# DESIGN FOR POSTPONEMENT: A COMPREHENSIVE CHARACTERIZATION OF ITS BENEFITS UNDER UNKNOWN DEMAND DISTRIBUTIONS


YOSSI AVIV

*Olin School of Business, Washington University, St. Louis, Missouri 63130, aviv@olin.wustl.edu*

AWI FEDERGRUEN

*Graduate School of Business, Columbia University, New York, New York 10027, af7@columbia.edu*


(Received September 1997; revision received July 1999; accepted March 2000)

Recent papers have developed analytical models to explain and quantify the benefits of delayed differentiation and quick response programs. These models assume that while demands in each period are random, they are independent across time and their distribution is perfectly known, i.e., sales forecasts do not need to be updated as time progresses. In this paper, we characterize these benefits in more general settings, where parameters of the demand distributions fail to be known with accuracy or where consecutive demands are correlated. Here it is necessary to revise estimates of the parameters of the demand distributions on the basis of observed demand data. We analyze these systems in a Bayesian framework, assuming that our initial information about the parameters of the demand distributions is characterized via prior distributions. We also characterize the structure of close-to-optimal ordering rules in these systems, for a variety of types of order cost functions.

## 1. INTRODUCTION AND SUMMARY

*Delayed product differentiation and quick response* rank among the most beneficial strategic mechanisms to manage the risks associated with product variety and uncertain sales.

The product portfolio offered by a company often consists of families of closely related products, which differ from each other in terms of a limited number of differentiating features only. Consider for example the apparel industry. A given design or style is usually offered in many distinct sizes and colors. Grocery and dry food products typically are sold in several package sizes, with a proliferation of differentiating features, e.g., different fragrances added to detergents. Automobile manufacturers offer a virtually endless variety of model configurations within a few basic product lines, while a given computer or printer model is distributed with a variety of accessories (e.g., power supply modules, key pads, or manuals written in different, languages). Delayed differentiation, or *postponement*, strategies attempt to reduce the risks associated with this product variety by exploiting the commonality between items and by designing the production and distribution processes to delay the point of differentiation. In the same vein, coordinating several geographically dispersed sales outlets via a regional distribution center delays the point of differentiation to the final points of sale.

In particular, under delayed product differentiation strategies one finds that the production and distribution process consists of several stages, each with a significant lead time. The quick response concept consists of introducing systematic reductions in the average value and variability of these lead times. Such reductions can be achieved by setup time reductions, the adoption of faster and more reliable production technology, electronic submission of purchase orders, and contractual agreements with suppliers, stipulating quick and reliable delivery times, to mention but a few possibilities.

Many companies have already reported major success stories that are attributed to the above pair of strategies: delayed differentiation and quick response. Examples include Benetton (see Signorelli and Heskett 1989), Sport Obermeyer (see Fisher et al. 1994a, b), Hewlett Packard (see Lee et al. 1993 and Lee and Billington 1994), Compaq (see *New York Times* 1997), Sun MicroSystems (see Mrena 1997), and Toyota (see Federgruen 1993). As an example of delayed *geographic differentiation*, General Motors recently announced the opening of regional distribution centers for its Cadillac product line, coordinating the previously independent replenishment processes of its GM dealers (see *Wall Street Journal* 1996).

The concept of delayed product differentiation was first introduced in the marketing literature by Alderson (1950). Several more recent papers have developed analytical models to explain and quantify the operational benefits of delayed differentiation; see Eppen and Schrage (1981), Federgruen and Zipkin (1984a, b, c), Jackson (1988), Schwarz (1989), Federgruen (1993), Lee (1993, 1996), and Garg and Tang (1997). These models, in the general tradition of inventory theory, assume that while demands in

*Subject classifications:* Inventory/Production: multi-item/echelon/stage. Inventory/Production: operating characteristics. forecasting.
*Area of review:* Logistics and Supply Chain Operations.


0030-364X/01/4904-0578 $05.00
1526-5463 electronic ISSN

each period are random, they are *independent* across time and their distribution is *perfectly* known, i.e., sales forecasts do not need to be updated as time progresses. Under this assumption, the benefits of delayed differentiation, in terms of inventory related performance measures, are restricted to *two* factors:

(i) *Statistical economies of scale*. Assume that $\tau$ denotes the total production or replenishment lead time. Product differentiation occurs after a first stage of $L \leqslant \tau$ periods, which terminates with a common intermediate product. The delayed differentiation permits one to specify only *aggregate* orders at the beginning of the replenishment process and to commit resources to individual products only at the end of the first stage of $L$ periods. It also allows one to observe the inventory status of the products at the end of the first stage and thus to make better informed allocations to the individual products. Eppen and Schrage (1981) coined the phrase "statistical economies of scale" for this effect. Others refer to it as *risk pooling*. Federgruen and Zipkin (1984a) and Garg and Tang (1997) consider a generalization with two (or more) points of differentiation, where a common product is differentiated into several families, and each family is differentiated into a set of end items. Statistical economies of scale continue to describe the benefits of extending the first (second) manufacturing stage, at the expense of the second (third) stage.

(ii) *Risk pooling via a common buffer*. To the extent inventories of the common intermediate product are stocked, these may serve as a *common* buffer from which all products can draw relatively quickly in case of need. The common buffer further reduces the magnitude of system-wide safety stocks to guarantee given service levels.

The key assumption of perfectly known demand distributions is in general rather restrictive, with few exceptions, such as staple goods, facing mature markets. Many products face a short life-cycle or are subject to dynamic and competitive market forces. Thus, even the most basic characteristics of the demand distributions (e.g., their means) might not be known with sufficient accuracy. However, estimates can be significantly improved on the basis of observed sales data. The fashion and high-technology industries represent extreme examples of this phenomenon. Fisher and Raman (1996) document dramatic improvement in forecast accuracy, which can be achieved after observing only 20% of initial sales in a sales season for fashion items. Delayed differentiation allows one to use observed sales data during the first common phase (of $L$ periods), not merely to get updated information about the products' inventory status (see (i) above) but also to exploit a third factor:

(iii) *The learning effect*. The generation of significantly more accurate forecasts of future demand distributions, further improving allocation to the individual products at the completion of the first phase. The learning effect arises even in settings where the demand distributions are known with accuracy but consecutive demands are correlated. Here, too, observed sales data during the first common phase can be exploited to revise forecasts for future demands affecting the allocations to individual products. Clearly, both types of learning effects often prevail simultaneously and their benefits compound on each other.

This paper characterizes the benefits of delayed differentiation in the presence of all three of the above factors. In particular, we assume that estimates of the parameters of the demand distributions are revised on the basis of observed sales data in a *Bayesian framework*. We show that the learning effect always results in increased benefits of delayed differentiation as well as lead time reductions through quick response programs, and that the incremental benefits can be very significant indeed. Our analysis continues to focus on benefits in terms of inventory related performance measures. Design for delayed differentiation and quick response have additional benefits and costs, which can be analyzed separately; see Lee (1996) and Lee and Tang (1997). Along with our characterization of the benefits of delayed differentiation, we develop close-to-optimal and easy-to-implement strategies to manage multi-item systems of this type.

The need to address settings in which some degree of uncertainty exists regarding the demand distributions was already recognized by the founding fathers of inventory theory. Dvoretsky et al. (1952) introduced a *single* item periodic review model where the distributions of the demand variables are of a known functional form but have one or more unknown parameters. The model assumes a prior distribution for these parameters, which in a Bayesian framework is updated as time progresses and demand realizations are observed. For the case of independent and identically distributed demands and a cost structure consisting of linear holding, backlogging, and ordering costs, Scarf (1959, 1960), Karlin (1960), and Iglehart (1964) showed that an adaptive order-up-to policy is optimal where the order-up-to level depends on the past history through a sufficient statistic $S$. These authors all consider a specific distributional form—in particular, exponential and range families. Azoury (1985) extended this result to a larger class of distributions and provided characterizations for the optimal adaptive order-up-to levels. Azoury and Miller (1984) show that major errors in the optimal ordering levels and expected costs can arise when uncertainty about the parameters is ignored, i.e., when the uncertain parameters are replaced by a single point estimate; see Azoury (1985) for additional references on this single-item model.

The case where demand distributions are unknown and a Bayesian approach is taken represents one setting in which consecutive demands are correlated. Intertemporal correlation can, however, arise even when the joint distribution of demands is perfectly known, e.g., when demands follow an autoregressive time-series model, a common assumption in many statistical forecasting systems (e.g., exponential smoothing, Box-Jenkins processes). For a variety of such single-item autoregressive models, and continuing to assume that ordering and inventory costs are linear, Veinott (1965), Johnson and Thompson (1975), Sobel

(1988), Miller (1986), and Lovejoy (1990) obtain the optimality of adaptive order-up-to policies.

Erkip et al. (1990) and Güllü (1997) revisit the two-stage production model of Eppen and Schrage (1981), assuming respectively that demands follow a specific first-order autoregressive time-series model or the more general forecast evolution model of Heath and Jackson (1994), which allows for demands to be correlated over time and across items. As in Eppen and Schrage (1981), an up-front restriction is made to a policy that increases the aggregate inventory position to a *constant*, nonadaptive, base-stock level and allocates incoming batches of the intermediate product myopically. The authors show how the correlation pattern impacts the effective standard deviation of the surrogate lead-time demand.

Fisher and Raman (1996) develop a *two*-stage model to represent *two* replenishment opportunities to cover the demand for a sales season of fashion items (in particular, skiwear): one before the start of the season and one after observing several weeks of sales. Demands in the two stages of the season are assumed to be correlated and follow a known joint distribution. Thus, demand observed during the first stage may be used to revise the demand distribution pertaining to the second stage, and hence the second replenishment orders. The model assumes multiple items with an upper and lower bound on the second stage purchase quantity, aggregated over all items. We refer to Fisher and Raman (1996) for a review of a handful of other similar models with two replenishment opportunities. Iyer and Bergen (1997) characterize the benefits of revising the demand distributions in a *single*-item, *single*-period model. As in our paper, it is assumed that the demand distribution is normal with an unknown mean, but with a given prior distribution.

In §2, we specify the basic model and its notation. Section 3 addresses the problem of allocating an incoming order of the intermediate product among the finished items and develops closed-form approximations for the optimal values of these allocation problems. Of interest by themselves, these approximations are needed to reduce the exact but intractable dynamic programming model to one that is two-dimensional and interpretable as a Bayesian version of a single-stage, single-item inventory model. The reduction is achieved in §4. It is well known that the performance of this type of single item inventory systems is determined by the characteristics of its lead-time demand. Section 5 studies these characteristics and its impact on the benefits of delayed differentiation as a function of all of the model parameters. In §6, we characterize the structure of an optimal ordering policy in the approximate model, for various types of order cost functions. Finally, in §7, we discuss extensions of the basic model in which the demand processes are correlated across the different items or where the period-by-period deviations from the mean demands are correlated across time.

## 2. MODEL AND NOTATION

Consider a company offering a product line of $J$ final products with exogenous, random demands. Inventories are reviewed and decisions are taken periodically. The $J$ items are produced in two stages. First, a common intermediate product is manufactured; this stage requires a lead time of $L$ periods. In the second stage of $l$ periods, the common intermediate product is differentiated into the finished goods.

In our basic model, we assume that no inventory is carried of the intermediate product, i.e., as soon as an order of the intermediate product is completed, it is allocated to the finished goods, and the second manufacturing stage commences. This assumption reflects many production processes; sometimes, it is physically impossible or highly expensive to store the intermediate product (perhaps because it is highly perishable or dangerous, as in smelting processes). In other settings, intermediate inventories are avoided, as a company policy, to reduce the lead time and minimize material handling costs. If, as mentioned in the introduction, our model is applied to two-stage distribution processes with a distribution center feeding several retail outlets, the assumption of no central stocks is often referred to as a *cross docking* policy, implemented by Wal-Mart and many other retail chains (see, e.g., Stalk et al. 1992). This assumption allows us to focus on two of the three benefits of delayed product differentiation, namely the statistical economies of scale and the learning effect.

The planning horizon consists of $N(\leqslant \infty)$ periods. Let $d_{jn}$ be the demand for item $j$ in period $n$ ($j = 1, \ldots, J$; $n = 1, \ldots, N$). For all $j = 1, \ldots, J$, we assume that the $d_{jn}$ variables are normally distributed with mean $\mu_j$ ($j = 1, \ldots, J$) and standard deviation $\sigma_j$ ($j = 1, \ldots, J$), for all $n = 1, \ldots, N$; i.e., the distributions are identical across time. However, contrary to standard inventory models, we assume that while the parameters $\{\sigma_j : j = 1, \ldots, J\}$ are known, the means $\{\mu_j : j = 1, \ldots, J\}$ are not. It is thus useful to write

$$d_{jn} = \mu_j + \epsilon_{jn}, \quad n = 1, 2, \ldots, \quad \text{and} \quad j = 1, \ldots, J, \quad (1)$$

where the $\epsilon_{jn}$-variables have a known normal distribution. Instead, our initial uncertainty about the mean demand $\mu_j$ is characterized by a prior distribution, itself assumed to be normal, with mean $\mu_{0j}$ and standard deviation $\sigma_{0j} = \eta\sigma_j$ for some $\eta \geqslant 0$ ($j = 1, \ldots, J$).

The additive demand model (1) applies to settings in which the average demand volume per period is unknown before the start of the sales season, but the variability of the period-by-period deviations from the mean is accurately predictable. This situation arises, for example, when a company sells its goods via two channels, both drawing from the same central inventory. The first channel may consist of export contracts or wholesale sales to department stores or other major retail organizations whose overall purchase volumes may not be known with certainty before the start of the season, but whose demands are predictably or evenly spread during the course of the season. The second channel may consist of an established network of retail outlets facing a mature consumer market, but with significant sales deviations from period to period because of weather conditions, promotional activities, the randomness of shopping patterns, etc. Iyer and Bergen (1997) describe how Model (1) is used as the basis of a forecasting system by quick response apparel distributors. Inputs by a team of buyers are aggregated into a prior distribution for the mean of any new item. The standard deviation of weakly sales, on the other hand, is adequately inferred from point-of-sales data of past seasons for similar garments and designs.

In other common settings, it is more reasonable to assume that the *coefficients of variation* of the demand variables $\{d_{jn}\}$ are known rather than their standard deviations per se. In §7.3, we outline a model based on this alternative assumption.

The assumed distributional form gives rise to a so-called *conjugate pair*. Posterior distributions for the demands in any given period, conditioned upon observed demands, are again normal; see Theorem 1 as well as DeGroot (1970), for example. The assumption that the standard deviations $\{\sigma_{0j}\}$ and $\{\sigma_j\}$ are proportional is equivalent to the assumption that the unconditional standard deviation of sales is proportional to the conditional standard deviation of sales, when conditioned upon the value of the means $\{\mu_j\}$. This assumption appears appropriate in most settings. Note that the above specifications allow for arbitrary and nonidentical coefficients of variation of both the unconditional and conditional demand distributions.

We initially assume that the demand processes for the different items are independent of each other, and that for any $j = 1, \ldots, J$, the conditional distributions $\{(d_{jn}|\mu_j)\colon n = 1, \ldots, N\}$ are independent as well. Note that while the conditional distributions $(d_{jn}|\mu_j)$ are independent across time, the unconditional distributions are not. To the contrary, observation of the realized demands $d_{jn} (j = 1, \ldots, J)$ in period $n$ permits us to revise our forecasts for all future periods, i.e., to revise the conditional distributions $\{d_{jn'}\colon n' > n\}$. In other words, uncertainty about some of the parameters in the distributions generates correlation across time. In §7, we discuss extensions of the basic model in which even the conditional distributions $(d_{jn}|\mu_j)$ are correlated across time or across items.

The sequence of events is as follows. At the beginning of each period, one decides on the magnitude of a new production order, if any. Any order of the intermediate product that has just been completed (i.e., that was ordered $L$ periods ago) is allocated to the final products. Thereafter, demands of the final products are observed and end-of-the-period inventories assessed.

Unfilled demand for each one of the items is backlogged. Linear costs are incurred for each item, both for holding inventory and for its backorders. Each time a production order is initiated, an ordering cost is assessed. We consider a variety of structures for the ordering cost functions; these can represent, for example, economies of scale resulting from fixed costs. Finally, we assume that the second stage differentiation costs are proportional with the specific production quantities per item. More specifically, we introduce the following notation.

Let

$x_{jn}$ = the inventory position of product $j$ at the beginning of period $n$, before this period's allocation, i.e., the inventory on hand − backlogs + common product units being transformed into product $j$.

$y_n$ = the size of a new production order for the intermediate product, initiated at the beginning of period $n$, if any.

$z_{jn}$ = the allocation to item $j$ in period $n$.

$I_{n-1} = (d_{11}, \ldots, d_{J1}, \ldots, d_{1,n-1}, \ldots, d_{J,n-1})$ the history of observed demand realizations at the beginning of period $n$.

We assume that the vector $(x_{11}, \ldots, x_{J1})$ of initial inventory positions for the $J$ items is exogenously given, as are any intermediate product orders that are to be completed at the beginning of periods $1, \ldots, L$. (To enable the availability of finished goods stocks in the first several periods, orders for the common intermediate product must be initiated from period $1-(L+l)$ on, i.e., $(L+l)$ periods prior to the start of the planning horizon. We denote these orders by $y_{1-(L+l)}, y_{2-(L+l)}, \ldots, y_0$ and assume they are exogenously given.) The following parameters describe the cost structure.

$h_{jn}$ = the holding cost per unit of item $j$ carried at the end of period $n$ $(j = 1, \ldots, J; n = 1, \ldots, N)$.

$p_{jn}$ = the cost for each unit of item $j$ that is backlogged at the end of period $n$ $(j = 1, \ldots, J; n = 1, \ldots, N)$.

To avoid pathological cases, we assume $h_{jn}, p_{jn} > 0$ for all $j$ and $n$.

$c_{jn}$ = the second-stage manufacturing cost per unit of item $j$, where production is initiated in period $n$ $(j = 1, \ldots, J; n = 1, \ldots, N)$.

$\gamma_n(y)$ = cost to place an order of size $y$ in period $n$. (This function can be nonlinear; e.g., it might contain a fixed cost component.)

Also, let $N(\mu, \sigma)$ denote the normal distribution, with mean $\mu$ and standard deviation $\sigma$.

Our objective is to minimize the expected discounted (or undiscounted) cost over the planning horizon with $0 < \alpha \leqslant 1$, the one-period discount factor. As in standard inventory models, we assign to period $n$, for each item $j = 1, \ldots, J$, the expected inventory holding and backlogging costs $l$ periods later. More specifically, the expected holding and backolgging costs for item $j$, assigned to period $n$, are given by

$$h_{j,n+l}\mathsf{E}\left[x_{jn} + z_{jn} - \sum_{i=n}^{n+l} d_{ji}\right]^+ + p_{j,n+l}\mathsf{E}\left[\sum_{i=n}^{n+l} d_{ji} - x_{jn} - z_{jn}\right]^+, \quad (2)$$

with the convention that $h_{j,n} = p_{j,n} = 0$ for $n > N$. Note that the holding and backlogging costs incurred in the first $l$ periods are independent of any of our decisions; thus, as in standard inventory models with lead-times, we exclude these cost terms from consideration. Theorem 1(c) shows that the distribution of the lead time demand $\sum_{i=n}^{n+l} d_{ji}$, conditional upon all observed demand data, remains normal, with an easily computable mean $\mu$ and standard deviation $\sigma$. The expression in Equation (2) can be written as $H(x_{jn}+z_{jn}|h_{j,n+l}; p_{j,n+l}; \mu; \sigma)$ (see, e.g., Federgruen and Zipkin 1984b, p. 100), where

$$H(y|h; p; \mu\colon \sigma) \equiv p\cdot(\mu-y)+(p+h) \cdot\left[(y-\mu)\Phi\left(\frac{y-\mu}{\sigma}\right)+\sigma\phi\left(\frac{y-\mu}{\sigma}\right)\right],$$

and

$\Phi(\cdot) =$ the cumulative distribution function of the standard normal distribution;

$\phi(\cdot) =$ the probability density function of the standard normal distribution.

In the appendix we derive various properties of the $H$ function, which will be used in the remainder.

The analysis below uses the following characterization of conditional future demand distributions, given a series of observed demand data.

THEOREM 1. *Fix $j = 1, \ldots, J$. Assume a sequence of n demand observations $(d_{j1}, d_{j2}, \ldots, d_{jn})$ with mean $\bar{d}_{jn} \equiv \frac{1}{n}\sum_{i=1}^{n} d_{ji}$. Then*:

(a) *The posterior distribution of the mean demand $\mu_j$ is*

$$N\left(\mu^*_{jn} \equiv \frac{\frac{n\bar{d}_{jn}}{\sigma_j^2}+\frac{\mu_{0j}}{\sigma_{0j}^2}}{\frac{n}{\sigma_j^2}+\frac{1}{\sigma_{0j}^2}}, \sigma^*_{jn} \equiv \sqrt{\frac{1}{\frac{n}{\sigma_j^2}+\frac{1}{\sigma_{0j}^2}}}\right).$$

(b) *The posterior demand distribution is* $N(\mu^*_{jn}, \sqrt{(\sigma^*_{jn})^2+\sigma_j^2})$.

(c) *For all* $t \geqslant 0$,

$$S \equiv \left(\sum_{i=n+t+1}^{n+t+l+1} d_{ji}|d_{j1,\ldots,d_{jn}}\right) \sim N((l+1)\mu^*_{jn,}\sqrt{(l+1)^2(\sigma^*_{jn})^2+(l+1)\sigma_j^2}).$$

PROOF. The proof is a special case of the proof of Theorem 5.

REMARK 1 REMARK.. For notational convenience, we set $\bar{d}_{j0} = 0$ and we extend the definition of $\mu^*_j$ and $\sigma^*_j$ to include the case $n = 0$.

## 3. AN ALLOCATION PROBLEM

In this section, we consider the myopic allocation problem, i.e., the problem of allocating an incoming order of the intermediate product to minimize expected costs in the very first period in which the allocation has an impact; i.e., $l$ periods later. While of interest by itself, the allocation problem plays a crucial role in the analysis of the entire $N$-period dynamic optimization problem; see §4. That, in period $n$, can be formulated as

$$(P_n)\colon \min\sum_{j=1}^{J}\left\{c_{jn}z_{jn}+h_{j,n+l}\mathsf{E}\left(\left[x_{jn}+z_{jn}-\sum_{i=n}^{n+l}d_{ji}\right]^+|I_{n-1}\right)\right. \left.+p_{j,n+l}\mathsf{E}\left(\left[\sum_{i=n}^{n+l}d_{ji}-x_{jn}-z_{jn}\right]^+|I_{n-1}\right)\right\}, \quad (3)$$

$$\text{s.t } \sum_{j=1}^{J} z_{jn} = y_{n-L}, \quad (4)$$

$$z_{jn} \geqslant 0, \quad j=1,\ldots,J. \quad (5)$$

Invoking Theorem 1, the objective in Equation (3) may be written as

$$\min\sum_{j=1}^{J}\left\{c_{jn}z_{jn}+H(x_{jn}+z_{jn}|h_{j,\,n+l}; p_{j,\,n+l}; (l+1)\mu^*_{j,\,n-1}; \sqrt{(l+1)^2\sigma^{*2}_{j,\,n-1}+(l+1)\sigma_j^2})\right\}, \quad (6)$$

and is separable and convex in the $z$-variables; see Theorem A part (d) in the appendix. As a consequence, several highly efficient solution methods for ($P_n$) prevail. An optimal integer solution can be found by the greedy procedure allocating each of the $y_{n-L}$ units sequentially to whichever item benefits most from this units's allocation; see Gross (1956) and Fox (1966). See Zipkin (1980) for an efficient method to compute the continuous optimum of ($P_n$).

As shown in Zipkin (1982), it is even possible to obtain a closed-form approximate expression for the solution value of ($P_n$). If all cost parameters are identical across items (i.e., $c_{jn} = c_n$; $h_{jn} = h_n$; $p_{jn} = p_n$ for all $j = 1, \ldots, J$), the following closed-form lower bound approximation is obtained by relaxing the nonnegativity constraints (5): $c_n y_{n-L} + \widehat{R}_n(X_n + y_{n-L}|I_{n-1})$, where

$$\widehat{R}_n(X_n+y_{n-L}|I_{n-1}) = H\left(X_n+y_{n-L}|h_{n+l}; p_{n+l}; (l+1)\sum_{j=1}^{J}\mu^*_{j,n-1}\colon \sum_{j=1}^{J}\sqrt{(l+1)^2\sigma^{*2}_{j,n-1}+(l+1)\sigma_j^2}\right), \quad (7)$$

and $X_n \equiv \sum_{j=1}^{J} x_{jn}$. If the cost parameters are item-dependent, Equation (7) continues to be usable as an extremely accurate approximation, provided $c_n$, $h_{n+l}$, and $p_{n+l}$ are specified as appropriate weighted averages of the parameters $\{c_{jn}\}$, $\{h_{j,n+l}\}$, and $\{p_{j,n+l}\}$; see Zipkin (1982) for details. Observe that

$$\sum_{j=1}^{J}\mu_{jn}^{*}=\sum_{j=1}^{J}\left[\frac{n\bar{d}_{jn}\sigma_{0j}^{2}+\mu_{0j}\sigma_{j}^{2}}{n\sigma_{0j}^{2}+\sigma_{j}^{2}}\right]=\sum_{j=1}^{J}\left[\frac{n\eta^{2}\bar{d}_{jn}+\mu_{0j}}{n\eta^{2}+1}\right]$$
$$=\frac{n\eta^{2}\overline{D}_{n}+\sum_{j}\mu_{0j}}{n\eta^{2}+1}, \tag{8}$$

where $\overline{D}_n \equiv \sum_j \bar{d}_{jn}$. Also,

$$\sigma_{jn}^{*}=\sqrt{\frac{\sigma_{0j}^{2}}{n\sigma_{0j}^{2}/\sigma_{j}^{2}+1}}=\sqrt{\frac{\eta^{2}}{n\eta^{2}+1}}\cdot\sigma_{j}. \tag{9}$$

Substituting Equations (8) and (9) in (7), we obtain

$$\widehat{R}_{n}(X_{n}+y_{n-L}|I_{n-1})=H\Bigg(X_{n}+y_{n-L}|h_{n+l};\,p_{n+l};\,(l+1)\left[\frac{(n-1)\eta^{2}\overline{D}_{n-1}+\sum_{j}\mu_{0j}}{(n-1)\eta^{2}+1}\right];\,\sqrt{\frac{(l+1)^{2}\eta^{2}}{(n-1)\eta^{2}+1}+(l+1)}\cdot\left(\sum_{j=1}^{J}\sigma_{j}\right)\Bigg). \tag{10}$$

$\widehat{R}_n$ depends on the history of observed demand values only via the *single* sufficient statistic $\overline{D}_{n-1}$, i.e., the mean of its aggregate demand, and on the vectors of a priori means $\{\mu_{0j}: j=1,\ldots,J\}$ and standard deviations $\{\sigma_j: j=1,\ldots,J\}$ only via their aggregates $(\sum_j \mu_{0j})$ and $(\sum_j \sigma_j)$. (We henceforth write $\widehat{R}_n(\cdot|\overline{D}_{n-1})$ instead of $\widehat{R}_n(\cdot|I_{n-1})$.)

## 4. THE DYNAMIC PROGRAM

An exact dynamic program has an overwhelmingly large state space because the state of the system at the beginning of period $n$ involves (i) the vector of inventory positions $x=(x_{1n}, x_{2n}, \ldots, x_{Jn})$, (ii) all outstanding orders for the intermediate product $y=(y_{n-L}, \ldots, y_{n-1})$, and (iii) the complete history of observed demands $I_{n-1}$. The state space is thus of dimension $(Jn+L)$ and is growing as time progresses. Fortunately, Theorem 1 shows that *all* demand distributions and cost measures pertaining to period $n$ and to later periods depend on the observed history of demand values $I_{n-1}$ only through the vector of historically observed averages $\bar{d}=\{\bar{d}_{j,n-1}=\frac{1}{n-1}\sum_{i=1}^{n-1}d_{ji};\, j=1,\ldots,J\}$ as sufficient statistics. Thus, for any $n \geqslant 1$, let

$V_n(x;y;\bar{d})=$ minimum expected total cost charged to periods $n, n+1, \ldots, N$ when at the beginning of period $n$, item $j$'s inventory position equals $x_{jn}$, its average observed demand $\bar{d}_{j,n-1}(j=1,\ldots,J)$, and where $y$ denotes the vector of outstanding orders for the common product.

The $V_n$-functions satisfy the following dynamic programming recursion:

$$\begin{aligned}&V_{n}(x;y;\bar{d})\\&=\min_{y_{n}}\Bigg\{\gamma_{n}(y_{n})+\sum_{j=1}^{J}c_{jn}z_{jn}\\&\quad+\sum_{j=1}^{J}H\Big(x_{jn}+z_{jn}|h_{j,n+l};\,p_{j,n+l};\,(l+1)\mu_{j,n-1}^{*};\\&\quad\sqrt{(l+1)^{2}\sigma_{j,n-1}^{*2}+(l+1)\sigma_{j}^{2}}\Big)\\&\quad+\alpha\mathsf{E}_{(d|I_{n-1})}V_{n+1}(x+z-d;\,\tilde{y};\\&\quad[(n-1)\bar{d}+d]/n)|y_{n}\geqslant 0;\,z\text{ s.t. }(4),(5)\Bigg\},\end{aligned} \tag{11}$$

where $V_{N+1}\equiv 0$, $d=(d_{1n},\ldots,d_{Jn})$, $z=(z_{1n},\ldots,z_{Jn})$ and $\tilde{y}=(y_{n-L+1},\ldots,y_{n-1},y_n)$.

A few observations are in order when $N=\infty$, i.e., when the planning horizon is infinite. Clearly, when $\alpha=1$, the total expected cost measure is no longer relevant; neither is the long-run *average* cost criterion because in the long run the exact values of the unknown means $\{\mu_j\}$ become perfectly known, as is immediate from the Law of Large Numbers, under the long-run average cost criterion, the model thus reduces to the model in Federgruen and Zipkin (1984a, c). When $\alpha<1$, for any $n\geqslant 1$ the $V_n$-function is best computed as $V_n=\lim_{\overline{N}\to\infty}V_{n,\overline{N}}$, with $\overline{N}$ an arbitrary finite horizon of length $>n$. The fact that $\{V_{n,\overline{N}}\}_{\overline{N}=n+1}^{\infty}$ converges to the infinite horizon value-function follows from Bertsekas and Shreve (1978, proposition 9.17).

While the state space of Equation (11) is of dimension $(2J+L)$, as opposed to $(nJ+L)$ in the original formulation, the recursions in (11) are still entirely intractable. We thus replace the exact model by an approximate one relaxing in (11) the constraints $z_{jn}\geqslant 0$, a technique successfully employed in Eppen and Schrage (1981). Federgruen and Zipkin (1984a, b, c), and Aviv and Federgruen (2001). Extensive numerical studies in Federgruen and Zipkin (1984a, c) and Aviv and Federgruen (2001) have demonstrated that the approximation resulting from the relaxation step is very accurate (i.e., the accuracy gap is on the order of a few percentage points only), as long as the coefficients of variation of the one-period demands are not too large (e.g., if they are less than one). The approximation is all the more accurate when replenishment orders arrive and are allocated frequently, enabling the prevention of major imbalances between the items' inventories. The same approximation is used "implicitly" in many other multiechelon models (e.g., Eppen and Schrage 1981, Jönsson and Silver 1987a, b, Jackson 1988, Schwarz 1989, Erkip et al. 1990, Güllü 1997, and Garg and Tang 1997), where it is referred to as the "allocation assumption," i.e., the assumption that in each allocation round the inventory position of each item can be elevated to the same, common fractile of its lead time demand distribution.

The approximate model represents a *lower bound* in the important special case where all variable cost parameters are identical across items, i.e., for all $n = 1, 2, \ldots, c_{jn} = c_n$, $h_{jn} = h_n$ and $p_{jn} = p_n$ for all $j = 1, \ldots, J$. Most importantly, the state space of the approximate dynamic program is *two*-dimensional only and as we see in §6, has optimal strategies of simple structure. With this relaxation, one easily verifies (see e.g., §4 in Federgruen and Zipkin 1984a) that

(a) the value-functions $V_n$ depend on the vector of initial inventory positions $x$ only via their aggregate sum $X_n$, and

(b) in each period $n$ it is optimal to select the vector $z$ as an optimal solution of the myopic allocation problem ($\text{P}_n$), without the Non-negativity Constraints (5).

As shown in the previous section, the values of the relaxed myopic allocation problems can be obtained in closed form via the function $\widehat{R}_n(\cdot|\overline{D}_{n-1})$. (Recall that this function is exact when the variable cost parameters are identical across items, and it is to be used as a close approximation with $c_n$, $h_{n+l}$ and $p_{n+l}$ appropriate weighted averages of the corresponding item-dependent parameters. See Federgruen and Zipkin 1984a, c and Zipkin 1982.) The one-step expected cost function in the approximate dynamic program thus depends on the history of observed demands only via the single statistic $\overline{D}_{n-1}$. The aggregate inventory position at the beginning of period $(n+1)$ is given by $X_n + y_{n+L} - D_n$, where $D_n \equiv \sum_{j=1}^J d_{jn}$, and it follows from Theorem 1 that the conditional distribution of $(D_n|I_{n-1})$ is

$$\text{N}\left(\sum_{j=1}^J \mu^*_{j,n-1}, \sqrt{\sum_{j=1}^J [\sigma^{*2}_{j,n-1} + \sigma_j^2]}\right)$$
$$= \text{N}\left(\frac{(n-1)\eta^2 \overline{D}_{n-1} + \sum_{j=1}^J \mu_{0j}}{(n-1)\eta^2 + 1},\right.$$
$$\left.\sqrt{\frac{\eta^2}{(n-1)\eta^2+1} + 1} \cdot \sqrt{\sum_{j=1}^J \sigma_j^2}\right).$$

More specifically, $(D_n|I_{n-1})$ once again depends on $I_{n-1}$ only via $\overline{D}_{n-1}$. Finally, $\overline{D}_n = [(n-1)\overline{D}_{n-1} + D_n]/n$. We conclude that the state of the approximate dynamic program is given by $(X_n; y; \overline{D}_{n-1})$ and its value-functions satisfy the recursion

$$\underline{\widehat{V}}_n(X_n, y; \overline{D}_{n-1})$$
$$= \min_{y_n \geqslant 0} \left\{ \gamma_n(y_n) + \widehat{R}_n(X_n + y_{n-L}|\overline{D}_{n-1}) + \alpha \mathsf{E}_{(D_n|\overline{D}_{n-1})}\right.$$
$$\left. \underline{\widehat{V}}_{n+1}\left(X_n + y_{n-L} - D_n; \tilde{y}; \frac{(n-1)\overline{D}_{n-1} + D_n}{n}\right)\right\}. \quad (12)$$

The dynamic programming recursion may be interpreted as that describing a *single* item inventory model, with an order lead time of $L$ periods, full backlogging, and one-period demands $\{D_n\}$, which are normally distributed with mean $\sum_j \mu_j$ and standard deviation $\sqrt{\sum_j \sigma_j^2}$. However, contrary to the standard single-item inventory model, the mean $\sum_j \mu_j$ is not known and itself a random variable whose a priori distribution is normal with mean $\sum_j \mu_{0j}$ and standard deviation $\eta\sqrt{\sum_j \sigma_j^2}$. Also, the conditional distributions $\{(D_n|\sum_j \mu_j); n \geqslant 1\}$ are independent. Finally, the one-step expected cost in period $n$ (exclusive of ordering costs) as a function of the inventory level after receipt of incoming orders, is given by Equation (10). Note that the function depends on the history of observed demands $D_1, D_2, \ldots, D_{n-1}$ only via its average $\overline{D}_{n-1}$, and so do the future demand distributions $\{D_i; i \geqslant n\}$.

We conclude that the dynamic program recursion in Equation (12) may be interpreted as that pertaining to a Bayesian single-item inventory model, as in Scarf (1959, 1960), Karlin (1960), and Azoury (1985). A further major simplification can be achieved by assigning to period $n$ the expected value of the one-step expected costs (exclusive of ordering costs) $L$ periods later. The shift permits us to collapse all inventory-related state variables into a single aggregate variable, the system-wide inventory position, defined as follows:

$X_n^S$ = the aggregate sum of all units in stock as final products, as well as those undergoing the first- or second-stage manufacturing process minus all backlogs:

$$X_n^S = X_n + y_{n-L} + \cdots + y_{n-1}. \quad (13)$$

This result is established by the following lemma, which is proved in the appendix:

Let $R_n(X_n^S + y_n|I_{n-1})$
$$\equiv \mathsf{E}\{\widehat{R}_{n+L}(X_{n+L} + y_n|\overline{D}_{n+L-1})|X_n^S + y_n, I_{n-1}\}.$$

LEMMA 1. *Fix $n = 1, 2, \ldots,$*

$$R_n(X_n^S + y_n|I_{n-1})$$
$$= H\left(X_n^S + y_n|h_{n+L+l}; p_{n+L+l}; (L+l+1)\frac{(n-1)\eta^2\overline{D}_{n-1} + \sum_j \mu_{0j}}{(n-1)\eta^2+1};\right.$$
$$\left.\sqrt{\left[\frac{(l+1)^2\eta^2}{(n-1+L)\eta^2+1} + (l+1)\right] . \left(\sum_{j=1}^J \sigma_j\right)^2 + \frac{L[(n+l+L)\eta^2+1]^2}{[(n-1)\eta^2+1][(n-1+L)\eta^2+1]} . \left(\sum_{j=1}^J \sigma_j^2\right)}\right).$$

*In particular, $R_n(X_n^S + y_n|I_{n-1})$ is a function of $(X_n^S + y_n)$ and $\overline{D}_{n-1}$ only!*

In view of Lemma 1, we write $R_n(\cdot|\overline{D}_{n-1})$, instead of $R_n(\cdot|I_{n-1})$.

The transformation of variables (13) thus gives rise to a dynamic program with a *two-dimensional* state space and the following recursion (with $X^S = X_n^S$ and $\overline{D} = \overline{D}_{n-1}$):

$$\underline{V}_n(X^S; \overline{D})$$
$$= \min_{y_n \geqslant 0} \left\{ \gamma_n(y_n) + c_{n+L} y_n + R_n(X^S + y_n|\overline{D}) + \right.$$
$$\left. \alpha \mathsf{E}\left[\underline{V}_{n+1}\left(X^S + y_n - D_n; \frac{(n-1)\overline{D} + D_n}{n}\right)\Big|\overline{D}\right]\right\}. \quad (14)$$

In §6 we show that the approximate dynamic program (14) allows for optimal policies of a simple structure, depending on the specific shape of the ordering cost functions $\gamma_n(\cdot)$.

## 5. THE SURROGATE LEAD TIME DEMAND AND ITS EFFECTIVE STANDARD DEVIATION

The approximate model, described by the dynamic program (14), may be viewed as a Bayesian version of a single-stage, single-item inventory model with one-step expected costs $R_n(\cdot \mid \overline{D}_{n-1})$ (exclusive of ordering costs), given by the expected (linear) holding and backlogging costs after a surrogate lead time demand that is normally distributed, after $n$ periods of demand observations, with

$$\text{mean} = (L+l+1)\frac{n\eta^2\overline{D}_n + \sum_j \mu_{0_j}}{n\eta^2+1},$$

and

$$\begin{aligned}\text{ESD}^2 &\equiv (\text{standard deviation})^2\\ &= \left(\frac{(l+1)^2\eta^2}{(n+L)\eta^2+1} + (l+1)\right)\cdot\left(\sum_{j=1}^{J}\sigma_j\right)^2\\ &\quad + \left[\frac{(l+1)\eta^2}{(n+L)\eta^2+1}+1\right]^2\\ &\quad \cdot\left(\frac{L^2\eta^2}{n\eta^2+1}+L\right)\left(\sum_{j=1}^{J}\sigma_j^2\right),\end{aligned} \tag{15}$$

or alternatively,

$$\begin{aligned}\text{ESD}^2 &= \left(\frac{(l+1)^2\eta^2}{(n+L)\eta^2+1} + (l+1)\right)\cdot\left(\sum_{j=1}^{J}\sigma_j\right)^2\\ &\quad + \frac{L[(n+L+l+1)\eta^2+1]^2}{[n\eta^2+1]\cdot[(n+L)\eta^2+1]}\cdot\left(\sum_{j=1}^{J}\sigma_j^2\right).\end{aligned} \tag{16}$$

Note that the distribution of this lead time demand is *different* from the aggregate demand in the original system over the complete manufacturing and review lead time of $(L+l+1)$ periods. As in the model with perfect knowledge of all demand distributions (see Eppen and Schrage 1981 and Federgruen and Zipkin 1984a), the surrogate lead time demand has the *same* mean but a significantly larger standard deviation. Following the terminology of Erkip et al. (1990), we refer to this standard deviation (in Equations (15) and (16)) of the surrogate lead time demand as the *effective standard deviation* (ESD).

It is well known that the effective standard deviation is the prime determinant of the cost performance of single-item stochastic inventory models. For example, under perfect knowledge regarding the demand distributions, the minimum achievable long-run average cost is directly proportional with this effective standard deviation, when the ordering cost function $\gamma(\cdot)$ is linear and the demand distributions normal; see Scarf (1958) and Gallego and Moon (1993). Similarly, when the ordering cost functions contain fixed components and $(s, S)$-type policies are optimal, it is known from Ehrhardt (1979) that the optimal safety stock is roughly proportional to $\text{ESD}^{1.706}$.

In this section, we investigate how the effective standard deviation depends on (i) the relative lead times $l$ and $L$—in particular, how it changes when the point of differentiation is postponed; (ii) the initial degree of uncertainty regarding the mean demands, which is characterized by $\eta$; (iii) $n$, the number of periods in which demands have been observed; (iv) the degree of demand dispersion over a variety of finished items, characterized by $J$; and (v) the standard deviations of the one-period demands, as characterized by $\{\sigma_j\}$. We first verify the following monotonicity properties.

COROLLARY 1. *ESD is increasing in $l$, $\eta$ and in each of the standard deviations $\sigma_j$, $(j = 1, \ldots, J)$. ESD is decreasing in $n$.*

PROOF. Monotonicity in $l$ and in each of the $\sigma_j$, $(j = 1, \ldots, J)$ is immediate. To verify monotonicity in $\eta$, note first that the second term in Equation (15) is increasing in $\eta$ because both $[(n+L+l+1)\eta^2+1]/[n\eta^2+1]$ and $[(n+L+l+1)\eta^2+1]/[(n+L)\eta^2+1]$ are increasing in $\eta$, as $(n+L+l+1) > (n+L) \geqslant n$. Monotonicity of the first term in Equation (15) follows by simple calculus. The fact that ESD is decreasing in $n$ follows immediately from (15). □

Perhaps surprisingly, ESD may fail to increase when $L$, the lead time of the first manufacturing stage increases. Note that the first term in Equation (15) (or (16)) is decreasing in $L$, and this term may dominate, in particular when $(\sum_j \sigma_j)^2$ is much larger than $(\sum_j \sigma_j^2)$. (For example, when all $\sigma_j = \sigma$, $(\sum_j \sigma_j)^2 = J(\sum_j \sigma_j^2)$ so that the first term may dominate the second term in (15) and (16) by an arbitrarily large factor as $J \to \infty$.) Intuitively, this phenomenon may arise because an extension of $L$, the first manufacturing stage of the common product, permits one to observe additional demand values prior to making detailed allocations to the individual final products. The potential for this phenomenon decreases as time progresses and clearly increases with $\eta$, the degree of uncertainty surrounding the mean demand values, vanishing in the traditional model where $\eta = 0$.

Observe also that the asymptotic growth rate of ESD is linear in $l$ and $L$. This phenomenon is in sharp contrast to standard models where ESD grows as the *square root* of the lead time when, as in our case, the period-by-period deviations from the means are independent of each other. This implies that in the absence of perfect knowledge about the demand distributions, the impact of quick response programs, reducing manufacturing lead times, can be far greater than what is suggested by standard inventory models. Moreover, given Corollary 1 and our discussion about the impact of changes in the *first stage* manufacturing lead time $L$, we conclude that quick response programs are best geared toward reductions of $l$, the lead time of the second stage.

Note that the mean of the effective lead time demand in the approximate model depends only on the total manufacturing lead time $\tau = L + l$, i.e., it is independent of the point of differentiation. The square of the effective standard deviation, as a function of $\tau$ and $L$, equals

$$\mathrm{ESD}^2 = \frac{[(n+\tau+1)\eta^2+1]^2}{\eta^2[(n+L)\eta^2+1]}\left[\left(\sum_{j=1}^{J}\sigma_j\right)^2-\left(\sum_{j=1}^{J}\sigma_j^2\right)\right] - \frac{(n+\tau+1)\eta^2+1}{\eta^2} \times\left[\left(\sum_{j=1}^{J}\sigma_j\right)^2-\frac{(n+\tau+1)\eta^2+1}{n\eta^2+1}\left(\sum_{j=1}^{J}\sigma_j^2\right)\right]. \quad (17)$$

Note that, given a fixed value of $\tau$, the second term in Equation (17) is independent of $L$.

COROLLARY 2. *Assume that $\tau \equiv L+l$ is fixed. ESD is monotonically decreasing with L. The magnitude of this decrease is by itself a decreasing function of the number of periods of observed demands, n, and an increasing function of $\eta$.*

PROOF. Observe that

$$\mathrm{ESD}(L+1)-\mathrm{ESD}(L) = \frac{-[(n+\tau+1)\eta^2+1]^2}{[(n+L+1)\eta^2+1][(n+L)\eta^2+1]} \cdot\left[\left(\sum_j\sigma_j\right)^2-\left(\sum_j\sigma_j^2\right)\right]<0. \quad (18)$$

Note that both $[(n+\tau+1)\eta^2+1]/[(n+L+1)\eta^2+1]$ and $[(n+\tau+1)\eta^2+1]/[(n+L)\eta^2+1]$ are decreasing in $n$ because $(\tau+1)\eta^2+1 \geqslant (L+1)\eta^2+1 > L\eta^2+1$, and increasing in $\eta$ because $(n+\tau+1) \geqslant (n+L+1) > (n+L)$. The same monotonicities therefore apply to $|\mathrm{ESD}(L+1)-\mathrm{ESD}(L)|$. □

In other words, Corollary 2 shows that delayed differentiation always reduces ESD and that the magnitude of the reduction is especially large at the beginning of the planning horizon, and monotonically increasing with the degree of uncertainty in our a priori knowledge regarding the mean demand values. Thus, the more differentiation is postponed, the more we can benefit from *statistical economies of scale* as well as the pooling benefits associated with the *learning effect;* moreover, the magnitude of this benefit of postponed differentiation is all the larger as we face increased uncertainty about the mean demand values $\{\mu_j\}$, either because of limited historical data ($n$) or because of large initial uncertainty in our a priori assessments of these means ($\eta$). Indeed, when $\eta \to 0$ or $n \to \infty$, the benefit associated with the learning effect disappears and the benefit of a postponement of the point of differentiation by a single period is given by $(\sum_j \sigma_j)^2 - (\sum_j \sigma_j^2) > 0$, as in the model with perfect knowledge, see Federgruen and Zipkin (1984a).

Corollary 2 may also be used to compare the performance of a two-stage manufacturing system with delayed product differentiation, with one with *immediate* product differentiation, in which all end items are produced separately in a *single* manufacturing stage. Corollary 2 shows that if the single lead time $\hat{\tau}$ of the system with immediate differentiation is equal to or larger than $\tau$, product differentiation is always cost-effective, where the magnitude of the benefits depends on the system parameters as described above. On the other hand, if $\hat{\tau} < \tau$, Equation (17) may be used to calculate a break-even lead time $\hat{\tau} = \tau^0$ under which the benefits of the reduced lead time in the single-stage system exactly offset the risk-pooling and the learning effect benefits in the system with delayed differentiation. The break-even value can be computed by calculating the positive root $T^0$ of the quadratic equation in the auxiliary variable $T = (n+\hat{\tau}+1)\eta^2+1$,

$$\frac{\left(\sum\sigma_j\right)^2}{\eta^2(n\eta^2+1)}T^2 - \frac{\left(\sum\sigma_j\right)^2}{\eta^2}T = \mathrm{ESD}^2.$$

Thus, $\tau^0 = \eta^{-2} * (T^0-1)-(n+1)$. As $n \to \infty$ or $\eta \to 0$, $\tau^0$ converges to the break-even value $l + L(\sum\sigma_j^2)/(\sum\sigma_j)^2$, which varies between $l+L/J$ and $l+L=\tau$, depending on how close the variance ratio $(\sum\sigma_j^2)/(\sum\sigma_j)^2$ is to its minimum value $1/J$ or its maximum value 1. (Note that the convex objective $\sum\sigma_j^2$ is minimized, subject to $\sigma_j = S$, if all $\sigma_j = S/J$, and that it is maximized if one of the $\sigma_j$ equals $S$ and the others equal zero.) For finite $n$ and $\eta$, the break-even value $\tau^0$ is closer to $l$ in view of the increased benefits of delayed differentiation under the learning effect; see Corollary 1.

We demonstrate the dependence of ESD on its determining factors, via a series of figures. All consider instances with $\sigma_j = \sigma$ for all $j = 1, \ldots, J$.

Figure 1 shows how ESD decreases as time progresses and more and more demand realization are observed; five curves are displayed for different values of $\eta$, i.e., representing different degrees of uncertainty about the mean demands. Observe that the magnitude of the reduction of ESD greatly varies with $\eta$: when $\sigma_0 = \sigma$, i.e., when the degree of uncertainty surrounding the mean demands is

**Figure 1.** Effective Standard Deviation (ESD) as a function of the number periods of observed demands $n$. ($J=2$ identical items, $L=l=2$ and $\sigma_j = 10$.).

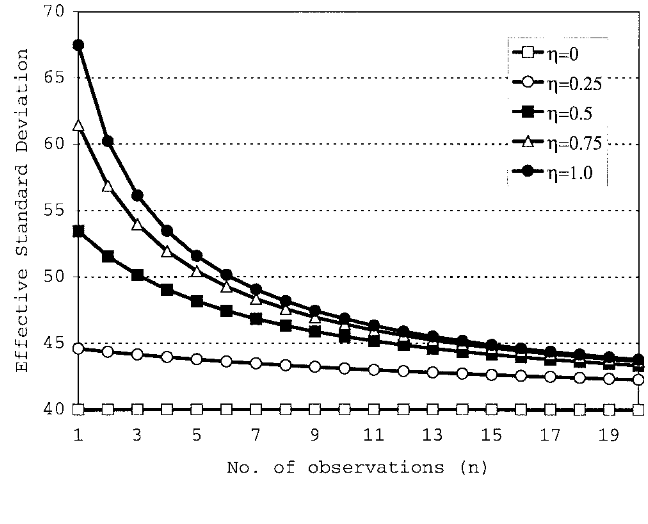
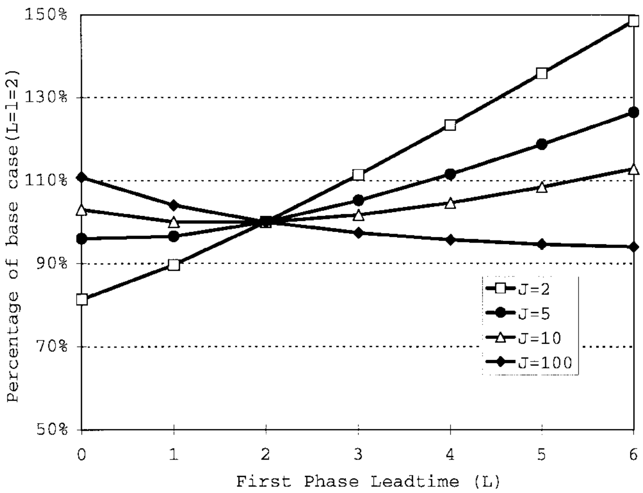


**Figure 2.** ESD, as a percentage of base case ($L = l = 2$), as a function of the first phase lead time $L$. ($n = 2$, $\eta = 1$ and $\sigma_j = 10$.).




identical to that of the period-by-period deviations from the mean, ESD decreases by 30.6% after 10 observations, compared to its values after the first period.

Figures 2 and 3 exhibit the dependence of ESD on the two lead times, $L$ and $l$; each figure consists of four curves, for different values of $J (J = 2, 5, 10$, and 100). We express ESD as a percentage of its value in the (base) case where $l = L = 2$. As Corollary 1 indicates and Figure 2 confirms, ESD is always increasing in $l$ and the rate of increase is increasing in the number of final products $J$. Figure 3 shows, as discussed above, that the dependency on $L$ is more complex. For $J = 2$ items, in this example, an increase in the first manufacturing lead time consistently results in an increase of ESD. On the other hand, with $J = 100$ final items, ESD consistently decreases with $L$ over the range $L = 0, 1, \ldots, 6$, where for the two intermediate cases, with

**Figure 3.** ESD, as a percentage of base case ($L = l = 2$), as a function of the second phase lead time $l$. ($n = 2$, $\eta = 1$ and $\sigma_j = 10$.).

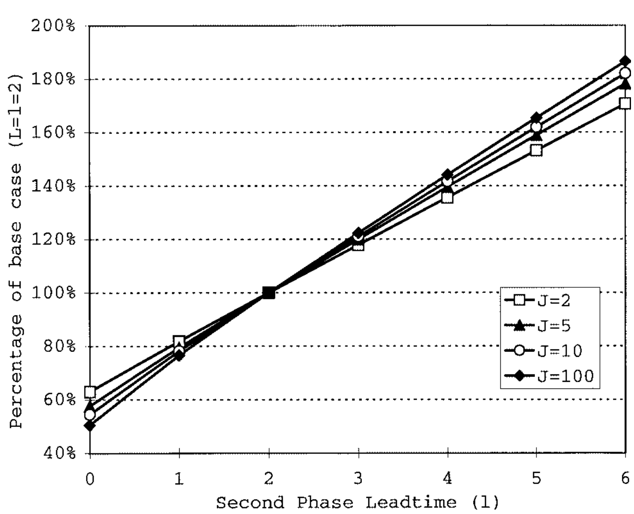
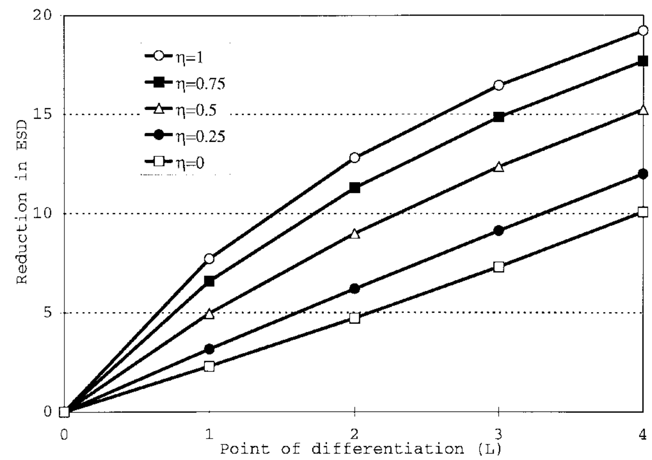



**Figure 4.** Reduction in ESD as a function of the point of differentiation $L$. ($J = 2$ identical items, $n = 2$, $\tau = L + l = 4$ and $\sigma_j = 10$.).




$J = 5$ and $J = 10$ items, ESD first decreases and then increases.

Figure 4 illustrates Corollary 2. For a fixed value of $\tau = 4$, we display the reduction in the ESD value when postponing the point of differentiation from $L = 0$ to 4. Once again, we display five curves for five values of $\eta$ Postponement is always beneficial (see Corollary 2), but the benefit of a 50% postponement ($L = l = 2$) is 2.7 times larger when $\eta = 1$ as when $\eta = 0$ (i.e., when the mean demands are perfectly known) and the benefit of *maximal* postponement ($L = 4$) is 1.9 times as large, comparing the same pair of $\eta$-values.

Finally, recall that our model assumes that the initial orders for the common product in the $(L + l)$ periods, as well as allocation decisions in the $l$ periods prior to the beginning of the planning horizon (i.e., $y_{1-(L+l)}, y_{2-(L+l)}, \ldots, y_0$ and $\{z_{jn};\ n = 1 - l, \ldots, 0,\ j = 1, \ldots, J\}$), are predetermined. Alternatively, these may be endogenously determined within the model by extending the planning horizon with periods $1 - (L + l),\ 2 - (L + l), \ldots, 0$, all with zero demands. One easily verifies that the effective lead time demand (in the approximate model) for each of these periods is again normal with an easily computable mean and standard deviation.

## 6. STRUCTURAL PROPERTIES

In this section, we characterize the structure of an optimal ordering policy (for the common intermediate product) in the approximate model, described by the recursion (14). As in standard inventory models, the structure depends heavily on the form of the order cost functions $\gamma_n(\cdot)$. We focus on the two most important types of order cost functions where they are linear and fixed-plus-linear, respectively. We also show that the results obtained for linear order costs carry over to general convex order cost functions.

Under linear order costs, we show that a base-stock policy is optimal in every period, where the base-stock level depends on the period index as well as the prevailing sufficient statistic $\overline{D}_{n-1}$ for the mean of aggregate demands. We also show that the optimal base-stock level is increasing in this sufficient statistic. Under fixed-plus-linear order costs, we show that the order policy is of an $(s, S)$-type, where both parameters $s$ and $S$ depend on the period index and the prevailing value of the estimator of the mean aggregate demand.

### 6.1. Linear Order Costs

Assume that $\gamma_n(\cdot)$ is linear for all $n = 1, \dots, N$. Let

$$U_n(y|X;\overline{D}) \equiv \gamma_n(y) + c_{n+L}y + R_n(X+y|\overline{D}) + \alpha \mathsf{E}_D\left[\underline{V}_{n+1}\left(X+y-D; \frac{(n-1)\overline{D}+D}{n}\right)\middle|\overline{D}\right]$$

denote the minimand of the recursion (14).

Theorem 2. *Assume all $\gamma_n(\cdot)$-functions are linear and $N < \infty$.*

(a) *There exist base-stock levels $\beta_n(\overline{D}_{n-1})$ such that in period $n$, it is optimal (in the approximate model) to increase the aggregate inventory position to $\beta_n(D)$ if $\overline{D}_{n-1} = D$ and $X_n^S < \beta_n(D)$, and not to order if $X_n^S \geqslant \beta_n(D)$.*

(b) *The functions $U_n(y|X;\overline{D})$ are convex in $y$ and $O(y + |X| + |\overline{D}|)$, for all $n = 1, \dots, N$.*

(c) *The functions $\underline{V}_n(X;\overline{D})$ are convex in $X$ and $O(|X| + |\overline{D}|)$, for all $n = 1, \dots, N$.*

Proof. First note that $R_n(\cdot|\overline{D})$ is of the form $H(\cdot|h, p, \mu, \sigma)$, for appropriate choices of $h, p, \mu$ and $\sigma$. By Theorem A, part (d), this function is convex and has a finite minimizer.

We prove the theorem by induction. For $n = N$, $\underline{V}_{N+1} \equiv 0$ is convex and so is $U_N(\cdot|X_N^S;\overline{D}_{N-1})$, by the convexity of $R_N(\cdot|\overline{D}_{N-1})$ and the linearity of $\gamma_N(\cdot)$. Also, a base-stock policy is optimal, where the base-stock level depends on $\overline{D}_{N-1}$. It follows that $\underline{V}_N(\cdot|\overline{D}_{N-1})$ is convex as well. Finally, $U_N(y|X_N^S;\overline{D}_{N-1})$ is $O(y + |X_N^S| + |\overline{D}_{N-1}|)$, and $\underline{V}_N(X_N^S|\overline{D}_{N-1}) = O(|X_N^S| + |\overline{D}_{N-1}|)$.

Now assume the theorem holds for some $n \leqslant N$. To show that the theorem holds for $n-1$, note that $E[\underline{V}_n(X_n^S + y - D_n; (n-1)D_{n-1}/n + D_n/n) \mid \overline{D}_{n-1}]$ is convex in $y$ by the induction assumption. Moreover, because $\underline{V}_n(X_n^S + y - D_n; (n-1)\overline{D}_{n-1}/n + D_n/n) = O(|X_n^S| + y + |D_n| + |(n-1)\overline{D}_{n-1}/n + D_n/n|)$, and since $\mathsf{E}[|D_n|] < \infty$, we have that $\mathsf{E}[\underline{V}_n(X_n^S + y - D_n; (n-1)\overline{D}_{n-1}/n + D_n/n) \mid \overline{D}_{n-1}]$ is finite and $O(|X_n^S| + y + |\overline{D}_{n-1}|)$. Because $R_{n-1}(X_{n-1}^S + y|\overline{D}_{n-2})$ is convex as well, and $\lim_{|w|\to\infty} R_{n-1}(w|\overline{D}_{n-2}) = \infty$, it follows that $U_{n-1}(y|X_{n-1}^S;\overline{D}_{n-2})$ is convex and has a finite minimizer. The remainder of the proof is identical to that given for the case $n = N$. □

We now show that in each period the optimal base-stock level is increasing in the prevailing estimator of the mean aggregate demand.

Theorem 3. *Assume that all $\gamma_n(\cdot)$-functions are linear. In each period $n$, the optimal base-stock level $\beta_n(\overline{D}_{n-1})$ is increasing in $\overline{D}_{n-1}$.*

Proof. The proof is based on results from *lattice programming*, see Topkis (1978). To apply the results there, it is useful to treat $W_n = X_n^S + y_n$, the aggregate inventory position *after* ordering as the action variable (instead of $y_n$, the order size). Because $W_n$ is real-valued, let $S = \Re$ denote its domain. Similarly, let $T = \Re^2$ denote the domain of the state spaces $\{(X_n^S, \overline{D}_{n-1})\}(n = 1, \dots, N)$ of the approximate model. Clearly, $S$ is a lattice (because if $y_1, y_2 \in S$, so is $\min(y_1, y_2)$ and $\max(y_1, y_2)$) and $T$ is a *partially ordered set* under the normal component-wise ordering. For any state $t = (X_n^S, \overline{D}_{n-1}) \in T$ define the set of *feasible* orders $S_t = \{W : W \geqslant X_n^S\} \subseteq S$. Note that the set $S_t$ is ascending in $t$ because for any $t^1 = (X^{(1)}, \overline{D}^{(1)}) \leqslant t^2 = (X^{(2)}, \overline{D}^{(2)})$ and $W^1 \in S_{t^1}$, $W^2 \in S_{t^2}$: $\min(W^1, W^2) \in S_{t^1}$ and $\max(W^1, W^2) \in S_{t^2}$. (To verify the latter, $W^1 \geqslant X^{(1)}$ and $W^2 \geqslant X^{(2)} \geqslant X^{(1)}$; hence, $X^{(1)} \leqslant \min(W^1, W^2) \in S_{t^1}$ and $X^{(2)} \leqslant W^{(2)} \leqslant \max(W^1, W^2) \in S_{t^2}$.)

In view of Topkis (1978, Theorem 6.1), it thus suffices to verify that

$$\begin{aligned}\widetilde{U}_n(W_n, X_n^S, \overline{D}_{n-1}) &\equiv U_n(W_n - X_n^S|X_n^S;\overline{D}_{n-1}) = \gamma_n(W_n - X_n^S) \\ &\quad + c_{n+L}\cdot(W_n - X_n^S) + R_n(W_n|\overline{D}_{n-1}) \\ &\quad + \alpha\mathsf{E}[\underline{V}_{n+1}(W_n - D_n; (n-1)\overline{D}_{n-1}/n + D_n/n)|\overline{D}_{n-1}]\end{aligned}$$

has antitone differences in all three of its arguments. We verify this by induction. For $n = N$, $\widetilde{U}_N$ is the sum of three terms, each of which is of the form $\varphi(x - \alpha y)$ with $\varphi(\cdot)$ a convex function, $\alpha \geqslant 0$ and $\{x, y\} \subseteq \{W_n, X_n^S, \overline{D}_{n-1}\}$; hence, all three terms have antitone differences in $(W_n, X_n^S, \overline{D}_{n-1})$. (To verify that the third term, $R_n(W_n|\overline{D}_{n-1})$, is of this form, invoke Lemma 1, the definition of the $H$-function and Theorem A part (a).)

Now assume $\widetilde{U}_n$ has antitone differences for some $n \leqslant N$. To verify the same property for $\widetilde{U}_{n-1}$, note first that

$$\underline{V}_n(X_n^S;\overline{D}_{n-1}) \text{ has antitone differences,} \tag{19}$$

by Topkis (1978, Theorem 4.3). It suffices to show that $\mathsf{E}[\underline{V}_n(W_n - D_n; (n-1)\overline{D}_{n-1}/n + D_n/n)|\overline{D}_{n-1}]$ has antitone differences in $(W_n, \overline{D}_{n-1})$, because the remainder of the proof is identical to that of the case $n = N$. Because $\mathsf{E}[\underline{V}_n(W_n - D_n; (n-1)\overline{D}_{n-1}/n + D_n/n)|\overline{D}_{n-1}] = \int \underline{V}_n(W_n - u; (n-1)u/n + D_n/n)dF(u|\overline{D}_{n-1})$, with $F(u|\overline{D}_{n-1})$ the conditional cdf of $D_n$ given $\overline{D}_{n-1}$, it suffices to show that $\underline{V}_n(W_n - u; (n-1)u/n + D_n/n)$ has antitone differences in $(W_n, \overline{D}_{n-1})$ for all $-\infty < W_n < \infty$. The latter follows from (19). If $\varphi : \Re^2 \to \Re : (x, y) \to \varphi(x, y)$ has antitone differences, then so does $\psi(x, y) \equiv \varphi(\alpha x + \beta, \gamma y + \delta)$ for any quadruple of constants $\alpha, \beta, \gamma, \delta$ with $\alpha, \gamma \geqslant 0$. □

Remark 2. Both Theorems 2 and 3 can be extended to the case where the order cost functions $\gamma_n(\cdot)$ are convex but polynomially or exponentially bounded, i.e., for all $n = 1, 2, \dots \gamma_n(y) = O(e^{\alpha y})$, $y \to \infty$ for some $\alpha > 0$. Both theorems continue to hold with immediate extensions of their proofs, provided the asymptotic order results in parts (b) and (c) of Theorem 2 are appropriately modified.

### 6.2. Fixed-Plus-Linear Order Costs

In this subsection, we assume that the order cost functions $\gamma_n(\cdot)$ consist of a fixed and a linear component, i.e.,

$$\gamma_n(y) = \begin{cases} K_n + c_n y, & \text{if } y > 0, \\ 0, & \text{if } y = 0. \end{cases} \tag{20}$$

As in the treatment of the standard inventory model, with perfect knowledge of all demand distributions, we assume (see, e.g., Scarf 1959, and Denardo 1982)

$$K_n \geqslant \alpha K_{n+1} \quad \text{for all } n = 1, \dots, N. \tag{21}$$

THEOREM 4. *Assume all order cost functions are of the form* (20), *where the fixed cost components satisfy Equation* (21). *Let* $N < \infty$. *For each* $n = 1, 2, \dots,$ *the optimal order policy in the approximate model is an* $(s, S)$*-policy where both* $s$ *and* $S$ *depend on* $\overline{D}_{n-1}$, *the prevailing value of the estimator of the mean aggregate demand and the period index. In other words, there exist functions* $s_n(\overline{D}_{n-1})$ *and* $S_n(\overline{D}_{n-1})$ *such that*

$$y_n^* = \begin{cases} S_n(\overline{D}_{n-1}) - X_n^S, & \text{if } X_n^S < s_n(\overline{D}_{n-1}), \\ 0, & \text{otherwise.} \end{cases}$$

PROOF. The approximate model may be viewed as a Markov-modulated, single-item, periodic review inventory model where the one-step expected costs and demand distributions depend on an underlying *environmental state*, which evolves according to a given Markov chain. (Note that the process $\{i_n\} = \{(n, \overline{D}_{n-1})\}$ is Markov.) In the approximate model, the modulating environmental state is given by the pair $(n, \overline{D}_{n-1})$. The theorem now follows from a straightforward extension of Sethi and Cheng (1997, Theorem 4.1), itself an adaptation of Song and Zipkin (1993). Sethi and Cheng (1997) assume that the state space of the environmental state is finite, while in our case the second component has the real line as its domain, thus requiring a simple extension of the proof of Sethi and Cheng (1997, Theorem 4.1). □

## 7. MORE GENERAL AND ALTERNATIVE DEMAND PROCESSES

Our basic model assumes that the demand processes for the different items are independent of each other and that the conditional distributions $(d_{jn}|\mu_j)$ are independent across time as well. In this section, we extend our results to settings in which the demand processes are correlated across the different items or the period-by-period deviations from the mean demands are correlated across time. We also develop an alternative to Model (1) in which the coefficients of variation of the demand variables $\{d_{jn}\}$ are assumed to be known, as opposed to their standard deviations. This alternative assumption suits many common settings.

### 7.1. Correlation Across Items

To allow for arbitrary correlation between the demands for the different items pertaining to the same period, assume that in each period $n = 1, \dots, N$, the *vector* of demands $(d_{1n,\dots,d_{Jn}})$ has a multivariate normal distribution with $\mu = (\mu_1, \dots, \mu_J)$ as the vector of means and $\Sigma$ as the variance-covariance matrix. As before, we assume that $\Sigma$ is known but that the vector $\mu$ is not. As before, our initial uncertainty about the mean demands $\mu$ is characterized by a prior distribution, itself assumed to be multivariate normal with vector of means $\mu_0 = (\mu_{01}, \dots, \mu_{0J})$ and variance-covariance matrix $\Sigma_o = \eta^2\Sigma$. We continue to assume that the conditional distributions $\{(d_{jn}|\mu_j) : n = 1, \dots, N\}$ are independent (across time). Note that this demand process model contains the basic model as a special case where the matrix $\Sigma$ is *diagonal*.

We first need the following generalization of Theorem 1: Let $\mathbf{N}_k(u, T)$ denote a $k$-variate normal distribution with $u$ as the vector of means and $T$ as the $(k \times k)$ variance-covariance matrix. Also, let $\mathbf{1}_k \equiv (1, \dots, 1)^T \in \Re^k$ and $\mathbf{0}_k \equiv (0, \dots, 0)^T \in \Re^k$.

THEOREM 5. *Assume a sequence of n demand observations* $(d_{j1}, \dots, d_{jn})$ *with mean* $\bar{d}_{jn} = \frac{1}{n}\sum_{i=1}^n d_{ji}$ *for all* $j = 1, \dots, J$. *Then*:

(a) *The posterior distribution of the vector of mean demands* $\mu$ *is multivariate normal*:

$$\mu|I_n \sim \mathbf{N}_J\left(\mu_n^* = \begin{pmatrix} \mu_{1n}^* \\ \vdots \\ \mu_{Jn}^* \end{pmatrix} \equiv \frac{1}{n\eta^2+1}\left[n\eta^2\begin{pmatrix} \bar{d}_{1n} \\ \vdots \\ \bar{d}_{Jn} \end{pmatrix} + \begin{pmatrix} \mu_{01} \\ \vdots \\ \mu_{OJ} \end{pmatrix}\right]; \quad \Sigma_n^* \equiv \frac{\eta^2}{n\eta^2+1}\Sigma\right).$$

(b) *For all* $j = 1, \dots, J$,

$$\left(\sum_{i=n+t+1}^{n+t+l+1} d_{ji}|I_n\right) \sim N\left((l+1)\mu_{jn}^*, \sqrt{(l+1)^2(\sigma_{jn}^*)^2 + (l+1)\sigma_j^2}\right),$$

*where* $\sigma_j = \sqrt{\Sigma_{jj}}$ *and* $\sigma_{jn}^* = \sqrt{\frac{\eta^2}{n\eta^2+1}\Sigma_{jj}} = \sqrt{\frac{\eta^2}{n\eta^2+1}}\sigma_j$.

PROOF. (a) Notice first that because $\Sigma$ and $\Sigma_0$ are positive definite matrices, they have inverses. DeGroot (1970, §9.9, Theorem 1) shows that the posterior distribution of $\mu|I_n$ is multivariate normal, with

$$\mu_n^* = \begin{pmatrix} \mu_{1n}^* \\ \vdots \\ \mu_{Jn}^* \end{pmatrix}$$

$$= (\Sigma_0^{-1} + n\Sigma^{-1})^{-1}\left[\Sigma_0^{-1}\mu_0 + n\Sigma^{-1}\begin{pmatrix} \bar{d}_{1n} \\ \vdots \\ \bar{d}_{Jn} \end{pmatrix}\right]$$

as the vector of means, and $\Sigma_n^* = (\Sigma_0^{-1} + n\Sigma^{-1})^{-1}$as the variance-covariance matrix. Substituting $\Sigma_0 = \eta^2\Sigma$, we obtain

$$\mu_n^* = (\eta^{-2}\Sigma^{-1} + n\Sigma^{-1})^{-1}\left[\eta^{-2}\Sigma^{-1}\mu_0 + n\Sigma^{-1}\begin{pmatrix}\bar{d}_{1n}\\ \vdots \\ \bar{d}_{Jn}\end{pmatrix}\right]$$

$$= \frac{1}{n+\eta^{-2}}(\Sigma\Sigma^{-1})^{-1}\left[\eta^{-2}\mu_0 + n\begin{pmatrix}\bar{d}_{1n}\\ \vdots \\ \bar{d}_{Jn}\end{pmatrix}\right]$$

$$= \frac{1}{n\eta^2+1}\left[n\eta^2\begin{pmatrix}\bar{d}_{1n}\\ \vdots \\ \bar{d}_{Jn}\end{pmatrix} + \mu_0\right],$$

and $\Sigma_n^* = \frac{\eta^2}{n\eta^2+1}\Sigma$.

(b) For all $j = 1, \ldots, J$ and $t = 0, 1, \ldots$:

$$\left(\sum_{i=n+t+1}^{n+t+l+1} d_{ji}|I_n\right) = (l+1)(\mu_j|I_n) + \sum_{i=n+t+1}^{n+t+l+1}\epsilon_{ji}, \tag{22}$$

in view of the independence of the $\epsilon$-variables with respect to the observed demand realizations. The second term in Equation (22) is $\mathrm{N}(0, \sqrt{l+1}\sigma_j)$, while the first term is $\mathrm{N}((l+1)\mu_{jn}^*, (l+1)\sigma_{jn}^*)$ by part (a). Part (b) follows from the independence of the two terms in (22). □

Repeating the analysis in §§3 and 4, observe first that for all $n = 1, 2, \ldots$, the allocation problem ($\mathrm{P}_n$) depends only on the *marginal* distributions of the random variables $(\Sigma_{i=n}^{n+l} d_{ji}|I_{n-1})$ for all $j = 1, \ldots, J$. Thus $\widehat{R}_n(X_n^S + y_{n-L}|I_{n-1})$, the lower bound closed-form approximation for the value of ($\mathrm{P}_n$) (excluding $c_n y_{n-L}$) continues to be given by Equation (7) and hence by (10). We conclude that, as in the basic model, $\widehat{R}_n$ depends on the entire history of observed demand values only via the *single* sufficient statistic $\overline{D}_{n-1}$, enabling us to write $\widehat{R}_n(\cdot|\overline{D}_{n-1})$ instead of $\widehat{R}_n(\cdot|\overline{I}_{n-1})$. Next, following the proof of Lemma 1, one observes that for all $n = 1, \ldots, N$ the function $R_n(X_n^S + y_n|I_{n-1})$ continues to depend on $(X_n^S + y_n)$ and $\overline{D}_{n-1}$ only and continues to be given by its expression in Lemma 1, merely replacing $(\sum_{j=1}^J \sigma_j^2)$ by $\mathrm{var}(\sum_{j=1}^J \epsilon_{jn}) = \mathbf{1}_J^T\Sigma\mathbf{1}_J$, the variance of the deviations of the *aggregate* one-period demand from its mean:

$$R_n\big(X_n^S + y_n|I_{n-1}\big) = H\Bigg(X_n^S + y_n|h_{n+L+l}; p_{n+L+l}; (L+l+1)\frac{(n-1)\eta^2\overline{D}_{n-1} + \sum_j \mu_{0j}}{(n-1)\eta^2+1}; \left(\left[\frac{(l+1)^2\eta^2}{(n-1+L)\eta^2+1} + (l+1)\right]\cdot\left(\sum_{j=1}^J\sigma_j\right)^2 + \frac{L\big[(n+l+L)\eta^2+1\big]^2}{\big[(n-1)\eta^2+1\big]\big[(n-1+L)\eta^2+1\big]}\cdot\big(\mathbf{1}_J^T\Sigma\mathbf{1}_J\big)\right)^{1/2}\Bigg). \tag{23}$$

We conclude that the approximate model (14) continues to apply with the $R_n$-functions specified as in Equation (23). In particular, the state space of the approximate model continues to be two-dimensional, and this model satisfies all of the structural properties identified in §6.

The expression for $\mathrm{ESD}^2$, after $n$ periods of demand observations, is now given by

$$\mathrm{ESD}^2 = (\text{standard deviation})^2 = \left(\frac{(l+1)^2\eta^2}{(n+L)\eta^2+1} + (l+1)\right)\cdot\left(\sum_{j=1}^J\sigma_j\right)^2 + \frac{L[(n+L+l+1)\eta^2+1]^2}{[n\eta^2+1]\cdot[(n+L)\eta^2+1]}\cdot\big(\mathbf{1}_J^T\Sigma\mathbf{1}_J\big) \tag{24}$$

(again merely replacing $(\sum_{j=1}^J\sigma_j^2)$ by $\mathbf{1}_J^T\Sigma\mathbf{1}_J$ in Equation (16)). An important special case arises when all $\sigma_j = \sigma$ and $\Sigma_{jj'} = \rho\sigma^2$ for all $j \neq j'$, i.e., all pairs of items have identical correlation:

$$ESD = \left(\frac{\sigma^2[(n+L+l+1)\eta^2+1]}{(n+L)\eta^2+1}\cdot\left[J^2(l+1) + \frac{L[J(J-1)\rho+J]\cdot[(n+L+l+1)\eta^2+1]}{n\eta^2+1}\right]\right)^{1/2}. \tag{25}$$

For this special case, and a given choice for $J, n, l$ and $L$, Figure 5 exhibits the dependency of ESD on $\eta$, the degree of uncertainty surrounding the means of single-period demands, for five distinct correlation values $\rho$. The graphs confirm that ESD is increasing in $\rho$ and, more importantly, that the benefit of a *reduction* in $\eta$ (perhaps because of better forecasting, or pre-season test marketing)

**Figure 5.** ESD, as a function of the degree of uncertainty $\eta$, for various values of $\rho$. ($J = 2$ items, $n = 2$, $L = l = 2$ and $\sigma_j = 10$.).

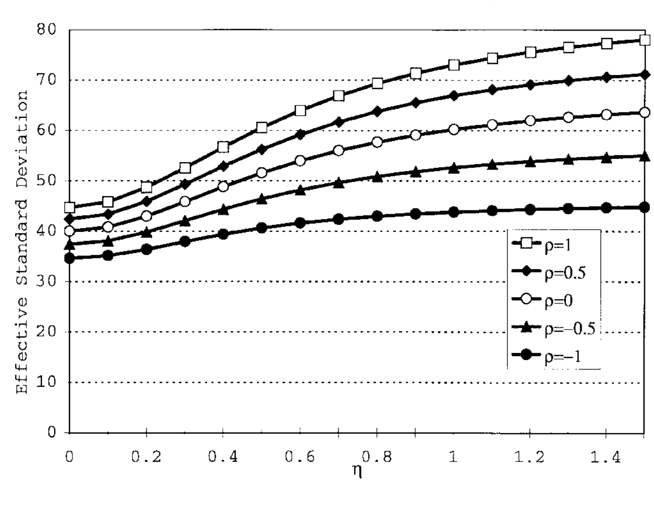
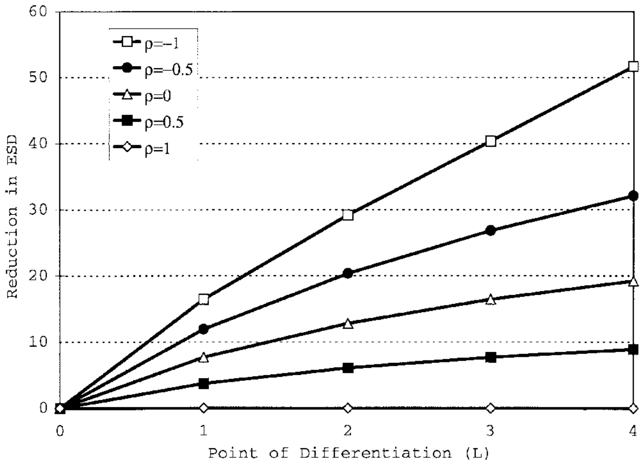


**Figure 6.** Reduction in ESD as a function of the point of differentiation $L$. ($J=2$ identical items, $n=2$, $\eta=1$, $\tau=L+l=4$ and $\sigma_j=10$.).




is increasing in $\rho$ as well. Figure 6 exhibits the reduction in ESD as a function of $L$, the point of differentiation, again in five separate graphs, for the same five values of $\rho$. Note that the benefits of postponement decrease with $\rho$, and even more strikingly, the magnitude of the incremental benefit resulting from an extension of the point of differentiation by one period, is itself decreasing in $\rho$. The first property is immediate from Equation (25) and the second follows from (18), replacing $J\sigma^2=\sum_j\sigma_j^2$ by $J\sigma^2+J(J-1)\rho\sigma^2$.

## 7.2. Correlation Across Time

The basic model assumes that in the general demand process (1), the variables $\{\epsilon_{jn}\}_{n=1}^{\infty}$ are independent, for all $j=1,\ldots,J$. We now extend our results to settings where the $\epsilon$-variables are correlated across time, in particular where they are generated by an *autoregressive time series* model. Autoregressive time series models underlie many standard demand forecasting systems, including the general class of Box-Jenkins methods. To facilitate the exposition, we confine ourselves to the simplest such model in which the time dependency is autoregressive of the *first* order, i.e.,

(i) $\epsilon_{j0}\sim \mathrm{N}(0,\sigma_j)$,

(ii) $\epsilon_{jn}=\theta\epsilon_{j,n-1}+\delta_{jn}$, $n=1,2,\ldots$, with $\{\delta_{jn}\}_{n=1}^{\infty}$ i.i.d. and independent of $\epsilon_{j0}$; $\theta<1$, and

(iii) $\delta_{j,n}\sim N(0,\sqrt{1-\theta^2}\sigma_j)$.

As with the basic model, we assume that the $\epsilon_{jn}$ variables are independent across items and that their distributions have *known* parameters $\sigma_j (j=1,\ldots,J)$ and $\theta$. Assuming perfect knowledge regarding the demand distributions, Erkip et al. (1990) address our multi-item model with aggregate demands following a (slight variant) of this first-order autoregressive pattern and demands for the individual items a fixed deterministic percentage of the aggregate. The authors restrict themselves up front to base-stock policies, combined with myopic allocations, and a *fixed* base-stock level. Our analysis shows that under linear or convex order costs, the order policies should be selected from the class of base-stock policies, with base-stock levels dependent on the last observed demand. In our setting, with unknown mean demands, the base-stock levels should be dependent on *this* as well as a *second* sufficient statistic of the observed history of demands.

It is easily shown (see, e.g., in Hamilton 1994, §3.4) that for all $j=1,\ldots,J$ and $n=1,2,\ldots$: (i) $\epsilon_{jn}\sim N(0,\sigma_j)$, and (ii) $\mathrm{cov}(\epsilon_{jn},\epsilon_{j,n+k})=\theta^k\sigma_j^2$, $k\geqslant 0$. For any $j=1,\ldots,J$, let $\epsilon_n^{(j)}\equiv(\epsilon_{j1},\ldots,\epsilon_{jn})^T$; its variance-covariance thus equals $\sigma_j^2\Theta_n$, where

$$\Theta_n=\begin{pmatrix}1 & \theta & \theta^2 & \cdots & \theta^{n-1}\\ \theta & 1 & \theta & \cdots & \theta^{n-2}\\ \theta^2 & \theta & 1 & \cdots & \theta^{n-3}\\ \vdots & \vdots & \vdots & & \vdots\\ \theta^{n-1} & \theta^{n-2} & \theta^{n-3} & \ldots & 1\end{pmatrix},$$

and

$$\Theta_n^{-1}=\frac{1}{1-\theta^2}\begin{pmatrix}1 & -\theta & 0 & 0 & \ldots & 0\\ -\theta & 1+\theta^2 & -\theta & 0 & \ldots & 0\\ 0 & -\theta & 1+\theta^2 & -\theta & \ldots & 0\\ 0 & 0 & -\theta & 1+\theta^2 & \ldots & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & \ldots & 0 & -\theta & 1\end{pmatrix}.$$

See Hamilton (1994, §3.9).

We start with an extension of Theorem 1.

THEOREM 6. *Fix $j=1,\ldots,J$. Assume a sequence of $n$ demand observations $(d_{j1},d_{j2},\ldots,d_{jn})$ with mean $\bar{d}_{jn}\equiv\frac{1}{n}\sum_{i=1}^{n}d_{j1}$. Then,*

(a) *the posterior distribution of the mean demand $\mu_j$ is normal with mean $\mu_{jn}^*$ and standard deviation $\sigma_{jn}^*=\psi(n)\sigma_j$, where*

$$\mu_{jn}^*\equiv\frac{\frac{1-\theta}{1+\theta}\left(\frac{\theta}{1-\theta}d_{j1}+n\bar{d}_{jn}\right)+\frac{\theta}{1+\theta}d_{jn}+\frac{\mu_{0j}}{\eta^2}}{\frac{2}{1+\theta}+\frac{1-\theta}{1+\theta}(n-2)+\frac{1}{\eta^2}},\tag{26}$$

*and*

$$\psi(n)\equiv\sqrt{\frac{1}{\frac{2}{1+\theta}+\frac{1-\theta}{1+\theta}(n-2)+\frac{1}{\eta^2}}}.\tag{27}$$

(b) *Let $S_{j,n+1}(t)\equiv\sum_{i=n+1}^{n+t}\alpha_{i-n}d_{ji}$, for some vector of coefficients $\alpha=(\alpha_1,\ldots,\alpha_t)$. Then, for all $t\geqslant 1$:*

$$(S_{j,n-1}(t)|d_{j1},\ldots,d_{jn})$$
$$\sim N\left(\mu_{jn}^*\sum_{i=1}^{t}\alpha_i(1-\theta^i)+d_{jn}\cdot\left(\sum_{i=1}^{t}\alpha_i\theta^i\right),\right.$$
$$\left.\sqrt{\left[\sum_{i=1}^{t}\alpha_i(1-\theta^i)\right]^2\psi^2(n)+\epsilon^2(t,\alpha)\sigma_j}\right),$$

*where*

$$\epsilon(t,\alpha)\equiv\sqrt{\sum_{p=1}^{t}\sum_{q=1}^{t}\alpha_p\alpha_q\theta^{|p-q|}-\sum_{p=1}^{t}\sum_{q=1}^{t}\alpha_p\alpha_q\theta^{p+q}}$$
$$=\sqrt{\sum_{p=1}^{t}\sum_{q=1}^{t}\alpha_p\alpha_q(\theta^{|p-q|}-\theta^{p+q})}.$$

PROOF. (a) Let $d_n^{(j)} \equiv (d_{j1}, \ldots, d_{jn})^T$. The vector representation of Model (1) is

$$d_n^{(j)} = \mu_j \mathbf{1}_n + \epsilon_n^{(j)}, \tag{28}$$

where the unconditional distribution of $\mu_j \mathbf{1}_n$ is $\mathbf{N}_n(\mu_{0j}\mathbf{1}_n, \sigma_{oj}^2 \mathbf{1}_n \mathbf{1}_n^T)$, and $\epsilon_n^{(j)} \sim \mathbf{N}_n(\mathbf{0}_n, \sigma_j^2 \Theta_n)$. Because $\Theta_n^{-1}$ is positive definite, there exists a matrix $L$ such that $L^T L = \Theta^{-1}$. Premultiply (28) with $L$ to obtain for $d_n' = L d_n^{(j)}$ and $\epsilon_n' = L\epsilon_n^{(j)}$:

$$d_n' = \mu_j L \mathbf{1}_n + \epsilon_n'. \tag{29}$$

The components of $\epsilon_n'$ are i.i.d. and normal with mean 0 and standard deviation $\sigma_j$. (By the general formula for linear transformations of multivariate normals, the variance-covariance matrix of $\epsilon_n'$ is $L\sigma_j^2 (L^T L)^{-1} L^T = \sigma_j^2 L L^{-1} (L^T)^{-1} L^T = \sigma_j^2 I_n$, where $I_n$ is the $n$-dimensional identity matrix.) Thus, (29) may be viewed as an ordinary *simple* regression equation with $d_n'$ as the vector of dependent variables, $(L\mathbf{1}_n)$ as the vector of values of the independent explanatory variable and $\mu_j$ as the regression coefficient. We thus invoke Hamilton (1994, proposition 12.2) to obtain the posterior distribution of $(\mu_j | d_{j1}, \ldots, d_{jn})$: it is normal with

$$\begin{aligned}
\text{mean} &= \left[\frac{\sigma_j^2}{\sigma_{oj}^2} + (L\mathbf{1}_n)^T (L\mathbf{1}_n)\right]^{-1} \left[\frac{\mu_{0j}\sigma_j^2}{\sigma_{oj}^2} + (L\mathbf{1}_n)^T d_n'\right] \\
&= \left[\eta^{-2} + \mathbf{1}_n^T \Theta_n^{-1} \mathbf{1}_n\right]^{-1} \left[\frac{\mu_{0j}}{\eta^2} + \mathbf{1}_n^T L^T L d_n'\right] \\
&= \left[\eta^{-2} + \mathbf{1}_n^T \Theta_n^{-1} \mathbf{1}_n\right]^{-1} \left[\frac{\mu_{0j}}{\eta^2} + \mathbf{1}_n^T \Theta_n^{-1} d_n'\right],
\end{aligned}$$

and variance $= \sigma_j^2 / [\eta^{-2} + \mathbf{1}_n^T \Theta_n^{-1} \mathbf{1}_n]$. Part (a) now follows by simple algebra noting that $\mathbf{1}_n^T \Theta_n^{-1} \mathbf{1}_n$ equals the sum of all entries of $\Theta_n^{-1}$.

(b) Let $S_{j,n+1}(t | d_{j1}, \ldots, d_{jn})$ denote the value of $S_{j,n+1}(t)$, conditional on the sequence of demand observations $(d_{j1}, \ldots, d_{jn})$. Note the the set of variables $\{\mu_j, (\epsilon_{j,n+1}, \ldots, \epsilon_{j,n+t}), (d_{j1}, \ldots, d_{jn})\}$ has a multivariate normal distribution. It follows that the partial set of variables $\{\mu_j, (\epsilon_{j,n+1}, \ldots, \epsilon_{j,n+t})\}$, *given* a value of $\{(d_{j1}, \ldots, d_{jn})\}$, continues to have a multivariate normal distribution (see, e.g., Tong 1990, Theorem 3.3.4) and hence that $S_{j,n+1}(t | d_{j1}, \ldots, d_{jn})$ has a normal distribution. By the same argument, the conditional distribution of $(\epsilon_{j,n+1}, \ldots, \epsilon_{j,n+t} | \epsilon_{jn})$ is multivariate normal.

We now show how to calculate the vector of means and the variance-covariance matrix of the conditional distribution of $(\epsilon_{j,n+1}, \ldots, \epsilon_{j,n+t} | \epsilon_{jn})$. First, observe that the joint unconditional distribution of $(\epsilon_{jn}, \epsilon_{j,n+1}, \ldots, \epsilon_{j,n+t})^T$ is $\mathbf{N}_{t+1}(\mathbf{m}, \mathbf{\Lambda} = \sigma_j^2 \Theta_{t+1})$, where we partition the vector $\mathbf{m}$ and the matrix $\mathbf{\Lambda}$ as follows:

$$\mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_{11} & \mathbf{\Lambda}_{12} \\ \mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{22} \end{pmatrix},$$

with $\mathbf{m}_1 = 0$, $\mathbf{m}_2 = \mathbf{0}_t$, $\mathbf{\Lambda}_{11} = \boldsymbol{\sigma}_j^2$, $\mathbf{\Lambda}_{12} = \mathbf{\Lambda}_{21}^T = \sigma_j^2(\theta, \theta^2, \ldots, \theta^t)$ and $\mathbf{\Lambda}_{22} = \sigma_j^2 \Theta_t$. It now follows from Tong (1990, Theorems 3.3.15 and 3.3.19) that

$$\begin{aligned}
\mathsf{E}[\epsilon_{j,n+1}, \ldots, \epsilon_{j,n+t} | \epsilon_{jn}] &= \mathbf{m}_2 + \mathbf{\Lambda}_{21}\mathbf{\Lambda}_{11}^{-1}(\epsilon_{jn} - \mathbf{m}_1) \\
&= \epsilon_{jn} \cdot (\theta, \theta^2, \ldots, \theta^t)^T,
\end{aligned} \tag{30}$$

and

$$\begin{aligned}
&\operatorname{Var}(\epsilon_{j,n+1}, \ldots, \epsilon_{j,n+t} | \epsilon_{jn}) \\
&\quad = \mathbf{\Lambda}_{22} - \mathbf{\Lambda}_{21}\mathbf{\Lambda}_{11}^{-1}\mathbf{\Lambda}_{12} \\
&\quad = \sigma_j^2 \Theta_t - \sigma_j^2 \begin{pmatrix} \theta^2 & \theta^3 & \cdots & \theta^{t+1} \\ \theta^3 & \theta^4 & \cdots & \theta^{t+2} \\ \vdots & \vdots & & \vdots \\ \theta^{t+1} & \theta^{t+2} & \cdots & \theta^{2t} \end{pmatrix}.
\end{aligned} \tag{31}$$

To calculate the mean and standard deviation of $S_{j,n+1}(t | d_{j1}, \ldots, d_{jn})$, observe that

$$\begin{aligned}
&\mathsf{E}[S_{j,n+1}(t | d_{j1}, \ldots, d_{jn})] \\
&\quad = \mathsf{E}_{\mu_j}\Bigg\{\mathsf{E}_{\epsilon_{j,n+1},\ldots,\epsilon_{j,n+t}}\Bigg[\Bigg(\mu_j \sum_{i=n+1}^{n+t} \alpha_{i-n} + \sum_{i=n+1}^{n+t} \alpha_{i-n}\epsilon_{ji}\Bigg)\Bigg| \\
&\qquad\qquad \mu_j, d_{j1}, \ldots, d_{jn}\Bigg]\Bigg| d_{j1}, \ldots, d_{jn}\Bigg\} \\
&\quad = \mathsf{E}_{\mu_j}\Bigg\{\mu_j \sum_{i=1}^{t} \alpha_i + \mathsf{E}_{\epsilon_{j,n+1},\ldots,\epsilon_{j,n+t}} \\
&\qquad\qquad \Bigg[\sum_{i=n+1}^{n+t} \alpha_{i-n}\epsilon_{ji} \,\Bigg|\, \epsilon_{jn} = d_{jn} - \mu_j\Bigg]\Bigg| d_{j1}, \ldots, d_{jn}\Bigg\} \\
&\quad = \mathsf{E}_{\mu_j}\Bigg\{\mu_j \sum_{i=1}^{t} \alpha_i + (d_{jn} - \mu_j) \sum_{i=n+1}^{n+t} \alpha_{i-n}\theta^{i-n} \Bigg| d_{j1}, \ldots, d_{jn}\Bigg\} \\
&\quad = \Bigg[\sum_{i=1}^{t} \alpha_i (1 - \theta^i)\Bigg] \cdot \mathsf{E}_{\mu_j}\{\mu_j | d_{j1}, \ldots, d_{jn}\} + d_{jn} \sum_{i=1}^{t} \alpha_i \theta^i \\
&\quad = \mu_{jn}^* \sum_{i=1}^{t} \alpha_i (1 - \theta^i) + d_{jn} \sum_{i=1}^{t} \alpha_i \theta^i,
\end{aligned}$$

where the third equality follows from Equation (30). Similarly,

$$\begin{aligned}
&\operatorname{Var}\big[S_{j,n+1}(t | d_{j1}, \ldots, d_{jn})\big] \\
&\quad = \operatorname{Var}_{\mu_j}\Bigg\{\mathsf{E}_{\epsilon_{j,n+1},\ldots,\epsilon_{j,n+t}}\Bigg[\mu_j \sum_{i=n+1}^{n+t} \alpha_{i-n} + \sum_{i=n+1}^{n+t} \alpha_{i-n}\epsilon_{ji} \Bigg| \mu_j, \\
&\qquad\qquad d_{j1}, \ldots, d_{jn}\Bigg]\Bigg| d_{j1}, \ldots, d_{jn}\Bigg\} \\
&\qquad + \mathsf{E}_{\mu j}\Bigg\{\operatorname{Var}_{\epsilon_{j,n+1},\ldots,\epsilon_{j,n+t}}\Bigg(\mu_j \sum_{i=n+1}^{n+t} \alpha_{i-n} \\
&\qquad\qquad + \sum_{i=n+1}^{n+t} \alpha_{i-n}\epsilon_{ji} \Bigg| \mu_j, d_{j1}, \ldots, d_{jn}\Bigg)\Bigg| d_{j1}, \ldots, d_{jn}\Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&= \left[\sum_{i=1}^{t} \alpha_i (1-\theta^i)\right]^2 \mathrm{Var}_{\mu_j}(\mu_j | d_{j1}, \ldots, d_{jn}) \\
&\quad + \mathsf{E}_{\mu_j}\left\{ \mathrm{Var}\left( \sum_{i=n+1}^{n+t} \alpha_{i-n}\epsilon_{ji} \middle| \epsilon_{jn} = d_{jn} - \mu_j \right) \middle| \mu_j, d_{j1}, \ldots, d_{jn} \right\} \\
&= \left[\sum_{i=1}^{t} \alpha_i (1-\theta^i)\right]^2 \psi^2(n)\sigma_j^2 \\
&\quad + \mathsf{E}_{\mu_j}\left\{ \sigma_j^2 \sum_{p=1}^{t}\sum_{q=1}^{t} \alpha_p\alpha_q \theta^{|p-q|} - \sigma_j^2 \sum_{p=1}^{t}\sum_{q=1}^{t} \alpha_p\alpha_q\theta^{p+q} \middle| \mu_j, d_{j1}, \ldots, d_{jn} \right\} \\
&= \left[\sum_{i=1}^{t} \alpha_i (1-\theta^i)\right]^2 \psi^2(n)\sigma_j^2 \\
&\quad + \left[ \sum_{p=1}^{t}\sum_{q=1}^{t} \alpha_p\alpha_q \theta^{|p-q|} - \sum_{p=1}^{t}\sum_{q=1}^{t} \alpha_p\alpha_q\theta^{p+q} \right]\sigma_j^2.
\end{aligned}
$$

The second equality is justified by the fact that the expected value, inside the curled brackets in the first term to its left, equals $\mu_j \sum_{i=1}^{t} \alpha_i(1-\theta^i)$, plus additional terms that are constant (in $\mu_j$). The third equality follows from Equation (31). Also, note that the $(p, q)$ entry of $\Theta_t$ (i.e., $[\Theta_t]_{p,q}$) equals $\theta^{|p-q|}$. $\square$

Remark 3. We thus conclude that for any $j = 1, \ldots, J$, after observing demands over $n$ periods, the posterior distribution of the mean $\mu_j$ as well as any functional of future demands, depend on *two* sufficient statistics: (i) an adjusted average of observed demands where a higher weight is attributed to the first demand observation, and (ii) the last observed demand. As mentioned above, only the second statistic is needed when the mean is perfectly known ($\eta = 0$), and only the first statistic when $\theta = 0$, i.e., in the absence of intertemporal correlation. Note that if $\theta = 0$, the first statistic reduces to the simple sample mean, as in the basic model.

We now show how to extend the results of §§3 and 4 to fit to this model. Note first that the expected holding and backlogging costs for item $j$, assigned to period $n$, are now given by

$$
H\left( \cdot | h_{n+l}; p_{n+l}; \mu^*_{j,n-1} \sum_{i=1}^{l+1}(1-\theta^i) + d_{j,n-1} \cdot \left(\sum_{i=1}^{l+1}\theta^i\right), \sqrt{\left[\sum_{i=1}^{l+1}(1-\theta^i)\right]^2 \psi^2(n-1) + \left[\epsilon(l+1, \mathbf{1}_{l+1})\right]^2}\ \sigma_j \right).
$$

Therefore, similar to our basic model, one easily verifies that the expression for $\widehat{R}_n$ needs to be adjusted to

$$
\begin{aligned}
&\widehat{R}_n(X_n + y_{n-L} | I_{n-1}) \\
&\quad = H\Bigg( X_n + y_{n-L} | h_{n+l}; p_{n+l}; \mu^*_{n-1}\sum_{i=1}^{l+1}(1-\theta^i) + D_{n-1}\cdot\left(\sum_{i=1}^{l+1}\theta^i\right), \\
&\qquad \sqrt{\left[\sum_{i=1}^{l+1}(1-\theta^i)\right]^2 \psi^2(n-1) + \left[\epsilon(l+1, \mathbf{1}_{l+1})\right]^2}\ \sum_{j=1}^{J}\sigma_j \Bigg),
\end{aligned}
$$

where $\mu^*_n$ is given by Equation (26), merely by replacing $d_{j1}$, $d_{jn}$, and $\bar{d}_{jn}$ by $D_1$, $D_n$ and $\overline{D}_n$, respectively. Moreover, $\widehat{R}_n(X_n^S + y_{n-L} | I_{n-1})$, as well as the distribution of $D_n$, depend on $I_{n-1}$ only via the *two* sufficient statistics $Q^{(1)}_{n-1} \equiv \frac{\theta}{(1-\theta)(n-1)} D_1 + \overline{D}_{n-1}$ and $Q^{(2)}_{n-1} \equiv D_{n-1}$; hence, we can write $\widehat{R}_n(\cdot | Q^{(1)}_{n-1}, Q^{(2)}_{n-1})$ instead of $\widehat{R}_n(\cdot | I_{n-1})$. The sufficient statistics evolve as simple linear functions of $D_n$: $Q^{(1)}_n = \frac{(n-1)Q^{(1)}_{n-1} + D_n}{n}$ and $Q^{(2)}_n = D_n$. Therefore, the approximate dynamic program in Equation (12) is now given by

$$
\begin{aligned}
&\underline{\widehat{V}}_n(X_n, y_{n-L}, \ldots, y_{n-1}; Q^{(1)}_{n-1}, Q^{(2)}_{n-1}) \\
&\quad = \min_{y_n \geqslant 0} \Bigg\{ \gamma_n(y_n) + \widehat{R}_n(X_n + y_{n-L} \mid Q^{(1)}_{n-1}, Q^{(2)}_{n-1}) \\
&\qquad + \alpha \mathsf{E}_{(D_n | Q^{(1)}_{n-1}, Q^{(2)}_{n-1})} \underline{\widehat{V}}_{n+1}\left( X_n + y_{n-L} - D_n; y_{n-L+1}, \ldots, y_{n-1}, y_n; \frac{(n-1)Q^{(1)}_{n-j} + D_n}{n}, D_n \right) \Bigg\}.
\end{aligned}
$$

As in §4, we further simplify the (approximate) dynamic program by collapsing all inventory-related state variables into the single, aggregate variable, $X_n^S$. The expected value $R(X_n^S + y_n | I_{n-1})$ is now given by

$$
\begin{aligned}
&R_n(X_n^S + y_n | I_{n-1}) \\
&\quad = \mathsf{E}\left\{ \widehat{R}_{n+L}(X_{n+L} + y_n | Q^{(1)}_{n+L-1}, Q^{(2)}_{n+L-1}) \middle| X_n^S + y_n, I_{n-1} \right\} \\
&\quad = \mathsf{E}\left\{ \widehat{R}_{n+L}\left( X_n^S + y_n - \sum_{i=n}^{n+L-1} D_i \,\middle|\, \frac{(n-1)Q^{(1)}_{n-1} + \sum_{i=n}^{n+L-1} D_i}{n+L-1},\ D_{n+L-1} \right) \middle| I_{n-1} \right\} \\
&\quad = \mathsf{E}\Bigg\{ H\Bigg( X_n^S + y_n - \sum_{i=n}^{n+L-1} D_i | h_{n+L+l}; p_{n+L+l}; \\
&\qquad \left[\sum_{s=1}^{l+1}(1-\theta^s)\right]\cdot\psi^2(n+L-1) \\
&\qquad \cdot\left[ \frac{1-\theta}{1+\theta}((n-1)Q^{(1)}_{n-1} + \sum_{i=n}^{n+L-1} D_i) + \frac{\theta}{1+\theta} D_{n+L-1} + \frac{\sum_j \mu_{0j}}{\eta^2} \right] + D_{n+L-1}\cdot\sum_{s=1}^{l+1}\theta^s;
\end{aligned}
$$

$$\sqrt{\left[\sum_{s=1}^{l+1}(1-\theta^s)\right]^2\cdot\psi^2(n+L-1)+\left[\epsilon(l+1,\mathbf{1}_{l+1})\right]^2}\cdot\left(\sum_{j=1}^{J}\sigma_j\right)\Bigg)\Bigg|Q_{n-1}^{(1)},Q_{n-1}^{(2)}\Bigg\},$$

where the last equality follows from the fact that the (joint) distribution of $(D_n,\ldots,D_{n+L-1}|I_{n-1})$ depends on $I_{n-1}$ solely via the two sufficient statistics $(Q_{n-1}^{(1)},Q_{n-1}^{(2)})$; thus, we can write $R_n(\cdot|Q_{n-1}^{(1)},Q_{n-1}^{(2)})$ instead of $R_n(\cdot|I_{n-1})$. One can easily verify that the approximate dynamic program (14) can now be written as follows:

$$\begin{aligned}&\underline{V}_n\left(X_n^S;Q_{n-1}^{(1)},Q_{n-1}^{(2)}\right)\\&=\min_{y_n\geqslant 0}\Bigg\{\gamma_n(y_n)+c_{n+L}y_n+R_n(X_n^S+y_n|Q_{n-1}^{(1)},Q_{n-1}^{(2)})\\&\quad+\alpha\mathsf{E}\Bigg[\underline{V}_{n+1}\Bigg(X_n^S+y_n-D_n;\\&\quad\frac{(n-1)Q_{n-1}^{(1)}+D_n}{n},D_n\Bigg)\Bigg|Q_{n-1}^{(1)},Q_{n-1}^{(2)}\Bigg]\Bigg\}.\end{aligned}$$

Invoking Theorem A, part (a), it can be easily shown that

$$\begin{aligned}&R_n\left(X_n^S+y_n|Q_{n-1}^{(1)},Q_{n-1}^{(2)}\right)\\&=\mathsf{E}\Bigg\{H\Bigg(X_n^S+y_n|h_{n+L+l};p_{n+L+l};\sum_{i=n}^{n+L-1}\alpha_{i-(n-1)}D_i\\&\quad+\left[\sum_{s=1}^{l+1}(1-\theta^s)\right]\psi^2(n+L-1)\\&\quad\cdot\left[\frac{1-\theta}{1+\theta}(n-1)Q_{n-1}^{(1)}+\frac{\sum_j\mu_{0j}}{\eta^2}\right]\\&\quad\sqrt{\left[\sum_{s=1}^{l+1}(1-\theta^s)\right]^2\cdot\psi^2(n+L-1)+\left[\epsilon(l+1,\mathbf{1}_{l+1})\right]^2}\\&\quad\cdot\left(\sum_{j=1}^{J}\sigma_j\right)\Bigg)\Bigg|Q_{n-1}^{(1)},Q_{n-1}^{(2)}\Bigg\},\end{aligned}$$

where

$$\alpha_i=\begin{cases}1+\psi^2(n+L-1)\frac{1-\theta}{1+\theta}\sum_{s=1}^{l+1}(1-\theta^s),\\ \qquad i=1,\ldots,L-1\\ 1+\psi^2(n+L-1)\frac{1}{1+\theta}\sum_{s=1}^{l+1}(1-\theta^s)+\sum_{s=1}^{l+1}\theta^s,\\ \qquad i=L.\end{cases}$$

Now note that by applying Theorem 6, part (b), to the aggregate demand process, we have

$$\begin{aligned}&\left(\sum_{i=n}^{n+L-1}\alpha_{i-(n-1)}D_i|Q_{n-1}^{(1)},Q_{n-1}^{(2)}\right)\\&\sim N\Bigg(\mu_{n-1}^*\sum_{s=1}^{L}\alpha_s(1-\theta^s)+D_{n-1}\cdot\left(\sum_{s=1}^{L}\alpha_s\theta^s\right),\\&\quad\sqrt{\left[\sum_{s=1}^{L}\alpha_s(1-\theta^s)\right]^2\psi^2(n-1)+[\epsilon(L,\alpha)]^2}\\&\quad\cdot\sqrt{\sum_{j=1}^{J}\sigma_j^2}\Bigg).\end{aligned}\tag{32}$$

We conclude that the approximate dynamic program reduces to one with a three-dimensional state space, i.e., its state is given by

(1) the aggregate system wide inventory position $X_n^S$;

(2) the sufficient statistic $Q_{n-1}^{(1)}$, a weighted average of the observed aggregate demands; and

(3) the sufficient statistic $Q_{n-1}^{(2)}=D_{n-1}$, the last observed aggregate demand.

Moreover, as in the basic model, this dynamic program can be viewed as a Markov modulated single-item inventory system in which the environmental state is now given by the triple $(n,Q_{n-1}^{(1)},Q_{n-1}^{(2)})$. Structural results similar to those in §6 can thus be obtained, depending on the type of order cost functions $\gamma_n(\cdot)$.

The above approach can be further extended to more sophisticated autoregressive time series patterns, except that as the order of the autoregressive time series pattern is increased by one, another state variable is needed in the resulting approximate dynamic program.

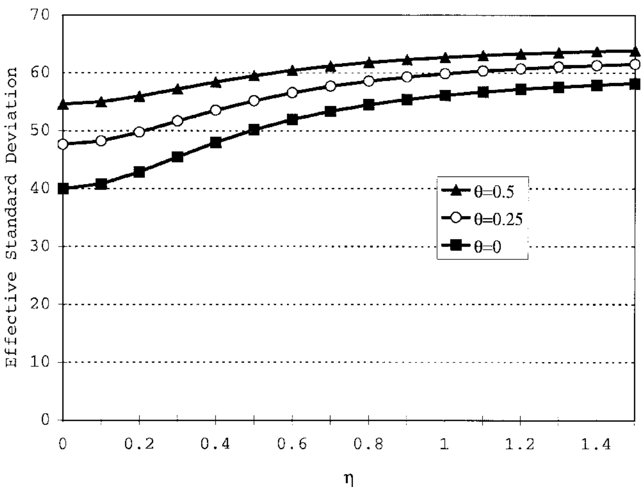
To calculate the effective standard deviation (ESD), after $n$ periods of demands are observed, we invoke Theorem A, part (c) with $\beta=1$ and

$$\tilde{\sigma}=\sqrt{\left[\sum_{s=1}^{L}\alpha_s(1-\theta^s)\right]^2\psi^2(n)+[\epsilon(L,\alpha)]^2}\cdot\sqrt{\sum_{j=1}^{J}\sigma_j^2}:$$

$$\begin{aligned}\mathrm{ESD}^2&=\left(\left[\sum_{s=1}^{l+1}(1-\theta^s)\right]^2\psi^2(n+L)+\left[\epsilon\big(l+1,\mathbf{1}_{l+1}\big)\right]^2\right)\\&\quad\cdot\left(\sum_{j=1}^{J}\sigma_j\right)^2+\left(\left[\sum_{s=1}^{L}\alpha_s(1-\theta^s)\right]^2\psi^2(n)+[\epsilon(L,\alpha)]^2\right)\\&\quad\cdot\left(\sum_{j=1}^{J}\sigma_j^2\right).\end{aligned}\tag{33}$$

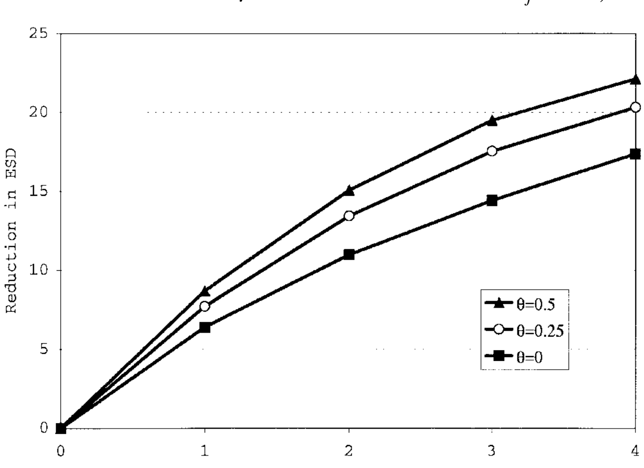
As in the case of correlation across items, we conclude with two figures. Figure 7 displays the dependency of ESD on $\eta$, the degree of uncertainty surrounding the means of single period demands, for three different correlation values $\theta(\theta=0,0.25,$ and $0.5)$. Observe that ESD increases

**Figure 7.** ESD, as a function of the degree of uncertainty $\eta$, for various values of $\theta$. ($J=2$ items, $n=3$, $L=l=2$ and $\sigma_j=10$.).




with $\eta$ and with $\theta$. However, the increase in ESD resulting from increased intertemporal correlation ($\theta$) diminishes as the degree of uncertainty about the mean demands ($\eta$) increases. Indeed, with perfectly known mean demands, ESD increases by no less than 36% when consecutive demands have a 0.5 correlation as compared with when they are independent! This exemplifies the severity of the intertemporal independence assumption made in most existing inventory models. Figure 8 exhibits the reduction of ESD as a function of $L$, the point of differentiation, again given in three graphs, for the same three values of $\theta$. Note that the benefits of postponement increase with $\theta$.

**Figure 8.** Reduction in ESD as a function of the point of differentiation $L$. ($J=2$ identical items, $n=3$, $\eta=1$, $\tau=L+l=4$ and $\sigma_j=10$.).




### 7.3. A Demand Model With Known Coefficients of Variation

In this subsection, we outline a demand model in which the coefficients of variation of the demand variables $\{d_{jn}\}$ are assumed to be known as opposed to their standard deviations. Assume, therefore, that for all $n=1,2,\ldots$, the demand variable $d_{jn}$ has a Gamma$(k_j,\lambda_j)$ distribution, where $\lambda_j$ is *unknown* with a prior distribution which is Gamma$(\kappa_j,\nu_j)$. (Recall that a Gamma$(k,\alpha)$ distribution has pdf $f(x)=e^{-\alpha x}\alpha^k x^{k-1}/\Gamma(k)$, $x\geqslant 0$, with mean $k/\alpha$, variance $k/\alpha^2$, and coefficient of variation $1/\sqrt{k}$, independent of $\alpha$.) As in the basic additive model (1), we assume that the conditional distributions $(d_{jn}|\lambda_j)$ are independent across items and time. Extensions to allow for correlation across items or time are similar to those discussed in §§7.1 and 7.2.

Note first that the posterior distribution of $(\lambda_j|d_{j1},\ldots,d_{jn})$ is again Gamma-distributed. More specifically, it is Gamma$(nk_j+k_j,\nu_j+\sum_{i=1}^n d_{ji})$-distributed. Unfortunately, this does not imply that the conditional distribution of future demands is of a simple form. However, it is possible to obtain closed-form expressions for the conditional mean and standard deviations of future lead time demands:

$$\mathsf{E}\left[\sum_{i=n+l+1}^{n+t+l+1} d_{ji}|d_{j1},\ldots,d_{jn}\right]=\psi_j(n,l+1)\left(\nu_j+\sum_{i=1}^{n}d_{ji}\right)$$

$$\text{Std}\left[\sum_{i=n+t+1}^{n+t+l+1} d_{ji}|d_{j1},\ldots,d_{jn}\right]=\theta_j(n,l+1)\left(\nu_j+\sum_{i=1}^{n}d_{ji}\right), \quad (34)$$

where $\psi_j(n,l+1)=(l+1)k_j/(nk_j+\kappa_j-1)$, and

$$\theta_j(n,l+1)=\frac{\sqrt{(l+1)k_j+\frac{(l+1)^2k_j^2}{nk_j+\kappa_j-1}}}{\sqrt{(nk_j+\kappa_j-1)(nk_j+\kappa_j-2)}}$$

are known coefficients that are independent of the demand observations. (The above expressions are easily verified by first conditioning on $\lambda_j$ and integrating out over the above conditional distribution of $(\lambda_j|d_{j1},\ldots,d_{jn})$.) To obtain a closed-form expression for the allocation problem $(P_n)$, we replace the conditional future lead time demand distributions by normal distributions with mean and standard deviations given by Equation (34). We first consider the following important special case.

**All $k_j=k$ and $\kappa_j=\kappa$, $j=1,\ldots,J$.** This condition is equivalent to assuming that the coefficients of variation of the individual distributions $(d_{jn}|\lambda_j)$ as well as those of the prior distributions of $\lambda_j$ are identical across all items, an important special case that applies to many product families; see, e.g., the discussion in Federgruen and Zipkin (1984c). Note that under this condition, $\psi_j(n,l+1)=\psi(n,l+1)$ and $\theta_j(n,l+1)=\theta(n,l+1)$ for all $j=1,\ldots,J$. We thus obtain a closed-form lower bound approximation for the cost value of the allocation problem

$(P_n)$, which is analogous to Equation (7):

$$\begin{aligned}\widetilde{R}_n\big(X_n+y_{n-L}|I_{n-1}\big)&\\ =H\bigg(&X_n+y_{n-L}|h_{n+l};\,p_{n+l};\,\psi(n-1,l+1)\\ &\cdot\bigg[\sum_j \nu_j+\sum_{i=1}^{n-1}D_i\bigg];\,\theta(n-1,l+1)\bigg[\sum_j \nu_j+\sum_{i=1}^{n-1}D_i\bigg]\bigg).\end{aligned}$$

Once again, $\widetilde{R}_n(\cdot|I_{n-1})$ depends on the history of observed demand observations via the single sufficient statistic $\overline{D}_{n-1}$ only. The relaxation technique of §4 once again results in a lower bound approximate dynamic program with the same two state variables as in the basic model, i.e., $X_n^S$ and $\overline{D}_{n-1}$, the aggregate system-wide inventory position and the average observed aggregate single-period demand. The one-step expected cost function in this dynamic program is again

$$\begin{aligned}&R_n(X_n^S+y_n|I_{n-1})\\ &\quad=\mathsf{E}\{\widetilde{R}_{n+L}(X_{n+L}+y_n|\overline{D}_{n-1})|X_n^S+y_n,\,I_{n-1}\},\end{aligned}$$

which is again convex in $(X_n^S+y_n)$. On the other hand, $R_n(X_n^S+y_n|I_{n-1})$ cannot be obtained in closed form and in particular cannot be expressed in terms of the $H$-function, even if the conditional distributions of future lead time demands are approximated by normal distributions. This implies, in particular, that the one-step expected cost function can no longer be interpreted as that of a single item with a specific "effective" lead time demand distribution. (On the other hand, an accurate closed-form approximation for $R_n(X_n^S+y_n|I_{n-1})$ can be obtained; we omit the details.)

**The general case of nonidentical $k_j$- and $\kappa_j$-parameters.** We now turn to the general case in which the coefficients of variation of the conditional distributions $(d_{jn}|\lambda_j)$, and those of $\lambda_j$ themselves are *item-dependent*. For this general case, we introduce a restriction to the prior distributions of the $\lambda_j$-parameters. Similar to the restriction $\sigma_{0j}=\eta\sigma_j$, $j=1,\ldots,J$, made in the basic model, let, for some constant $\eta>0$,

$$\kappa_j=1+\eta k_j,\qquad j=1,\ldots,J.$$

Under this restriction, we have

$$\begin{aligned}\psi_j(n,l+1)&=\frac{(l+1)k_j}{(n+\eta)k_j}=\frac{l+1}{n+\eta}\\ &\equiv\psi(n,l+1),\quad j=1,\ldots,J,\end{aligned}$$

while the coefficients $\theta_j(n,l+1)$ are very closely approximated (from below) by the coefficients

$$\tilde{\theta}_j(n,l+1)=\frac{\sqrt{(l+1)+\frac{(l+1)^2}{n+\eta}}}{n+\eta}\cdot\frac{1}{\sqrt{k_j}}.$$

We now have

$$\begin{aligned}\widetilde{R}_n\big(X_n+y_{n-L}|I_{n-1}\big)&\\ =H\bigg(&X_n+y_{n-L}|h_{n+l};\,p_{n+l};\,\psi(n-1,l+1)\\ &\cdot\bigg[\sum_j \nu_j+\sum_{i=1}^{n-1}D_i\bigg];\,\frac{\sqrt{(l+1)+\frac{(l+1)^2}{n-1+\eta}}}{n-1+\eta}\\ &\cdot\bigg[\sum_j\frac{\nu_j}{\sqrt{k_j}}+\sum_{i=1}^{n-1}\Delta_i\bigg]\bigg),\end{aligned}$$

where $\Delta_i\equiv\sum_{j=1}^J d_{ji}/\sqrt{k_j}$. Thus, $\widetilde{R}_n(\cdot|I_{n-1})$ now depends on the history of demand observations via *two* sufficient statistics, $\overline{D}_{n-1}$ and $\overline{\Delta}_{n-1}\equiv\frac{1}{n-1}\sum_{i=1}^{n-1}\Delta_i$. As a consequence, the approximate dynamic program, similar to that in §7.2, now has three state variables: $X_n^S$, $\overline{D}_{n-1}$, and $\overline{\Delta}_{n-1}$.

## APPENDIX

In this appendix, we derive a number of properties of the $H$ functions, which we use throughout the paper. We also provide the proof of Lemma 1. To shorten the notation, we write without possible confusion $H(y|\mu,\sigma)$ for $H(y|h;p;\mu;\sigma)$.

Theorem A (Properties of the Function H). *The function $H$ satisfies the following properties*:

(a) $H(y-\alpha|\mu,\sigma)=H(y|\mu+\alpha,\sigma)$.

*Consider the function $H(\cdot|\Lambda,\sigma)$ where $\sigma$ is a constant but $\Lambda$ is a normally distributed random variable, with $\Lambda\sim N(\tilde{\mu},\tilde{\sigma})$. Then,*

(b) $\mathsf{E}=[H(y|\Lambda,\sigma)]=H(y|\tilde{\mu},\sqrt{\sigma^2+\tilde{\sigma}^2})$

(c) *For any given constants $\alpha$ and $\beta$*:

$$E[H(y|\alpha+\beta\Lambda,\sigma)=H\big(y|\alpha+\beta\tilde{\mu},\sqrt{\sigma^2+(\beta\tilde{\sigma})^2}\big).$$

(d) *$H(y|\mu,\sigma)$ is strictly convex in $y$, and its minimum is obtained at the value $y^*=\mu+\sigma\Phi^{-1}(\frac{p}{p+h})$.*

(e) $H(y|\mu,\sigma)=O(|y|)$: *i.e., there exist constants $\kappa_1,\kappa_2\geqslant 0$ such that $H(y|\mu,\sigma)\leqslant\kappa_1+\kappa_2\cdot|y|$.*

Proof. (a) This part is immediate because $H(y|\mu,\sigma)$ can be written as a function of $y-\mu$.

(b) Let $\epsilon$ denote a random variable, independent of $\Lambda$, which is normally distributed with mean 0 and standard deviation $\sigma$. $\mathsf{E}_\Lambda H(y|h,p,\Lambda,\sigma)=\mathsf{E}_\Lambda\mathsf{E}_\epsilon\{h[y-\Lambda-\epsilon]^++p[\Lambda+\epsilon-y]^+|\Lambda\}=\mathsf{E}_{\Lambda,\epsilon}\{h[y-(\Lambda+\epsilon)]^++p[(\Lambda+\epsilon)-y]^+\}=H(y|h,p,\tilde{\mu},\sqrt{\tilde{\sigma}^2+\sigma^2})$, because $\Lambda+\epsilon\sim N(\tilde{\mu},\sqrt{\tilde{\sigma}^2+\sigma^2})$.

(c) Let $\Lambda'\equiv\alpha+\beta\Lambda\sim N(\alpha+\beta\tilde{\mu},\beta\tilde{\sigma})$ and apply part (b).

(d) First observe that

$$\frac{\partial}{\partial y}H(y|\mu,\sigma)=-p+(p+h)\Phi\Big(\frac{y-\mu}{\sigma}\Big),$$

$$\text{and}\quad\frac{\partial^2}{\partial y^2}H(y|\mu,\sigma)=\frac{(p+h)}{\sigma}\phi\Big(\frac{y-\mu}{\sigma}\Big)>0.$$

Strict convexity follows from the positivity of the second derivative, and therefore the (unique and global) minimum is obtained at the point $y^*$ in which the first derivative is zero (i.e., $y^*=\mu+\sigma\Phi^{-1}(\frac{p}{p+h})$).

(e) Let $u \sim N(\mu, \sigma)$. $H(y|\mu, \sigma) = \mathsf{E}\{h[y-u]^+ + p[u-y]^+\} \leqslant (h+p)\mathsf{E}\{|y|+|u|\} = O(|y|)$. □

PROOF OF LEMMA 1. By Equation (11) and the definition of $\overline{D}_{n+L-1}$, we have

$$
\begin{aligned}
&R_n(X_n^S + y_n | I_{n-1})\\
&\quad = \mathsf{E}\{\widehat{R}_{n+L}(X_{n+L} + y_n | \overline{D}_{n+L-1}) | X_n^S + y_n, I_{n-1}\}\\
&\quad = \mathsf{E}\Bigg\{\widehat{R}_{n+L}\Bigg(X_n^S + y_n - \sum_{i=n}^{n+L-1} D_i\\
&\qquad\qquad \cdot \Bigg| \frac{\sum_{i=n}^{n+L-1} D_i + (n-1)\overline{D}_{n-1}}{n+L-1}\Bigg)\Bigg| I_{n-1}\Bigg\}\\
&\quad = \mathsf{E}\Bigg\{H\Bigg(X_n^S + y_n - \sum_{i=n}^{n+L-1} D_i | h_{n+L+l}; p_{n+L+l};\\
&\qquad (l+1)\Bigg[\frac{\eta^2(\sum_{i=n}^{n+L-1} D_i + (n-1)\overline{D}_{n-1}) + \sum_j \mu_{0j}}{(n+L-1)\eta^2+1}\Bigg];\\
&\qquad \sqrt{\frac{(l+1)^2\eta^2}{(n+L-1)\eta^2+1} + (l+1)\cdot\left(\sum_{j=1}^{J}\sigma_j\right)}\Bigg)\Bigg|\overline{D}_{n-1}\Bigg\},
\end{aligned}
$$

where the last equality follows from Equation (10) and the fact that $(\sum_{i=n}^{n+L-1} D_i)$, as well as $\overline{D}_{n-1}$, both depend on $I_{n-1}$ only via $\overline{D}_{n-1}$; see Theorem 1, part (c), applied to the sequence $\{D_n\}_{n=1}^{\infty}$, which as demonstrated above satisfies all the required properties similar to the sequences $\{d_{jn}\}_{n=1}^{\infty} (j = 1, \ldots, J)$. In particular,

$$
\sum_{i=n}^{n+L-1} D_i | \overline{D}_{n-1} \sim N\left(L\frac{(n-1)\overline{D}_{n-1}\eta^2 + \sum_j \mu_{0j}}{(n-1)\eta^2+1},\ \sqrt{\frac{L^2\eta^2}{(n-1)\eta^2+1} + L\cdot\sqrt{\sum_j \sigma_j^2}}\right). \quad (35)
$$

Invoking Theorem A, part (a), we obtain:

$$
\begin{aligned}
&R_n(X_n^S + y_n | I_{n-1})\\
&\quad = \mathsf{E}\Bigg\{H\Bigg(X_n^S + y_n | h_{n+L+l}; p_{n+L+l};\\
&\qquad \left[1 + \frac{(l+1)\eta^2}{(n+L-1)\eta^2+1}\right]\\
&\qquad \cdot \sum_{i=n}^{n+L-1} D_i + (l+1)\frac{\eta^2(n-1)\overline{D}_{n-1} + \sum_j \mu_{0j}}{(n+L-1)\eta^2+1};\\
&\qquad \sqrt{\frac{(l+1)^2\eta^2}{(n+L-1)\eta^2+1} + (l+1)\cdot\left(\sum_{j=1}^{J}\sigma_j\right)}\Bigg)\Bigg|\overline{D}_{n-1}\Bigg\}.
\end{aligned}
$$

We now invoke Theorem A, part (c), with $\Lambda \sim (\sum_{i=n}^{n+L-1} D_i | \overline{D}_{n-1})$:

$$
\alpha = (l+1)\frac{\eta^2(n-1)\overline{D}_{n-1} + \sum_j \mu_{0j}}{(n+L-1)\eta^2+1};
$$

$$
\text{and } \beta = \left[1 + \frac{(l+1)\eta^2}{(n+L-1)\eta^2+1}\right],
$$

using Equation (35). The lemma now follows by simple algebra. □

## REFERENCES


Alderson, W. 1950. Marketing efficiency and the principle of postponement. *Cost and Profit Outlook* **3** (Sep).

Aviv, Y., A. Federgruen. 2001. Capacitated multi-item inventory systems with random and seasonally fluctuating demands: implications for postponement strategies. *Management Sci.* **47** 512–531.

Azoury, K. S. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Sci.* **31** 1150–1160.

——, B. L. Miller. 1984. A comparison of the optimal ordering levels of Bayesian and non-Bayesian inventory models. *Management Sci.* **30** 993–1003.

Bertsekas, D. P., S. E. Shreve. 1978. *Stochastic Optimal Control: The Discrete Time Case.* Prentice Hall, Englewood Cliffs, NJ.

DeGroot, M. H. 1970. *Optimal Statistical Decisions.* McGraw Hill, Inc., New York.

Denardo, E. V. 1982. *Dynamic Programming: Models and Applications.* Prentice-Hall, Englewood Cliffs, NJ.

Dvoretzky, A., J. Kiefer, J. Wolfowitz. 1952. The inventory problem: II. Case of unknown distributions of demand. *Econometrica* **19**.

Eppen, G., L. Schrage. 1981. Centralized ordering policies in a multiwarehouse system with leadtimes and random demands. *Multi-Level Production/Inventory Control Systems: Theory and Practice.* L. Schwarz (ed.), North-Holland, Amsterdam. 51–69.

Erhardt, R. 1979. The power approximation for computing $(s, S)$ inventory policies. *Management Sci.* **25** 777–786.

Erkip, N., W. H. Hausman, S. Nahmias. 1990. Optimal centralized ordering policies in multi-echelon inventory systems with correlated demands. *Management Sci.* **36** 381–392.

Federgruen, A. 1993. Recent advances in production and distribution management. *Perspectives in Operations Management.* R. Sarin (ed.), Kluwer Academic Publishers, Norwell, MA.

——, P. Zipkin. 1984a. Approximations of dynamic, multilocation production and inventory problems. *Management Sci.* **30** 69–84.

——, ——. 1984b. Allocation policies and cost approximations for multilocation inventory systems. *Naval Res. Logist. Quart.* **31** 97–129.

——, ——. 1984c. Computational issues in an infinite horizon multi-echelon inventory model. *Oper. Res.* **32** 818–832.

Fisher, M. L., A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* **44** 87–99.

——, W. R. Obermeyer, J. Hammond, A. Raman. 1994a. Accurate response: the key to profiting from quick response. *Bobbin Mag.* February 48–62.

——, J. Hammond, W. R. Obermeyer, A. Raman. 1994b. Making supply meet demand in an uncertain world. *Harvard Bus. Rev.* May–June, 83–93.

Fox, B. 1966. Discrete optimization via marginal analysis. *Management Sci.* **4** 36–153.

Gallego, G., I. Moon. 1993. The distribution free newsboy problem: review and extensions. *J. of Oper. Res. Soc.* **44** 825–834.

Garg, A., C. S. Tang. 1997. On postponement strategies for product families with multiple points of differentiation. *IIE Trans.* **29** 641–650.

Gross, O. 1956. A class of discrete-type minimization problems. Working Paper RM-1644-PR, The RAND Corporation, Santa Monica, CA.

Güllü, R. 1997. A two-echelon allocation model and the value of information under correlated forecasts and demands. *Eur. J. Oper. Res.* **99** 386–400.

Hamilton, J. D. 1994. *Time Series Analysis.* Princeton University Press, Princeton, NJ.

Heath, C. D., P. L. Jackson. 1994. Modeling the evolution of demand forecasts with applications to safety stock analysis in production/distribution systems. *IIE Trans.* **26** 17–30.

Iglehart, D. 1964. The dynamic inventory model with unknown demand distribution. *Management Sci.* **10** 429–440.

Iyer, A. V., Bergen M. E. 1997. Quick response in manufacturer-retailer channels. *Management Sci.* **43** 559–570.

Jackson, P. L. 1988. Stock allocation in a two echelon distribution system or "what to do until your ship comes in" *Management Sci.* **34** 880–895.

Johnson, G., H. Thompson. 1975. Optimality of myopic inventory policies for certain dependent demand processes. *Management Sci.* **21** 1303–1307.

Jönsson, H., E. A. Silver. 1987a. Analysis of a two-echelon inventory control system with complete redistribution. *Management Sci.* **33** 215–227.

——, ——. 1987b. Modification of demand distributions on the basis of aggregate information. *IIE Trans.* **19** 379–384.

Karlin, S. 1960. Dynamic inventory policy with varying stochastic demands. *Management Sci.* **6** 231–258.

Lee. H. 1993. Design for supply chain management: concepts and examples. *Perspectives in Operations Management.* R. Sarin (ed.), Kluwer Academic Publishers, Norwell, MA.

——. 1996. Effective management of inventory and service through product and process redesign. *Oper. Res.* **44** 151–159.

——, C. Billington. 1994. Designing products and processes for postponement. *Management of Design: Engineering and Management Perspectives.* S. Dasu, C. Eastman (eds.), Kluwer Academic Publishers, Norwell, 105–122.

——, ——, B. Carter. 1993. Hewlett-Packard gains control of inventory and service through design for localization. *Interfaces* August, 1–11.

——, C. S. Tang. 1997. Modeling the costs and benefits of delayed product differentiation. *Management Sci.* **43** 40–53.

Lovejoy, W. S. 1990. Myopic policies for some inventory models with uncertain demand. *Management Sci.* **36** 724–738.

Miller, B. 1986. Scarf's state reduction method, flexibility, and a dependent demand inventory model. *Oper. Res.* **34** 83–90.

Mrena, C. 1997. Supply chain strategies at Sun Microsystems. Presentation at the *Supply Chain Management Conference*, Santa Clara University, CA, May 15.

*New York Times.* 1997. Compaq Plans A Big Change in its PC Sales. L. Zuckerman, May 5.

Scarf, H. 1958. A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production.* Chapter 2. Stanford University Press, Stanford, CA.

——. 1959. Bayes solutions of the statistical inventory problem. *Ann. Math. Statist.* **30** 490–508.

——. 1960. Some remarks on Bayes solutions to the inventory problem. *Naval Logist. Res. Quart.* **7** 591–596.

Schwarz, L. B. 1989. Model for assessing the value of warehouse risk-pooling: risk pooling over outside-supplier leadtimes. *Management Sci.* **35** 828–842.

Sethi, S. P., F. Cheng. 1997. Optimality of $(s, S)$ policies in inventory models with Markovian demand. *Oper. Res.* **45** 931–939.

Signorelli, S., J. L. Heskett. 1989. Benetton (A). *Harvard Bus. School.*

Sobel, M. 1988. Dynamic Affine Logistics Models. Report. SUNY at Stony Brook, NY.

Song, J. S., P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Oper. Res.* **41** 351–370.

Stalk, G., Jr., P. Evans, L. E. Shulman. 1992. Competing on capabilities: the new rules of corporate strategy. *Harvard Bus. Rev.* March.

Tong, Y. L. 1990. *The Multivariate Normal Distribution.* Springer-Verlag, New York.

Topkis, D. M. 1978. Minimizing a submodular function on a lattice. *Oper. Res.* **26**(2) 305–321.

Veinott, A., Jr. 1965. Optimal policy for a multi-product, dynamic non-stationary inventory problem. *Management Sci.* **12** 206–222.

*Wall Street Journal.* 1996. GM expands plan to speed cars to buyers. G. Stern and R. Blumenstein, October 21.

Zipkin, P. 1980. Simple ranking methods for allocation of one resource. *Management Sci.* **26** 34–43.

——. 1982. Exact and approximate cost functions for product aggregates. *Management Sci.* **28** 1002–1012.