A DYNAMIC MODEL FOR REQUIREMENTS PLANNING WITH APPLICATION TO SUPPLY CHAIN OPTIMIZATION

STEPHEN C. GRAVES

Massachusetts Institute of Technology, Cambridge, Massachusetts

DAVID B. KLETTER

Booz, Allen & Hamilton Inc., New York, New York

WILLIAM B. HETZEL

Merck & Co., Inc., Whitehouse Station, New Jersey

(Received March 1994; revisions received July 1995, January 1996; accepted June 1996.)

This paper develops a new model for studying requirements planning in multistage production-inventory systems. We first characterize how most industrial planning systems work, and we then develop a mathematical model to capture some of the key dynamics in the planning process. Our approach is to use a model for a single production stage as a building block for modeling a network of stages. We show how to analyze the single-stage model to determine the production smoothness and stability for a production stage and the inventory requirements. We also show how to optimize the tradeoff between production capacity and inventory for a single stage. We then can model the multistage supply chain using the single stage as a building block. We illustrate the multistage model with an industrial application, and we conclude with some thoughts on a research agenda.

Most discrete parts manufacturing firms plan their production with MRP (materials requirements planning) systems, or at least, with logic based on the underlying assumptions of MRP. A typical planning system starts with a multiperiod forecast of demand for each finished good or end item. The planning system then develops a production plan (or master schedule) for each end item to meet the demand forecast. These production plans for the end items, after offsetting for lead times, then act as the requirement forecasts for the components needed to produce the end items. The requirements forecast for each component gets translated into production plans for the component, similar to how the production plan for the end items was created. The planning system continues in this way, developing requirement forecasts and production plans for each level of the bill of materials.

Implicit in this planning process are assumptions about the production and demand process. The production plan is developed assuming that the forecast is accurate and will not change. Within the production process, requirements are generated assuming that there are deterministic production lead times and deterministic yields. Needless to say, these assumptions of a benevolent world do not match reality. Inevitably, the forecast changes, and uncertainties in the production process arise that result in deviations from the plan. To respond to these changes, most planning systems will completely revise their plan after some time period, say a week or a month. Again, the planning process starts with the (new) forecast and repeats the steps necessary to regenerate a plan for each level in the products' bills of materials.

The intent of this paper is to present a model that captures the basic flavor of this planning process, and does so in such a way that it can be used to look at various tradeoffs within the production and planning systems. In particular, we model the forecasts for the planning system as a stochastic process. In this way, we try to represent a dynamic input to the planning system, namely, how forecasts change and evolve over time. The forecast process is a key input for the model. Another key is how the forecasts get converted into production plans or master schedules. We model this process as a linear system, with which we can represent the logic for MRP systems and from which we get significant analytical tractability. Finally, the model is structured so that it can describe multistage production-inventory systems.

We are not aware of very much work that is directly related to the dynamic modeling of requirements planning. Baker (1993) provides a nice survey and critique of the literature relevant to the general topic of requirements planning. However, most of the work deals with specific issues like lot sizing or determination of buffer levels. Karmarkar (1993) discusses tactical issues of lot sizing, order release, and lead times in the context of dynamic planning systems. But neither of these papers reports on work that attempts to model a dynamic forecast process. One exception is Graves et al. (1986), in which we modeled a two-stage production-inventory system with a dynamic forecast

Subject classifications: Inventory/production: multi-item, multi-stage supply chain with uncertain demand and dynamic forecast revisions; application to film production; dynamic requirements planning and supply chain optimization.
Area of review: Manufacturing Operations.

0030-364X/98/4603-S035 $05.00

process. In contrast with the present paper, Graves et al. (1986) focused on issues of how to disaggregate an aggregate plan in the two-stage context. Although this paper does not consider the disaggregation issue, it does provide a more powerful model that is applicable to general multistage systems.

Another exception is Heath and Jackson (1994), who considered the same dynamic forecast process as this paper as part of a simulation model that was used to analyze safety stock levels in a multiproduct production/distribution system.

The model for converting the forecast into a production plan is related to earlier work by the first coauthor, in that it uses linear systems for a production-inventory context. (See Graves 1986, 1988a, 1988b, 1988c, and Fine and Graves 1989.)

Lastly, we note Lee and Billington (1993), who develop a model for supply chain optimization and describe its application at Hewlett Packard. Our work complements their work but differs as we try to model the process of requirements planning.

In the next section we develop the model for a single production-inventory stage. As part of the development, we present our model for the forecast process, and we develop the analyses to generate three performance measures for the stage: production smoothness, production stability, and inventory requirements. In the second section we examine an optimization for the tradeoff between production capacity and inventory for a single stage. Although the development is somewhat involved, the final results are surprisingly simple and, we believe, of interest. This section can be omitted by the reader without loss of continuity. In the third section, we show how the model for the single stage can serve as a building block in modeling a general acyclic network of multiple stages. We report on an application of the model to a supply-chain study in the fourth section. The application demonstrates the value of a system-wide perspective for optimizing the supply chain. In the final section we briefly summarize the paper, and then lay out a research agenda for further work.

1. SINGLE-STAGE MODEL

In this section we present the model for a single production stage that produces one (aggregate) product and serves demand from a finished good inventory. The single-stage model serves as a building block for creating models of multistage, multiitem systems. We first describe the forecast process and state our assumptions about how the forecast evolves over time. We then give a model for determining the schedule for production outputs from the production stage, and we show how to manipulate this model to obtain three measures of interest: (1) the production variance as a measure of production smoothing, (2) the inventory variance as a measure of safety stock, and (3) the stability of the production schedule as a measure for the forecast process passed on to any upstream stages.

1.1. Forecast Process

We assume that there is a forecast horizon H such that in each time period t we have forecasts for the requirements for the next H periods. Let $f_t(t + i)$ be the forecast made at time t for the requirements in period $t + i$, $i = 1, 2, \ldots H$. We denote the demand observed in period t by $f_t(t)$, the forecast made in period t for requirements in period t. Beyond the forecast horizon, there is no specific information about requirements. In effect, for $i > H$ we assume that $f_t(t + i) = \mu$, where μ equals the long-run average demand rate.

We propose a stochastic model of this forecast process and show that the forecasts are unbiased, the forecasts improve as they are revised, and the forecast error over the forecast horizon matches the inherent variability in the demand process.

We assume that, each period, we generate a new set of forecasts $f_t(t + i)$ that incorporates new information about future demand. We define the updates of the forecasts from period to period by the *forecast revision*, $\Delta f_t(t + i)$:

$$\Delta f_t(t + i) = f_t(t + i) - f_{t-1}(t + i) \quad \text{for } i = 0, 1, \ldots H, \tag{1}$$

where $f_{t-1}(t + H) = \mu$ by assumption.

Let $\underline{\Delta f_t}$ be the vector for the revisions to the forecast process, where $\Delta f_t(t + i)$ is the $i + 1$st element, $i = 0, 1, 2, \ldots H$. We assume that $\underline{\Delta f_t}$ is an i.i.d. random vector with $E[\underline{\Delta f_t}] = 0$ and $\mathrm{Var}[\underline{\Delta f_t}] = \Sigma$, the covariance matrix. Thus, for a fixed index i, $\Delta f_t(t + i)$ is an i.i.d. random variable over time t with zero mean, and the forecast process is a martingale. We note that if we can observe the forecast process, then we can assess whether or not the forecasts are unbiased (i.e., $E[\underline{\Delta f_t}] = 0$) with independent revisions and we can estimate the covariance matrix Σ.

This model of the forecast process is the same as that of Graves et al. (1986) and Heath and Jackson (1994). We have validated this model as part of field studies at AT&T and at Kodak. And this forecast model is descriptive of the forecast process at nearly all of the discrete-part manufacturing contexts we have encountered.

The *i-period forecast error* is the difference between the actual demand in period t and the forecast of this demand made i periods earlier:

$$f_t(t) - f_{t-i}(t) = \Delta f_t(t) + \Delta f_{t-1}(t) + \cdots + \Delta f_{t-i+1}(t).$$

We can now demonstrate the following properties for this model of the forecast process:

1. The i-period forecast, $f_{t-i}(t)$, is an unbiased estimate of demand in period t.
2. The variance of the i-period forecast error is no greater than the variance of the $(i + 1)$-period forecast error, for $i = 1, 2, \ldots H$.
3. The trace of the covariance matrix Σ equals the variance of the demand process.

We see that the first property must be true by observing that the expectation of the i-period forecast error is zero, since $E[\Delta f_{t-s}(t)] = 0$, for $s = 0, \ldots i - 1$.

We now prove the second property. Since $\Delta f_{t-s}(t)$ for $s = 0, 1, \ldots i - 1$ are independent random variables, the variance of the i-period forecast error is given by:

$$\begin{aligned}\operatorname{Var}[f_t(t) - f_{t-i}(t)] &= \operatorname{Var}(\Delta f_t(t)) + \operatorname{Var}(\Delta f_{t-1}(t)) \\ &\quad + \cdots + \operatorname{Var}(\Delta f_{t-i+1}(t)) \\ &= \sigma_0^2 + \sigma_1^2 + \cdots + \sigma_i^2,\end{aligned}$$

where $\sigma_j^2 = \operatorname{Var}(\Delta f_{t-j}(t))$ is the $j + 1$st element on the diagonal of the covariance matrix Σ, for $j = 0, 1, \ldots H$. Thus, since $\sigma_j^2 \geqslant 0$ for $j = 0, 1, \ldots H$, each forecast revision improves the forecast, in that it reduces the variance of the forecast error.

For the third property, we observe from the above expression that the variance of the $(H + 1)$-period forecast error equals $\sigma_0^2 + \sigma_1^2 + \cdots + \sigma_H^2$, i.e., the trace of Σ. Since by assumption $f_{t-H-1}(t) = \mu$, we have,

$$\operatorname{Var}[f_t(t) - f_{t-H-1}(t)] = \operatorname{Var}[f_t(t)],$$

which proves the third property.

Since the demand variance is an exogenous parameter, this imposes a constraint on the forecast process: namely, the variance of the forecast error over the forecast horizon must equal the demand variance.

1.2. Schedule for Production Outputs

Given the forecast vector for period t, we need to convert it into a schedule or plan for production. This is often termed the master schedule. We focus on production outputs from the production stage. Later we will discuss how to translate a plan for production outputs into production starts. Production starts will be of interest, since they serve as the requirements forecast for the next upstream production stage.

Let $F_t(t + i)$ equal the *planned production outputs* for period $t + i$ as of period t, where $F_t(t)$ is the actual production completed in period t. We assume that the production plan extends out only for the next H periods, and that beyond this horizon the plan is just to produce the average demand, that is, $F_t(t + i) = \mu$ for $i > H$.

Each period, after we obtain the new forecast, we update or revise the plan for production outputs. We define $\Delta F_t(t + i)$ as the *plan revision*:

$$\Delta F_t(t + i) = F_t(t + i) - F_{t-1}(t + i).$$

From this definition and the fact that $F_t(t + i) = \mu$ for $i > H$, we see that:

$$F_t(t + i) = \mu + \Delta F_{t+i-H}(t + i) + \cdots + \Delta F_t(t + i) \qquad \text{for } i = 0, 1, \ldots H. \quad (2)$$

Thus to model the production plan, we need to model the plan revision $\Delta F_t(t + i)$. To do this, we first define the inventory process. For I_t being the inventory at time t, the inventory balance equation is:

$$I_t = I_{t-1} + F_t(t) - f_t(t). \quad (3)$$

The *planned inventory* at time $t + i$ is the expected level of inventory in a future period given the current forecast and the current production plan as of time t:

$$\begin{aligned}I_t(t + i) = I_t &+ F_t(t + 1) + \cdots + F_t(t + i) \\ &- f_t(t + 1) - \cdots - f_t(t + i). \quad (4)\end{aligned}$$

We assume that for each time t, we set the production plan $F_t(t + i)$, $i = 0, 1, \ldots H$, so that the planned inventory at the end of the planning horizon, $I_t(t + H)$, is a given constant. That is, we will set the production plan and maintain it from period to period so that the end-of-horizon inventory neither grows nor decreases, but remains constant. We term the level to which the inventory is targeted as the *safety stock*. In a later section we will discuss how to set this level. For now, all we need to know is that this level remains constant.

From (3) and (4), we obtain by equating $I_{t-1}(t - 1 + H)$ and $I_t(t + H)$ that:

$$\begin{aligned}&\Delta F_t(t) + \Delta F_t(t + 1) + \cdots + \Delta F_t(t + H) \\ &\quad = \Delta f_t(t) + \Delta f_t(t + 1) + \cdots + \Delta f_t(t + H). \quad (5)\end{aligned}$$

That is, to assure that the end-of-horizon inventory remains constant, we require that the cumulative revision to the production plan should equal the cumulative forecast revision in each period.

Each period we revise the production schedule to ensure (5). To do this, we model the schedule update as a linear system:

$$\Delta F_t(t + i) = \sum_{j=0}^{H} w_{ij} \Delta f_t(t + j) \quad \text{for } i = 0, 1, \ldots H, \quad (6)$$

where w_{ij} denotes how the forecast revision affects the schedule. In particular, w_{ij} is the proportion of the forecast revision for period $t + j$ that is added to the schedule of production outputs for period $t + i$.

We expect that $0 \leqslant w_{ij} \leqslant 1$. To ensure that (5) is true, we require that for each j:

$$\sum_{i=0}^{H} w_{ij} = 1.$$

We refer to w_{ij} as a *weight* or proportion. We can interpret these weights either as decision variables in a prescriptive model or as parameters in a descriptive model. On the one hand, we can view these weights as control or smoothing parameters and use the model for prescription. To smooth production we set the weights w_{ij} for a fixed j to be as nearly constant as possible (e.g., $w_{ij} = 1/(H + 1)$ for $i = 0, 1, \ldots H$). To minimize inventory, we set the weights so that the production plan tracks the forecast as closely as possible (e.g., for fixed j, $w_{ij} = 1$ for $i = j$ and $w_{ij} = 0$ otherwise). In this way, specification of the weights permits one to balance the tradeoff between production smoothing and inventory requirements, as will be seen.

On the other hand, we can view the weights as parameters for a descriptive model of an existing planning system.

In particular, we can use (6) to model how most implementations of MRP systems translate forecast revisions into schedule revisions. For instance, in the simplest case at time t the schedule is frozen for periods $t + j$, $j = 0, 1, 2 \ldots k$ for some value of $k < H$, and is totally free to change for periods $t + j$, $j = k + 1, \ldots H$. Then, any revision to the forecast within the frozen zone results in a schedule revision for the first period beyond the frozen zone; i.e., for $0 \leqslant j \leqslant k$, $w_{ij} = 0$ for $i \neq k + 1$ and $w_{ij} = 1$ for $i = k + 1$. Any revision to the forecast beyond the frozen zone results in a one-for-one schedule revision in the same period: for $k + 1 \leqslant j \leqslant H$, $w_{ij} = 1$ for $i = j$ and $w_{ij} = 0$ otherwise. Occasionally there is an intermediate zone (between the frozen and free zones) in which changes to the schedule are permitted but are restricted in size, e.g., no more than 10% increase or decrease in the scheduled amount. The model given by (6) cannot exactly capture this policy, but it can approximate its behavior by using fractional weights.

In matrix notation, we can rewrite (6) as:

$$\underline{\Delta F}_t = \mathbf{W}\, \underline{\Delta f}_t, \tag{7}$$

where $\mathbf{W} = \{w_{ij}\}$ is an $(H + 1) \times (H + 1)$ matrix, and $\underline{\Delta F}_t$ and $\underline{\Delta f}_t$ are column vectors with elements $\Delta F_t(t + i)$ and $\Delta f_t(t + i)$, for $i = 0, 1, \ldots H$. From this, we observe that $\underline{\Delta F}_t$ is an independent random vector, has zero mean, and has a covariance matrix $\mathbf{W}\Sigma\mathbf{W}'$. (We will see later that this is an important observation for the extension to multiple stages: we will derive the forecast revision for upstream stages from $\underline{\Delta F}_t$.)

We can express the production plan in matrix notation by:

$$\underline{F}_t = \mathbf{B}\, \underline{F}_{t-1} + \mu \underline{U}_{H+1} + \underline{\Delta F}_t, \tag{8}$$

where $F_t(t + i)$ is the $i + 1$st element of the vector $\underline{F}_t$ for $i = 0, 1, \ldots H$; $\underline{U}_{H+1}$ is a unit vector with $u_i = 0$ for $i = 1, \ldots H$ and $u_{H+1} = 1$; and $\mathbf{B}$ is a matrix with elements $b_{ij} = 1$ for $j = i + 1$, and $b_{ij} = 0$ else. Premultiplying a column vector by $\mathbf{B}$ replaces the ith element in the vector with the $i + 1$st element and replaces the last element with a zero.

From (7) and (8) and repeated substitution, we obtain:

$$\begin{aligned}\underline{F}_t &= \mathbf{B}\underline{F}_{t-1} + \mu \underline{U}_{H+1} + \mathbf{W}\underline{\Delta f}_t \\ &= \mathbf{B}^{H+1}\underline{F}_{t-H-1} + \underline{\mu} + \sum_{i=0}^{H} \mathbf{B}^i\mathbf{W}\underline{\Delta f}_{t-i},\end{aligned} \tag{9}$$

where $\underline{\mu}$ is the vector with each element equal to μ, and the superscript i in $\mathbf{B}^i$ denotes the ith power of $\mathbf{B}$. We can simplify (9) by noting that premultiplying an $(H + 1) \times 1$ vector by $\mathbf{B}^{H+1}$ gives the null vector:

$$\underline{F}_t = \underline{\mu} + \sum_{i=0}^{H} \mathbf{B}^i\mathbf{W}\underline{\Delta f}_{t-i}. \tag{10}$$

1.3. Measures of Interest

There are three categories of measures for the single-stage model: the *smoothness* of the production outputs, the *safety stock* for the end-item inventory, and the *stability* of the production plan.

The *smoothness* of the production outputs is of interest because more variable (less smooth) production is expected to require more production resources or capacity. Furthermore, we can influence the smoothness of production via our inventory and control policies.

An output of the model is the variability of the inventory process, which will dictate how much *safety stock* is needed to ensure an acceptable service level. If the inventory process is more variable, more safety stock will be needed.

The *stability* of the production output plan is of interest, since the output plan determines the plan for production starts, which determines the requirements forecast for upstream stages. We will, in effect, equate the stability of the production plan to the accuracy of the forecast process for the upstream stages. More stability means a more accurate forecast process upstream. This measure is critical as we try to understand the workings of a multistage system, since the inventory requirements and the variability of the production outputs for a stage will depend heavily on the accuracy of the forecast process.

We first develop the measures for production smoothing and for the stability of the production plan. We will need a more extensive development to obtain the variability of the inventory process in order to set the safety stock.

Production Smoothing. A common measure of production smoothing is the variance of the production output, $\mathrm{Var}[F_t(t)]$. From (10) we see immediately that the random vector $\underline{F}_t$ has mean $\underline{\mu}$ and has a covariance matrix given by:

$$\mathrm{Var}\,(\underline{F}_t) = \sum_{i=0}^{H} \mathbf{B}^i\mathbf{W}\Sigma\mathbf{W}'\mathbf{B}'^{\,i}. \tag{11}$$

We can use the covariance matrix to obtain the first measure of the production smoothing, $\mathrm{Var}[F_t(t)]$. Indeed, one can show that:

$$\mathrm{Var}\,[F_t(t)] = \mathrm{tr}(\mathbf{W}\Sigma\mathbf{W}'), \tag{12}$$

where $\mathrm{tr}(\mathbf{A})$ is the trace of matrix $\mathbf{A}$.

A second measure of production smoothing is given by $F_t(t) - F_{t-1}(t - 1)$, the change in production outputs from one period to the next. In matrix notation we see from (9) that:

$$\underline{F}_t - \underline{F}_{t-1} = \mu\underline{U}_{H+1} + \mathbf{W}\underline{\Delta f}_t - [\mathbf{I} - \mathbf{B}]\underline{F}_{t-1},$$

where $\mathbf{I}$ is the identity matrix. Since $\underline{\Delta f}_t$ and $\underline{F}_{t-1}$ are independent of each other, we find that the covariance matrix for $\underline{F}_t - \underline{F}_{t-1}$ is given by:

$$\begin{aligned}\mathrm{Var}\,(\underline{F}_t - \underline{F}_{t-1}) &= \mathbf{W}\Sigma\mathbf{W}' \\ &\quad + \sum_{i=0}^{H} [\mathbf{I} - \mathbf{B}]\mathbf{B}^i\mathbf{W}\Sigma\mathbf{W}'\mathbf{B}'^{\,i}[\mathbf{I} - \mathbf{B}]' \\ &= (\mathbf{I} - \mathbf{B})\ \mathrm{Var}\,(\underline{F}_t) + \mathrm{Var}\,(\underline{F}_t)(\mathbf{I} - \mathbf{B})'.\end{aligned} \tag{13}$$

From this covariance matrix, we can determine the second measure of production smoothing, namely Var $[F_t(t) - F_{t-1}(t-1)]$.

Production Stability. For the stability of the production plan, we use $\underline{\Delta F}_t$: the random vector for the one-period revision to the production plan, which is the basis for the revision to the forecast of requirements for upstream stages. (The production starts, as described earlier, would generate the actual forecast seen by the upstream stages; but since the starts are usually just the production plan offset by the lead time, we can use the revision to the production plan for defining stability.) From (7) we obtain its expectation and covariance matrix:

$$\begin{aligned} E(\underline{\Delta F}_t) &= \underline{0} \\ \text{Var}\,(\underline{\Delta F}_t) &= \mathbf{W}\Sigma\mathbf{W}'. \end{aligned} \tag{14}$$

We propose the covariance matrix $\mathbf{W}\Sigma\mathbf{W}'$ as a measure of the stability of the production plan. A more stable production plan will have a smaller covariance matrix, and will yield more accurate forecasts for the upstream stages. When analyzing the upstream stages, the dynamics of the requirements depend upon this covariance matrix. In this sense, for the upstream stages, the covariance matrix in (14) is analogous to Σ for the downstream stage, namely it is the covariance matrix for the relevant requirements forecast process.

One measure of the size of a covariance matrix is its trace. We note that with this interpretation the $\text{tr}(\mathbf{W}\Sigma\mathbf{W}')$ signifies not only the stability of the production plan, but also the variance of the requirements forecast for the upstream stages over the planning horizon. Furthermore, we see that, according to the proposed measures (12) and (14), smoothing production is essentially equivalent to stabilizing the production plan and requirements forecast for the upstream stages.

Inventory. We focus on the end-item inventory for the single stage, namely the random variable I_t given in (3). We assume that the requirements for the single stage are to be met from the end-item inventory and that typical service expectations apply, e.g., the inventory should stock out in no more than 2% of the periods, or that the inventory should provide a 97% fill rate. We will find the expectation $E(I_t)$ and variance $\text{Var}(I_t)$, from which we can determine the safety stock required to achieve a desired service level, under suitable distributional assumptions. For instance, if the forecast errors are normally distributed, then we will see that I_t has a normal distribution. For a desired service level expressed as the stockout probability, we need to set the safety stock level so that:

$$E(I_t) > k\sigma(I_t), \tag{15}$$

where k is such to ensure the service level, and $\sigma(\,)$ denotes the standard deviation.

Recall that in (4) we defined $I_t(t+i)$ to be the planned inventory level in period $t+i$ as of time t; that is, $I_t(t+i)$ is the expected inventory in period $t+i$, where the expectation is as of period t. For notational convenience, $I_t(t)$ denotes the actual inventory in period t, i.e., $I_t(t)$ is the same as I_t. As stated in the earlier development of (5), we assume that the end-of-horizon inventory $I_t(t+H)$ is targeted to equal some constant, which we call the safety stock and denote by ss.

The inventory flow equation for the planned inventory is:

$$\begin{aligned} I_t(t+i) = I_t(t) &+ F_t(t+1) + \cdots + F_t(t+i) \\ &- f_t(t+1) - \cdots - f_t(t+i). \end{aligned} \tag{16}$$

Define $\Delta I_t(t+i) = I_t(t+i) - I_{t-1}(t+i)$. From (3) and (16), we find that

$$\begin{aligned} \Delta I_t(t+i) = \Delta F_t(t) &+ \cdots + \Delta F_t(t+i) \\ &- \Delta f_t(t) - \cdots - \Delta f_t(t+i), \end{aligned}$$

for $i = 0, 1, \ldots H-1$. By assumption, since we keep the inventory constant at ss beyond the horizon, we have:

$$\Delta I_t(t+H) = I_t(t+H) - I_{t-1}(t+H) = ss - ss = 0.$$

In matrix notation, let $\underline{I}_t$ be an $(H+1) \times 1$ column vector with $I_t(t+i)$ as its $i+1$st element. Then

$$\underline{\Delta I}_t = \mathbf{T}[\underline{\Delta F}_t - \underline{\Delta f}_t],$$

where $\mathbf{T}$ is a matrix with element $t_{ij} = 1$ for $i \geq j$ and $t_{ij} = 0$ else. We can now write the inventory random vector as

$$\underline{I}_t = \mathbf{T}[\underline{\Delta F}_t - \underline{\Delta f}_t] + \mathbf{B}\underline{I}_{t-1} + ss\underline{U}_{H+1}, \tag{17}$$

where we use the fact that $I_{t-1}(t+H) = ss$. We can simplify (17) by repeated substitution, by substitution of (7), and by noting that premultiplication of an $(H+1) \times 1$ vector by $\mathbf{B}^{H+1}$ gives the null vector:

$$\underline{I}_t = \sum_{i=0}^{H} \mathbf{B}^i\mathbf{T}[\mathbf{W} - \mathbf{I}]\underline{\Delta f}_{t-i} + \underline{ss}, \tag{18}$$

where $\underline{ss}$ denotes the column vector with each element equal to ss. From (18) we see that the random vector $\underline{I}_t$ has mean equal to $\underline{ss}$, and has a covariance matrix given by:

$$\text{Var}\,(\underline{I}_t) = \sum_{i=0}^{H} \mathbf{B}^i\mathbf{T}[\mathbf{W} - \mathbf{I}]\Sigma[\mathbf{W} - \mathbf{I}]'\mathbf{T}'\mathbf{B}'^{\,i}. \tag{19}$$

We can use (19) to find $\text{Var}[I_t(t)]$, which is necessary to determine how to set the safety stock level ss. From (19), we can show with some effort that

$$\text{Var}\,[I_t(t)] = \text{tr}(\mathbf{T}[\mathbf{W} - \mathbf{I}]\Sigma[\mathbf{W} - \mathbf{I}]'\mathbf{T}') = \sum_{k=0}^{H}\sum_{i=0}^{k}\sum_{j=0}^{k} q_{ij}, \tag{20}$$

where

$$\mathbf{Q} = \{q_{ij}\} = [\mathbf{W} - \mathbf{I}]\Sigma[\mathbf{W} - \mathbf{I}]'.$$

Now from (15) we set the safety stock by $ss = k\,\sigma[I_t(t)]$, where $\sigma[I_t(t)]$ is obtained from (20) and k is such to provide the desired service level from the inventory.

2. OPTIMAL WEIGHTS FOR SINGLE-STAGE MODEL

For a single stage it is natural to wonder how to choose the weights in (6) that determine how a forecast revision is converted into a revision of the production plan. To gain some insight into this question, we pose and solve an optimization problem for choosing the weights for the simple case of uncorrelated demand. The tradeoff between production smoothing in the stage and the end-item inventory requirements should govern the choice of weights. This tradeoff is the basis for stating the optimization problem:

Min $\sigma[F_t(t)]$,

subject to:

$\sigma[I_t(t)] \leqslant K$,

$$\sum_{i=0}^{H} w_{ij} = 1 \quad \forall j. \tag{21}$$

The optimization problem minimizes production smoothing, as given by the standard deviation of the production output, subject to a constraint on the standard deviation of the inventory and the requirement that the weights sum to one. We interpret the objective as minimizing required production capacity. We view the nominal capacity required at the stage as being the expected production requirements, plus some number of standard deviations. (See Graves 1988a for further discussion.) The constraint on the standard deviation of the inventory is effectively a constraint on the amount of safety stock required, where we assume that the safety stock is a multiple of $\sigma[I_t(t)]$. An alternative formulation would be to minimize the standard deviation of the inventory, equivalently minimize the safety stock, subject to a constraint on the standard deviation of the production output.

There are no restrictions in the optimization on the weights, other than the convexity constraint. We have not imposed any nonnegativity constraints, nor any restrictions on the weights due to a fixed production lead time. Rather, we allow the weights to be totally free. In this sense, the optimization will produce a lower bound for the case with fixed lead times.

To develop some insights on the optimal weights, we transform the original optimization problem (21) into an equivalent form by restating it in terms of the variances of the production and inventory variables:

$$\text{Min Var}\,[F_t(t)], \tag{21a}$$

subject to:

$\text{Var}\,[I_t(t)] \leqslant K^2$,

$$\sum_{i=0}^{H} w_{ij} = 1 \quad \forall j.$$

To analyze this equivalent problem, we consider the Lagrangian relaxation:

$$L(\lambda) = \text{Min Var}\,[F_t(t)] + \lambda \text{ Var}\,[I_t(t)] - \lambda K^2, \tag{21b}$$

subject to:

$$\sum_{i=0}^{H} w_{ij} = 1 \quad \forall j.$$

By solving this problem over a range of positive values for the Lagrange multiplier λ, we can find the tradeoff surface between production smoothing and inventory requirements for a single stage. We will also obtain some intuition for the form of the optimal weighting function.

In the remainder of this section we will focus on solving (21b). To solve (21a), and equivalently (21), we would need to search over λ until the solution to (21b) satisfies the relaxed constraint.

We only consider the case when the covariance matrix for the forecast revision process is diagonal. That is, the forecast revisions are uncorrelated, and $\text{Var}[\underline{\Delta f_t}] = \Sigma = \{\sigma_i^2\}$, where $\sigma_i^2 = \text{Var}[\Delta f_t(t+i)]$ is the $i+1$st element on the diagonal, $i = 0, \ldots H$.

For this case, we can simplify (12) and (20) to be:

$$\text{Var}\,[F_t(t)] = \sum_{i=0}^{H} \sum_{j=0}^{H} (w_{ij}\sigma_j)^2, \tag{12*}$$

and

$$\text{Var}\,[I_t(t)] = \sum_{i=0}^{H} \sum_{j=0}^{H} (b_{ij}\sigma_j)^2, \tag{20*}$$

where

$$\begin{aligned} b_{ij} &= w_{1j} + \cdots + w_{ij} && \text{for } i < j, \\ &= w_{1j} + \cdots + w_{ij} - 1 && \text{for } i \geqslant j. \end{aligned} \tag{22}$$

By substituting (12*) and (20*) into (21b), we observe that the minimization problem separates into $H+1$ subproblems, one for each period j:

$$L(\lambda) = \sum_{j=0}^{H} L_j(\lambda) - \lambda K^2, \tag{21c}$$

where

$$L_j(\lambda) = \text{Min} \sum_{i=0}^{H} (w_{ij}\sigma_j)^2 + \lambda \sum_{i=0}^{H} (b_{ij}\sigma_j)^2, \tag{23}$$

subject to:

$$\sum_{i=0}^{H} w_{ij} = 1.$$

We now characterize the solution to $L_j(\lambda)$ with a series of propositions.

Proposition 1. *The optimal weights in* (23) *are independent of* σ_j^2.

Proof. Each term in the objective function of $L_j(\lambda)$ in (23) is proportional to σ_j^2, which can then be factored out. □

Thus, we can determine the optimal weights in the Lagrangian (21b) without knowing the covariance matrix for the forecast revision. We only need to know that the covariance matrix is diagonal. However, to solve the original problem, (21) or (21a), does require knowledge of the covariances to ensure satisfaction of the inventory constraint.

The Kuhn-Tucker conditions for (23) consist of the convexity constraint over the weights, plus the following set of equations:

$$w_{ij} + \lambda \sum_{k=i}^{H} (w_{0j} + \cdots + w_{kj} - u_{kj}) = \gamma \text{ for } i = 0, \ldots H, \tag{24}$$

where $u_{kj} = 1$ if $k \geq j$, $u_{kj} = 0$ if $k < j$, and γ is the (scaled) dual variable for the single convexity constraint in (23). Since (23) is a convex program, the Kuhn-Tucker conditions are both sufficient and necessary, and they identify a unique solution.

To find the solution, we equate (24) for $i - 1$ and i to obtain:

$$w_{ij} = w_{i-1,j} + \lambda(w_{0j} + \cdots + w_{i-1,j} - u_{i-1,j}) \quad \text{for } i = 1, \ldots H. \tag{25}$$

We can construct a solution to (24) by selecting a value for w_{0j} and repeatedly applying (25). To satisfy the convexity constraint, we could search over values for w_{0j}. Alternatively, we describe in the next two propositions how to find w_{0j} analytically.

Proposition 2. *For a given value of* λ, *the solution to* (25) *for* w_{ij} *is a linear function of* w_{0j} *given by*:

$$w_{ij} = P_i(\lambda) w_{0j} \qquad \text{for } i = 0, 2 \ldots j, \tag{26a}$$

$$w_{ij} = P_i(\lambda) w_{0j} - R_{i-j}(\lambda) \qquad \text{for } i = j+1, \ \ldots H, \tag{26b}$$

where $P_i(\lambda)$ *is a polynomial in* λ *of degree i, and* $R_{i-j}(\lambda)$ *is a polynomial in* λ *of degree* $i - j$. *In particular, we can show by induction that for* $n = 0, 1, \ldots H$,

$$P_n(\lambda) = \sum_{i=0}^{n} \frac{(n+i)!}{(2i)!(n-i)!} \lambda^i,$$

and that for $n = 1, 2, \ldots H - j$,

$$R_n(\lambda) = \sum_{i=1}^{n} \frac{(n+i-1)!}{(2i-1)!(n-i)!} \lambda^i.$$

Proposition 3. *The optimal choice for* w_{0j} *that solves* (23) *is given by*:

$$w_{0j} = \frac{1 + \sum_{i=j+1}^{H} R_{i-j}(\lambda)}{\sum_{i=0}^{H} P_i(\lambda)} = \frac{P_{H-j}(\lambda)}{\sum_{i=0}^{H} P_i(\lambda)}, \tag{27}$$

which simplifies to:

$$w_{0j} = \frac{\sum_{i=0}^{H-j} \frac{(H-j+i)!}{(2i)!(H-j-i)!} \lambda^i}{\sum_{i=0}^{H} \frac{(H+i-1)!}{(2i+1)!(H-i)!} \lambda^i}.$$

Proof. From Proposition 2 we can rewrite the convexity constraint as follows:

$$1 = \sum_{i=0}^{H} w_{ij} = \sum_{i=0}^{H} P_i(\lambda) w_{0j} - \sum_{i=j+1}^{H} R_{i-j}(\lambda).$$

We can now use this to express w_{0j} in terms of $P_i(\lambda)$ and $R_i(\lambda)$, as given in the proposition. We simplify the expression for w_{0j} by substituting the following for $R_i(\lambda)$:

$$\sum_{i=1}^{n} R_i(\lambda) = \sum_{i=1}^{n} \frac{(n+i)!}{(2i)!(n-i)!} \lambda^i = P_n(\lambda) - 1,$$

which is found by an induction argument. Similarly, we can simplify (27) by noting that

$$\sum_{i=0}^{n} P_i(\lambda) = \sum_{i=0}^{n} \frac{(n+i+1)!}{(2i+1)!(n-i)!} \lambda^i. \quad \square$$

Having found the optimal choice of w_{0j}, we obtain the remaining weights by iteratively solving (25). We see immediately from Proposition 3 that for positive λ, w_{0j} is positive; we can similarly show that w_{Hj} is positive. From these facts, we can obtain the following proposition by examining the first differences for the optimal weights.

Proposition 4. *The optimal weights* w_{ij} *are positive, increasing, and strictly convex over the range* $i = 0, 1, \ldots j$. *The optimal weights* w_{ij} *are positive, decreasing, and strictly convex over the range* $i = j, j + 1, \ldots H$.

Proposition 5. *The matrix of optimal weights is symmetric about the off-diagonal, i.e.,* $w_{ij} = w_{H-j,H-i}$.

Proof. This can be shown by substitution of (27) into (26). $\square$

Proposition 6. *The optimal weights are such that* $w_{ij} = w_{H-i,H-j}$.

Proof. Since the optimal weights satisfy the convexity constraint, we can substitute the convexity constraint into (25) and rewrite, after some rearrangement, as:

$$w_{i-1,j} = w_{ij} + \lambda(w_{ij} + \cdots + w_{Hj} - (1 - u_{i-1,j})) \quad \text{for } i = 1, \ldots H. \tag{28}$$

From (28), by a similar development as used to find (26), we can express the weights as linear functions of w_{Hj}:

$$w_{ij} = P_{H-i}(\lambda) w_{Hj} \qquad \text{for } i = j, \ldots H, \tag{29a}$$

$$w_{ij} = P_{H-i}(\lambda) w_{Hj} - R_{j-i}(\lambda) \qquad \text{for } i = 0, 1, \ldots j - 1. \tag{29b}$$

In order for the weights to sum to one, we then find that:

$$w_{Hj} = \frac{P_j(\lambda)}{\sum_{i=0}^{H} P_i(\lambda)}. \tag{30}$$

From (29) and (30), we establish the result. $\square$

Proposition 7. *The matrix of optimal weights is symmetric about the diagonal*; *i.e.,* $w_{ij} = w_{ji}$.

Proof. This follows immediately from Propositions 5 and 6. $\square$

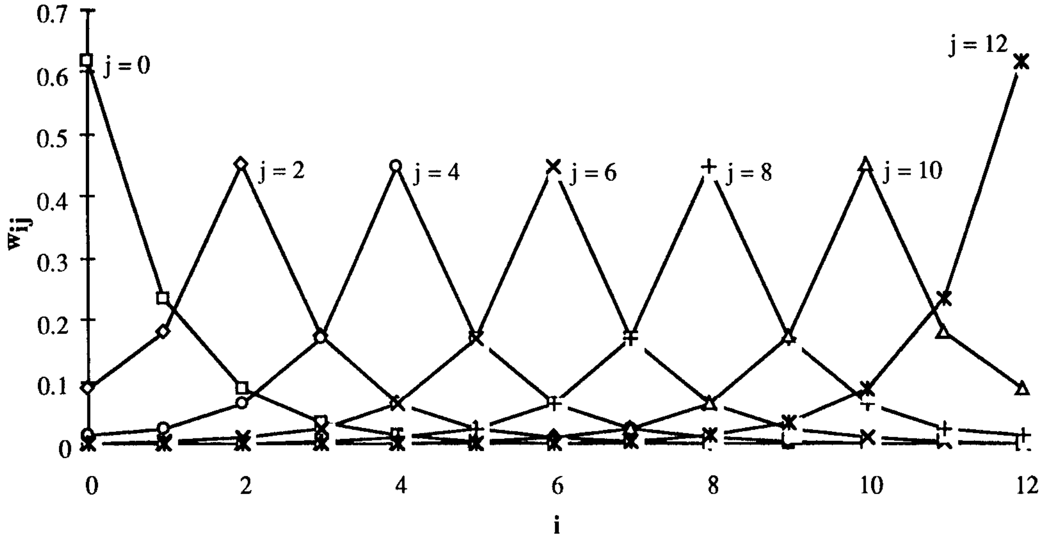

Figure 1. Optimal weights for $\lambda = 1$ and various j.

Figure 1 shows the form of the optimal weights for various values of j for $\lambda = 1$ and $H = 12$. Table I lists the actual values for the optimal weights. From the table we observe that the matrix of optimal weights is symmetric about both diagonals, as stated in the propositions above. Furthermore, for a fixed index j, the weights increase geometrically to a maximum at w_{jj}, and then decay geometrically over the rest of the column.

Figures 2 and 3 show the form of the optimal weights for $\lambda = 4$ and $\lambda = 0.25$ at $H = 12$. Intuitively, we would expect that as λ increases to ∞, w_{jj} goes to 1 and w_{ij} goes to 0 for $i \neq j$ (no production smoothing), and as λ decreases to 0, w_{ij} goes to $1/(H + 1)$ (maximum production smoothing). At $\lambda = 4$ and $\lambda = 0.25$ we already begin to observe this behavior.

Proposition 8. *The optimal objective value for the Lagrangian function in* (23) *is given by* $L_j(\lambda) = w_{jj}\sigma_j^2$ *for* $j = 0, 1, \ldots H$.

Our proof of Proposition 8 involves quite a bit of unattractive and nonintuitive algebra. (See Kletter 1994 for the details.) The basic structure of the proof is as follows: we rewrite the right-hand side of (23) strictly in terms of w_{0j} and λ for a given j by repeatedly applying (26) and factoring out σ_j^2. We then show that this expression equals w_{jj}, where w_{jj} is also expressed in terms of w_{0j} and λ. This is achieved by replacing w_{0j} with the expression given in (27), expressing all terms as polynomials in λ, and then manipulating the binomial coefficients until they are shown to be equal.

The value of Proposition 8 is that it provides a relatively quick way to evaluate the objective function of the Lagrangians, namely (21b) and (23). Also, we show next how to get a good approximation of w_{jj}, which will then yield an analytic expression for the objective function of the Lagrangian.

Suppose we define the first difference $\Delta w_{ij} = w_{ij} - w_{i-1,j}$; we can use (25) to express Δw_{ij} by:

$$\Delta w_{ij} = \Delta w_{i-1,j} + \lambda w_{i-1,j}$$

$$\text{for } i = 1, 2, \ldots H \text{ and } i \neq j + 1,$$

$$\Delta w_{j+1,j} = \Delta w_{jj} + \lambda w_{jj} - \lambda.$$

Table I
Optimal Weights for $\lambda = 1$, $H = 12$

	j												
	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.6180	0.2361	0.0902	0.0344	0.0132	0.0050	0.0019	0.0007	0.0003	1.1E-04	4.1E-05	1.6E-05	8.2E-06
1	0.2361	0.4721	0.1803	0.0689	0.0263	0.0101	0.0038	0.0015	0.0006	0.0002	8.2E-05	3.3E-05	1.6E-05
2	0.0902	0.1803	0.4508	0.1722	0.0658	0.0251	0.0096	0.0037	0.0014	0.0005	0.0002	8.2E-05	4.1E-05
3	0.0344	0.0689	0.1722	0.4477	0.1710	0.0653	0.0250	0.0095	0.0036	0.0014	0.0005	0.0002	1.1E-04
4	0.0132	0.0263	0.0658	0.1710	0.4473	0.1709	0.0653	0.0249	0.0095	0.0036	0.0014	0.0006	0.0003
5	0.0050	0.0101	0.0251	0.0653	0.1709	0.4472	0.1708	0.0653	0.0249	0.0095	0.0037	0.0015	0.0007
i 6	0.0019	0.0038	0.0096	0.0250	0.0653	0.1708	0.4472	0.1708	0.0653	0.0250	0.0096	0.0038	0.0019
7	0.0007	0.0015	0.0037	0.0095	0.0249	0.0653	0.1708	0.4472	0.1709	0.0653	0.0251	0.0101	0.0050
8	0.0003	0.0006	0.0014	0.0036	0.0095	0.0249	0.0653	0.1709	0.4473	0.1710	0.0658	0.0263	0.0132
9	1.1E-04	0.0002	0.0005	0.0014	0.0036	0.0095	0.0250	0.0653	0.1710	0.4477	0.1722	0.0689	0.0344
10	4.1E-05	8.2E-05	0.0002	0.0005	0.0014	0.0037	0.0096	0.0251	0.0658	0.1722	0.4508	0.1803	0.0902
11	1.6E-05	3.3E-05	8.2E-05	0.0002	0.0006	0.0015	0.0038	0.0101	0.0263	0.0689	0.1803	0.4721	0.2361
12	8.2E-06	1.6E-05	4.1E-05	1.1E-04	0.0003	0.0007	0.0019	0.0050	0.0132	0.0344	0.0902	0.2361	0.6180

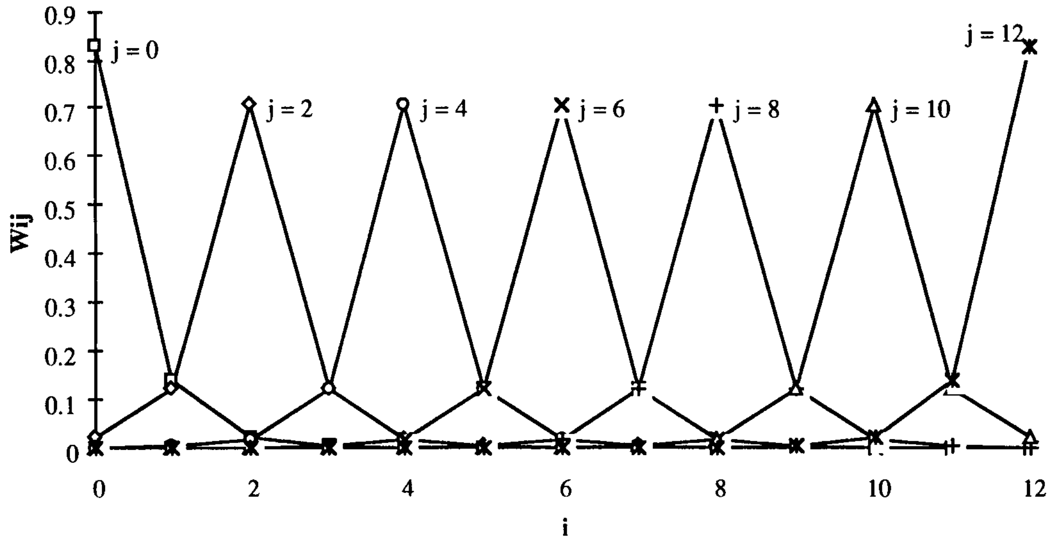

Figure 2. Optimal weights for $\lambda = 4$ and various j.

To get an approximate solution to these first difference equations, suppose we look at a limiting case where we allow both H and j to grow. In effect, we let the range be $i = \ldots -2, -1, 0, 1, 2, \ldots$, except for $i = j + 1$. Then in the limit, the solution to these difference equations is:

$$w_{j+k,j} = w_{j-k,j} = \alpha[(1-\alpha)/(1+\alpha)]^k \quad \text{for } k = 0, 1, 2, \ldots, \tag{31}$$

where $\alpha = \sqrt{\lambda/(\lambda + 4)}$.

Furthermore, this solution satisfies the convexity constraint over the weights. From (31) we see that in the limit:

- the optimal weights are symmetric about w_{jj};
- the optimal weights decline geometrically on either side of w_{jj};
- the value of the maximum weight w_{jj} is independent of j; and
- the maximum weight w_{jj} is a simple monotonic function of λ, that approaches 1 as λ increases.

We can see from Figure 1 and Table I that for $\lambda = 1$, the optimal weights already begin to approach the limit at $H = 12$. In particular, we observe that, except at the end points $j = 0$ and $j = H$, $w_{jj} \approx \alpha = \sqrt{\lambda/(\lambda + 4)} = \sqrt{1/5} \approx 0.4472$ and $w_{j+1,j} = w_{j-1,j} = \alpha[(1-\alpha)/(1+\alpha)] \approx 0.1708$.

The limit provides a simple approximation to the objective function of the Lagrangian relaxation. Using Proposition 8 and (31), we find that for large values of H we can approximate (21b) by:

$$L(\lambda) = \text{Min Var}\,[F_t(t)] + \lambda\ \text{Var}\,[I_t(t)] - \lambda K^2$$
$$\approx \text{tr}(\Sigma)\sqrt{\lambda/(\lambda + 4)} - \lambda K^2.$$

This simplification is helpful for finding the value of λ that maximizes the Lagrangian, and thus solves the original optimization problem (21).

We end this section with an interesting and perhaps useful result.

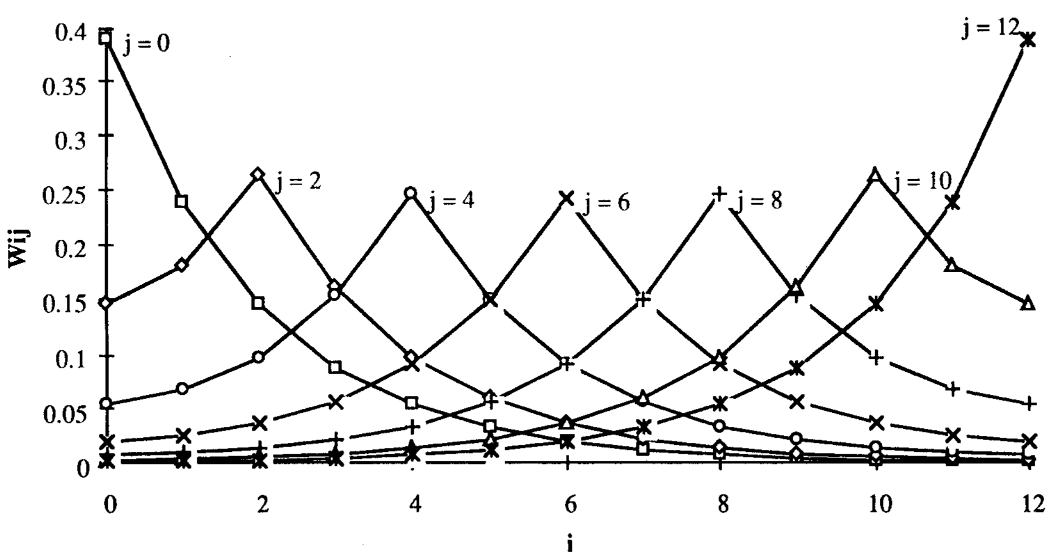

Figure 3. Optimal weights for $\lambda = 0.25$ and various j.

Proposition 9. *The optimal weight matrix is the inverse of a tridiagonal matrix* $\mathbf{C}$, *with* $c_{00} = c_{HH} = (\lambda + 1)/\lambda$, $c_{01} = c_{10} = c_{H,H-1} = c_{H-1,H} = -1/\lambda$, *and with* $(c_{i,i-1}, c_{ii}, c_{i,i+1})$ *given by* $(-1/\lambda, (\lambda + 2)/\lambda, -1/\lambda)$ *for* $i = 1, 2, \ldots H - 1$.

Proposition 9 can be proved by construction through a series of careful matrix operations. (See Kletter 1994 for details.) Our proof simply shows that inverting the matrix $\mathbf{C}$ gives $\mathbf{W}$, as specified in (29). This is accomplished by first factoring $\mathbf{C}$ into $\mathbf{LDL}'$, where $\mathbf{L}$ is bidiagonal, since $\mathbf{C}$ is symmetric and tridiagonal, and then inverting to obtain $(\mathbf{LDL}')^{-1} = (\mathbf{L}')^{-1}\,\mathbf{D}^{-1}\,\mathbf{L}^{-1}$. Since the diagonal matrix $\mathbf{D}$ and the bidiagonal matrix $\mathbf{L}$ are both easily inverted, we then compute the product and simplify to show that $c_{ij}^{-1} = w_{ij}$ for all i and j.

One significance of Proposition 9 is that it makes the computation of the optimal weight matrix even easier.

3. EXTENSION TO MULTISTAGE SYSTEMS

In the previous sections we developed a single-stage model of requirements planning. We now discuss how this single-stage model can serve as a building block in modeling a general acyclic network of multiple stages.

To begin, we state the assumptions and introduce some additional notation that will be necessary for our discussion.

Assumption 1. *The production system is an acyclic network with* n *distinct stages,* m *of which produce end-items, where* $m < n$. *We index the stages so that if stage* i *is downstream from stage* j, *then* $i < j$. *In addition, the end item stages are numbered* $1, 2, \ldots m$.

Assumption 2. *The forecast processes at the end-item stages are mutually independent.*

Assumption 3. *Each downstream stage is effectively decoupled from the upstream stages, i.e., there is always adequate (raw material) inventory for a stage to make its production starts. This is an approximation that is likely to be reasonable if each stage operates with an inventory policy in which stockouts are rare.*

Assumption 4. *Each stage operates according to the assumptions for the single-stage model.*

Namely, let $\underline{f}_i$, $i = 1, \ldots n$, be the forecast vector for each stage i; for simplicity, we will omit the subscript t in this section. Note that for $i = 1, \ldots m$, $\underline{f}_i$ is an exogenous random vector, whereas for $i = m + 1, \ldots n$, $\underline{f}_i$ will be a derived forecast. Let $\underline{F}_i$, $i = 1, \ldots n$, be the output plan for each stage i. Thus, by Assumption 4, there is a weight matrix $\mathbf{W}_i$, and $\underline{\Delta F}_i = \mathbf{W}_i\,\underline{\Delta f}_i$ for each stage i.

To link the requirements of a downstream stage to an upstream stage, we need to model the production starts or releases into each stage. We assume that in each period each stage i, $i = 1, 2, \ldots n$, must translate its planned production outputs $\underline{F}_i$ into a plan of production starts, call it $\underline{G}_i$, over some planning horizon.

Assumption 5. *At each stage, we model production starts as a linear system of production outputs:* $\underline{G}_i = \mathbf{A}_i\underline{F}_i$ *for some matrix* $\mathbf{A}_i$.

We can set $\mathbf{A}_i$ to model a variety of real-world considerations as well as production policies. For instance, we might use the matrix $\mathbf{A}_i$ to model production leadtimes, where production starts are just the production outputs offset by the leadtime, to model yield factors within the production stage (e.g., need to start 1.2 units to get output 1.0), or to model the fact that production starts occur on a different time scale (biweekly rather than weekly) from the production outputs. We can also model a constant work-in-process policy where production starts for the period exactly equal production outputs. Indeed, in this way, for general multistage systems we can use this general approach to compare push policies—where starts equal planned output L periods from now—with pull policies, where starts "replace" the outputs produced in the current period.

Assumption 6. *At each stage we know how many units of input are required for one unit of output. Without loss of generality, we assume that one unit of input is required for one unit of output at each stage.*

The single-stage model that we wish to use as a building block takes as input a dynamic forecast process of the requirements for the stage. We now show that, given the assumptions above, the forecast process at each stage in the multistage network satisfies the assumptions of the single-stage model. In particular, we show the following proposition:

Proposition 10. *At each stage* i, *the forecast revision* $\underline{\Delta f}_i$ *can be expressed as a linear combination of* $\underline{\Delta f}_1, \ldots \underline{\Delta f}_m$: $\underline{\Delta f}_i = \sum_{j=1}^{m} M_{ij}\,\underline{\Delta f}_j$ *for some matrices* M_{ij}. *By Assumption* 2, *this implies that* $\underline{\Delta f}_i$ *is an i.i.d. random vector.*

We will demonstrate this proposition by an induction argument. The proposition is true by assumption for the end-item stages $1, \ldots m$. Suppose this proposition is true for stages $i = 1, \ldots j - 1$; we will now show that it is true for $\underline{\Delta f}_j$. Let S_j be the index set of immediate successors to stage j. The forecast process for outputs of an upstream stage $j > m$ is

$$\underline{f}_j = \sum_{k \in S_j} \underline{G}_k.$$

Accordingly, we can write

$$\underline{\Delta f}_j = \sum_{k \in S_j} \underline{\Delta G}_k.$$

We note by Assumption 6 that $\underline{\Delta G}_k = \mathbf{A}_k\,\underline{\Delta F}_k = \mathbf{A}_k\,\mathbf{W}_k\,\underline{\Delta f}_k$, and by the induction hypothesis that each $\underline{\Delta f}_k$ is a linear combination of $\underline{\Delta f}_1, \ldots \underline{\Delta f}_m$. Thus, we can see that each $\underline{\Delta G}_k$ for $k \in S_j$ is a linear combination of $\underline{\Delta f}_1, \ldots \underline{\Delta f}_m$, and hence so is $\underline{\Delta f}_j$. This completes the induction argument, showing that each $\underline{\Delta f}_j$ is an i.i.d. random vector. □

We have thus shown that at each stage we have preserved the essential requirement that the forecast revisions are i.i.d. random vectors, and thus, that the assumptions for the forecast process of the single-stage model are satisfied at each stage in the multistage network. *This is an important result because it means that we can now model an acyclic multistage system by just replicating the single-stage model.* In this sense, the single-stage model serves as a building block.

4. CASE STUDY

In this section we describe an industrial application of the Dynamic Requirements Planning (DRP) model from a thesis internship performed by one coauthor (Hetzel) at the Eastman Kodak Company. The internship was conducted as part of MIT's Leaders for Manufacturing Program and ran from June 1992 to December 1992. (See Hetzel 1993 for more details on the application.)

The general charge for the thesis was to investigate cycle time reduction within the context of the film manufacturing processes at Kodak. As part of the internship, Hetzel joined an internal supply chain optimization team that was investigating opportunities for better coordination over a specific supply chain, including issues of cycle time and inventory reduction. One open issue facing the team was that of strategic inventory placement: how much inventory was needed, and where should it be placed across a multistage supply chain. Hetzel identified this as an opportunity to apply the DRP model, and the team agreed that it was an appropriate tool for their task of strategic inventory placement. The only alternative considered was to develop a simulation: since the DRP model was already available from the authors in a software package, developing a simulation would have required extensive additional work.

The goal of the supply chain analysis was to determine the optimal safety stock levels between each stage in the film making supply chain. The underlying concept is that looking at one stage of the supply chain in isolation is inherently suboptimal. All the stages in the supply chain are interconnected by information flows. In short, the inventory and production policies that are best for one stage may not be optimal for the supply chain as a whole.

In the case study, the team was able to address this situation by using the DRP model to consider all stages in the supply chain. Their recommendations challenged the conventional targets and performance measures for individual divisions (stages). For example, an upstream stage, roll coating, faced a corporate-wide mandate to lower inventories. However, by using the DRP model, the team discovered that roll coating needed to increase inventories to provide the desired service to the next stage. When roll coating holds sufficient inventory to provide a high level of service, downstream stages can hold less, resulting in a net savings for the corporation. Overall, the analysis determined that inventories for the products of the case study could be reduced by 20%. This example highlights the importance of considering the entire supply chain when setting inventory and production policies.

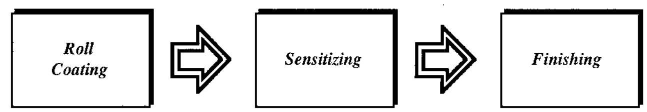

Figure 4. Simplified version of film manufacturing supply chain.

The rest of this section will describe the supply chain for the case study, provide the results from the DRP model, and comment on implementation issues.

4.1. Supply Chain for Case Study

In Figure 4 we give a simplified version of the process for film making. Roll coating transforms raw chemicals into a roll of film base. Sensitizing coats the film base with a silver halide emulsion. Then finishing cuts and packages the sensitized rolls into finished products. The structure of this supply chain has three interesting characteristics. First, the number of items grows dramatically from stage to stage; one film base might result in 5 to 10 different sensitized rolls, which might lead to a hundred or more finished goods. Second, there is a rapid growth in the value of the product due to added material (e.g., silver) and nature of the processes. Third, there is a gradual decrease in the leadtimes across the supply chain.

For the case study, the supply chain optimization team focused on a single film base (called a support). That single base becomes three different sensitized film codes because it can be coated with three different emulsions. The three film codes can be finished (slit, chopped, and packaged) into 24 different finished good items. Figure 5 illustrates the supply chain for the case study. This particular product "tree" was chosen because it is high volume, it has relatively few end items (24 total), and it represents a "typical product" that the team felt would make a useful pilot program.

It is important to note that the case study does establish arbitrary bounds on the supply chain. The case study starts with the creation of a film base in roll coating and excludes the upstream raw material stages such as chemical, gelatin, and polymer production. The case study ends with the finishing process and arrival at the Central Distribution Center, and ignores the rest of the distribution system. Besides being bounded at both ends, the case study's supply chain is also simplified. In reality, the sensitizing and finishing stages have materials flowing into them such as emulsion and packaging components. Even though these materials require inventory management, they are assumed to be available with 100% service, and were not explicitly incorporated into the model.

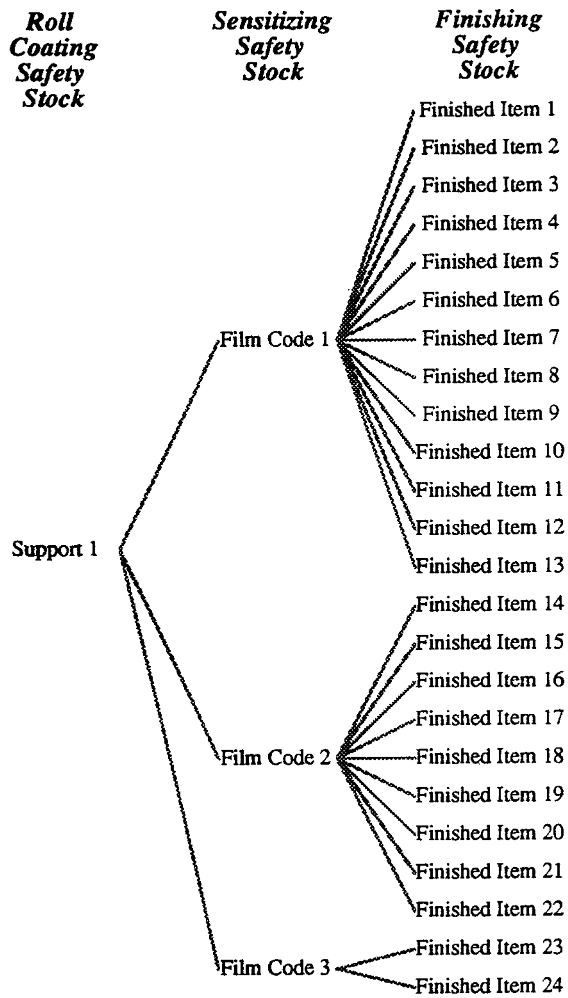

Figure 5. Supply chain analysis case study.

4.2. Data Collection

Parameterizing the DRP model required an extensive data collection effort. For each item in the chain, the team gathered data on the item's leadtime, unit cost, inventory holding cost, manufacturing frequency, and desired service level. For each end item, they needed the planning horizon, the average demand level, and a time history of the forecast process.

The leadtime and manufacturing frequency were modeled through the weight matrix (**W**). Since the team did not consider production smoothing, in the absence of leadtime and production frequency considerations, the weight matrix is simply the identity matrix. A leadtime of L periods is then captured by forcing the first L rows of **W** to be zero. To represent a production frequency of once every two weeks, **W** would then be modified so that every other row was zero. It should be noted that this method of capturing production frequency is only an approximation.

From the forecast histories, the team estimated the diagonal elements of the covariance matrix (Σ) for the forecast revision process $\underline{\Delta f}_t$; the off-diagonal elements were assumed to be zero. Associated with each "branch" linking different stages they calculated a historical "goes into factor" to capture any yield loss or conversion factors. This information was used to construct the matrix **A**, as described in Section 3.

A side benefit of applying the DRP model was that the data collection effort identified some potential issues along the supply chain. For example, in the course of reviewing the forecast data, the team discovered that the forecasts varied in a systematic way that led to a reevaluation of the forecasting process. In addition, collecting data enhanced supply chain communication and allowed the team to resolve a discrepancy in the annual planned volumes between two of the stages.

4.3. Results

The team used the DRP model to develop a base case recommendation on inventory placements. The 24 finished items were grouped into nine product aggregates, where the product aggregates shared common production processes and had similar demand histories. Service levels were set at 95% for each stage. The weight matrices were not optimized and were set to reflect each stage's leadtime and manufacturing frequency. In order to reflect Kodak's current scheduling systems, there was no production smoothing across weeks.

The DRP model showed the potential to lower inventory across the case study product "tree" by 20%, as shown in Figure 6. Note that, in general, inventories can be pushed upstream where they are in a strategic position because: (1) the inventory is common to the greatest number of finished end items desired by the customer, and (2) the inventory is at its lowest value added and thus at its lowest carrying cost. In fact, the inventory levels of roll coating's "Support 1" actually need to increase to provide savings for the supply chain as a whole.

The definition of "inventory" as it is used in this results section is important. The inventory changes and the comparisons in Figure 6 represent average inventories. Average inventory for each item includes the safety stock calculated by the DRP model, plus the cycle stock due to production batching, plus the pipeline stock from transport needs.

Besides the required safety stocks, the DRP model also provided information on the variance of the production requirements at each stage of the supply chain. The supply chain optimization team used this variance to determine the "surge" production capability needed for any stage. For instance, they might set the surge capability to be the production level that would cover the production requirements 95% (1.645 standard deviations above the mean, assuming normal forecast errors) of the time.

4.4. Validation

Before the DRP model recommendations could be implemented, the team needed to develop confidence in the results. Therefore, multiple scenarios were run to test the

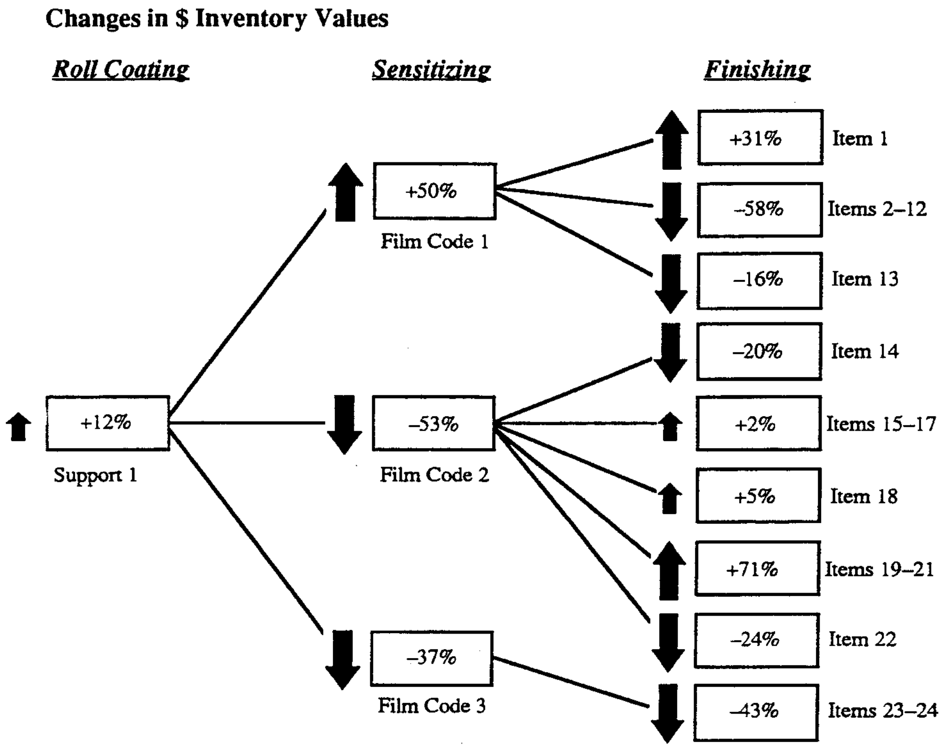

Figure 6. Results from the supply chain analysis—strategic inventory placement. Note: Excludes emulsion and chemicals inventories. Excludes regional distribution center (RDC) inventories. Includes all WIP, cycle, and in-transit stocks. Finished goods inventories are gross CDC averages. Service levels are at 95% for all stages. There is no production smoothing.

sensitivity to various parameters, including the service levels, the leadtimes, and the size (variance) of the forecast errors. (See Hetzel 1993 for details.)

The main barrier that the team had to overcome was understanding how the DRP model works. This was accomplished by exercising the model for different scenarios, especially conservative ones; by displaying all the input data and its sources for validation; by keeping the model (relatively) simple, e.g., assuming a diagonal covariance matrix and limiting the size of the explosion; by comparing the model results versus current inventory levels; and by acknowledging the model's shortcomings. Finally, a key success factor was that the model was implemented on a personal computer, provided a graphical interface for representing and visualizing the supply chain, and provided an almost instantaneous response. In addition to the analytic model, a Monte Carlo simulation was used that simply worked through the mechanics of the analytic model for a randomly generated demand stream, reporting on performance measures of interest. The simulation allowed assessment of model assumptions, thereby validating the analysis. For example, constraints were added to the simulation that enforced production capacities and prohibited production from beginning if no raw material was onhand.

No model is perfect, and no description of a model is complete without a list of shortcomings. The supply chain optimization team identified three weaknesses: (1) the DRP model does not account for lead time variability, (2) it assumes stationary average demand over time, and (3) it cannot accommodate a large product explosion. Whereas the first two are inherent assumptions for the model, the latter concern was due to a limitation in the software that could be easily overcome. However, we expect that in most practical situations a team should probably not be working at any greater level of detail than the case study, say, less than 25 items. Keeping the model at an aggregate level both reinforces the fundamental guiding principles and also makes implementation simpler.

4.5. Implementation

Once all of the supply chain analysis requirements were complete, the supply chain optimization team added local intelligence about specific customers and manufacturing issues for each item that could not be captured by the model. After reaching an understanding about how all of the model's proposed changes would impact the supply chain, the team decided to implement a pilot program, with the intention of moving to the more aggressive "base case" if there were no service problems.

The pilot program only involved the inventories of three items. The plan raised the one roll coating item's inventory by 20%, and it lowered two sensitizing items' inventories, each by 60%. The plan was implemented in early 1993. The savings were captured in the 1993 Annual Operating Plan for the case study's line of business. As of the end of April 1993, not a single end-customer order had been missed on the pilot product due to stockouts or inventory shortages. The team then implemented the remaining recommendations over the course of 1993.

5. CONCLUSIONS

This paper presents a new model of the requirements planning process. We first describe in detail how to model a single production-inventory stage as a linear system, and provide the analysis for determining performance measures on production smoothness, production stability, and inventory requirements. We also show how to optimize the tradeoff between production smoothness and inventory for a single stage.

To model a multistage system, we can use the single-stage model as a building block. The structure of the single-stage model makes it very easy to link single-stage models together to represent the multistage system. In particular, each single-stage model takes as input a forecast of demand requirements and converts this forecast into a production plan. In the context of a network of production stages, the production plan from a downstream stage acts as the demand forecast for an upstream stage. In this way, we can cascade the single-stage models to model a multistage system.

We also report on an application of the model within the context of a supply chain study. The DRP model was used as a tool to help determine inventory placement across a multistage supply chain. This illustration provides some evidence of the value of taking a corporate-wide view by optimizing the supply chain rather than suboptimizing each of the pieces.

One outgrowth from the case study is a better understanding of industry needs, and where the DRP model is weak. Based on this experience, as well as observations from industry, we identify the following research topics.

- *Nonstationary demand.* A stationary demand process is not an accurate model for the demand experienced by many products. Common nonstationary effects include seasonal effects, end-of-quarter or end-of-year effects (the "hockey stick"), and short-product life cycles. Some of these nonstationarities get masked when products are aggregated into families or product groups. Nevertheless, an important enhancement to the model would be to capture, in some way, nonstationary demand processes.

- *Service-Level Assumptions.* In extending the single-stage model to a multistage setting, we assume that there will be sufficient inventory to decouple the stages. In effect, we assume that the service levels will be set to assure a high level of service, and in the model analysis, we ignore the downstream consequences of an upstream stockout; i.e., starvation of inputs. These assumptions raise two questions. One is, what are the consequences of ignoring the internal stockouts, and the second is, what should the internal service levels be. Graves (1988a) provides some justification for these assumptions in a related setting. And simulation tests that we have done confirm that ignoring the internal stockouts in the analysis, when service levels are high, does not distort the results of the model. But the issue remains as to how to set the service levels. The literature on multiechelon distribution systems (e.g., Jackson 1988, Schwarz 1989, Graves 1995) suggests that, from a system perspective, it often may be better to have low levels of internal service.

- *Guidelines for Consolidating Stages.* On a related note, we conjecture that, in some instances, the best policy may be to remove the inventory between an upstream and downstream stage, and thus consolidate these stages for planning purposes (Simpson 1958). Rather than have two stages separated by an inventory buffer, we would have one (combined) stage, albeit with a longer leadtime. Within a multistage system, depending on the leadtimes and holding costs, it may be optimal to consolidate some of the stages. We expect it would be helpful to have guidelines for determining what stages are good candidates for consolidation.

- *Multistage Optimization.* The paper describes the optimization of the tradeoff between capacity and inventory in a single stage for a diagonal covariance matrix: It would be interesting to explore how this development extends to nondiagonal covariance matrices, as well as to a multistage system. In particular, we would like to develop guidelines for setting the weight matrix $\mathbf{W}$ for each stage. Furthermore, one could explore how to choose among alternative production release policies, such as pull versus push, in a multistage setting.

- *Production Assumptions.* The model has a highly-simplified model of the production process. The model sets the production outputs, and these outputs are translated into production starts (e.g., by a leadtime offset). With this model, we can represent fixed lead times, yield loss factors, batch setup frequencies, as well as uncertainty that can be modeled as an additive factor. Nevertheless, there are issues as to the validity or appropriateness of this representation and the sensitivity of the model results to these assumptions. It would certainly be useful to have a richer model of the production process. For instance, it would be useful to capture the nonlinear congestion effects due to multiple items competing for a shared resource.

ACKNOWLEDGMENT

We wish to thank Chris Athaide for his contributions in the initial stages of this research, the IBM Thomas J. Watson Research Center and the NSF Strategic Manufacturing Initiative for financial support for this research, MIT's Leaders for Manufacturing Program for its support and resources to complete and apply this research, and the referees for their helpful and constructive comments on an earlier draft.

REFERENCES

BAKER, K. R. 1993. Requirements Planning. In *Handbooks in Operations Research and Management Science, Vol. 4, Logistics of Production and Inventory*. S. C. Graves, A. H. Rinnooy Kan and P. H. Zipkin (eds.), North-Holland, Amsterdam.

FINE, C. H. AND S. C. GRAVES. 1989. A Tactical Planning Model for Manufacturing Subcomponents in Mainframe Computers. *J. Manuf. and Opns. Mgmt.* **2,** 1, 4–34.

GRAVES, S. C., H. C. MEAL, S. DASU, AND Y. QIU. 1986. Two-Stage Production Planning in a Dynamic Environment. In *Multi-Stage Production Planning and Inventory Control*. S. Axsäter, C. Schneeweiss and E. Silver, (eds.), Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, **266,** 9–43.

GRAVES, S. C. 1986. A Tactical Planning Model for a Job Shop. *Opns. Res.* **34,** 4, 522–533.

GRAVES, S. C. 1988a. Safety Stocks in Manufacturing Systems. *J. Manuf. and Opns. Mgmt.* **1,** 1, 67–101.

GRAVES, S. C. 1988b. Determining the Spares and Staffing Levels for a Repair Depot. *J. Manuf. and Opns. Mgmt.* **1,** 2, 227–241.

GRAVES, S. C. 1988c. Extensions to a Tactical Planning Model for a Job Shop. *Proceedings of the 27th IEEE Conference on Decision and Control*, Austin, Texas, December.

GRAVES, S. C. 1996. A Multiechelon Inventory Model with Fixed Replenishment Intervals. *Mgmt. Sci.* **42,** 1–18.

HEATH, D. C. AND P. L. JACKSON. 1994. Modeling the Evolution of Demand Forecasts with Application to Safety Stock Analysis in Production/Distribution Systems. *IIE Trans.* **26,** 3, 17–30.

HETZEL, W. B. 1993. Cycle Time Reduction and Strategic Inventory Placement Across a Multistage Process. MIT Master's Thesis.

JACKSON, P. L. 1988. Stock Allocation in a Two-Echelon Distribution System or 'What to Do Until Your Ship Comes In'. *Mgmt. Sci.* **34,** 7, 880–895.

KARMARKAR, U. S. 1993. Manufacturing Lead Times, Order Release and Capacity Loading. In *Handbooks in Operations Research and Management Science, Vol. 4, Logistics of Production and Inventory*. S. C. Graves, A. H. Rinnooy Kan and P. H. Zipkin (eds.), North-Holland, Amsterdam.

KLETTER, D. B. 1994. Proofs of P8 and P9. Technical Appendix.

LEE, H. L. AND C. BILLINGTON. 1993. Material Management in Decentralized Supply Chains. *Opns. Res.* **41,** 5, 835–847.

SCHWARZ, L. B. 1989. A Model for Assessing the Value of Warehouse Risk-Pooling: Risk-Pooling over Outside-Supplier Leadtimes. *Mgmt. Sci.* **35,** 828–842.

SIMPSON, K. F. 1958. In-Process Inventories. *Opns. Res.* **6,** 863–873.