: From the computational results (summarized in Table 1), we noted that, for the current supply model, a 10% change in supply service level (or supply *cv*) resulted, on average, in a 4% (respectively 1%) change in the total inventory investment. For the proposed L, $L+1$ supply model the corresponding change in inventory investment was under 0.2%. Thus, the effect of supply service level γ and supply *cv* is relatively insignificant for the proposed L, $L+1$ supply model. This reaffirms the intuition that *guaranteed supply lead time (even with a single period delay permitted) goes far in mitigating the detrimental effects of highly variable, random supply capacity.*

Table 1: Impact on Inventory Investment of Problem Parameters for Current and Proposed Supply Models.

Supply Model	C^p/C^c	l	L	α	γ	demand *cv*	supply *cv*
Proposed	Medium	Medium	Low	High	V.Low	High	Negligible
Current	Medium	Medium	Low	High	High	High	Medium

For a 10% change in the parameter value, Low implies of the order of $<$ 2.5 units (1%) change in inventory investment, High is $>$ 10 units (4%) change, intermediate impact is Medium; V.Low implies much smaller than Low but statistically significant impact; Negligible implies statistically insignificant impact.

On average, the inventory investment for the proposed supply model was 5% lower than that for the current supply model. Over the tested problem instances, the relative benefits of the proposed supply environment, measured as a percentage reduction in inventory investment, ranged from nearly 0% up to 63%. In our application environment, benefits exceeding 50% was considered very significant. The benefits were greater than 50% only when the supply *cv* and service level were small (0.1 and 0.7), demand *cv* and customer service level were high (*cv*=1 and $\alpha \geq 0.85$) and lead times were low (2, or 4 with $C^p = 2$). The fact that the benefits are higher when supply *cv* is lower is somewhat counter-intuitive and possibly because under high supply *cv* both supply models carry higher inventories, consequently, the relative benefit[3] relative to

[3]Note that, as supply *cv* increases, the mean supply capacity $E[\eta]$ data is also increased to keep the

the higher inventory investment in the current system may be lower. The results, summarized in Table 2, indicate that, in negotiating new contracts with the supplier, *the manufacturer's inventory investment would be most reduced when the supply cv and service level are low and when demand cv and customer service level α are high.*

Table 2: % Benefits of Proposed Supply Model

Magnitude of Benefits	Supply & Demand Parameters and their Values			
	Supply *cv*	Supply Service γ	Demand *cv*	Customer Service α
High (>50%)	Low	Low	High	High
Low (<0.5%)	Medium	High, or Medium (with high l, L)	Low	Low, or Medium (with high supply *cv*)

The Low, Medium, and High parameter values correspond to the three tested values for each parameter. For instance, the Low, Medium, and High supply *cv* values were respectively 0.1, 0.4 and 1.0; High l, L corresponds to $l = L = 8$.

2. *Stock Levels and Positioning*: In a manner analogous to the above analysis of inventory investment, we studied the optimal base stock levels of components and end-products. As expected, *lead times had the most dominant impact on the corresponding stock levels.* Typically, *customer service level α and demand cv had the next most dominant impact on stock levels. This impact was different in the two supply models*: For the current capacitated supply model, demand *cv* had a much bigger impact on component inventory than on end-product inventory. For the proposed system, demand *cv* had a similar impact on component and end-product inventories and the magnitude of this impact was smaller than that in the current system. The change in end-product stock levels with α was quantitatively similar in both supply models. However, in the proposed system, α had about a 50% smaller impact on component stocks than on the end-product. Finally, it appears that due to the lead time guarantee, *the proposed supply model's optimal stock levels were fairly insensitive to changes in supply parameters such as supply service level γ and supply cv* since the impact was statistically insignificant (except for a small impact of supply *cv* on component stocks). In contrast, for the current supply system, γ had the second most dominant impact

supply service level at the desired level. This change in $E[\eta]$ sometimes has a more significant effect in reducing inventories in the current supply system than in the proposed supply system, and as a result the percentage benefit of the proposed system may even decrease with *cv*.

on component stock levels (in magnitude only smaller than the effect of supply lead times).

In comparing stock levels for the two models, we noted that component (end-product) stock levels in the current supply model were, on average, about 10% (2.5%) higher than in the proposed supply model, with a range of -21% to 68% (-7% to 71%). The general behavior of these differences with changes in problem parameters is similar to that of inventory investment. We note that the negative percentages above imply that *component inventory in the current model may sometimes be less than in the proposed model.* While this may seem counter-intuitive, it results from the fact that in all these cases the end-product inventory in the current system was larger (as was the total inventory investment).

To summarize, we note that *a supply lead time guarantee has a significant impact on the total inventory as well as its distribution in the system.* The computational testing permits us to address questions, such as, should the manufacturer prefer a supplier with high service level (mean capacity) but high supply *cv* or a low service level but low supply *cv*? Our experiments suggest that, for the current supply model, a supplier with greater mean capacity should be favored, even if the supply is more variable, since mean capacity (supply service level) has a more significant effect on stock levels. For the proposed supply model, this question is moot, since the effect of these supply parameters vanishes when the supplier offers and successfully implements a lead time guarantee. In the remainder of this section we investigate the effect of problem parameters in the proposed $L, L+1$ supply model.

5.2 Effect of Customer Parameters in the Proposed Supply Model

In this section, (1) we briefly discuss the effects of demand parameters (shape w and scale λ), (2) we explore the effects of different target customer service levels, α, on the optimal inventory distribution between components and end-products; (3) we investigate the effects of problem parameters on the *optimal* internal service level, β. Note that the internal service level (Type-1), β, is defined as the probability that the demand from the downstream stage (echelon) is satisfied using on-hand inventory at the upstream stage. Thus $\beta = P\{IL^c(S^c) \geq 0\}$, where $IL^c(S^c)$ is the steady state inventory level of the component stage after demand materializes. Given the optimal component and end-product base-stock levels, the corresponding *optimal* internal service level, β^*, is easily computed using Monte Carlo simulation. When a service requirement of β^* is communicated to the upstream echelon, the supplier will select a globally optimal component base-stock level and the manufacturer can then pick end-product stock levels optimally (by estimating the component shortfall distribution). Thus, under base-stock policies, the internal service level serves as a means to monitor and coordinate the two echelons of the supply chain (see Cachon 1999).

We generate all test problem instances starting with a base set of three problems (differing in the target service level α) with parameter values shown below:

- Cost: $C^c = 1$, $C^p = 2.0$.
- Lead-time: $L = 5$, $l = 4$.
- Demand: Shape parameter $w = 4$, scale parameter $\lambda = 0.4$.
- Service level: $\alpha = \{0.90, 0.95, 0.98\}$.
- Supplier service level: 0.90. For simplicity, we assume that the supplier's capacity is deterministic and fixed at C such that $P\{\xi \leq C\} = 0.90$. Based on preliminary experiments with Gamma distributed capacity, we saw that the deterministic capacity model provided qualitatively similar results without the confounding effects of supply uncertainty. Further, the emphasis in this section is on customer parameters, not supplier data, so we do not explore different levels of supplier service. The effects of random (gamma distributed) supply capacity have been considered in Section 5.1.

The data values above were considered representative of industrial problems; for instance, in our experience, customer service levels of 0.9 and 0.95 are commonly used. The results, summarized below, are illustrated using problem instances corresponding to the base problem set with the parameter of interest changed incrementally. However, our empirical observations are consistent with results obtained by solving over a thousand other problem instances.

Effects of Demand Parameters, shape w and scale λ. The effect of $cv = 1/\sqrt{w}$ was studied by assigning w values ranging from 0.1 to 4.0, in steps of 0.1, with the mean demand fixed at 10. Scale effects were studied by increasing λ from 0.1 to 2.0 in steps of 0.1, with $w = 4.0$. Our experiments confirm that items with low w (high cv) require substantially higher safety stock to provide the same level of service. However, *increase in optimal inventory investment was less-than-proportional to the increase in cv*. This may be explained by noting that the total inventory investment includes a portion corresponding to mean demand, which is not affected by changes in cv. We noted that the *optimal cost-weighted stock levels (component, end-product, and their total) are linear in the inverse of the scale parameter*, λ. This is consistent with previous research (Lariviere and Porteus 1999) in which demand scale drops out of the safety stock calculations.

Effects of External Service Level, α. We focus on how the service level constraint (in particular, the choice of customer service target α) affects inventory investment and its positioning under capacitated supply. We solved 24 problem instances generated from the base problem by varying the external service level α from 0.85 to 0.99 in steps of 0.01, and from 0.99 to .999 in steps of 0.001. We also solved other sets of 24 problems, for instance, changing the per unit cost of the end-product to 4 or the component lead-time to 5, and obtained similar results. Figure 1 demonstrates the effects of the external service level, α, on the total inventory cost (and its breakdown into components and end-products). We see that *as α increases, the total inventory cost is monotone increasing in a near-convex manner*; the marginal increase in the total inventory investment increases rapidly as α approaches 1. This result also applies to the traditional inventory holding + shortage penalty cost model, when parameterized by unit shortage cost, π. Although the component and end-product base-stock levels also show an increasing trend, this increase is *not* monotone, and is higher for the end-product. For $\alpha = 0.88$, 0.99, and 0.999, the end-product costs as a percentage of total inventory investment is 66%, 69% and 76% respectively. *For high α, the end-product becomes more important for satisfying customers' demand*, which may be a consequence of the fact that component supply is capacity constrained and component availability can not adequately protect against shortages during the assembly lead time. In other words, for high enough α, storing end-products provides more protection against shortages.

Consistent with Theorem 2, *inventory of components and of end-products exhibits a complementary role in maintaining the external service level.* As α increases, the increase in the optimal end-product (component) inventory is sometimes large enough to allow component (end-product) inventory to decrease. From Figure 1, we conclude that, *rather than focusing on individual items in the supply chain, inventory managers should set target levels for the total inventory investment* since the breakdown of this inventory into components and end-products is complex and not characterizable by simple rules of thumb based on input parameters such as the external service level. Note that there is a monotonic relationship between the optimal total inventory investment and the external service level, consequently, once the curve in Figure 1 is known, inventory managers may choose the appropriate cus-

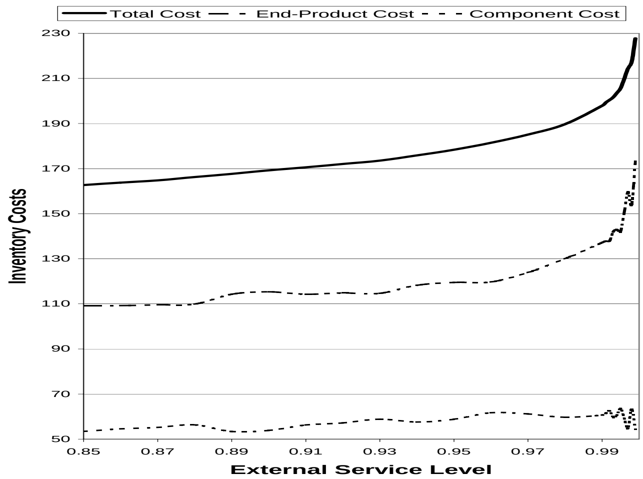

Figure 1: Effects of external service level on inventory costs

tomer service level as a means of setting a target on total inventory investment. In general, a quantitative model such as the one presented in this paper could be used as a guideline for setting external service level targets and for positioning inventory optimally.

The Optimal Internal Service Level, β. As discussed earlier, the inventory levels for components and end-products is, in practice, set independently. Component base-stock levels are calculated based on a pre-specified internal service level (which is set using the judgment and experience of inventory managers). In this context, we first study the sensitivity of the total inventory cost to the internal service level, β, then we study the effects of different problem parameters (such as unit cost, lead-time, demand parameters and external service level) on the *optimal* internal service level, β^* (corresponding to component stock levels minimizing inventory investment).

Figure 2 shows how the total inventory cost changes with β for the base problem with external service levels $\alpha = 0.9$, 0.95, and 0.98. For $\alpha = 0.9$, the optimal internal service level β^* is 0.64 and the corresponding inventory cost is 169.25. This β^* value of 64% is much lower[4] than the desired 90% external service level. From the figure we see that the total inventory cost is insensitive to the internal service level in the neighborhood of the

[4]In practice, we have observed that inventory managers set an internal service level higher than the desired external service level so as to "ensure" component availability. Our results suggest that such a conservative stocking policy can sometimes substantially increase inventory investment. Instead, we advocate communicating a higher service level to the supplier *for service within lead time*.

optimal internal service level. In fact, the total inventory cost for $\alpha = 0.9$ is within 1% of the minimum inventory cost for any internal service level between 0.51 and 0.85. As α increases, the "boat-shaped" inventory investment vs. β^* curve shifts upwards and the zone of insensitivity shrinks slightly and moves rightward (with higher β^* values). Similar results were obtained for other problems with different parameters (e.g., $C^p = 4$, $L = 2$ or 6): The width of the zone of insensitivity is not affected significantly by problem parameters. Typically, within this insensitivity zone, the right half of the boat-shaped trough is flat, the left half decreases with β in a gradual, near-linear manner. The whole boat-shaped curve moves up significantly as L increases; while the change with C^p is less perceptible. Overall, *the internal service level, β, appears to be a robust measure for decomposing a serial system* (specifying β determines component stock levels, then, the corresponding end-product stock level may be easily determined).

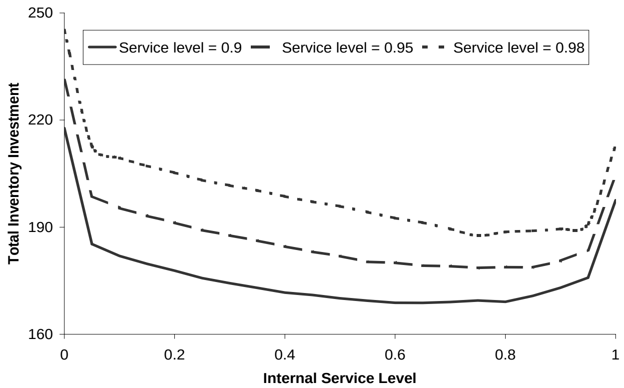

Figure 2: Effects of internal service level on total inventory investment

One explanation for the "boat-shaped" form of the inventory investment vs. β curve is the following: As β approaches 1.0, from elementary single-item, service-level constrained inventory models, we expect that the corresponding component stock level should increase dramatically. (This is akin to how the expected number of customers in a system increases dramatically when the system utilization approaches 1.0. It has been demonstrated for inventory systems using simulation and tail approximations of the shortfall distribution – see, for instance, Glasserman and Tayur 1995.) The corresponding reduction in the end-product stock level is relatively small, hence the total inventory investment increases dramatically as seen in the right portion of the curve in Figure 2. When β is very small, the corresponding component stock level S^c is small. From Theorem 2, we see that the end-product stock level is *convex* decreasing in S^c. Hence, when β approaches 0.0, a small decrease in β (and in S^c) translates into a relatively large increase in the corresponding end-product stock level which results in a higher inventory investment. For intermediate values of β, the end-product stock level decreases with S^c in a near-linear manner (whose slope is governed by factors such as the C^p/C^c ratio). Consequently, an increase in S^c with β is often compensated by an equivalent decrease in S^p so that the inventory investment value does not change significantly (and hence, this portion of the curve in Figure 2 remains flat).

We also investigated how the optimal β changes with α for the base problem instance and others (with different component lead-time L and unit cost C^p). From these experiments we concluded that *the external service level α is not a good indicator of the optimal β^* since the relation between α and β^* is not monotone.* This non-intuitive behavior may be because, as α increases, the end-product inventory generally shows a greater increasing trend, while the component inventory is often more stable. Hence significant increases in α produce little or no increase in β^*; sometimes, due to the nonlinearity in the customer service constraint, the increase in end-product inventory with α could correspond to decreasing β^*. We also observed that β^* increased with C^p/C^c according to a near step-functional form. An implication of this is that for a wide range of C^p/C^c values, the optimal internal service level β^* may not need to be changed. However, since a step change can occur sometimes, we advocate that the serial system problem be re-solved to evaluate the change in β^* caused by any changes in data.

Overall, our experiments refuted management's belief that it would be easy to develop simple heuristic rules for setting near-optimal component inventory levels (or equivalently choosing β) using input parameters such as α, lead times, and unit costs. However, the flatness of the curve in Figure 2 near the optimal β^* indicates that, although a suboptimal choice of β may adversely affect component inventory this will often be compensated by appropriate end-product stocks, consequently, the impact of suboptimal β on the total inventory investment is often likely to be small. However, the magnitude of this impact would depend on problem parameters such as the C^p/C^c ratio, the supply capacity η, and the degree of uncertainty.

The insensitivity of the objective function to the choice of β near β^* has managerial implications. From a manager's point of view, the internal service level is easier to track and manage compared to the base-stock levels which are typically set by the plant managers. Therefore, when the component and end-product are made at different facilities, the internal service level may be used to coordinate the production activities at the component level and at the end-product level. However when the two facilities belong to two independent companies, there may be incentive incompatibility issues arising from the fact that the supplier would prefer to have a small β while the manufacturer prefers a large β. Resolving this issue is beyond the scope of our research since it would require a new model where the transfer price of the component is an increasing function of β. This provides the supplier (manufacturer) a counterbalancing monetary incentive to agree to higher (lower) β.

6 Summary, Discussion, and Extensions

In this paper, we compute base-stock inventory levels under both supply and demand uncertainty in a two-echelon system with a customer service level constraint (instead of a shortage penalty cost). We developed two supply models and showed that, for the general class of quasi-concave demand distributions, the resulting problems are instances of convex programming; we characterized the convex decreasing relationship between the base-stock levels of components and end-products.

By comparing the current and proposed supply environments computationally (see Tables 1 and 2), we verified that guaranteed supply lead time (with a single period delay permitted) goes far in mitigating the detrimental effects of highly variable, random supply capacity. Further, the manufacturer's inventory investment would be most reduced with the proposed supply model when the supply cv and service level are low and when demand cv and α are high and it is not possible to reduce the assembly lead time. For the proposed $L, L+1$ supply environment, the optimal total inventory investment increases with the external service level in a near-convex monotone manner. Component and end-product inventories complement each other (one is convex-decreasing in the other). The total inventory investment is not sensitive to internal service level over a range near the optimal, however it increases rapidly beyond this range (particularly at higher customer service levels). Our parametric sensitivity analysis suggests that solution algorithms such as those developed in this paper may be necessary to cost-effectively trade-off component inventory versus end-product inventory.

Note that, throughout this paper, we restricted attention to base-stock policies, which may not be optimal under our model of supply. An alternative policy could be to use a more complicated order-up-to level that is a function of the installation stock (determining such optimal functions requires dynamic programming, but a simpler approximations could be used). Other options include using a policy that tracks more than the installation stock, for instance, the echelon stock level plus pipeline stocks and their expected delivery epochs, or policies incorporating batching, risk-pooling, and anticipated information on supplier capacity η_t. A complete analysis of such policies would be a challenging area for future research. We conclude this section by considering two simple extensions to our model, followed by potential future work.

6.1 Type-2 Service Level

In addition to Type-1 service level, another commonly used service level measure is Type-2 service level (or *fill rate*), which is the long-run average proportion of demands met from on-hand inventory. For the two-echelon serial system we discussed in Section 4.1, the expected backlog each period is $E[(\xi - I^p(S^c, S^p))^+]$, and hence the fill rate is

$$1 - E[(\xi - I^p(S^c, S^p))^+]/D,$$

where D is the mean demand per period. If δ denotes the desired Type-2 service level, the problem may be formulated the as

$$\min \quad U(S^c, S^p) \;\; s.t. \;\; E[(\xi - I^p(S^c, S^p))^+] \leq (1-\delta)D.$$

Similar to the case when a Type-1 service level is used, the feasible region for the above problem is a closed convex set. This follows from the fact that $I^p(S^c, S^p)$ is a concave function. Therefore the left hand side of the constraint is convex in (S^c, S^p). Having established the convexity of the feasible region, we can prove results similar to Theorems 2 and 3. It follows that, even with a Type-2 service level, we have a convex programming problem for which the solution approach presented in this paper remains valid.

6.2 Inventory Holding Cost Model

Our results apply to the case when inventory holding costs are considered instead of inventory investment. Let h^c and h^p denote the per unit holding cost per period for the component and the end-product respectively. Assume inventory costs are incurred after demand materializes. Then the problem may be formulated as,

$$\min\ U(S^c, S^p) \ \equiv\ h^c E[IL^c(S^c)]^+ + h^p E[IL^p(S^c, S^p)]^+ \ \ s.t.\ \ P\{IL^p(S^c, S^p) \geq 0\} \geq \alpha.$$

It is easy to verify that all our results, except Theorem 3(ii), are valid for the inventory holding cost objective. Further, $h^c E[IL^c(S^c)]^+ + h^p E[IL^p(S^c, S^p)]^+$ is increasing in (S^c, S^p) and $h^c E[IL^c(S^c(S^p))]^+ + h^p E[IL^p(S^c(S^p), S^p)]^+$ is unimodal in S^p, where $S^c(S^p)$ and $S^p(S^c)$ are defined in Theorem 2. Hence, a convex programming model and solution approach is applicable.

6.3 Future Work

A two-stage assembly system may be modeled in a manner similar to the serial system problem. Rosling (1989) has shown that certain assembly systems may be converted to an equivalent serial system. However, due to supply uncertainty, Rosling's conversion does not apply. Furthermore, finding the optimal stocking levels for all components and the end-product is likely to be computationally intensive. In Bollapragada, Rao, and Zhang (2001) we extend the two-echelon serial system results from Section 4.1 to two-stage assembly systems. We also show how multi-echelon systems (with more than two stages) may be converted to an equivalent two-stage assembly system. We are currently working on extending the models and approaches presented in this paper to more general assembly environments where common components are used to produce different products (the component commonality problem with product-specific service level constraints). The identification of quick and effective network decomposition approaches or good rules of thumb for inventory positioning in general supply chains would also constitute worthwhile and challenging research. Another interesting issue is to analyze the supplier's problem of satisfying demand from two customer classes, one with a $L, L+1$ delivery lead time guarantee and another without.

A Appendix

Proof of Theorem 2. We only show that when S^c increases, $S^p(S^c)$ is decreasing and convex in S^c. The monotonicity follows trivially from the fact that $I^p(S^c, S^p)$ is monotone in each argument. Convexity may be proven as follows: For any S^{c1} and S^{c2}, let $S^{p1} = S^p(S^{c1})$ and $S^{p2} = S^p(S^{c2})$. Hence $(S^{c1}, S^{p1}) \in A(S^c, S^p)$ and $(S^{c2}, S^{p2}) \in A(S^c, S^p)$.

For any θ, $0 \le \theta \le 1$, let

$$S^{p\theta} = S^p(\theta S^{c1} + (1-\theta)S^{c2}).$$

Since $A(S^c, S^p)$ is a convex set, $(\theta S^{c1} + (1-\theta)S^{c2}, \theta S^{p1} + (1-\theta)S^{p2}) \in A(S^{c1}, S^{c2})$. That is, $P\{I^p(\theta S^{c1} + (1-\theta)S^{c2}, \theta S^{p1} + (1-\theta)S^{p2}) \ge 0\} \ge \alpha$. Recall that $S^{p\theta} = \min\{S^p | P\{I^p(\theta S^{c1} + (1-\theta)S^{c2}, S^p) \ge 0\} = \alpha\}$. Combining the previous two equations, it follows that $S^{p\theta} \le \theta S^{p1} + (1-\theta)S^{p2}$. That is,

$$S^p(\theta S^{c1} + (1-\theta)S^{c2}) \le \theta S^p(S^{c1}) + (1-\theta)S^p(S^{c2}).$$

Hence $S^p(S^c)$ is convex in S^c. ∎

Proof of Theorem 3. Let $\Phi(.)$ be the CDF of $\xi - \xi \wedge \eta + \xi^{(L)}$ and $\Psi(.)$ be the CDF of $\xi^{(l)}$. By Lemma 2,

$$\begin{aligned} P\{I^p(S^c, S^p) \ge 0\} &= \Phi(S^c)\int_0^{S^p} d\Psi(y) + \int_0^{S^c}\int_{S^c}^{S^c+S^p-y} d\Phi(x)d\Psi(y); \\ &= \Phi(S^c)\Psi(S^p) + \int_0^{S^p} (\Phi(S^c+S^p-y) - \Phi(S^c))d\Psi(y); \\ &= \int_0^{S^p} \Phi(S^c+S^p-y)d\Psi(y). \end{aligned}$$

Thus, all feasible base-stock levels satisfy $\int_0^{S^p} \Phi(S^c+S^p-y)d\Psi(y) = \alpha$. Taking the derivative of both sides with respect to S^p, we get:

$$\Phi(S^c)\psi(S^p) + (1 + \frac{\partial S^c}{\partial S^p})\int_0^{S^p} \phi(S^c+S^p-y)d\Psi(y) = 0.$$

Since both $\Phi(S^c)\psi(S^p)$ and $\int_0^{S^p} \phi(S^c+S^p-y)d\Psi(y)$ are nonnegative and at least one of them must be positive (otherwise, $S^c = S^p = 0$), we must have $\frac{\partial S^c}{\partial S^p} \le -1$, which yields the first result.

To prove the second result, note that SSP is equivalent to

$$\min_{S^p} U(S^p) = C^c \cdot S^c(S^p) + C^p \cdot S^p.$$

Since $S^c(S^p)$ is convex in S^p, $U(S^p)$ is convex in S^p. Since $\lim_{S^p \to \infty} U(S^p) = \infty$ and $U(0) = \infty$ (the service level cannot be satisfied at $S^p = 0$), the minimum cost will be achieved at an S^p value such that $\frac{\partial U}{\partial S^p} = 0$, or $C^c \frac{\partial S^c}{\partial S^p} + C^p = 0$. Therefore, at optimality, $\frac{\partial S^c}{\partial S^p} = -\frac{C^p}{C^c}$. ∎

Acknowledgements: We thank the department editor Candi Yano and the anonymous referees; their comments helped improve the content and presentation of this paper.

References

Anupindi, R. and R. Akella (1993). Diversification under supply uncertainty. *Management Science 39*, 944–963.

Axsater, S. (1993). Continuous review policies for multi-level inventory systems with stochastic demand. In S. Graves, A. R. Kan, and P. Zipkin (Eds.), *Logistics of Production and Inventory*. North Holland.

Bashyam, S. and M. C. Fu (1998). Optimization of (s, S) inventory systems with random lead times and a service level constraint. *Management Science 44*, S243–S256.

Bollapragada, R., U. S. Rao, and J. Zhang (2001). A decomposition approach for managing assembly systems with demand and supply uncertainty. Working Paper, under second review at Management Science.

Boyaci, T. and G. Gallego (2001). Serial production/distribution systems under service constraints. *Manufacturing & Service Operations Management 3*, 43–50.

Cachon, G. (1999). Competitive supply chain inventory management. In S. Tayur, R. Ganeshan, and M. Magazine (Eds.), *Chapter 5, Quantitative methods for supply chain management*. Kluwer Academic Publishers.

Ciarallo, F. W., R. Akella, and T. E. Morton (1994). A periodic review production planning model with uncertain capacity and uncertain demand-optimality of extended myopic policies. *Management Science 40*, 320–332.

Clark, A. and H. Scarf (1960). Optimal policies for a multi-echelon inventory problem. *Management Science 6*, 475–490.

Diks, E., A. de Kok, and A. Lagodimos (1996). Multi-echelon systems: A service measure perspective. *European Journal of Operational Research 95*, 241–263.

Federgruen, A. (1993). Centralized planning models for multi-echelon inventory systems under uncertainty. In S. Graves, A. R. Kan, and P. Zipkin (Eds.), *Logistics of Production and Inventory*. North Holland.

Gallego, G. and P. Zipkin (1999). Stock positioning and performance estimation in serial production-transportation systems. *Manufacturing & Service Operations Management 1*, 77–88.

Glasserman, P. and S. Tayur (1995). Sensitivity analysis for base-stock levels in multi-echelon production-inventory systems. *Management Science 41*, 263–281.

Henig, M. and Y. Gerchak (1991). The structure of periodic review policies in the presence of random yield. *Operations Research 38*, 634–643.

Lariviere, M. and E. Porteus (1999). Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science 45*, 346–363.

Moinzadeh, K. and H. L. Lee (1989). Approximate order quantities and reorder points for inventory systems where orders arrive in two shipments. *Operations Research 37*, 277–287.

Rao, U. S., A. Scheller-Wolf, and S. Tayur (2000). Development of a rapid-response supply chain at caterpillar. *Operations Research 48*(2), 189–204.

Rosling, K. (1989). Optimal inventory policies for assembly systems under random demands. *Operations Research 37*, 565–578.

Song, J.-S., C. Yano, and P. Lerssrisuriya (2000). Contract assembly: Dealing with combined supply lead time and demand quantity uncertainty. *Manufacturing & Service Operations Management 2*(3), 287–296.

Wets, R. J. B. (1989). Stochastic programming. In G. Nemhauser, A. R. Kan, and M.J.Todd (Eds.), *Handbooks in Operations Research and Management Science, Vol 1 Optimization*, Chapter 8, pp. 573–630. Elsevier Science Publisher B.V.

Yano, C. A. and H. L. Lee (1995). Lot sizing with random yields: A review. *Operations Research 43*, 311–335.