Effective Dual Sourcing with a Single Index Policy

Alan Scheller-Wolf • Senthil Veeraraghavan • Geert-Jan van Houtum

Tepper School of Business, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

OPIM Department, Wharton School, University of Pennsylvania, 3740 Walnut St, Philadelphia, PA 19104, USA

Faculty of Technology Management, Eindhoven University of Technology, Eindhoven, The Netherlands 5600MB.

awolf@andrew.cmu.edu • senthilv@wharton.upenn.edu • G.J.v.Houtum@tm.tue.nl

Abstract

We consider a single stage periodic-review inventory system with backorders and dual sourcing: Replenishment can occur through a regular channel and/or a more expensive expedited channel with a smaller lead time. Globally optimal policies for these problems are in general highly complex; we therefore introduce a simple class of heuristic base stock policies. We call our class of policies the *single index* policies, as they require maintaining only one system statistic. We derive simple expressions for optimal single index parameters, enabling us to analytically find the optimal single index policy (by solving a newsvendor problem) in less than a second. Despite its simplicity, we establish computationally that the single index policy can be highly effective, especially on large problems. In this case it is comparable in performance to the more complex dual index heuristic, which has been shown in the literature to perform within a few percent of the optimal solution. Moreover, we demonstrate when dual sourcing can achieve significant savings compared to single sourcing, and establish that for our model single sourcing with the expedited supplier is *never* optimal.

Subject Classifications: Inventory: dual sourcing, options, leadtimes. Inventory: policies, heuristics, approximations.

1. Introduction

Many firms are trying to reduce costs while maintaining high levels of customer service by constructing supply chains with sourcing alternatives – either by using different suppliers, or different modes of delivery with one supplier. For example, Hewlett Packard uses this strategy for manufacturing servers (Beyer and Ward 2000), as does Caterpillar for compact worktools in their North American market (Rao, Scheller-Wolf and Tayur 2000). This latter paper, in fact, describes how Caterpillar uses single index policies as proposed here, combining trucking with air-shipping. Océ, a leading

manufacturer of document printing systems, faces a situation similar to Caterpillar and Hewlett Packard; they provide a good illustration of some of the driving factors behind contemporary dual sourcing strategies.

Océ produces office, production, and wide format printing systems in Europe for sale in the European, American and Asian markets. A key component of Océ's business relationship with their customers involves repair service: It is crucial that Océ be able to swiftly respond to (and remedy) any problems a customer is experiencing with one of their products. As such, spare parts management plays an important role in Océ's business strategy. Spare parts for Océ's American and Asian markets are stocked in warehouses in Chicago and Singapore, respectively. These stocks are replenished weekly (from Europe); replenishments for items with a high value density and/or a low demand rate travel by air, while items with a relatively low value density and sufficiently high demand rate travel by ocean. For these latter items, the pipeline holding cost for the longer travel time by sea (three weeks versus one week for air freight) and the costs for larger safety stocks are more than compensated for by the lower transportation costs. Occasionally though, air transport is used for this latter class of items, for example when unexpected peaks in demand occur. Océ thus uses dual transportation modes to control costs while meeting customer service targets.

These three companies, Hewlett Packard, Caterpillar and Océ, provide examples of the alternative sourcing problem as it confronts industry today. Unfortunately, whereas optimal inventory policies are known for quite general single-source models (see Tayur, Ganeshan and Magazine 1999), results are limited when even two sourcing alternatives are present. Thus new management techniques for chains with sourcing options are required: simple yet effective ways of choosing how much to source, when, and from whom.

Possibly motivated by this need, variants of the multiple supplier problem have seen much renewed interest in the literature. Zhang (1996), Gallego and Zhang (2003) and Feng, et al. (2003,2005) are all concerned with systems with more than two *consecutive* supply modes. For two supplier systems, Lawson and Porteus (2000) show "top-down base stock policies" to be optimal under the (somewhat restrictive) assumption that previously shipped units can be expedited or delayed at will. Moinzadeh and Nahmias (1988) approximate the optimal (Q, R) policy for a dual

sourcing inventory system under continuous review, with the condition that there will only be a single outstanding order of each type. Çakanyildirim and Luo (2005) extend the continuous review model, while allowing for postponement of deliveries of orders. Tagaras and Vlachos (2001), Teunter and Vlachos (2001), Vlachos and Tagaras (2001) and Groenevelt and Rudi (2003) focus on the model where the review period is much greater than the expedited lead time. This assumption prevents order crossing which facilitates analysis but does not accurately model many practical situations. (For a discussion of environments where order crossing is important, and its effects, see Robinson et al. 2001.) Veeraraghavan and Scheller-Wolf (2008) propose and evaluate the class of dual index policies, discussed briefly here as well, and Kiesmüller (2003) compares the use of single and dual index policies in a remanufacturing environment. Yazlali and Erhun (2005) model the dual supply contracting problem by combining options contracts with index policies for dual suppliers with consecutive lead times. Finally, Yi and Scheller-Wolf (2003), Sethi, Yan and Zhang (2003), and Fox, Metters and Semple (2005) find optimal policies for different models when setup costs are present.

In this paper we consider a periodic-review inventory system under full back-ordering, with no setup costs, a linear holding cost, and either a linear penalty cost or a γ-service level constraint. The system faces stochastic demand over an infinite horizon, with replenishment available from *two* supply modes. The two modes' lead times may be *any* integral multiples of the period (taken to be unit length, coinciding with the review epochs). These are the *only* restrictions placed on leadtimes. As globally optimal policies for problems of our type may be highly complex, we focus on finding optimal policies within a subclass of stationary *base stock policies*: If stocks fall below the *expedited level* an *expedited* order is placed to return them there. Then a *regular* order is placed to bring the total inventory up to the *regular level*(greater than equal to the *expedited level*). As the names indicate, the expedited order has a shorter lead time, and presumably a greater cost.

The base stock variant we analyze – the *single index policy* – makes both the regular and expedited ordering decisions based on *all* goods on hand, owed to customers, and on order. This is the optimal policy when leadtimes differ by one time unit (Fukuda 1964), but like all base stock policies is sub-optimal for the general leadtime case (Whittemore and Saunders 1977). Nevertheless,

we focus on this class of policies for two reasons:

1. When the difference in leadtimes is greater than one, expedited orders may arrive before regular orders that were placed earlier: *order crossing* may occur. Optimal policies for problems with order crossing are notoriously state-dependent; it is in this case that globally optimal policies for our model become complex, requiring the use of dynamic programming for their solution. These globally optimal policies depend on the entire vector of ordered, but undelivered items, and in general not suited for industrial implementation. Thus we seek an appropriate class of simple, yet effective heuristics.

2. Slightly more complex base stock policies than those we consider here, so-called dual index policies, make expedited ordering decisions based only on items that will arrive within the expedited lead time, and are known to perform well (Veeraraghavan and Scheller-Wolf, 2008). But finding optimal dual index parameters is non-trivial; for example it requires establishing a shortfall distribution via simulation, which is much more time consuming than the analytical procedure we derive for finding optimal single index parameters. In addition, our computational studies indicate that for problems with large state spaces (e.g. continuous demand), single index and dual index costs are comparable, with single index policy performance improving relatively as demand variability increases (see Table 4).

As alluded to above, in contrast to both the globally optimal policy and the dual index policy, finding optimal single index parameters is straightforward: single index policy costs decompose, admitting swift, sequential optimization via a newsvendor fractile. This is elementary for discrete demands. For continuous demands, specifically to address the need to easily convolve demands over multiple periods, we model demand as a mixture of Erlang distributions, which are closed under such convolutions. This enables us to find optimal single index parameters *in less than a second*. We note that this is a very general technique, as any continuous distribution on $(0, \infty)$ can be approximated arbitrarily closely by a mixture of Erlang distributions (see Schassberger, 1973; see also Tijms, 1986).

In summary, this paper makes the following contributions: (i) We formally define the class of

single index policies (in Section 2); (ii) We derive expressions for, and properties of, the optimal single index parameters for the penalty cost and γ-service level models, respectively (in Section 3); (iii) We develop a method to solve both the penalty cost and the service level problems with demand comprised of a mixture of Erlang distributions (in Section 4); (iv) After describing the dual index policy (in Section 5), for both the penalty cost and service level problems we computationally compare our single index policy against the dual index policy, single sourcing, and the globally optimal policy (found via dynamic programming, when tractable) (in Section 6). We discuss extensions in Section 7, and conclude in Section 8.

2. Model Definitions and Recursions

We consider a discrete time inventory system with expedited and regular sourcing. We use a single index inventory policy: one measure of inventory is tracked, the inventory position over the *entire* leadtime horizon. Each period, target levels are compared with quantity on hand, plus all outstanding regular or expedited orders, minus any items owed to customers. If the inventory position is below the expedited target level, an expedited order is placed to bring the inventory position to this level. Then a regular order is placed to bring the final inventory position up to the regular target level. We assume excess demand is backordered and the model allows for a linear penalty cost (Section 3.1) or a γ-service level constraint (Section 3.2).

We define problem parameters:

$\underline{n}$: Period index, $n \geq 0$.

$\underline{d_n, F}$: New customer demand in period n, $\{d_n : n \geq 0\}$, form a stationary and iid sequence. This family of random variables is generically referred to as d, with distribution function F. We assume $\mathrm{E}[d] \stackrel{\text{def}}{=} \mu < \infty$ and the standard deviation of $d \stackrel{\text{def}}{=} \sigma < \infty$.

$\underline{l_r, l_e, l}$: l_r and l_e are the nonnegative deterministic lead times for regular and expedited orders, respectively. We define $l \stackrel{\text{def}}{=} l_r - l_e \geq 0$, with the convention that vacuous summations (when $l = 0$) return zero.

c_r, c_e, c: c_r and c_e are the nonnegative unit ordering costs for regular and expedited orders, respectively. We define $c \stackrel{\text{def}}{=} c_e - c_r > 0$. If $c_r \geq c_e$ expediting all orders is optimal.

h, p: Strictly positive per-period unit holding and backorder cost, respectively.

B: Maximum permitted average backlog at the end of a period, $B \in (0, \mu)$. This defines a γ-service level: $\gamma = 1 - (B/\mu) \in (0, 1)$.

The γ-service level behaves very similarly to the more common β-service level (the average fill rate, or proportion of customers served immediately). These two service levels differ only along sample paths in which the same item stocks out for more than one consecutive period; the β-service level counts this once, the γ-service level counts it once for each period it is backlogged. Thus the significantly easier-to-analyze γ-service level provides a lower bound for the β-service level, and under FIFO service, the measures diverge only when a period starts with a backlog and then the *entire* next period's demand is left unsatisfied. At high service levels, as we consider here, such events occur very rarely, making these two service measures very similar. When the objective was 99% γ-service, we found the maximum deviation of the β-service level was 0.2% higher than γ-service level. In general, over all of our instances at 95% and 99%, β-service level was on average 0.6% higher, with a maximum deviation of 2%. Thus aiming for high γ-service level ensures higher β-service levels.

And our decision variables:

z_r, z_e, Δ: z_r and z_e are the regular and expedited order-up-to parameters, respectively. We let $\Delta \stackrel{\text{def}}{=} z_r - z_e \geq 0$, as any policy with $z_r < z_e$ is equivalent to policy $(\tilde{z}_r, z_e)$ with $\tilde{z}_r = z_e$.

And finally our system variables:

I_n: Inventory level at the start of period n; the amount on hand or on back-order.

IP_n: Inventory position at the start of period n; inventory level plus all goods on order.

X_n^r, X_n^e: Regular and expedited orders placed in period n, respectively.

In period n orders X_n^e, X_n^r are placed, a shipment of $X_{n-l_e}^e + X_{n-l_r}^r$ is received, demand d_n is revealed, and customers are satisfied. Costs are then assessed, and any unsatisfied customers or excess inventory is carried to the next period.

2.1 System Evolution

Recall that under the single index policy, in period $n-1$, X_{n-1}^e and X_{n-1}^r bring IP_{n-1} up to z_r, and then d_{n-1} occurs. Thus, at the start of period n, for $n > 0$, $IP_n = z_r - d_{n-1} = z_e + \Delta - d_{n-1}$.

If $IP_n < z_e$, an expedited order to make up this difference is placed:

$$X_n^e = (z_e - IP_n)^+ = (z_e - (z_e + \Delta - d_{n-1}))^+ = (d_{n-1} - \Delta)^+. \tag{1}$$

Then a regular order is placed to bring the inventory position up to z_r:

$$X_n^r = (z_r - (IP_n + X_n^e))^+ = (z_r - (z_r - d_{n-1} + (d_{n-1} - \Delta)^+)^+ = \min(\Delta, d_{n-1}). \tag{2}$$

Formulae (1) and (2) specify that in any period the portion of demand that exceeds Δ will be reordered by expedited delivery, and the rest – no more than Δ – will be reordered by regular delivery; large single period demands trigger expediting, see Figure 1. Note that substituting $\Delta = 0$ or $\Delta = \infty$ into (1) and (2) shows that in these cases the single index policy reduces to one with a single expedited or regular supplier, respectively.

Under a single index policy, once the problem is recast using Δ the behavior of the complex dual sourcing system – with order crossing – is much easier to understand. Specifically, the effect of expediting is made clear: the reaction time to unusually large demands (those greater than Δ) is reduced from l_r to l_e for those units exceeding Δ. Moreover, the long-run proportion of demand filled via the expedited channel can be obtained analytically:

$$\mathrm{E}[X^e] = \frac{\mathrm{E}[(d - \Delta)^+]}{\mu}. \tag{3}$$

Thus, if the percentage of expedited sourcing is bounded contractually, this is easily translated into a constraint on Δ.

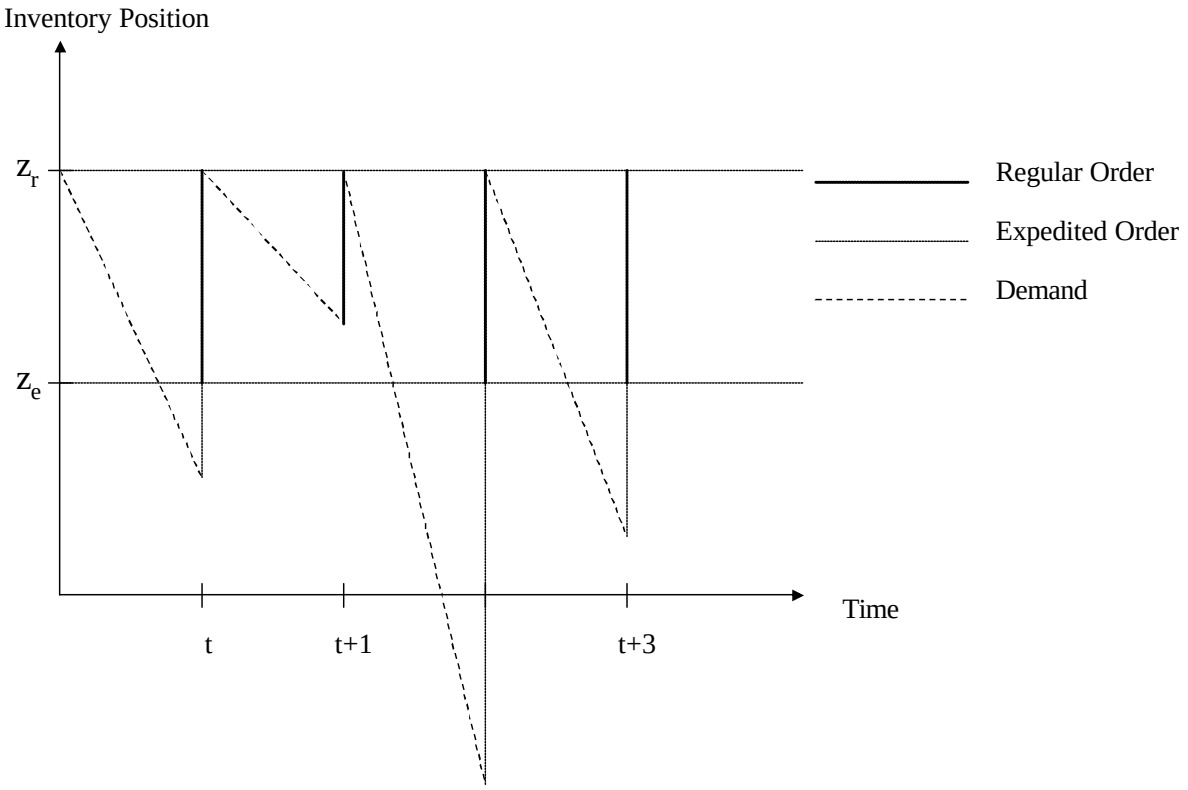

Figure 1: Demand stream with corresponding regular and expedited orders.

3. Analysis

Having simplified the single index policy through the use of Δ, we derive expressions for optimal solutions for the penalty cost problem in Section 3.1, and the service level problem in Section 3.2. Building on these, we provide limiting and bounding behavior for both cases, in Section 3.3.

3.1 Penalty Cost Problem

In this subsection we consider the linear penalty cost model; a fixed amount is charged at the end of each period for each unit of customer demand outstanding.

We begin with an explicit cost formulation. Assume without loss of generality that $I_0 = z_r$; then for periods $0 \leq i \leq N$ an arbitrary (Δ, z_r) pair has cost $\sum_{i=1}^{N} g_i(\Delta, z_r) \stackrel{\text{def}}{=} Y_N + Z_N$, where

Y_N captures ordering costs and Z_N captures holding and penalty costs:

$$Y_N \stackrel{\text{def}}{=} c_r \sum_{i=0}^{N-1} d_i + (c_e - c_r) \sum_{i=1}^{N} X_i^e; \qquad Z_N \stackrel{\text{def}}{=} h \sum_{i=1}^{N} I_i^+ + p \sum_{i=1}^{N} I_i^-.$$

Thus our infinite horizon problem is:

$$\min_{\Delta, z_r} \left\{ \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} g_i(\Delta, z_r) \right\} = \min_{\Delta, z_r} \left\{ \lim_{N\to\infty} \frac{Y_N + Z_N}{N} \right\},$$

with the system evolving as in (1) and (2). For simplicity, we define the time average ordering and inventory costs, $\lim_{N\to\infty} \frac{Y_N}{N}$ and $\lim_{N\to\infty} \frac{Z_N}{N}$, as $\mathrm{E}[Y]$ and $\mathrm{E}[Z]$ respectively. A coupling argument can be used to show that these costs converge over the infinite horizon:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} g_i(\Delta, z_r) = c_r\mu + (c_e - c_r)\mathrm{E}[X^e] + h\mathrm{E}[I^+] + p\mathrm{E}[I^-] = \mathrm{E}[Y] + \mathrm{E}[Z] \tag{4}$$

where random variables without subscripts refer to stationary versions.

To facilitate derivation of our analytical results, using $I_0 = z_r$, (1) and (2) we can establish:

$$\begin{aligned} I_n &= z_r - \sum_{i=n-l_r-1}^{n-1} d_i + \sum_{j=n-l_r}^{n-l_e-1} X_j^e \\ &= z_r - \sum_{i=n-l_e-1}^{n-1} d_i - \sum_{i=n-l_r-1}^{n-l_e-2} \min(d_i, \Delta) \\ &\stackrel{\text{dist}}{=} z_r - D(\Delta), \quad \forall n > l_r \end{aligned} \tag{5}$$

for the random variable $D(\Delta)$:

$$D(\Delta) \sim \sum_{i=1}^{l_e+1} d_i + \sum_{i=l_e+2}^{l_r+1} \min(d_i, \Delta). \tag{6}$$

Equation (6) illustrates the effect of expediting: In exchange for the marginal expediting cost it enables the manager to "truncate" (to Δ) the outstanding orders in each period between the expedited and regular leadtimes. Thus the value of the expediting option becomes more powerful as the efficacy of this truncation grows, i.e. as the difference between the leadtimes becomes larger or as demand becomes more variable. We will see this illustrated in our experiments in Section 6.

Having established the above relations, we are now ready to present the main results of this subsection: Defining our policy in terms of z_r and Δ we can decompose our problem, transforming it such that *optimal* single-index parameters can be found by applying standard newsvendor results.

Lemma 3.1 *Along any sample path, for all n:*

(i) Y_n is determined solely by Δ, independent of z_r.

(ii) For any fixed Δ, Z_n is determined solely by z_r.

(iii) For fixed Δ, the minimum value satisfying $z_r(\Delta) \stackrel{\text{def}}{=} F^{-1}_{D(\Delta)}\left(\frac{p}{p+h}\right)$ yields an optimal z_r, where $F_{D(\Delta)}$ is the cumulative distribution function of the random variable $D(\Delta)$ in (6).

Proof :

(i) From the definition of Y_n, along any sample path the first term $c_r \sum_i d_i$, is fixed. This leaves $(c_e - c_r)\sum_i X_i^e$, where $X_n^e = (d_{n-1} - \Delta)^+$, due to (1).

(ii) Due to its definition, Z_n is solely a function of $\{I_i : 1 \leq i \leq n+1\}$. Given a fixed Δ and a fixed sample path of demands, (5) shows that I_n is a function of z_r.

(iii) Defining $D(\Delta)$ as in (6), (5) shows that $I_n \sim z_r - D(\Delta)$ for all n. Our problem thus reduces to an infinite horizon news-vendor problem, the solution to which is the given "critical fractile." (See for example Tayur, Ganeshan and Magazine 1999). ■

Part (i) of Lemma 3.1 implies that for any fixed Δ, minimizing the expected cost for a single index policy reduces to minimizing $\mathrm{E}[Z]$. Parts (ii) and (iii) show that this minimization is achieved by finding the optimal $z_r(\Delta)$ via the critical fractile of $D(\Delta)$. Following from Lemma 3.1 we have the following Theorem:

Theorem 3.1 *The following procedure may be used to find the optimal single index policy for the dual sourcing problem with linear penalty cost:*

1. *Choose an initial Δ.*
2. *For the given Δ find an optimal $z_r(\Delta)$: $z_r(\Delta) = F^{-1}_{D(\Delta)}\left(\frac{p}{p+h}\right)$, where $D(\Delta)$ is given by (6).*
3. *Find the cost of the $(\Delta, z_r(\Delta))$ pair, using Equations (3) - (5).*
4. *Update Δ according to some search procedure, and go to 2.*

3.2 Service Level Problem

Now we consider the service level problem, i.e. one having a maximum permissible level for average customer backlog, B. To facilitate analysis, we will assume strictly continuous demand, $0 < F(x) < 1$ for all $x \in (0, \infty)$ and $F(0) = 0$, but unless otherwise noted, the results of the next two subsections hold under *general* demand distributions.

By repeating the arguments preceding (4), and then using (1), the fact that $I \sim z_r - D(\Delta)$, and $\mathrm{E}[D(\Delta)] = (l_r + 1)\mu - l\mathrm{E}[(d - \Delta)^+]$, the service-level problem can be formulated as the following non-linear minimization problem:

$$\begin{aligned} \min \quad & g_s(\Delta, z_r) \stackrel{\text{def}}{=} c_r\mu + c\mathrm{E}[(d - \Delta)^+] + h\mathrm{E}[(z_r - D(\Delta))^+] \\ s.t \quad & \mathrm{E}[(D(\Delta) - z_r)^+] \leq B \ . \end{aligned} \tag{7}$$

We sharpen this formulation and investigate its elementary properties below:

Lemma 3.2 *For the service level problem, given a fixed Δ:*

(i) The objective function g_s is constant for $z_r \leq 0$ and strictly increasing for $z_r > 0$.

(ii) The average backlog $E[(D(\Delta) - z_r)^+]$ is strictly decreasing in z_r; $E[(D(\Delta) - z_r)^+] \uparrow \infty$ as $z_r \to -\infty$; $E[(D(\Delta) - z_r)^+] \downarrow 0$ as $z_r \to \infty$.

(iii) There is a unique finite positive value, $z_r(\Delta)$, for which (7) is satisfied at equality. At optimality this equality holds: $z_r = z_r(\Delta)$.

Proof : Parts (i) and (ii) follow from our demand assumptions. The first statement of part (iii) follows from part (ii), continuous demand, $B \in (0, \mu)$, and the continuity of $\mathrm{E}[(D(\Delta) - z_r)^+]$ with respect to z_r. This, and part (i) imply the rest of (iii). ■

Given part (iii) of Lemma 3.2, the problem that remains is to minimize

$$\begin{aligned} g_s(\Delta, z_r(\Delta)) \quad &= \quad c_r\mu + c\mathrm{E}[(d - \Delta)^+] + h\mathrm{E}[(z_r(\Delta) - D(\Delta))^+] \ , \qquad \Delta \geq 0 \\ s.t \quad & \mathrm{E}[(D(\Delta) - z_r(\Delta))^+] = B \ . \end{aligned} \tag{8}$$

3.3 Properties of Optimal Single Index Policies

In this section we derive some properties of optimal single index policies: We establish a lower bound on the optimal Δ; using this bound we illustrate some limiting behavior of single index policies; we then detail relations between the single index policy and single sourcing; and finally we illustrate relations between z_r and Δ.

We first derive our lower bound on the optimal value of Δ for the service level model. We can then apply Theorem 1 of van Houtum and Zijm (2000) to extend these results to the penalty cost model when p is sufficiently large to ensure $B < \mu$, (which practically speaking is always).

Lemma 3.3 *Any* $\Delta < \Delta_{min} \stackrel{\text{def}}{=} F^{-1}\left(\frac{c}{c+hl}\right)$ *cannot be an optimal solution to the service level problem; an optimal* Δ *must fall within the range* $[\Delta_{min}, \infty)$.

Proof : Follows from Lemma 9.4, which appears in the Appendix. ■

We also have the following properties which follow from Lemma 3.3.

Corollary 3.1 *For an optimal single index policy:*

(i) For $c > 0$, $\Delta^* > 0$*; using the expedited supplier alone will <u>never</u> be optimal.*

(ii) $c \to \infty \;\Rightarrow\; \Delta_{min} \to \infty \;\Rightarrow\; \Delta^* \to \infty$*; as the relative cost of using the expedited supplier grows, the proportion of expedited goods approaches zero.*

(iii) $l \to 0 \;\Rightarrow\; \Delta_{min} \to \infty \;\Rightarrow\; \Delta^* \to \infty$*; as the difference between leadtimes shrinks the proportion of expedited goods approaches zero.*

Leveraging the results of Lemmas 3.1 and 3.2, and using (6), we can state the following fundamental properties of single index policies:

Lemma 3.4 *For an optimal single index policy:*

(i) As $\Delta \to \infty$, $D(\Delta) \to D_r$ *from below, where*

$$D_r \sim \sum_{j=1}^{l_r+1} d_j; \tag{9}$$

$z_r(\Delta)$ *converges from below to* $z_r(\infty) = z_r^{max} \stackrel{\text{def}}{=} F_{D_r}^{-1}\left(\frac{p}{p+h}\right)$ *for the penalty cost problem, and to* $z_r(\infty) \stackrel{\text{def}}{=} z_r^{max} = \{x \in \mathbf{R} \mid E[(D_r - x)^+] = B\}$, *for the service level problem. This corresponds to single sourcing via the regular supplier.*

(ii) Setting $\Delta = 0$, $D(\Delta) \sim D_e$, *where*

$$D_e \sim \sum_{j=1}^{l_e+1} d_j; \tag{10}$$

$z_e(\Delta)$ *converges from above to* $z_r(0) = z_r^{min} = F_{D_e}^{-1}\left(\frac{p}{p+h}\right)$ *for the penalty cost problem, and to* $z_r(0) \stackrel{\text{def}}{=} z_r^{min} = \{x \in \mathbf{R} \mid E[(D_e - x)^+] = B\}$, *for the service level problem. This corresponds to single sourcing via the expedited supplier.*

(iii) $z_r(\Delta)$ *is a non-decreasing function; as* Δ *increases expediting becomes less common and* $z_r(\Delta)$ *increases to compensate.*

Note that properties (i) and (ii) prove that the costs from the single source models serve as upper bounds for the optimal single index policy cost.

4. Solution Procedure

We now describe our exact solution procedure for the γ-service level problem under mixed Erlang demand. Our procedure extends straightforwardly to the penalty cost problem, and a comparable exact procedure can, in principle, be developed for the β-service level, although the current procedure is a very good approximation to the β-service level problem at high service levels.

Section 4.1 provides a detailed expression for (8), Section 4.2 describes how to evaluate this expression, and finally Section 4.3 combines these results into our general solution procedure.

4.1 Determination of $z_r(\Delta)$

We first focus on the determination of $z_r(\Delta)$ for a given $\Delta \geq 0$. Lemma 3.2 implies that $z_r(\Delta)$ can be computed by a bisection search provided that $\mathrm{E}[(D(\Delta) - x)^+]$ can be evaluated for any $x \geq 0$. Thus we describe how to evaluate $\mathrm{E}[(D(\Delta) - x)^+]$.

Recall that the random variable $D(\Delta)$ is equal to the sum of $l_e + 1$ demands d and l *truncated demands* $\hat{d} \stackrel{\text{def}}{=} \min\{d, \Delta\}$. We show in Lemma 4.1 that the distribution function of such a sum may be written as a linear combination of shifted distribution functions of sums of regular demands d and *residual demands* $\tilde{d} \stackrel{\text{def}}{=} (d - \Delta | d \geq \Delta)$. To support this result, we define the random variable $\hat{Y}_{m,n}$ as the sum of $m \in \mathbf{N}_0 \stackrel{\text{def}}{=} \mathbf{N} \cup \{0\}$ truncated demands $\hat{d}$ and $n \in \mathbf{N}_0$ regular demands d. Similarly, $\tilde{Y}_{m,n}$ is defined as the sum of $m \in \mathbf{N}_0$ residual demands $\tilde{d}$ and $n \in \mathbf{N}_0$ regular demands d. The corresponding distribution functions are defined by $\hat{G}_{m,n}$ and $\tilde{G}_{m,n}$.

Lemma 4.1 *For all $m \in \mathbf{N}_0$ and $n \in \mathbf{N}_0$:*

$$\hat{G}_{m,n}(x) = \sum_{s=0}^{m} (-1)^s \binom{m}{s} \overline{p}^s \sum_{i=0}^{s} (-1)^i \binom{s}{i} \tilde{G}_{s-i,m+n-s}(x - s\Delta) \ , \qquad x \in \mathbf{R},$$

where $\overline{p} \stackrel{\text{def}}{=} \mathbf{P}\{d \geq \Delta\}$.

The proof of this lemma is presented in the unabridged version of the paper.

Now we move to $\mathrm{E}[(D(\Delta) - x)^+]$. We first define, for each $m, n \in \mathbf{N}_0$, $\hat{K}_{m,n}(x) = \mathrm{E}[(\hat{Y}_{m,n} - x)^+]$ and $\tilde{K}_{m,n}(x) = \mathrm{E}[(\tilde{Y}_{m,n} - x)^+]$, $x \in \mathbf{R}$. Then, since $\hat{Y}_{0,0} = \tilde{Y}_{0,0} = 0$,

$$\hat{K}_{0,0}(x) = \tilde{K}_{0,0}(x) = \begin{cases} -x & \text{if } x < 0; \\ 0 & \text{if } x \geq 0; \end{cases}$$

and for all $m, n \in \mathbf{N}_0$ with $m \geq 1$ or $n \geq 1$, we find by Lemma 4.1 that

$$\hat{K}_{m,n}(x) = \sum_{s=0}^{m} (-1)^s \binom{m}{s} \overline{p}^s \sum_{i=0}^{s} (-1)^i \binom{s}{i} \tilde{K}_{s-i,m+n-s}(x - s\Delta) \ , \qquad x \in \mathbf{R}. \tag{11}$$

Clearly, $D(\Delta) = \hat{Y}_{l,l_e+1}$ and thus $\mathrm{E}[(D(\Delta) - x)^+] = \hat{K}_{l,l_e+1}(x)$, $x \geq 0$. Hence, by (11), evaluating $\mathrm{E}[(D(\Delta) - x)^+]$ for any $x \geq 0$ is reduced to evaluating the functions $\tilde{K}_{m,n}(x)$. We describe this operation below for mixed Erlang demand, in (12).

4.2 Closed-form Approximations via Erlang Mixtures

Henceforth we assume the demand distribution F is a mixture of Erlang distributions with the same scale parameter (demand rate per stage); i.e. there is a discrete distribution on $\mathbf{N}$, $\{q_k\}_{k \in \mathbf{N}}$, and a $\lambda > 0$ such that $F(x) = \sum_{k=1}^{\infty} q_k E_{k,\lambda}(x)$, $x \in \mathbf{R}$, where $E_{k,\lambda}$ denotes the Erlang distribution

with $k \in \mathbf{N}$ phases and scale parameter $\lambda > 0$. This assumption is motivated by the property that the class of mixtures of Erlang distributions with the same scale parameter is dense in the class of all distributions on $[0, \infty)$ (cf. Schassberger 1973). In general a mixture of an infinite number of Erlangs is needed for an exact approximation, but for practical purposes it is common to approximate a distribution by fitting the first two moments. We assume the latter if the demand distribution is not given as a mixed Erlang.

When $c_{var} \stackrel{\text{def}}{=} \sigma/\mu \leq 1$, a mixture of an Erlang$(k_0 - 1, \lambda)$ and an Erlang(k_0, λ) distribution is fitted (cf. Tijms 1986). We choose k_0 such that $\frac{1}{k_0} < c_{var}^2 \leq \frac{1}{k_0 - 1}$ (notice that $k_0 \geq 2$). Next we choose $q_{k_0-1} = \frac{1}{1+c_{var}^2}\left[k_0 c_{var}^2 - \{k_0(1+c_{var}^2) - k_0^2 c_{var}^2\}^{0.5}\right]$, $q_{k_0} = 1 - q_{k_0-1}$, and $q_k = 0$ for all other k. Finally, $\lambda = \frac{k_0 - q_{k_0-1}}{\mu}$.

When $c_{var} > 1$, we fit a mixture of an Erlang$(1, \lambda)$ and an Erlang(k_0, λ) distribution. We choose k_0 such that $k_0 \geq 3$ and $\frac{k_0^2+4}{4k_0} \geq c_{var}^2$, taking the smallest k_0 which satisfies this inequality. Next we choose $q_1 = \frac{2k_0 c_{var}^2 + k_0 - 2 - (k_0^2 + 4 - 4kc_{var}^2)^{0.5}}{2(k_0-1)(1+c_{var}^2)}$, $q_{k_0} = 1 - q_1$, and $q_k = 0$ for all other k. Finally, $\lambda = \frac{q_1 + k_0(1-q_1)}{\mu}$.

When F is mixed Erlang, $\tilde{F}$ is also a mixture of Erlang distributions with scale parameter λ, but with mixture probabilities $\{\tilde{q}_k\}_{k \in \mathbf{N}}$. Define for all $k \in \mathbf{N}$, $j = 0, 1, \ldots, k$,

$$
\begin{aligned}
r_{k,j} &\stackrel{\text{def}}{=} \mathbf{P}\{j \text{ phases left after } \Delta \text{ time units} \mid d \text{ is Erlang}(k, \lambda) \text{ distributed}\};\\
r_{k,j} &= \frac{(\lambda\Delta)^{k-j}}{(k-j)!} e^{-\lambda\Delta}\ , \qquad j = 1, \ldots, k,\\
r_{k,0} &= \sum_{i=k}^{\infty} \frac{(\lambda\Delta)^i}{i!} e^{-\lambda\Delta} = 1 - \sum_{i=0}^{k-1} \frac{(\lambda\Delta)^i}{i!} e^{-\lambda\Delta} = E_{k,\lambda}(\Delta)\ ,\\
\tilde{q}_j &= \mathbf{P}\{j \text{ phases left after } \Delta \text{ time units} \mid \text{at least 1 phase left}\}\\
&= \frac{\mathbf{P}\{j \text{ phases left after } \Delta \text{ time units}\}}{\mathbf{P}\{d > \Delta\}} = \frac{1}{\overline{p}} \sum_{k=j}^{\infty} q_k r_{k,j}\ , \qquad j \in \mathbf{N}.
\end{aligned}
$$

Since both F and $\tilde{F}$ are mixtures of Erlang distribution with scale parameter λ, all distributions $\tilde{G}_{m,n}$, $m, n \in \mathbf{N}_0$ with $m \geq 1$ or $n \geq 1$, are also mixtures of Erlang distributions with scale parameter λ. Let their mixture probabilities be denoted by $\{\tilde{q}_k^{(m,n)}\}_{k \in \mathbf{N}}$. Then $\tilde{q}_k^{(1,0)} = \tilde{q}_k$ for all $k \in \mathbf{N}$, and for each $m \geq 2$: $\tilde{q}_k^{(m,0)} = \sum_{i=1}^{k-1} \tilde{q}_i \tilde{q}_{k-i}^{(m-1,0)}$, $k \in \mathbf{N}$, reading zero for the sum when $k = 1$. Further, $\tilde{q}_k^{(0,1)} = q_k$ for all $k \in \mathbf{N}$, and for each $n \geq 2$: $\tilde{q}_k^{(0,n)} = \sum_{i=1}^{k-1} q_i q_{k-i}^{(0,n-1)}$, $k \in \mathbf{N}$. For each

$m \geq 1$ and $n \geq 1$: $\tilde{q}_k^{(m,n)} = \sum_{i=1}^{k-1} \tilde{q}_i^{(m,0)} q_{k-i}^{(0,n)},\ k \in \mathbf{N}$.

Since all $\tilde{G}_{m,n}$ are mixed Erlangs, the following closed-form expression is found for the functions $\tilde{K}_{m,n}$ for all $m, n \in \mathbf{N}_0$ with $m \geq 1$ or $n \geq 1$:

$$\tilde{K}_{m,n}(x) = \sum_{k=1}^{\infty} \tilde{q}_k^{(m,n)} \left(\frac{k}{\lambda}(1 - E_{k+1,\lambda}(x)) - x(1 - E_{k,\lambda}(x)) \right), \qquad x \in \mathbf{R}. \tag{12}$$

This expression in combination with (11) is sufficient to determine $z_r(\Delta)$ for a given $\Delta \geq 0$. Once $z_r(\Delta)$ has been found, the corresponding average costs $g_s(\Delta, z_r(\Delta))$ are easily obtained - via (14) in the Appendix – and the property that $\mathrm{E}[(d - \Delta)^+] = \overline{p}\tilde{\mu} = \overline{p} \sum_{k=1}^{\infty} \tilde{q}_k \frac{k}{\lambda}$.

The approach described in this subsection is easily extended to the more general class of phase-type distributions; closure under truncated and residual demands appears to be a very broad and powerful property within this class.

4.3 Determination of Optimal Parameters for the γ-Service Level Problem

To determine the optimal single index policy for the γ-service level problem, we do the following:

1. Let δ be a small positive number. (In all of our examples the curves about the optimal were quite flat, making $\delta = 0.1$ a good choice.)

2. For each $\Delta = \Delta_{min} + k\delta,\ k = 0, 1, \ldots$, compute $z_r(\Delta)$ and the average costs $g_s(\Delta, z_r(\Delta))$, storing $\tilde{g}$, the lowest average cost among these points.

3. If $c_r\mu + hz_r(\Delta) - h(l_r + 1)\mu + hB < \tilde{g}$ go to 2. (From (8), $g_s(\tilde{\Delta}, z_r(\tilde{\Delta})) \geq c_r\mu + hz_r(\tilde{\Delta}) - h(l_r + 1)\mu + hB \geq \tilde{g}$ implies that $g_s(\tilde{\Delta}, z_r(\tilde{\Delta})) \geq g_s(\Delta, z_r(\Delta)$ for all $\tilde{\Delta} > \Delta$.)

4. Search the neighborhood of this best point to find a local optimum among all points $\Delta \geq 0$.

We generated $g_s(\Delta, z_r(\Delta))$ numerically for many instances, all of which had a unique local minimum; thus we conclude that our procedure returns the optimal single index parameters.

The function $g_s(\Delta, z_r(\Delta))$ typically behaves in one of two ways, as depicted in Figure 2, depending upon which supplier is least expensive to use as a sole source. It appears that $g_s(\Delta, z_r(\Delta))$ is a combination of a convex and then a concave function, making the form of the optimal policy difficult to analyze in general.

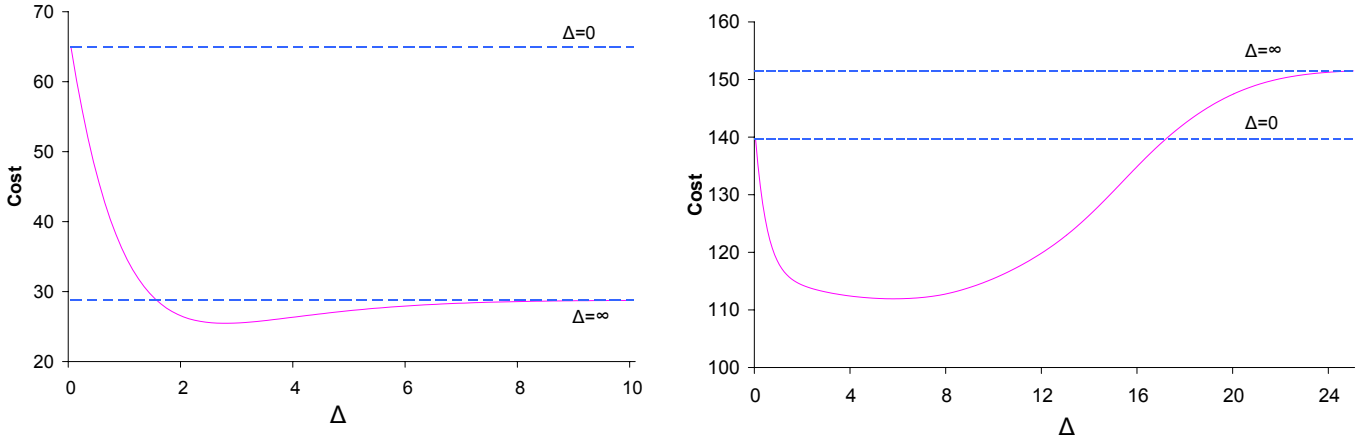

Figure 2: Average costs $g_s(\Delta, z_r(\Delta))$ (excluding $c_r\mu$) as a function of Δ for 2 instances: $\mu = 1$, $c_r = 1000$, $c_e = 1050$, $l_r = 6$, $l_e = 1$, $\gamma_0 = 0.95$, $h = 5$, and $\sigma = 1$ (lefthand side) and $\sigma = 3$ (righthand side). The $\Delta = 0$ solution is equivalent to single sourcing via the expedited mode, and $\Delta = \infty$ via the regular mode. Note the unimodality and irregular shapes.

5. Dual Index Policies

In this section we briefly discuss properties of the class of dual index policies. For a detailed discussion, please see Kiesmüller (2003), or Veeraraghavan and Scheller-Wolf (2008).

In the dual index policy, the expedited orders are based only on the outstanding orders that will arrive within the expedited lead time, i.e. the next l_e periods. The expedited order, if placed, restores the *expedited inventory position* (inventory level plus those goods on order that will arrive within the expedited leadtime) to the target z_e. The regular order is then placed based on the regular inventory position (sum of on-hand inventory and all outstanding orders, including any expedited orders just placed), and restores it to the target z_r. Thus, in the dual index policy two inventory positions are carried, one for expedited ordering and another for regular ordering.

Note that items already on order but just outside the expedited ordering horizon ($l_e + 1$ periods from delivery) are *not* included in the expedited inventory position, but will be in the next period. Thus with the arrival of this outstanding order, the expedited inventory position may up-cross the target expedited inventory level, z_e. Thus z_e forms a *lower* bound for the expedited inventory position just prior to demand, and the exact expedited inventory position depends on the regular

orders and demands over the indefinite past.

Since the dual index policy tracks more information, it seems intuitive that the optimal dual index policy would be superior to the optimal single index policy. Surprisingly, computational work indicates that this is not true in general (see Tables 2 and 4). How the additional information is used appears to be crucial; for example a dual index policy *must* order up to the expedited level, even if there is a large order just outside of the expedited leadtime, which may make placement of such an order inadvisable.

6. Computational Results

In all experiments we report average costs excluding the (fixed) purchasing costs $c_r\mu$ incurred if all goods are bought via the regular channel. We study first the penalty cost model in Section 6.1, and then the γ-service level model in Section 6.2.

6.1 Penalty Cost Model

In this section we present the results of 24 experiments with discrete demand, over parameters shown in Table 1. We use $U[0, B]$ to denote discrete demand uniformly distributed over the support $[0, B]$. We compare the single index policy (solution found as in Theorem 3.1), with the dual index policy (solution found as in Veeraraghavan and Scheller-Wolf, 2008), single sourcing (solution found via a simple Newsvendor fractile), and the globally optimal policy (solution found via dynamic programming). Finding globally optimal policies is possible only because these problems are small – they have a limited discrete state space as they have short leadtimes and their demand is drawn from a restricted set of values. Even with this fact solution times for dynamic programming range from 45 minutes to over eight hours, as compared to dual index times which average 25 seconds, and single index times which were less than a second each.

c_r	h	Demand	c_e	p	l	l_r
1000	5	U[0,4], U[0,8]	1020, 1050, 1100	95, 495	2, 3	0, 1

Table 1: Parameters for set of 24 experiments for penalty cost model with discrete demand.

Looking at Table 2, we see that for these problems the dual index policy is comparable to optimal; it is always within 3% of optimal, and on average within 1% of optimal. (The dual index cost is at times below the DP cost due to simulation effects.) The single index policy, while superior to single sourcing, is nonetheless inferior overall to the dual index – it may be more than 25% above optimal, and on average is almost 8% above. This is because for problems such as these, with limited demand support, the single index policy has a much smaller action space than the dual index (or the optimal) policy – for example for uniform demand between zero and four the single index policy can expedite only 0, 25, 50, 75 or 100% of demand. Thus the single index policy has less fine control, and its performance suffers. This problem appears to become less critical as l_e increases, and we would expect it to be greatly reduced as the demand distribution grows finer. In the limit, as demand becomes continuous (and globally optimal policies impossible to find in general), we would anticipate that the single index performance would improve relative to the dual index (and the optimal). We will see that this is indeed the case in the next section.

6.2 Service Level Model

We now turn to the service level model – experimenting with the performance of the single and dual index policies with continuous demand (mixed Erlang, as described in Section 4.2) and larger leadtimes. We examine 36 instances, having parameters shown in Table 3.

As the problem sizes have grown and demand is continuous we can no longer compute optimal policies. It has been established that as demand becomes finer, or approaches continuity, the cost of the dual index solution becomes comparable with the globally optimal policy (see Veeraraghavan and Scheller-Wolf, 2008). And, following the discussion above, we would expect the single index policy performance to improve relative to the dual index in this environment.

The choice of the parameter values in Table 3 is motivated by examples faced by Hewlett Packard and Océ. The value $l_e = 1$ represents 1 week, the time needed to transport goods by air, including preparation and receiving times. The values for l_r represent leadtimes when other transport modes are chosen (e.g. ocean). The values for μ and c_r can be set arbitrarily; they do not affect the solution behavior or computation times. Here we have chosen 10 and 1000. The

				Single Index			Dual Index		Single Source		Opt
l_e, l_r	c_e	Dem	p	z_e^*, z_r^*	Cost	Exp	Cost	Exp	Reg	Exp	(DP)
0,2	1020	U[0,4]	95	6,10	24.00	0%	23.25	6.1 %	24.00	50.00	22.82
0,2	1020	U[0,4]	495	7,10	26.00	25%	23.32	13.9 %	29.00	50.00	23.07
0,2	1050	U[0,4]	95	6,10	24.00	0%	23.98	0 %	24.00	110.00	24.00
0,2	1050	U[0,4]	495	7,11	29.00	0%	26.99	1.7 %	29.00	110.00	26.75
0,2	1100	U[0,4]	95	6,10	24.00	0%	23.98	0 %	24.00	210.00	24.00
0,2	1100	U[0,4]	495	7,11	29.00	0%	28.65	1.7 %	29.00	210.00	28.67
0,3	1020	U[0,4]	95	8,11	27.76	25%	25.00	14.30 %	28.36	50.00	24.29
0,3	1020	U[0,4]	495	8,10	31.00	50%	25.00	14.30 %	34.80	50.00	24.33
0,3	1050	U[0,4]	95	9,13	28.36	0%	27.71	1.45 %	28.36	110.00	27.51
0,3	1050	U[0,4]	495	10,14	34.80	0%	30.06	3.90 %	34.80	110.00	29.56
0,3	1100	U[0,4]	95	6,10	28.36	0%	28.32	0 %	28.36	210.00	28.36
0,3	1100	U[0,4]	495	7,11	34.80	0%	33.33	1.45 %	34.80	210.00	32.98
0,3	1020	U[0,8]	95	16,22	50.60	25%	45.53	9.35 %	52.04	100.00	44.44
0,3	1020	U[0,8]	495	17,22	58.21	37.5%	47.60	12.58 %	64.27	100.00	46.20
0,3	1050	U[0,8]	95	17,25	52.04	0%	50.56	1.71 %	52.04	220.00	50.34
0,3	1050	U[0,8]	495	20,28	64.27	0%	56.78	4.53 %	64.27	220.00	55.93
0,3	1100	U[0,8]	95	15,25	52.04	0%	51.93	0.14 %	52.04	420.00	51.95
0,3	1100	U[0,8]	495	20,28	64.27	0%	61.43	0.91 %	64.27	420.00	61.10
1,4	1020	U[0,4]	95	11,14	31.67	25%	29.78	8.11 %	31.72	59.00	29.55
1,4	1020	U[0,4]	495	13,16	38.28	25%	35.00	14.30 %	39.48	60.00	34.30
1,4	1050	U[0,4]	95	11,15	31.72	0%	31.45	1.45 %	31.72	119.00	31.43
1,4	1050	U[0,4]	495	13,17	39.48	0%	37.70	3.90 %	39.48	120.00	37.66
1,4	1100	U[0,4]	95	11,15	31.72	0%	31.69	0 %	31.72	219.00	31.72
1,4	1100	U[0,4]	495	13,17	39.48	0%	39.06	0.32 %	39.48	220.00	39.11
Average					37.29		34.92		37.96		34.59

Table 2: Optimal single index and dual index policies and costs, optimal single sourcing cost by the regular and expedited channel, and globally optimal cost. The values for c_e, l_e, $l_r$and p vary as indicated; the other input parameters are fixed at $c_r = 1000$, $h = 5$.

values for σ have been chosen such that we have a range of coefficients of variation. The values of c_e are such that the unit price when ordering by the expedited channel is 2%, 5%, and 10% more expensive than when ordering by the regular channel; it is in this region, where the expedited costs c_e are somewhat larger than the regular costs c_r that it most is attractive to combine both modes. The inventory holding costs are fixed at $h = 5$. This corresponds to a yearly rate of 25 % for interest and storage costs ($h = 1000 \cdot 0.25/50$). Finally, the target service level γ_0 values have been chosen to mimic those typically used in industry.

In Table 4, we compare the costs of the single index policy, dual index policy, and single sourcing, for the 36 instances listed in Table 3. We show the relative improvement of each of the dual sourcing

Constants				Variables		
μ	c_r	l_e	h	$\frac{\sigma}{\mu}$	c_e	γ_0
10	1000	1	5	$\frac{1}{3}$, 1, 3	1020, 1050, 1100	0.95, 0.99

Table 3: Parameters for set of 36 experiments for penalty cost model with continuous demand.

policies as compared to single sourcing as well, along with the proportion of items expedited under both dual sourcing policies. The optimal dual index parameters were found via simulation and search in C++ on a PC with Pentium III processor; all other values are analytical computations. The total computation time for all 36 single index policies was about 33 seconds, or less than a second for each instance, programmed in Delphi. The computational time for the dual index policies are much longer; about a minute per instance. Although not reported, single index policy costs were also simulated to validate our analytical method. For all instances the difference between the analytical and simulated single index costs was less than 1%, the mean absolute deviation was 0.28%, with the simulated costs being on average 0.18% lower.

Table 4 shows that dual sourcing may lead to savings of 25% or more in comparison to the best single sourcing option, and in most cases the performance of the single and dual index policies are comparable. The maximal difference between the policies was under 3% of the optimal single sourcing solution, and the mean difference was less than 0.5%. The dual index policy usually performed better in situations with low or moderate coefficient of variation of demand, while the single index policy was often slightly better in cases with high coefficients of variation. Notice also that both policies almost always expedite very similar proportions of items.

Given the comparable performance of the single index and dual index policies, the other advantages of the single index policy may become deciding factors in its favor: (i) The optimal single index policy is simpler – it relies on a single control parameter, rather than two; (ii) Calculation of optimal single index parameters is 25-60 times faster than calculation of optimal dual index parameters. This is partially due to the fact that the single index solution method is purely analytical, while the dual index method relies on simulation; (iii) Use of the single index policy results in smoothing for the regular supplier, by bounding the maximum regular order (at Δ). This is an

				Single Index		Dual Index		Single Source		Savings
$\frac{\sigma}{\mu}$	c_e	l_r	γ_0	Costs	%Exp	Costs	%Exp	Reg	Exp	SI/DI
$\frac{1}{3}$	1020	3	0.95	39.8	1 %	40.1	1 %	40.7	225.3	2/1 %
$\frac{1}{3}$	1020	3	0.99	63.3	2 %	62.3	2 %	67.2	245.6	6/7 %
$\frac{1}{3}$	1020	6	0.95	52.9	4 %	52.3	5 %	59.0	225.3	10/11 %
$\frac{1}{3}$	1020	6	0.99	77.6	6 %	75.8	6 %	92.0	245.6	16/18 %
$\frac{1}{3}$	1050	3	0.95	40.7	0 %	40.7	0 %	40.7	525.3	0/0 %
$\frac{1}{3}$	1050	3	0.99	66.1	0 %	66.2	0 %	67.2	545.6	2/1 %
$\frac{1}{3}$	1050	6	0.95	58.0	1 %	57.9	1 %	59.0	525.3	2/2 %
$\frac{1}{3}$	1050	6	0.99	87.0	1 %	85.8	2 %	92.0	545.6	5/7 %
$\frac{1}{3}$	1100	3	0.95	40.7	0 %	40.7	0 %	40.7	1025.3	0/0 %
$\frac{1}{3}$	1100	3	0.99	67.0	0 %	66.9	0 %	67.2	1045.6	0/0 %
$\frac{1}{3}$	1100	6	0.95	59.0	0 %	59.0	0 %	59.0	1025.3	0/0 %
$\frac{1}{3}$	1100	6	0.99	90.6	0 %	89.9	1 %	92.0	1045.6	2/2 %
1	1020	3	0.95	192.5	8 %	192.9	12 %	214.6	349.1	10/10 %
1	1020	3	0.99	284.5	8 %	282.0	9 %	320.9	439.4	11/12 %
1	1020	6	0.95	221.7	18 %	215.8	17 %	287.7	349.1	23/25 %
1	1020	6	0.99	313.9	20 %	302.3	14 %	411.2	439.4	24/26 %
1	1050	3	0.95	204.9	2 %	203.9	1 %	214.6	649.1	5/5 %
1	1050	3	0.99	298.8	3 %	295.6	3 %	320.9	739.4	7/8 %
1	1050	6	0.95	254.7	6 %	252.4	6 %	287.7	649.1	11/12 %
1	1050	6	0.99	350.9	8 %	351.3	6 %	411.2	739.4	15/15 %
1	1100	3	0.95	210.9	1 %	209.4	1 %	214.6	1149.1	2/2 %
1	1100	3	0.99	307.6	1 %	304.8	2 %	320.9	1239.4	4/5 %
1	1100	6	0.95	273.0	2 %	267.2	2 %	287.7	1149.1	5/7 %
1	1100	6	0.99	375.8	3 %	371.7	3 %	411.2	1239.4	9/10 %
3	1020	3	0.95	986.7	35 %	991.8	36 %	1159.1	1097.6	10/10 %
3	1020	3	0.99	1268.2	42 %	1271.0	41 %	1640.6	1372.9	8/7 %
3	1020	6	0.95	1001.5	44 %	1008.5	46 %	1517.1	1097.6	9/8 %
3	1020	6	0.99	1280.0	46 %	1286.1	44 %	1933.0	1372.9	7/6 %
3	1050	3	0.95	1060.7	20 %	1060.6	10 %	1159.1	1397.6	8/8 %
3	1050	3	0.99	1383.7	35 %	1381.8	33 %	1640.6	1672.9	16/16 %
3	1050	6	0.95	1119.3	35 %	1121.2	38 %	1517.1	1397.6	20/20 %
3	1050	6	0.99	1410.2	41 %	1413.2	39 %	1933.0	1672.9	16/16 %
3	1100	3	0.95	1132.0	7 %	1130.7	6 %	1159.1	1897.6	2/2 %
3	1100	3	0.99	1532.7	23 %	1531.4	24 %	1640.6	2172.9	7/7 %
3	1100	6	0.95	1268.0	25 %	1269.1	25 %	1517.1	1897.6	16/16 %
3	1100	6	0.99	1600.3	34 %	1597.9	30 %	1933.0	2172.9	17/17 %
Average				529.9		529.2		Min. 598.4		8.5/8.9 %

Table 4: *Optimal single index and dual index policies, including costs; optimal costs under single sourcing by the regular and expedited channel, respectively; and the relative savings obtained by dual sourcing in comparison with the best of the two single sourcing options. The values for* $\frac{\sigma}{\mu}$, c_e, l_r, *and* γ_0 *vary as indicated; the other input parameters are fixed at* $\mu = 10$, $c_r = 1000$, $l_e = 1$, $h = 5$.

attractive feature for a regular supplier who might only be willing to supply limited quantities at the lower price; and (iv) Use of the single index policy provides an explicit analytical expression (3) for the average expedited order. This is an attractive feature for a manager who requires an explicit bound on expediting orders (and costs).

In Table 5 we explore the behavior of the single index policy in greater detail, listing Δ_{min}, the optimal single index parameters Δ^* and $z_r(\Delta^*)$, the percentage ordered by the expedited channel, and breaking down the average costs for the 36 instances in Table 4 . These experiments demonstrate that:

1. Optimal average costs are increasing as a function of σ, c_e, γ_0 and l_r. *As expected, higher demand variability, expediting costs, service levels and longer leadtimes drive up costs.*

2. The value of Δ^* is decreasing in l_r and γ_0, and increasing as a function of c_e. Thus, via (3) expediting increases with longer regular leadtimes and higher service levels, and decreases with expediting cost. It is also apparent that expediting increases with demand variability.

3. The value of $z_r(\Delta^*)$ increases significantly with an increase in γ_0, l_r and σ. It increases more gradually with an increase in c_e. *The regular base stock level, and thus the total inventory in system, is significantly effected by service level, leadtime and demand variability.* The effects of these factors can be moderated by the (less significant) effect of expediting cost.

4. The lower bound Δ_{min} is best for small values of σ, and grows weaker as the coefficient of variation grows. *Our bound does not grow as fast as Δ^* as variation in the demand grows.*

5. In keeping with our analytical results, using the regular supplier as a single source is sometimes optimal ($\Delta = \infty$), but using the expedited supplier alone never is. *It is always beneficial to be able to utilize a lower cost, longer leadtime, sourcing option.*

A crucial task when considering any sort of dual sourcing is determining in which settings having the second sourcing option will prove most valuable. To explore this, we illustrate in Figure 3 the optimal single sourcing costs as well as the optimal single index costs for one specific instance as a function of c_e, together with the percentage of demand that is ordered by the expedited channel

$\frac{\sigma}{\mu}$	c_e	l_r	γ_0	Δ_{min}	Δ^*	$z_r(\Delta^*)$	% Exp.	Average Costs Ord.	Hold.	Total
$\frac{1}{3}$	1020	3	0.95	11.1	16.0	46.8	1 %	2.1	37.8	39.8
$\frac{1}{3}$	1020	3	0.99	11.1	14.5	51.3	2 %	4.3	59.0	63.3
$\frac{1}{3}$	1020	6	0.95	9.2	13.0	76.3	4 %	8.3	44.6	52.9
$\frac{1}{3}$	1020	6	0.99	9.2	12.0	79.8	6 %	12.6	65.0	77.6
$\frac{1}{3}$	1050	3	0.95	13.1	22.1	47.6	0 %	0.2	40.5	40.7
$\frac{1}{3}$	1050	3	0.99	13.1	18.0	52.7	0 %	1.9	64.2	66.1
$\frac{1}{3}$	1050	6	0.95	11.1	17.3	80.3	1 %	2.7	55.3	58.0
$\frac{1}{3}$	1050	6	0.99	11.1	15.3	85.1	1 %	7.4	79.7	87.1
$\frac{1}{3}$	1100	3	0.95	14.7	∞	47.6	0 %	0.0	40.7	40.7
$\frac{1}{3}$	1100	3	0.99	14.7	21.2	53.2	0 %	0.6	66.4	67.0
$\frac{1}{3}$	1100	6	0.95	12.6	22.1	81.2	0 %	0.4	58.6	59.0
$\frac{1}{3}$	1100	6	0.99	12.6	18.5	87.3	0 %	2.9	87.7	90.6
1	1020	3	0.95	11.0	25.8	73.4	8 %	15.2	177.3	192.5
1	1020	3	0.99	11.0	24.8	91.8	8 %	16.7	267.8	284.5
1	1020	6	0.95	5.9	16.9	97.3	18 %	36.8	184.9	221.7
1	1020	6	0.99	5.9	16.3	115.1	20 %	39.1	274.8	313.9
1	1050	3	0.95	17.9	38.1	77.8	2 %	11.1	193.8	204.9
1	1050	3	0.99	17.9	35.6	96.3	3 %	14.2	284.6	298.8
1	1050	6	0.95	11.0	27.4	110.8	6 %	32.2	222.5	254.7
1	1050	6	0.99	11.0	25.4	128.2	8 %	39.6	311.3	350.9
1	1100	3	0.95	24.0	50.4	80.3	1 %	6.5	204.4	210.9
1	1100	3	0.99	24.0	45.0	99.0	1 %	11.1	296.5	307.6
1	1100	6	0.95	16.1	39.1	119.1	2 %	20.0	253.0	273.0
1	1100	6	0.99	16.1	34.6	137.3	3 %	31.2	344.6	375.8
3	1020	3	0.95	5.8	58.1	215.9	35 %	69.7	917.0	986.7
3	1020	3	0.99	5.8	33.9	268.5	42 %	83.6	1184.6	1268.2
3	1020	6	0.95	3.1	27.2	230.3	44 %	87.7	913.8	1001.5
3	1020	6	0.99	3.1	21.1	284.4	46 %	92.3	1187.7	1280.0
3	1050	3	0.95	9.8	110.1	227.8	20 %	99.3	961.4	1060.7
3	1050	3	0.99	9.8	56.2	274.2	35 %	176.9	1207.8	1383.7
3	1050	6	0.95	5.8	57.7	240.0	35 %	174.7	944.6	1119.3
3	1050	6	0.99	5.8	35.7	290.1	41 %	206.3	1203.9	1410.2
3	1100	3	0.95	13.7	153.3	247.5	7 %	83.9	1048.1	1132.0
3	1100	3	0.99	13.7	99.9	296.1	23 %	229.0	1303.7	1532.7
3	1100	6	0.95	8.7	91.6	260.0	25 %	252.6	1015.4	1268.0
3	1100	6	0.99	8.7	59.5	303.9	34 %	344.4	1255.9	1600.3

Table 5: *Values for Δ_{min}, the parameters Δ^* and $z_r(\Delta^*)$ of the optimal single index policy, and the percentage ordered by the expedited channel and average costs under the optimal single index policy. Demand is mixed Erlang, the values for $\frac{\sigma}{\mu}$, c_e, l_r, and γ_0 vary as indicated; the other input parameters are fixed at $\mu = 10$, $c_r = 1000$, $l_e = 1$, $h = 5$.*

under the optimal single index policy. For small (large) price differences $c_e - c_r$, almost everything (nothing) is ordered by the expedited channel under the optimal single index policy, and thus the

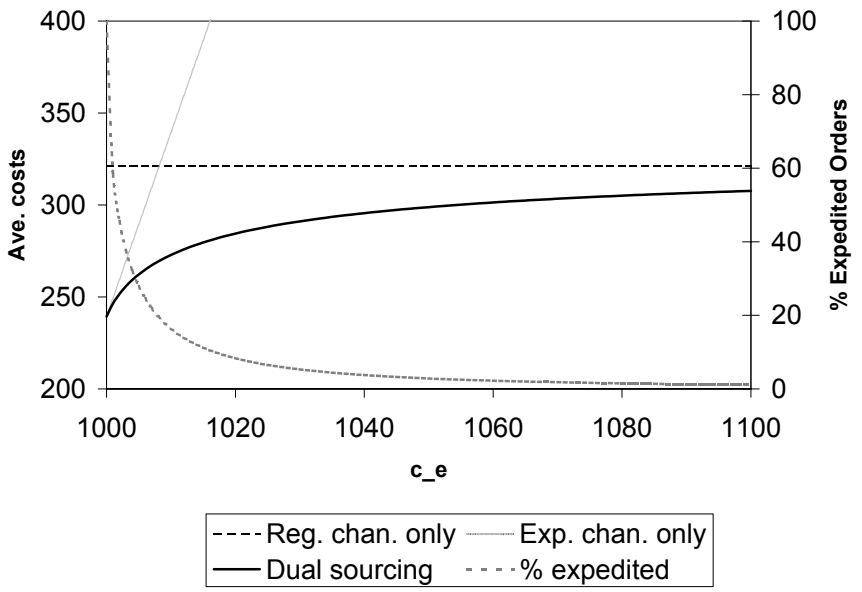

Figure 3: Average costs when using the regular channel only, when using the expedited channel only, and under a single index policy with the percentage ordered by the expedited channel for varying values of c_e. Other input parameters are fixed at $\mu = 10$, $\sigma = 10$, $c_r = 1000$, $l_r = 4$, $l_e = 1$, $\gamma_0 = 0.95$, $h = 5$. Percentage of orders expedited is shown by decreasing curve that is plotted against secondary axis on the RHS of the graph.

dual sourcing costs are close to the optimal single sourcing costs. In between there is a region where dual sourcing leads to significantly lower costs than either single sourcing option; the largest savings are available when the average costs of the two single sourcing options are close to each other. The magnitude of this intermediate region where dual sourcing is most worthwhile depends on the coefficient of variation of the demand (σ/μ) and the leadtime difference ($l_r - l_e$). The higher σ/μ and $l_r - l_e$, the larger this region. *Thus whether it is attractive for a company to consider dual sourcing depends fundamentally on (i) Whether they have alternate supply options with comparable single sourcing costs; and (ii) Whether they have many items in this intermediate region.*

7. Extensions

Here we discuss extensions to our model, to capacitated suppliers and to three supply modes.

7.1 Capacitated Suppliers

Corollary 7.1 *If there is a capacity C_e on the expedited orders, Lemma 3.1 and Lemma 3.2 hold with the modification that $D(\Delta) \sim \sum_{i=1}^{l_e+1} d_i + \sum_{j=l_e+2}^{l_r+1} M_j$, where*

$$M_j = \begin{cases} d_j & d_j \leq \Delta, \\ \Delta & \Delta < d_j \leq C_e + \Delta, \\ d_j - C_e & C_e + \Delta < d_j. \end{cases}$$

The single index policy with a capacitated regular supplier is even simpler (assuming in every period we wish to return the regular inventory position to z_r rather than running a backlog):

Corollary 7.2 *If there is a capacity C_r on the regular orders all the results above hold, with the additional constraint $\Delta \leq C_r$ added.*

7.2 Multiple Suppliers

If there are more delivery options (as in Zhang 1996, Gallego and Zhang 2003, and Feng et al. 2003, 2005) a single index policy again yields a dimensional reduction. For the case of three options:

Corollary 7.3 *If there are three delivery modes, for each fixed Δ_1 and Δ_2, given $l_1 < l_2 < l_3$ and $z_3 = z_2 + \Delta_2$, $z_2 = z_1 + \Delta_1$, Lemma 3.1 and Lemma 3.2 hold with the modification that :*

$$D(\Delta_1, \Delta_2) \sim \sum_{i=1}^{l_1+1} d_i + \sum_{j=l_1+2}^{l_2+1} \min(d_j, \Delta_1 + \Delta_2) + \sum_{k=l_2+2}^{l_3+1} \min(d_k, \Delta_2).$$

8. Conclusion

We have defined the class of single index inventory replenishment policies for use when there is a choice of two supply options having different unit costs and different fixed lead times. If the lead times differ by exactly one unit our policy is optimal, otherwise it serves as a heuristic; one of the first effective heuristics for this general leadtime case. We show how to swiftly calculate *optimal* single index parameters for both the penalty cost and γ-service level versions of the problem, deriving properties of Erlang mixtures when necessary (for the continuous demand case). We then

demonstrate the effectiveness of the single index policy via computational experiments. These show that for continuous demand distributions the performance of the optimal single index policy is comparable to the more complex dual index policy (known itself to perform near-optimally), and that the relative performance of the single index policy improves as demand becomes more variable (a setting for which dual sourcing is often considered). Moreover, calculating an optimal single index policy is 25-60 times faster than finding an optimal dual index policy. (Calculating the globally optimal policy requires dynamic programming, and is in general not possible in continuous demand settings.) Our experiments also show that either of the dual sourcing schemes can offer significant savings over single sourcing, especially when combining two single sources of supply with comparable total costs.

Thus we have provided an easily implementable framework for quickly making ordering decisions from dual suppliers. This line of research offers a number of avenues for future work: Exploration of the extensions in Section 7, determination of further properties of the globally optimal policies, lost sales models, and the analysis and performance of other simple ordering policies (i.e. dual index).

9. Appendix

To avoid the issue of whether derivatives exist almost everywhere, we consider finite differences: For an arbitrary function $f(x)$ and positive ϵ, we define $f^{\epsilon}(x) \stackrel{\text{def}}{=} \frac{f(x+\epsilon)-f(x)}{\epsilon}$.

Lemma 9.1 *For all $\Delta \geq 0$ and $\nu > 0$ there exists an $\epsilon > 0$ such that*

$$|\{-(c+hl)\mathbf{P}(d \geq \Delta + \epsilon) + hz_r^{\epsilon}(\Delta)\} - g_s^{\epsilon}(\Delta, z_r(\Delta))| \leq \nu. \tag{13}$$

Proof : Substituting (8) into $g_s(\Delta, z_r(\Delta))$ and rewriting yields:

$$g_s(\Delta, z_r(\Delta)) \;=\; c_r\mu + (c+hl)\mathrm{E}[(d-\Delta)^+] + hz_r(\Delta) - h(l_r+1)\mu + hB \;,\; \Delta \geq 0. \tag{14}$$

Applying finite differences:

$$\begin{aligned} g_s^{\epsilon}(\Delta, z_r(\Delta)) &= \left\{(c+hl)(\mathrm{E}[(d-\Delta-\epsilon)^+ - (d-\Delta)^+]) + h(z_r(\Delta+\epsilon) - z_r(\Delta))\right\}/\epsilon, \\ &= \left\{(c+hl)(-\epsilon\mathbf{P}(d \geq \Delta+\epsilon) - \mathrm{E}[(d-\Delta)I\{d \in (\Delta, \Delta+\epsilon)\}]) + h\epsilon z_r^{\epsilon}(\Delta)\right\}/\epsilon. \end{aligned}$$

Therefore

$$-(c+hl)\mathbf{P}(d \geq \Delta+\epsilon) + h(z_r^\epsilon(\Delta)) - (c+hl)\mathbf{P}(d \in (\Delta, \Delta+\epsilon)) \leq$$

$$g_s^\epsilon(\Delta, z_r(\Delta) \leq \quad -(c+hl)\mathbf{P}(d \geq \Delta+\epsilon) + hz_r^\epsilon(\Delta).$$

As we have a continuous demand distribution, we can choose ϵ small enough such that for any $\Delta \geq 0$ and a given $\nu > 0$, $\mathbf{P}(d \in (\Delta, \Delta+\epsilon)) \leq \nu/(c+hl)$, completing the proof. ■

Lemma 9.2 *For any positive* ϵ, $D^\epsilon(\Delta) \in [0, l]$.

Proof : Recalling (6), $D(\Delta)$ is nondecreasing in Δ, but on any sample path no more than $l = (l_r+1) - (l_e+2) + 1$ elements of $D(\Delta)$ can increase (at rate 1) as Δ increases. ■

Lemma 9.3 *For any positive* ϵ, $z_r^\epsilon(\Delta) \in [0, l]$.

Proof : As $z_r(\Delta)$ is nondecreasing in Δ, we know that $z_r^\epsilon(\Delta) \geq 0$.

Assume that for some Δ and ϵ, $z_r^\epsilon(\Delta) > l$. Part (iii) of Lemma 3.2 establishes that for any given B and Δ, constraint (7) will hold as an equality at optimality. In particular:

$$\mathrm{E}[(D(\Delta) - z_r(\Delta))^+] = B = \mathrm{E}[(D(\Delta+\epsilon) - z_r(\Delta+\epsilon))^+]. \tag{15}$$

We divide the sample paths for the demand realizations into three disjoint sets:

ω_1: Sample paths where $D(\Delta) > z_r(\Delta)$ and $D(\Delta+\epsilon) > z_r(\Delta+\epsilon)$. These will be included in both expectations in (15). From Lemma 9.2 and our assumption that $z_r^\epsilon(\Delta) > l$, we see that $D(\Delta+\epsilon) - D(\Delta) \leq l\epsilon < z_r(\Delta+\epsilon) - z_r(\Delta)$, or

$$D(\Delta+\epsilon) - z_r(\Delta+\epsilon) < D(\Delta) - z_r(\Delta). \tag{16}$$

Relation (16) implies that the contribution of the sample paths in ω_1 to the left-hand expectation in (15) is strictly greater than their contribution on the right-hand side.

ω_2: Sample paths where $D(\Delta) > z_r(\Delta)$ and $D(\Delta+\epsilon) \leq z_r(\Delta+\epsilon)$. These make a strictly positive contribution to the left-hand side of (15) but zero on the right.

ω_3: Sample paths where $D(\Delta) \leq z_r(\Delta)$. Formula (16) implies the value of both expectations in (15) is zero.

Unless $D(\Delta) \leq z_r(\Delta)$ almost surely, which violates $B < 1$, taking expectation over $\omega_1 \cup \omega_2 \cup \omega_3$ yields $\mathrm{E}[(D(\Delta) - z_r(\Delta))^+] > \mathrm{E}[(D(\Delta+\epsilon) - z_r(\Delta+\epsilon))^+]$, contradicting (15). ■

This leads to

Lemma 9.4 *If $(\Delta^*, z_r(\Delta^*))$ is an optimal solution, then for all $\nu > 0$ there exists a $\zeta > 0$ such that $\zeta \downarrow 0$ if $\nu \downarrow 0$ and $\mathbf{P}(d \geq \Delta^* + \zeta) \leq \frac{hl+\nu}{c+hl}$.*

Proof : For there to be a minimum at $(\Delta^*, z_r(\Delta^*))$, continuity dictates that for all $\epsilon > 0$ small enough, $g_s(x, z_r(x)) = g_s(x+\epsilon, z_r(x+\epsilon))$ for some $x < \Delta^* < x + \epsilon$ (equivalently $g^\epsilon(x, z_r(x)) = 0$). From (13) this implies that for all $\nu > 0$ we can find an $\epsilon > 0$ and an $x < \Delta^* < x + \epsilon$ such that: $|-(c+hl)\mathbf{P}(d \geq x+\epsilon) + hz_r^\epsilon(x)| \leq \nu$.

Letting $x \stackrel{\text{def}}{=} \Delta^* - \eta$, and using $z_r^\epsilon \leq l$ from Lemma 9.3:

$$(c+hl)\mathbf{P}(d \geq \Delta^* - \eta + \epsilon) \leq \nu + hl \quad \Rightarrow \quad \mathbf{P}(d \geq \Delta^* - \eta + \epsilon) \leq \frac{hl+\nu}{c+hl}.$$

Defining $\zeta \stackrel{\text{def}}{=} \epsilon - \eta$ completes the proof. ■

Bibliography

Beyer, D. and J. Ward, "Network Server Supply Chain at HP: A Case Study", Hewlett Packard Software Technology Lab Report HPL 2000-84, Palo Alto, (2000).

Çakanyildirim, M. and S. Luo, "(R,Q) Policies with Lead Time Options", School of Management, UT Dallas, Working paper SOM200538, (2005).

Feng, Q, G. Gallego, S. Sethi, H. Yan, and H. Zhang, "A Periodic Review Inventory Model with Three Delivery Modes and Forecast Updates", *Journal of Optimization Theory and Applications*, 124 1 (2003) 137–155.

Feng, Q, G. Gallego, S. Sethi, H. Yan and H. Zhang, "Are Base Stock Policies Optimal in Inventory Problems with Multiple Delivery Modes?", *Operations Research*, 54 4 (2005) 801–807.

Fox, E., R. Metters and J. Semple, "Optimal Inventory Policy with Two Suppliers", *Operations Research* 54 2 (2005) 389–393.

Fukuda, Y., "Optimal Policy for the Inventory Problem with Negotiable Leadtime", *Management Science*, 10 4 (1964) pp. 690-708.

Gallego, G., and W. Zhang, "Optimal Stationary Policies for Multiple Procurement Modes", Working paper, IEOR Dept., Columbia University, New York (2003).

Groenevelt, H. and N. Rudi, "A Base-stock Inventory Model with Possibility of Rushing Part to Order", Working paper, Simon School, University of Rochester, Rochester, NY (2003).

Kiesmüller, G.P. "A New Approach for Controlling a Hybrid Stochastic Manufacturing / Remanufacturing System with Inventories and Different Leadtimes", *European Journal of Operational Research* 147 1 (2003) 62-71.

Lawson, D., and Porteus, E., "Multistage Inventory Management with Expediting", *Operations Research*, 46 6 (2000) 878-893.

Moinzadeh, K. and S. Nahmias,"A Continuous Review Model for an Inventory System With Two Supply Modes.", *Management Science*, 34 6 (1988), 483-494.

Rao, U., Scheller-Wolf, A., and S. Tayur, "Development of a Rapid-Response Supply Chain at Caterpillar", *Operations Research*, 48 2 (2000) 189–204.

Robinson, L., J. Bradley and J. Thomas, "Consequences of Order Crossover under Order-up-to Inventory Policies", *Manufacturing & Service Operations Management* 3 3 (2001) 175-188.

Schassberger, R., *Warteschlangen*, Springer-Verlag, Berlin, 1973.

Sethi, S. H. Yan, and H. Zhang, "Inventory Models with Fixed Costs, Forecast Updates, and Two Delivery Modes", *Operations Research*, 51 2 (2003) 321–328.

Tagaras, G., and D. Vlachos, "An Periodic Review Inventory System with Emergency Replenishments", *Management Science*, 47 3 (2001) 415–429.

Tayur, S., R. Ganeshan, and M. Magazine, *Quantitative Models for Supply Chain Management*, Kluwer, Newel, Massachusetts, 1999.

Teunter, R., and D. Vlachos, "An Inventory System with Periodic Regular Review and Flexible Emergency Review", *IIE Transactions* 33 8 (2001) 625–635.

Tijms, H.C., *Stochastic Models; An Algorithmic Approach*, Wiley, New York, 1986.

van Houtum, G.J., and W.H.M. Zijm, "On the Relation Between Cost and Service Models for

General Inventory Systems", *Statistica Neerlandica*, 54 2 (2000) 127–147.

Veeraraghavan, S. and A. Scheller-Wolf, "Now or Later: Simple Policy for Effective Dual Sourcing in Capacitated Systems.", *Operations Research*, 56 4 (2008) 850-864.

Vlachos, D. and G. Tagaras, "An Inventory System with Two Supply Modes and Capacity Constraints", *Int. Journal of Production Economics*, 72 1 (2001) 41–58.

Whittemore, A.S., and S. C. Saunders, "Optimal Inventory Under Stochastic Demand with Two Supply Options", *SIAM J. of Applied Math*, 32 2 (1977) 293-305.

Yazlali, O. and F. Erhun, "Managing Demand Uncertainty under Dual Supply Contracts", Working paper, Stanford University, (2008).

Yi, J., and A. Scheller-Wolf, "Dual Sourcing from a Regular Supplier and a Spot Market", GSIA Working Paper # 2003-E53, Pittsburgh, (2003).

Zhang, V. L., "Ordering Policies for an Inventory System with Three Supply Modes", *Naval Research Logistics*, 43 5 (1996) 691-708.