Extensions To a Model For Tactical Planning In a Job Shop Environment

By

Scott Howell Parrish

A. B., Grinnell College, 1968.
M. B. A., Loyola College, 1981.

Submitted in Partial Fulfillment of
the Requirements of the Degree of
Master of Science in Operations Research

at the
Massachusetts Institute of Technology
June 1987

Signature of Author ______________________________
Interdepartmental Program in Operations Research
June 1987

Certified by ______________________________
Stephen C. Graves
Associate Professor of Management Science
Thesis Supervisor

Accepted by ______________________________
Thomas L. Magnanti
Chairman, Interdepartmental Committee on Operations Research

Extensions To a Model For Tactical Planning In a Job Shop Environment

By

Scott Howell Parrish

Submitted in Partial Fulfillment of
the Requirements of the Degree of
Master of Science in Operations Research

Abstract

We develop three extensions to a previously published model which assists managerial planning in a multi-sector job shop [Graves, 1986]. This model provides insight into the trade-off between two aspects of system behavior, production smoothing and size of work-in-process inventories, which will be affected by management's choice of production lead times. For all three extensions we provide numerical examples of application.

First, we develop a model of a production release rule which coordinates demand forecasts over a planning horizon with knowlege of the projected completion times of items currently in work-in-process inventory in order to determine the number of items to start into production. This release rule model is necessary in order to apply Graves' model to a production line which produces in response to demand forecasts rather than in response to order arrivals, as is typical of a job shop.

Second, we develop two measures of the service level of a shop and show how these are affected by the choice of production lead times. The service measures considered are the probability that demand during a random period is met and average length of failure runs (by a failure run we mean several periods in succession during which the facility fails to meet demand). We discuss how management can evaluate the impact of lead time choice on the expanded set of trade-offs among service measures, production smoothing, and size of inventories.

Third, we model the dynamic behavior of a shop in response to various types of change in demand. The changes considered are a one-period increase, permanent increase, linear growth, and cyclical demand. We show how to assess the time the system will take to achieve its new equilibrium and the nature of the path followed during the transition and how these responses depend upon lead time choices.

These three extensions should enhance the usefulness of Graves' model by widening the scope of facilities to which it may be applied and by extending knowledge about the impact of lead time choices on additional important aspects of system behavior.

Thesis Supervisor: Dr. Stephen C. Graves

Title: Associate Professor of Management Science

Acknowledgments

This research was conducted with the support of a research assistantship administered by the Sloan School of Management.

I wish to express my sincere appreciation to my advisor, Stephen C. Graves, whose guidance, wisdom and encouragement have been instrumental in the completion of this work. His patience with my frequent mental wanderings into absurd and irrelevant areas has constantly amazed me.

My wife Kathleen has been a constant source of love and encouragement while I performed the research, and she contributed in a very tangible way by conquering a word processor which was filled with astounding behavioral quirks to type the several drafts and the final copy.

To her I dedicate this thesis:

A wife of noble character who can find?
She is worth far more than rubies.
Her husband has full confidence in her
and lacks nothing of value.
She brings him good, not harm,
all the days of her life.

Many women do noble things,
but you surpass them all.
Give her the reward she has earned,
and let her works bring her praise at the city gate.

Proverbs 31: 10-12, 29, 31.

Contents

Chapter 1
Introduction

A. Purpose

The purpose of this thesis is to develop and apply three extensions to a tactical planning model in a job shop environment. This work grew out of a research project during which a team of professors and students (one of whom was the author) applied the underlying model to a shop which manufactures electronic components.

The work contained herein is a combination of model development and model application. Chapter 2 records development and application work which was done during the course of the research project and was essential to the project's ability to use the underlying model for the electronic component line. Chapters 3 and 4 record work which is primarily theoretical development and which occurred after termination of the research project to follow up on additional modeling issues that were identified during the project.

We believe this work will enhance the usefulness of the underlying model by first providing some additional practical guidelines for use in its implementation to a specific situation and second by enhancing its ability to answer questions of importance to manufacturing management.

B. Organization

In the next section we briefly describe the underlying model, which this thesis extends. Following this, we provide an overview of the manufacturing facility to which the model was applied, and then we give a synopsis of the model extensions contained in Chapters 2, 3 and 4.

C. The Underlying Model

The underlying model is one proposed by Graves in his article, "A Tactical Planning Model for a Job Shop" [Graves, 1986]. Graves develops a model which provides the steady-state distribution of production levels and of queue lengths at each work center, or sector, of a job shop which result from a proposed control rule for setting production levels of each sector. He shows how to use the model to evaluate choices of controls in order to produce acceptable shop behavior.

We present here a very abbreviated overview of the Tactical Planning Model (referred to hereafter as the TPM) and its key equations. We refer the reader to Graves' article for a more extended treatment of the TPM and a discussion of the rationale underlying its development.

The TPM is a discrete time model and identifies all flow variables as occurring during a period, which may be specified to be an hour, shift, day, or whatever unit is convenient in applying the model. In this thesis we will refer simply to "periods".

Graves proposes to control production by a rule that at each sector a period's production be established as a chosen fraction of the work-in-process queue (WIP) at the sector:

(1.1) $$p_{it} = \xi_i q_{it} ,$$

where p_{it} is the production during period t for sector i, q_{it} is the WIP at sector i at the beginning of period t, and the control parameter ξ_i, $0 < \xi_i \leqq 1$, is the selected fraction. Choice of ξ_i is equivalent to choice of a lead time n_i at the sector. By "lead time" we mean the number of periods, on the average, an item will take to be processed through the sector. Clearly, ξ_i is the inverse of n_i: choosing $\xi_i = 0.50$ is equivalent to choosing a two-period lead time.

The units in which p_{it} and q_{it} are measured is a critical issue in applying the model to a given shop as the units must be uniform across all products. In a job shop which produces a variety of products, p_{it} and q_{it} cannot be stated in items since each item may have a substantially different work content. Graves suggests stating these in terms of hours of work for the sector. In the remainder of this thesis we will think of these quantities as units of product since this interpretation was appropriate for the manufacturing line to which the model was applied. The reader should keep this difference from Graves in mind and carefully make the correct choice of units when using the TPM and these extensions.

The WIP queue at each sector is governed by an inventory balance equation,

$$q_{it} = q_{i,t-1} - p_{i,t-1} + a_{it}, \tag{1.2}$$

where a_{it} is the amount of work that arrives at sector i at the beginning of period t. These arrivals may come from many other sectors, and the flow from sector j to sector i is modeled by

$$a_{ijt} = \Phi_{ij}\, p_{j,t-1} + \varepsilon_{ijt}, \tag{1.3}$$

where Φ_{ij} is the expected number of hours of work generated for sector i by every hour of work completed by sector j, and ε_{ijt} is a random variable with zero mean. The term ε_{ijt} is a noise term introducing uncertainty in the arrival stream. It is assumed that the terms of the time series $\{\varepsilon_{ijt}\}$ are i.i.d. The total arrivals to a sector in equation (1.2) are the sum over all preceding sectors of (1.3):

$$a_{it} = \sum_j \Phi_{ij}\, p_{j,t-1} + \varepsilon_{it}, \tag{1.4}$$

where

$$\varepsilon_{it} = N_{it} + \sum_j \varepsilon_{ijt}, \tag{1.5}$$

where N_{it} is a random variable representing new jobs which enter the shop directly to sector i. The elements of each time series $\{N_{it}\}$ are assumed to be i.i.d. The element ε_{it} can be thought of as an innovation to the system at time t, consisting of new items introduced plus a random noise.

By substituting (1.4) into (1.2), and (1.2) into (1.1), solving for p_{it}, and then restating the system of equations in vector and matrix form, we have

(1.6) $$\mathbf{P}_t = (\mathbf{I} - \mathbf{D} + \mathbf{D\Phi})\mathbf{P}_{t-1} + \mathbf{D}\varepsilon_t ,$$

where $\mathbf{P}_t$ is the vector of elements p_{it}, ε_t is the vector of elements ε_{it}, $\mathbf{D}$ is a diagonal matrix with the control parameters $\xi_i = 1/n_i$ on the diagonal, and $\mathbf{\Phi}$ is a matrix whose ijth element is ϕ_{ij}. The identity matrix is $\mathbf{I}$. By successively substituting into (1.6) for $\mathbf{P}_{t\text{-}1}$, $\mathbf{P}_{t\text{-}2}$, etc., and by assuming an infinite history to the system, $\mathbf{P}_t$ can be written as a weighted sum of all past innovations

(1.7) $$\mathbf{P}_t = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D} \varepsilon_{t-s} , \qquad \textit{where } \mathbf{B} = (\mathbf{I} - \mathbf{D} + \mathbf{D\Phi}).$$

The expected production levels are, taking the expectation of (1.7):

(1.8) $$\mathrm{E}(\mathbf{P}_t) = \rho = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D} \mu ,$$

where the vector μ is the expected value of the innovation vector ε_t. Graves shows that the geometric series converges provided that the spectral radius of the matrix $\mathbf{\Phi}$ is less than 1, which is necessary and sufficient for the spectral radius of $\mathbf{B}$ to be less than 1. With convergence guaranteed, (1.8) can be written in the equivalent form:

(1.9) $$\mathrm{E}(\mathbf{P}_t) = \rho = (\mathbf{I} - \mathbf{\Phi})^{-1} \mu .$$

The variance of production is

(1.10) $$\mathrm{Var}(\mathbf{P}_t) = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D} \Sigma \mathbf{D} \mathbf{B}'^s ,$$

where Σ is the variance-covariance matrix of the innovation vector ε_t. Equilibrium levels of $\mathbf{Q}_t$ may be stated from the equilibrium levels of $\mathbf{P}_t$, since, from (1.1), $q_{it} = p_{it}/\xi_i$. Thus, in matrix form we have:

(1.11) $$\mathbf{Q}_t = \mathbf{D}^{-1}\mathbf{P}_t ,$$

(1.12) $$\mathbf{E}(\mathbf{Q}_t) = \mathbf{D}^{-1}\boldsymbol{\rho} , \qquad \textit{and}$$

(1.13) $$\mathbf{Var}(\mathbf{Q}_t) = \mathbf{D}^{-1}\,[\mathbf{Var}(\mathbf{P}_t)]\,\mathbf{D}^{-1} .$$

From the equilibrium levels for $\mathbf{E}(\mathbf{P}_t)$, $\mathbf{Var}(\mathbf{P}_t)$, $\mathbf{E}(\mathbf{Q}_t)$ and $\mathbf{Var}(\mathbf{Q}_t)$ given by (1.9), (1.10), (1.12) and (1.13), Graves gives prescriptions for management choice of control parameters, ξ_i. First, as can be seen from (1.9), the choice of ξ_i's has no effect on equilibrium production levels. These depend only on the required work flows of the system, as contained in the matrix $\mathbf{\Phi}$. The variance of production at each sector is, however, affected by the choice of ξ_i, as can be seen from (1.10). In general, ξ_i operates as a smoothing parameter: decreasing ξ_i, which is the same thing as increasing lead time n_i, will smooth production and reduce variance.

In most circumstances management will desire to reduce variance of production levels as much as possible in order to reduce the costs of shifting labor to respond to variations and the extra capital necessary to assure sufficient capacity to respond to the peaks of production demands. There is a cost, however, of reducing production variance. Increasing lead time obviously creates a larger work-in-process inventory. This effect is evident intuitively, and its effect is seen directly in equation (1.12). Equation (1.13) demonstrates that increased smoothing of production levels is accomplished at the expense of increased variance of WIP levels.

The equations (1.9), (1.10), (1.12) and (1.13) are the primary results of the TPM. They permit management to obtain a characterization of the shop behavior for various choices of control parameters. By using these results, management can assess the impact and costs of different lead time choices and hopefully identify an appropriate level of trade-off between the benefits of production smoothing and the cost of extra inventory.

D. The Electronic Component Manufacturing Line

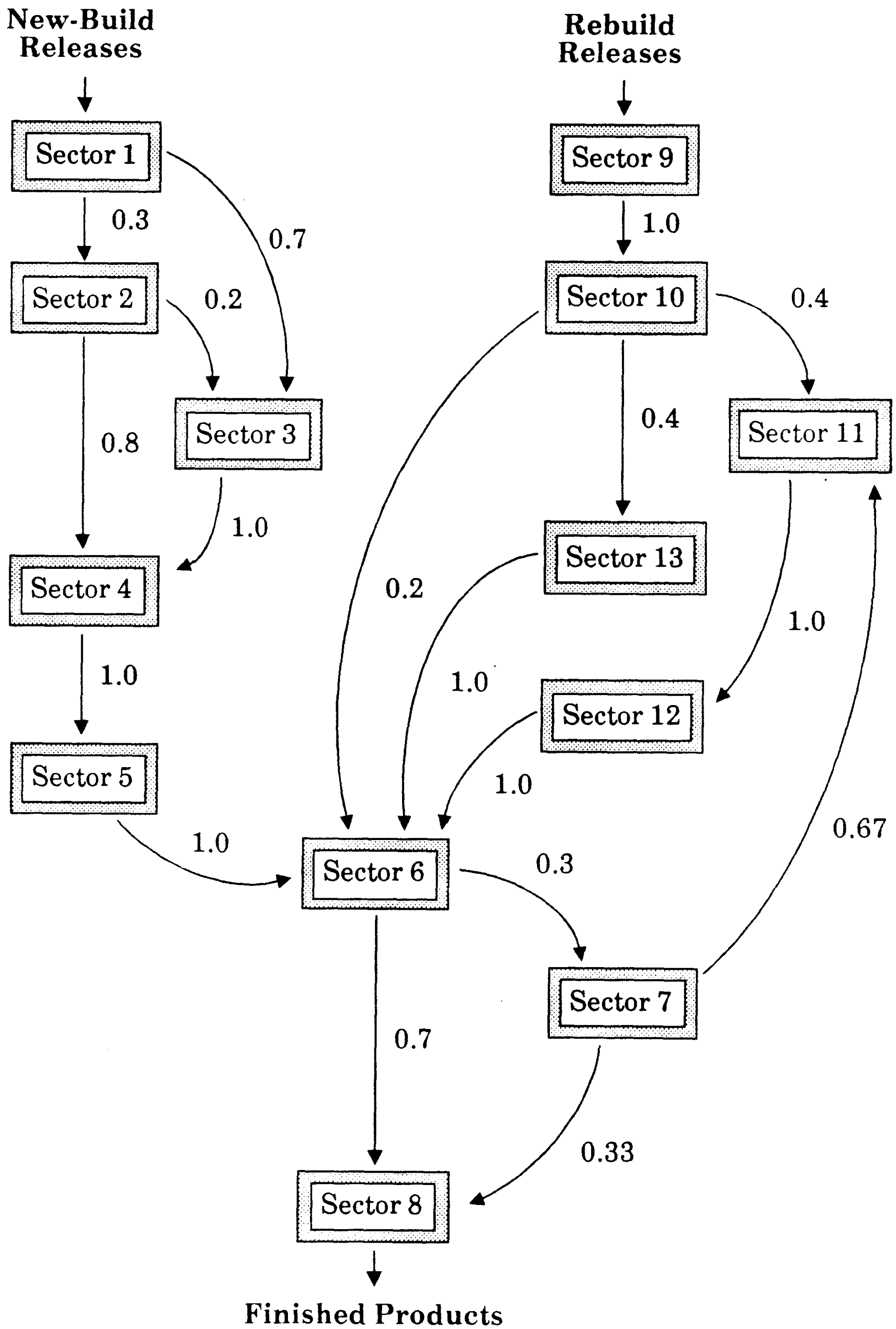
During the course of the research project, the TPM was applied to a production line which manufactures components for electronic equipment. To preserve confidentiality of the manufacturer, we will refer to this facility only as the "electronic component manufacturing line", or as "the ECM line". In this section we provide only a concise view of the ECM line since a full understanding of all its characteristics is not important in following the balance of this work.

The project team identified thirteen work sectors into which it was convenient to divide the various operations performed in the ECM line. Figure 1.1 shows the flows of work among sectors, and the numbers on the flows are the ϕ_{ij} elements from the matrix Φ. The product units flowing through the line have nearly uniform work content at each sector, so it was convenient and appropriate to state and interpret the ϕ_{ij} as proportions or probabilities of flow. That is, we can say that an item leaving sector 1 flows to sector 2 with 30% probability, and to sector 3 with 70% probability. As we discussed above, this interpretation would not be appropriate in a job shop which had a variety of products flowing through it.

The line produces new components, which start at sector 1 and flow through several assembly steps until they are tested at sector 6. As is not unusual in electronics manufacturing, a moderately high number of components fail the test. The test results are analyzed in sector 7 and here it is determined that some of the items which failed the test are indeed good and that others need to

Figure 1.1

The ECM Line



be reworked. Those which need rework are sent to sector 11 to begin that process, and those which do not are sent to sector 8 for final assembly.

In addition to reworking items which failed the test, the line also rebuilds units which have failed in service. These items enter through sector 9 and undergo cleaning and disassembly processes before they join the items shunted from the new-build side for reworking.

Even though the ECM line is not a job shop in the sense that we usually use this term there are two features it has in common with a job shop and these led to the decision to model it using the TPM. The existence of many alternate paths an item can take through the line is one of these features, and this is one which the TPM was designed to capture. Second, variability in the flows (which is increased by the existence of multiple paths) causes large variations in production requirements at each sector in the line, creating capacity problems and difficulties in making daily staffing decisions. This flow variability is, again, a characteristic for which the TPM was designed.

The TPM was applied and a program to run on a personal computer was developed which allows the line management to test the effect of different choices of sector lead times. Management inputs the level of demands and hypothetical choices of lead times. In response, the program reports expected production and WIP levels and their variances. If the results indicate unacceptably high production variance, management can try an alternate choice of lead times. Thus, using the program management can explore the trade-offs between WIP levels and production variations.

E. Preview of Chapters 2, 3 and 4

In the process of implementing the TPM on the ECM line, three significant questions arose. These are the subjects of the model extensions developed in these chapters.

First, as a prerequisite to implementing the TPM it was necessary to develop a model of the method of releasing new-build units into the ECM line at sector 1. In a job shop, releases at a sector are determined externally to the model since they simply result from the jobs the shop receives. Such releases are incorporated into the TPM by simply placing the mean of their distribution in the vector μ. The rebuild side of the ECM line operates in this job-shop-like manner since the arrival of a unit for rebuild generates a release. On the new-build side, in contrast, releases are based on a forecast of demand in future periods, and it was essential that a model of this process be incorporated into the TPM.

Chapter 2 discusses development of the release rule to accomplish this end. The release rule model, documented in section B of Chapter 2, was developed by Graves. The author is responsible for developing the analysis, documented in section C, for calculating exit probabilities which are required to implement the control rule.

The second issue which arose is the question of what impact, if any, the choice of lead times might have on the ability of the ECM line to meet demand. It was clear that increasing the lead times would increase the correlation between the finished goods output of successive periods, so that if output were low in one period, the likelihood of low output in the following period was increased. Increasing the lead times might generate increased probabilities of longer runs of low output. Thus it was conceivable that by increasing lead times to reduce production variances, management might actually harm performance as measured by response to demand.

It was thus obvious that an extension to the TPM which gives measures of the service provided by the system and the response of this service to the control parameter would be useful. In Chapter 3 we propose and develop service measures and then provide an integrated discussion of the impact of control parameter choice on the many aspects of system performance.

The third question which arose was about dynamic behavior of the system. The TPM solves for the equilibrium levels of $\mathbf{E}(\mathbf{P}_t)$, $\mathbf{E}(\mathbf{Q}_t)$, $\mathbf{Var}(\mathbf{P}_t)$ and $\mathbf{Var}(\mathbf{Q}_t)$ based on the assumption that demand is reasonably constant over a period long enough for the system to adjust to its level. The ECM line, however, like many other production lines, is subject to many substantial short and long-term variations in demand. In this dynamic environment it is natural to ask whether the ECM line will ever adjust to the equilibrium solution produced by the TPM, how rapidly it will approach a new equilibrium in response to a change in demand, and whether the path of adjustment to a new equilibrium will be monotonic and asymptotic or whether there will be oscillations and overshoot which may require temporary sectoral capacities in excess of those implied by the new equilibrium levels.

Chapter 4 contains the development of a model of the TPM's transient behavior in response to several types of demand changes, including a cyclical pattern which may be used to model response to seasonal demand cycles.

Chapter 2
Production Release Rule

A. The Need for a Release Rule

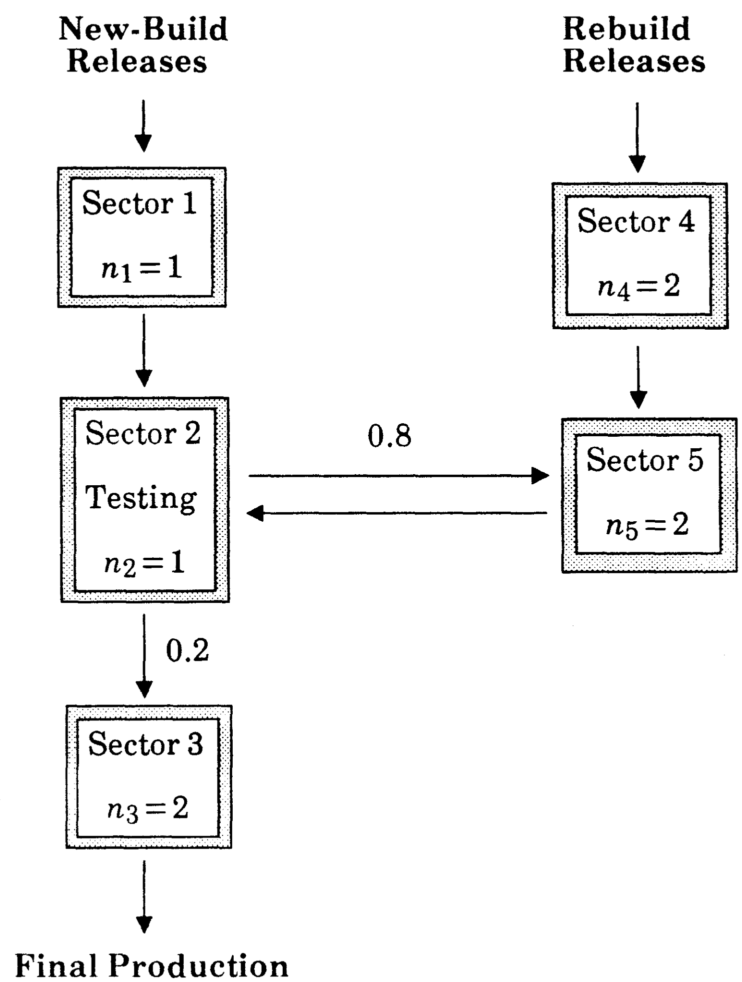
In applying the TPM to the ECM line, it was necessary to develop a model of the decision-making process which determines the number of new items to be started into the production line during each period. In terms of the model presentation in Chapter 1, we are looking for a way to characterize the arrivals to sector 1, a_{1t}, as seen in equation (1.4). The original specification of the TPM assumed that the requirement for releasing items into the line is independent of the shop status and is a result of new orders that arrive randomly to the shop. In the model in Chapter 1, these releases are modeled by setting N_{it} to the number of items (or hours of work, depending on the units used) started in sector i (equation (1.5)).

A production line which produces for inventory in response to forecasts of sales, rather than to specific orders, will determine its releases in a different fashion. Demand must be forecast and production must be started in relation to that forecast. The demand forecasts must be for a time horizon which is at least as long as the system's total lead time, which we denote by L, since an item will take that long to progress through the system. In determining today's releases from the demand forecast, the current WIP must be netted out from projected requirements since WIP consists of items which will be completed during the next L periods and are thus available to fill demand over the planning horizon. The excess of WIP over projected demand for these L periods is items that will be available, based on what is currently in the system, to meet demand in the L^{th} period in the future. Thus the release process should start into the line today only enough items to fill the shortfall between total demand forecast over the next L periods and current WIP in the line.

The above is a simplified description of the required release process. In the case of the ECM line, and most production systems, the decision-making process (and any model of it) is complicated by uncertainties in the demand forecasts and in the lead times which will be experienced by items in the current WIP and by those which are to be released. Uncertainty in lead times can arise from two major sources, only one of which is applicable in the TPM. First, the lead time for each sector may be a stochastic quantity. Since the TPM prescribes use of sectoral lead times as control variables, it is anticipated that management would place great emphasis on realizing planned lead times, and thus sectoral lead times would not vary greatly. Second, even in a system managed using the TPM's control rule, variability in total lead time may exist due to the presence of many alternate paths through the system. This source of uncertainty is present in the ECM line, and its essential nature is illustrated in Figure 2.1, which is a simplification of Figure 1.1. New-build items are released into sector 1. If they successfully pass the test in sector 2 (the probability is 0.2) they flow on to sector 3 and out of the line with a total lead time of four periods ($n_1 + n_2 + n_3$). If the test is failed (the probability is 0.8), they move to the rework sector 5 and, if they pass the test on return to sector 2, they exit the line with a lead time of seven periods ($n_1 + n_2 + n_5 + n_2 + n_3$). Some proportion (the same 80%), of course, may fail the test a second time and be sent back to sector 5 for additional rework, thus increasing those items' lead times. So, for an item currently located in WIP at sectors 1, 2, 4 and 5, its remaining lead time until exit is a random variable whose distribution depends upon the coefficients which describe the work flow from one sector to another.

Figure 2.1

Schematic of Flows in a Manufacturing Line



Another feature of the ECM line which must be taken into account is the difference between new-build and rebuild lead times and the impact of this difference on calculating WIP which will exit during the planning horizon. How this affects computation of releases is illustrated by the rebuild releases in Figure 2.1. It was decided to model rebuild releases as job-shop-type releases. That is, units are returned to the line for rebuilding according to a random process and are immediately released into the line. A rebuild item which passes the sector 2 test the first time has a total lead time of seven periods, whereas a new-build release which passes the test has a total lead time of four periods. Because the rebuild item has a longer total lead time, today's rebuild releases cannot be included in the WIP which is expected to meet demand over the immediate planning horizon of four periods. Of course, some of the WIP in the rebuild side can be expected to exit the line within the four period new-build lead time. For example, some fraction of the WIP at sector 5 will probably exit within this time and can thus be subtracted from demand forecast for the next four periods to determine today's new-build releases.

B. The Release Rule Model

With this general description of the release rule problem, let us now turn to development of the model. We build upon the equations of the TPM from Chapter 1. Our problem is to use the philosophy of computing new-build releases discussed above to characterize a_{1t}, the arrivals to sector 1 at time t, in accordance with the equations of the TPM. Specifically, we want to be able to write a_{1t} as (from equation (1.4)),

(2.1) $$a_{1t} = \sum_{j=1}^{m} \Phi_{1j}\, p_{j,\,t-1} + \varepsilon_{it,}$$

where m is the number of sectors in the line. Let m be the final sector from which finished items exit the line. We need to define Φ_{1j}, and we assume the random variables $\{\varepsilon_{1t}\}$ form a time series with i.i.d. elements.

We now define $R_t(u,v)$ as the forecast, made at time t, of requirements which must be met during periods $u, u+1, \ldots, v$. The change in forecast from period t-1 to t for the same interval $u, u+1, \ldots, v$ is $\zeta_t\,(u,v)$; that is,

(2.2) $$R_t(u,v) = R_{t-1}(u,v) + \zeta_t(u,v,).$$

We assume that $\zeta_t(u, v)$ is a random variable, and the elements of the time series $\{\zeta_t(u, v)\}$ are i.i.d., with mean zero. Let I_t be an inventory of finished goods which is governed by the inventory balance equation

$$I_t = I_{t-1} + p_{m,\,t-1} - R_{t-1}(t-1), \tag{2.3}$$

where $p_{m,\,t\text{-}1}$ is the output of finished goods and $R_{t\text{-}1}(t\text{-}1)$ is actual demand, both during period t-1.

Following the philosophy of release computation discussed in section A above, we see that releases to sector 1, a_{1t}, need to be calculated so that the WIP (including a_{1t}) which will exit within the planning horizon plus the inventory of final goods available at time t will meet the requirements over the horizon. The planning horizon at time t consists of periods t, $t+1$, ..., $t+L_N$, where L_N is the total lead time of all sectors in the shortest path through the new-build side. Note that period t is the period between time points t and $t+1$. A release into sector 1 at time t has a positive probability of exiting the line at time $t+L_N$ and thus of being available to satisfy demand during period $t+L_N$. Demand during period t can only be filled from the existing final goods inventory I_t. Demand during periods $t+1$, $t+2$, ..., $t+L_N$ can only be satisfied by the balance of I_t plus items which are currently in WIP and within L_N periods of exiting from the system.

Using the notation defined above, we see that $R_t(t, t+L_N)$ is the forecast at time t of demand in periods t, $t+1$, ..., $t+L_N$. Note that this includes the forecast for demand in period t, $R_t(t)$, which we assume to be known with certainty. We let θ_i denote the probability that an item in WIP at sector i will yield a usable final product within L_N periods (i.e. by time $t+L_N$). Thus $\theta_i q_{it}$ is the expected number of items currently in WIP at sector i which will exit during the planning horizon, and the total number of finished items expected to be available over the horizon is the current I_t plus the sum over all sectors of $\theta_i q_{it}$. That is, we must set releases, which are part of q_{it}, so that this relationship is satisfied:

(2.4)
$$I_t + \sum_{i=1}^{m} \theta_i q_{it} = R_t(t, t + L_N) .$$

We rewrite this and the similar expression for t-1 as:

$$\sum_{i=1}^{m} \theta_i q_{it} = R_t(t, t+L_N) - I_t$$

$$\sum_{i=1}^{m} \theta_i q_{i,t-1} = R_{t-1}(t-1, t+L_N-1) - I_{t-1} .$$

Subtracting the second from the first gives:

$$\sum_{i=1}^{m} \theta_i (q_{it} - q_{i,t-1}) = R_t(t, t+L_N) - R_{t-1}(t-1, t+L_N-1) - I_t + I_{t-1} .$$

We recall the following:

$$q_{it} - q_{i,t-1} = a_{it} - p_{i,t-1} \qquad (\textit{from}\ (1.2)),$$

$$R_t(t, t+L_N) - R_{t-1}(t-1, t+L_N-1) = R_t(t+L_N) + \zeta_t(t, t+L_N-1) - R_{t-1}(t-1),$$

$$I_{t-1} - I_t - R_{t-1}(t-1) = -p_{m,t-1} \qquad (\textit{from}\ (2.3)) .$$

Substituting these into the difference we have:

(2.5)
$$\sum_{i=1}^{m} \theta_i (a_{it} - p_{i,t-1}) = R_t(t+L_N) + \zeta(t, t+L_N-1) - p_{m,t-1} .$$

Operating on the left-hand side, we remove $\theta_1(a_{1t} - p_{1,\,t-1})$ from the summation and substitute from equation (1.4) for a_{it}, $i=2, ..., m$:

$$(\theta_1 a_{1t} - \theta_1 p_{1,t-1}) + \sum_{i=2}^{m} \theta_i \left(\sum_{j=1}^{m} \phi_{ij} p_{j,t-1} - p_{i,t-1} + \varepsilon_{it} \right) .$$

Now solving this for a_{1t}, we get:

(2.6)
$$a_{1t} = \frac{1}{\theta_1} \left[R_t(t+L_N) + \zeta_t(t, t+L_N-1) - p_{m,t-1} + \theta_1 p_{1,\,t-1} \right.$$

$$\left. + \sum_{i=2}^{m} \theta_i p_{i,\,t-1} - \sum_{i=2}^{m} \theta_i \sum_{j=1}^{m} \phi_{ij} p_{j,\,t-1} - \sum_{i=2}^{m} \theta_i \varepsilon_{it} \right] .$$

Comparing this to (2.1) we can solve for ε_{1t} and the coefficients ϕ_{1j} in terms of the other parameters and variables, as:

(2.7)
$$\varepsilon_{1t} = \frac{1}{\theta_1}\left[R_t(t+L_N) + \zeta_t(t, t+L_N-1) - \sum_{i=2}^{m} \theta_i \varepsilon_{it}\right]$$

(2.8)
$$\phi_{1j} = \frac{1}{\theta_1}\left[\theta_j - \sum_{i=2}^{m} \theta_i \phi_{ij}\right] \qquad j=1, \ldots, m-1,$$

(2.9)
$$\phi_{1m} = \frac{1}{\theta_1}\left[\theta_m - \sum_{i=2}^{m} \theta_i \phi_{im} - 1\right].$$

If there are several sectors from which finished items exit the line, as there indeed may be in a job shop, there will be several sectors m whose ϕ_{1m} is expressed in the form of (2.9).

The release rule of interest is equation (2.7). This says that releases into sector 1 need to consist of three components: first, the current forecast of demand for period $t+L_N$; second the current change in forecast for the current and future periods t, $t+1$, ..., $t+L_N-1$; third, we subtract the items which will exit within the planning horizon that may be generated by external injections to other sectors. Finally, the resulting quantity must be increased to account for the fact that a proportion $(1-\theta_1)$ of the releases into sector 1 will not complete processing within the horizon.

In order to use equation (2.7) in implementing the TPM, one will want to take the expectation in order to determine μ_1, the first component of the vector μ of equation (1.9). Under the assumptions that $E(\zeta_t)=0$ and $E(\varepsilon_{it})=\mu_i$ $(i \neq 1)$, we have:

(2.10)
$$E(\varepsilon_{1t}) = \mu_1 = \frac{1}{\theta_1}\left(\overline{R} - \sum_{i=2}^{m} \theta_i \mu_i\right)$$

where we have assumed also that the requirements process is stationary, so that

$$E(R_t) = \overline{R}, \qquad (t= \ldots, -1, 0, 1, \ldots).$$

C. Calculation of Exit Probabilities

In order to use (2.7), (2.8) and (2.9) in application of the TPM we must calculate the parameters θ_i, the probability that an item in WIP at sector i will yield a usable finished output within the planning horizon given by the total lead time, L_N.

We calculate θ_i by first defining a bivariate state (i, t) for an item in WIP, where the element i identifies the sector in which the item is currently located and the element t identifies the length of time (in periods) that the item has been in sector i. For example, suppose sector 3 has a four day lead time. To say that an item is in state (3,2) means that it is in its second day in sector 3. In other words, it has, on the average, two more days to spend at sector 3 before moving on to a subsequent sector.

Employing the Markov assumption in Graves' identification of the intersectoral flow rates, ϕ_{ij}, we define a Markov transition matrix Ω, whose elements are given by:

$$\omega_{(i,s)(j,t)} = \begin{cases} 1 \text{ if } i=j \text{ and } s=t+1 \\ 0 \text{ if } i=j \text{ and } s\neq t+1 \\ \lambda_{ij} \text{ if } i\neq j \text{ and } t=n_j\ , s=1 \end{cases}$$

Here λ_{ij} is the probability that an item leaving sector j moves to sector i. Note that the matrix Ω is the transpose of the more conventional transition matrix: the columns are source states and the rows are destination states. The matrix is presented in this transposed form simply to be consistant with Graves' definition of the Φ matrix.

The transition probabilities $\omega_{(i,\ s)\ (j,\ t)}$ have the following meanings. If an item is in a sector's WIP and has been there less than the sector's planned lead time, then at the end of a period it makes a transition to the same sector and to a time slot one period closer to exit from the sector. If an item is currently in the last period of a sector's lead time, then its transition is to another sector according to the flow possibilities identified in the Φ matrix. It cannot move to another sector until it has been in the current sector for the planned lead time.

To complete the formation of the matrix Ω, we must add one last row and column with zeros in all elements except a 1 in the southwest element. This represents a final absorbing state indicating an item has exited from the system. The Ω matrix is now square, of dimension

$$\sum_{i=1}^{m} n_i + 1 .$$

Now Ω is a transition probability matrix for a Markov chain with one absorbing state (namely exit from the system). By computing powers of Ω, i.e. Ω^{L}, we can find the probability of exit from the system from each state for any number of transitions L.

In applying the model to the ECM line, the probabilities λ_{ij} were given directly by the elements ϕ_{ij} of the Φ matrix, since production was defined in terms of units of product and the elements of the Φ matrix could be interpreted as probabilities. This correspondence will not be true in general when using the TPM and the λ_{ij} will have to be estimated in some other fashion. This is because, in a different application, the production of each sector, p_{it}, may be measured in hours of work, and the ϕ_{ij} elements will denote the hours of work generated at subsequent sectors by an hour of work completed in sector j.

The next step in calculating the probabilities θ_i is to raise the Ω matrix to the power L_N. The cells of the last row will accumulate the probabilities that an item which started in the state of that column will have exited within L_N periods. Finally, θ_i is the average of the entries in the cells representing the different time periods in sector i. The averaging can be accomplished by postmultiplying the last row of the of Ω^{L_N} matrix by a

$$\left(\sum_{i=1}^{m} n_i + 1\right) \; x \; m \; matrix$$

whose entries are the appropriate averaging elements.

D. Example of Exit Probability Calculations

We illustrate the calculations of θ_i using the data from Figure 2.1. The system states are (1,1), (2,1), (3,1), (3,2), (4,1), (4,2), (5,1), (5,2) and Exit. An item in (3,2), for example, has already spent one period in sector 3.

The transition matrix Ω, with states in the above order, is

$$\Omega = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Recall that columns represent source states and the rows represent destination states, so the 0.2 in row 3, column 2 is the 0.2 probability of transition from state (2,1) to state (3,1), and the 0.8 in column 2 is the 0.8 probability of transition from state (2,1) to state (5,1).

To calculate θ_i, we first raise Ω to the fourth power since $L_N = 4$.

$$\Omega^4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.8 & 0 & 0 & 0 & 1 & 0 & 0 & 0.8 & 0 \\ 0 & 0.16 & 0 & 0 & 0 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.64 & 0 & 0 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 0 \\ 0.2 & 0.2 & 1 & 1 & 0 & 0 & 0 & 0.2 & 1 \end{bmatrix}$$

Now we take the final row of Ω^4 and multiply it by a matrix that averages the probabilities over the states belonging to each sector

$$\begin{bmatrix} 0.2 \\ 0.2 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0.2 \\ 1 \end{bmatrix}' \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \theta = (0.2, 0.2, 1.0, 0, 0.10)$$

(The ' on the column vector above indicates the transpose.)

So we see that the probability that an item currently located in sector i will exit the line within the next four periods is

i	Probability of Exit
1	0.2
2	0.2
3	1.0
4	0
5	0.10

E. Conclusion

By following a similar procedure, exit probabilities may be calculated for any system, and then (2.7) may be used to calculate the required releases into sector 1. Finally, (2.8) and (2.9) can be applied to calculate the elements which must be inserted in the first row of the Φ matrix. Once this has been accomplished equations (1.9), (1.10), (1.12) and (1.13) may be used to calculate the behavior of the system.

Chapter 3

Service Measures

A. The Need for Service Measures

The Tactical Planning Model gives management insight into the choice of sector lead times by calculating the impact on two key aspects of line performance: production variances and size of WIP. These performance characteristics are of great interest to management since they are indicative of some of the major costs of running the system, and the TPM gives insight into the trade-offs available among the associated costs. Management is, of course, also interested in the service provided by the system as measured by the extent to which it produces the right quantities of finished goods at the correct times. The choice of lead times affects this responsiveness to demand in addition to the cost trade-offs, but the TPM as presented in Chapter 1 provides no insight into this effect. The purpose of this chapter is to enhance the TPM by adding a model of the impact of lead time choices on service.

The fact that lead time choices can affect service is easily seen by deriving from equations (1.1) and (1.2) the relationship between the production levels at a sector in two successive time periods. Substituting from (1.2) into (1.1), recognizing that $p_{i,\,t\text{-}1} = \xi_i\, q_{i,\,t\text{-}1}$ and that $\xi_i = 1/n_i$, we see that production of two successive periods is positively correlated, and that the correlation will increase with n_i:

$$p_{it} = \left(1 - \frac{1}{n_i}\right) p_{i,t-1} + \frac{1}{n_i} a_{it} . \tag{3.1}$$

If n_i is 1, then all arrivals at time t are produced during period t, there is no smoothing of production, and production in successive periods is uncorrelated. If n_i is very large, only a small portion of periods t's arrivals are produced during period t, there is much production smoothing, and the correlation between production in successive periods approaches perfect positive correlation. Since these comments apply to any sector, the implication for the system's response to demand can be seen by considering the final sector m, out of which finished products flow. Positive correlation among the terms of the time series $\{p_{mt}\}$ implies that if output is low during period t, the probability of low output during period $t+1$ is increased. This implies that increasing n_m may have the effect of creating runs, consisting of several consecutive periods of low output. A run of low output may generate a run of several periods in which the system fails to meet the demands.

In a multi-sector system the lead time at other sectors will also affect the ouput correlation. The matrix equation (1.6) shows that the period-to-period correlation of production at every sector in the system will be affected to some degree by the lead times of all sectors. For example, if in the ECM line the lead time of sector 1 is increased while lead times at all other sectors (including the output sector, 13) are left at $n_i = 1$, some positive correlation will be introduced into the output.

The rest of this chapter is organized into six sections. In section B we give a conceptual definition of the service level of a system. In C, we pose a simplified model of a manufacturing system which will be used to develop the service measures. Sections D, E, F and G contain the development of the service level measures, some numerical examples and some numerical indications which can be generalized to other manufacturing facilities. Finally, in section H we discuss the implications for management of a shop, taking into account effect of lead time choices on the many aspects of system behavior.

B. Definition of Service Level

There are many facets to the concept of service provided by a facility and many different causes of good or bad performance in each of these areas. In this section we clarify and limit the definition of service level to the two measures with which we will be concerned.

The most obvious basic measure of service is the probability with which the system will meet the requirements during a period from the total of that period's production and the available inventory of finished goods. We will use this measure and call it the "shipping level",

(3.2) $$S = Pr\{\textit{demand is met in a random period}\}.$$

If $S = 75\%$, then the probability that the total number of units required to be shipped on a random day are available is 75%. The other 25% of the time, less than the full number of required units are available and, presumably, a partial shipment is made. We do not attempt to capture in S (or any other measure), the extent to which the system falls short of meeting requirements in those periods in which shortfalls occur.

Secondly, recognizing the fact that there will be positive correlation in the output time series, we wish to have some measure of failure run length, where a "failure run" refers to several successive periods during which the system fails to meet demand. Management will have a keen interest in failure run lengths since they will usually be interested in assuring that the line will fall behind schedule for no more than a specified number of periods. A facility which provides adequate service as measured by the shipping level S may still occasionally run behind schedule for an unacceptably long time. Management will want to avoid choosing lead times which might cause this behavior.

We define the random variable F_t as the length of a failure run whose first period is t. We would like to be able to determine the distribution of failure run lengths, given by

(3.3) $$Pr\{F_t = i\}, \qquad i = 0, 1\ 2, \ldots$$

The impact of output correlation on system service is seen directly when we consider how we might compute the distribution of run lengths. If successive period's outputs are independent random variables, and S is the probability that demand is met, then F_t has the geometric distribution, given by

(3.4) $$Pr\{F_t = i\} \quad = \quad (1\text{-}S)^i S\,, \qquad i = 1, 2, \ldots\ .$$

Because the output is correlated rather than independent, however, the conditional probability,

$$Pr\,\{\textit{Demand is met in period } t \mid \textit{demand was not met in periods } t\text{-}1, t\text{-}2, \ldots\}$$

is not the same as the unconditional probability S, and (3.4) is not a correct statement for the failure run distribution. If each probability S in the product (3.4) were replaced by a conditional probability, each conditioned upon the entire past history of the system, and if (3.4) were calculated over every possible past state of the system, then a failure run length distribution could be generated. Performing this calculation requires knowledge of an infinite number of conditional probabilities and calculation of products (3.4) over an infinite number of possible paths of the system. Clearly this is an intractable problem. Thus the existence of output correlation first introduces a new concern by creating the possibility of nonrandom failure runs and then confounds the situation by preventing us from calculating a distribution of run lengths.

Even though we cannot develop a complete distribution of failure run lengths, we will discover that we can derive the expected value of run lengths. We will do this in sections E and F.

We must comment on one other source of discrepancy between production and demand which is encountered in a manufacturing system. In a system in which production decisions are based on forecasts of demand quantity and timing, incorrect forecasts will result in lower service levels than would perfect forecasts. In our treatment, we wish to focus attention only on the discrepancies which may be imparted by the structure of the system and the correlation generated within it, so we assume perfect demand forecasts. The effect of this assumption is that production, on the average, equals the average demand, and that the discrepancies which appear are ones of timing of output in relation to demand and are not created by incorrect forecasts.

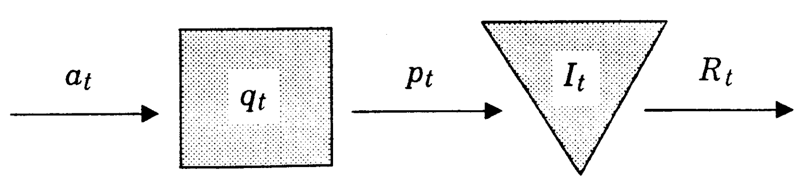
C. Simplified Model of a Manufacturing System

It is convenient to work with a simple model of a one sector system rather than the multi-sector TPM. It is possible to use equations (1.6), (1.7) and (1.8) of the TPM to solve for the covariance of p_{mt} over time as a function of all sector lead times, but the resulting equations are extremely cumbersome and hard to interpret. The following one sector model captures the essential nature of output correlation which occurs in the full multi-sector TPM, but is much easier to interpret. Thus we believe this step is a reasonable modeling simplification.

We represent a manufacturing system as one sector and an inventory of finished goods, as diagrammed in Figure 3.1. New work consisting of a_t items are released into the system each period and then go directly into the work-in-process inventory q_t. Production control is established using the TPM control

Figure 3.1

The One Sector System



rule, and $1/n$ of the items in q_t are produced each period, where n is management's choice of lead time. Finished goods are placed in an inventory, I_t, from which items are drawn to meet the period's demand. The equations describing the flow of production and inventory balances are:

(3.5) $$p_t = \frac{1}{n} q_t,$$

(3.6) $$q_t = q_{t-1} + a_t - p_{t-1}$$

(3.7) $$I_t = I_{t-1} + p_{t-1} - R_t$$

where R_t is the demand or requirements for the period.

We assume that each period's demand is a random variable, independent of demands in other periods, and that it is normally distributed with mean $\bar{R}$ and standard deviation σ_R. As discussed above, we assume that releases into the system are governed by a perfect forecast of demand, so that $a_t = R_t$, and a_t has the same distribution as R_t. It might seem more natural to assume releases are a forecast of future demand, $a_t = R_t(t+h)$, where $R_t(t+h)$ is the forecast made at time t of demand in period $(t+h)$. This refinement can be included in the model, but it simply complicates the analysis with no essential effect on the outcome, so we do not employ it.

We now substitute (3.6) into (3.5) and use the facts that $p_{t-1} = (1/n)(q_{t-1})$ and $a_t = R_t$ to write p_t as a function of p_{t-1}:

(3.8) $$p_t = \left(1 - \frac{1}{n}\right) p_{t-1} + \frac{1}{n} R_t.$$

Here we can see the essential nature of the inter-period correlation we discussed above.

By repeated substitution of (3.8) into itself and by assuming an infinite history of the system, we write p_t as a weighted sum of all past requirements:

(3.9) $$p_t = \sum_{s=0}^{\infty} \left(\frac{n-1}{n}\right)^s \frac{1}{n} R_{t-s}.$$

Since the R_t ($t=$ ... -1, 0, 1, ...) are independent normally distributed random variables, the p_t will also be normally distributed with expectation and variance

(3.10) $$E(p_t) = E(R_t) = \bar{R}.$$

(3.11) $$Var(p_t) = \left(\frac{1}{2n-1}\right)\sigma_R^2$$

In (3.9), we see again confirmation of the fact that the p_t are not independent random variables since each is a weighted sum of all past demands.

In order to answer questions about the system's service level, we focus on the final goods inventory, I_t, and its distribution. If $I_t<0$, then the requirements in period t were not met and some of period t's demand is backlogged. If $I_t<0$ for a sequence of periods, then we have a failure run. We can now quantify the service measures we discussed in generalities in section B in terms of the probability distribution of I_t:

(3.12) $$S = Pr\{\textit{demand is met in a random period}\} = Pr\{I_t \geq 0\},$$

(3.13) $$Pr\{F_t=i\} = Pr\{I_{t-1}\geq 0,\ I_t<0,\ I_{t+1}<0,\ \ldots,\ I_{t+i-1}<0,\ I_{t+i}\geq 0\}.$$

We now can express the moments of I_t and its covariance as functions of the moments of the R_t distribution. Note that the total inventory in the system, q_t+I_t, is constant for all t. This can clearly be seen by adding (3.6) and (3.7) and using the fact that $a_t = R_t$, which gives

(3.14) $$q_t + I_t = q_{t-1} + I_{t-1} = K.$$

We denote this constant inventory level by K and call it the "base stock".

It will be useful to call the components of base stock, q_t and I_t, the "in-process stock" and "safety stock" respectively. The size of q_t is related directly to the production process and the smoothing requirement dictated by management's choice of n, as can be seen from (3.5). The size of I_t is related, as we shall see later, to management's choice of S, the probability that demand is met. We can now express I_t as a function of n and the requirements by substituting for q_t in $I_t = K - q_t$, using (3.5) and (3.9):

(3.15) $$I_t = K - \sum_{s=0}^{\infty} \left(\frac{n-1}{n}\right)^s R_{t-s} .$$

Since I_t is a linear combination of the independent normally distributed random variables R_t, I_t is itself normally distributed with expectation and variance given by

(3.16) $$E(I_t) = K - n\bar{R} ,$$

(3.17) $$Var(I_t) = \left(\frac{n^2}{2n-1}\right)\sigma_R^2 = \sigma_I^2 .$$

As is true for p_t, successive values of I_t are not independent but are correlated with

(3.18) $$Cov(I_t, I_{t+h}) = \left(\frac{n-1}{n}\right)^{|h|}\sigma_I^2 .$$

With this information about I_t and its distribution, we can now turn our attention to describing how the shipping level S and failure run lengths are affected by n.

D. Shipping Level

Our first measure of system service is the shipping level

$$S = Pr\ \{demand\ is\ met\ in\ a\ random\ period\},$$

defined by (3.12).

Management can adjust system performance to any desired level S simply by selecting a value of base stock K to satisfy

$$K = n\bar{R} + \Psi^{-1}(S)\sigma_I \tag{3.19}$$

$$= n\bar{R} + \Psi^{-1}(S)\left[\sigma_R\sqrt{\frac{n^2}{2n-1}}\right]$$

where $\Psi(\cdot)$ is the cumulative standard normal distribution.

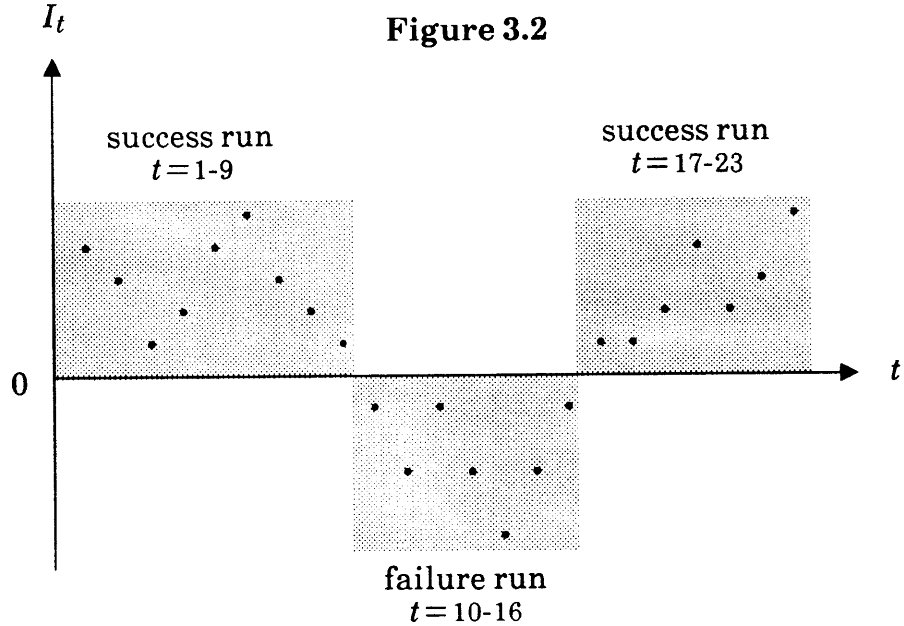
The first part of (3.19), $n\bar{R}$, is the in-process stock and is determined entirely by n. The second term, the safety stock, is that which is influenced by a choice of S. The safety stock is the amount of finished goods kept in inventory to protect against period-to-period fluctuations in demand. For example, at $S=50\%$, $\Psi^{-1}(S)=0$, thus requiring no safety stock. With no safety stock, $K=n\bar{R}$ so $E(I_t)=0$ and, since I_t is distributed normally, $Pr\{I_t<0\} = Pr\{I_t\geqq 0\} = 0.5$, verifying the choice of $S=50\%$. To establish the shipping level higher than 50%, (3.19) simply says that the safety stock needed is some multiple of the standard deviation of I, where the multiple is a function of S.

E. Failure Runs and Expected Crossings

The shipping level S is certainly an important measure of system performance but, as we have seen above, S does not address the problem of positive correlation in p_t and I_t. Even when S is set at a high level, since the conditional probability $Pr\ (I_t\geqq 0|I_{t-1}<0)<S$, the system may generate failure runs of greater length than desired by management.

Previously we pointed out that we cannot develop a distribution of run lengths. However we can derive an expression for expected run lengths by focusing on the associated level-crossing problem. The random variable I_t describes a sample path through time and occasionally crosses the level $I_t=0$. As long as I_t remains below $I_t=0$, a failure run is in process, while a success run is in process when $I_t\geqq 0$ for several successive periods. Figure 3.2 shows a possible sample path. In this figure we see that $I_t \geqq 0$ for $t=1$ through $t=9$ and for $t=17$ through $t=23$. These periods are success runs of various lengths. A failure run occurs from $t=10$ through $t=16$. Note that the number of times the path of I_t crosses the zero level during the interval shown is related to the average run length. We will exploit this property to derive expected run lengths.

Figure 3.2



Two facts about the stochastic process formed by the sequence of random variables I_t ($t=\ldots, -1, 0, 1, \ldots$) are important. First we note that it is a covariance stationary (or weakly stationary) stochastic process as it satisfies the three requirements for such a process: (1) finite second moment, (2) constant mean, and (3) a covariance function which depends on the lag between two elements but not on the actual value of the time index [Cramer and Leadbetter, pg. 121]. These properties can be verified from (3.16), (3.17) and (3.18). Second, since the elements I_t are normally distributed we can state the stronger result, that I_t is a strictly stationary process [Cramer and Leadbetter, pg. 123].

We define a new stochastic process by the random variable $C(s, t]$, which is the number of times I_t crosses the level zero in the half-closed interval $(s, t]$. In the next section we will prove that $C(s, t]$ is a stationary point process and we will find an expression for the expected number of crossings in a unit interval, $E[C(t, t+1]]$, which we denote as $E(C)$ for simplicity.

Associated with the I_t process are three processes of runs, which we define by the random variables W_t = length of a run beginning at time t, V_t = length of a success (think Victory) run beginning at time t, and F_t = length of a failure run beginning at time t.

Since $C(t, t+h]$ is a stationary point process, as $h \to \infty$ the average run length will approach the inverse of the average number of crossings which occur during h periods [Cox and Miller, pgs. 356-358]. So we will use $E(C)$ to find $E(W)$ by

$$E(W) = \frac{1}{E(C)}. \tag{3.20}$$

Once we have $E(C)$ and $E(W)$, we will be able to find expressions for the expected failure run length $E(F)$ and the expected success run length $E(V)$, as follows. Since as $h \to \infty$ the numbers of success runs and failure runs are equal,

$$E(W) = \frac{1}{2}[E(V) + E(F)]. \tag{3.21}$$

Also, the probability, S, of a successful period will equal the proportion of time the system is experiencing success,

$$S = \frac{E(V)}{E(F)+E(V)} \tag{3.22}$$

From (3.20), (3.21) and (3.22) we can solve for $E(F)$ and $E(V)$:

$$E(V) = \frac{2S}{E(C)}, \tag{3.23}$$

$$E(F) = \frac{2(1-S)}{E(C)}. \tag{3.24}$$

F. Formula for Expected Number of Crossings

The literature on stationary processes contains an expression for the expected number of crossings of a level for a continuous process with a covariance function that has a finite second derivative at $h=0$ [Cramer and Leadbetter, pg. 194; Karlin and Taylor, pg. 522]. Unfortunately, this expression does not work for the I_t process since it is a discrete process and its covariance function does not satisfy this property.

Although the formula itself does not apply, we can adapt the process used by Cramer and Leadbetter to develop the formula to produce an expression which does apply to the I_t process. The following development is most closely patterned after Karlin and Taylor's report of the derivation [Karlin and Taylor, pgs. 510-522].

Let I_t be a discrete time stochastic process as defined above with expectation, variance and covariance function given by (3.16), (3.17) and (3.18). Let $R_t (t= \ldots -1, 0, 1, \ldots)$ be distributed normally with mean $\bar{R}$, variance σ^2_R and $Cov(R_t, R_{t+h}) = 0$. Let $C(s, t]$ be an integer-valued random variable equaling the number of times I_t crosses the level zero in the interval $(s, t]$. In particular, we will denote $C(t, t+1]$ (crossings in a unit interval) by C alone.

First we prove that $C(s, t]$ is a stationary point process:

Theorem 1: Under the above assumptions $C(s, t]$ is a stationary point process. That is, the k-dimensional vector

$$\{C(s_1, t_1], \ldots, C(s_k, t_k]\}$$

has the same joint distribution as the vector

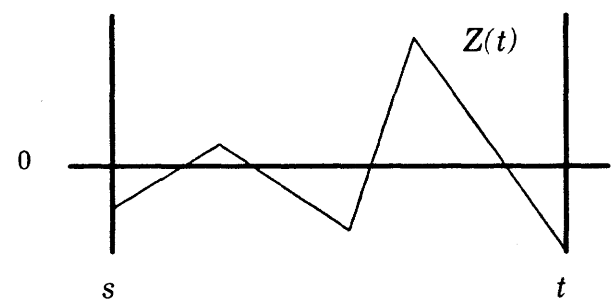
$$\{C(s_1+h, t_1+h], \ldots, C(s_k+h, t_k+h]\}$$

for every set of intervals $(s_1, t_1], \ldots, (s_k, t_k]$, for every positive integer k, and every integer h.

Proof: We prove this by assuming the I_t process is generated by sampling from a continuous time Gaussian process, whose counting process is known to be a stationary point process. The elements of the $C(s, t]$ process are then shown to be equivalent to the elements of this stationary point process, proving the desired result.

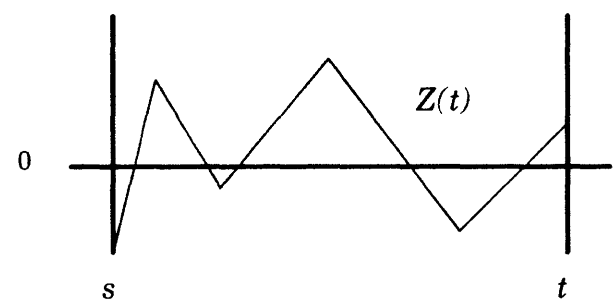
Assume $Z(t)$ is a continuous time Gaussian process having continuous sample functions. We sample the process only at discrete times, $t = 0, 1, 2, \ldots$, yielding the derived discrete time Gaussian process I_t. Define $N(s,t]$ = the number of times the trajectory of $Z(t)$ crosses the level $Z(t) = 0$ in the interval $(s, t]$.

Example 1



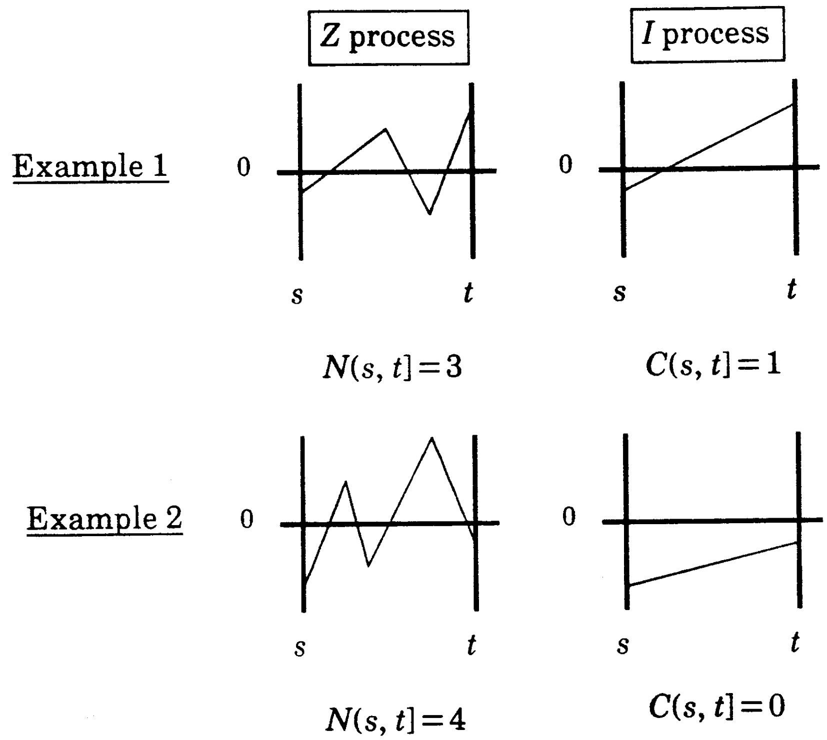
$N(s, t] = 4$

Example 2



$N(s, t] = 5$

Let $C(s, t]$ = the number of times the process I_t crosses $I_t = 0$ in the interval $(s,t]$, where a crossing occurs whenever $I_s > 0 > I_t$ or $I_s < 0 < I_t$. Note that if $C(s,t] = 1$, then $N(s, t] \geqq 1$ and is odd. If $C(s, t] = 0$, then $N(s, t] = 0$ or $N(s, t] > 0$ and is even. The relationship between $N(s, t]$ and $C(s, t]$ is shown in these examples:



Since $Z(t)$ is a Gaussian process, it is a stationary process, and $N(s, t]$ is then a stationary point process [Karlin and Taylor, pg. 516]. Now the event $\{C(s, t]=1\}$ in the I_t process occurs if and only if the event $\{N(s, t]=2i+1,\ i=0, 1, 2, \ldots\}$ occurs in the $Z(t)$ process. From this one-to-one equivalence of events and from the fact that $N(s, t]$ is a stationary point process, it follows that $C(s, t]$ is one also. This completes the proof.

Now we are ready to prove the main theorem which defines the expression for $E(C)$:

__Theorem 2__: Under the assumptions stated for Theorem 1, the expected number of crossings of the level zero in a unit interval by the I_t proccess is

$$(3.25)\qquad E(C) = \frac{\sqrt{2n-1}}{n\pi}\int_0^\infty \int_{-w}^0 exp\left\{-\frac{1}{n}\left[\left(x-\Psi^{-1}(S)\sqrt{n^2/(2n-1)}\right)^2 + w\left(x-\Psi^{-1}(S)\sqrt{n^2/(2n-1)}\right) + n\,w^2/2\right]\right\}dx\,dw$$

Proof: Define events $A = \{I_0 < 0 < I_1\}$ and $B = \{I_0 > 0 > I_1\}$. A crossing occurs if either A or B occurs.

Define a new random variable Y as the increment in I_t from one period to the next:

(3.26) $$Y = I_1 - I_0 .$$

Since I_t is normally distributed, Y is also normally distributed with

(3.27) $$E(Y) = 0,$$

(3.28) $$Var(Y) = \frac{2}{n}\left(\frac{n^2}{2n-1}\right)\sigma_R^2,$$

(3.29) $$Cov(Y, I_0) = -\left(\frac{n}{2n-1}\right)\sigma_R^2$$

The statements for variance and covariance are developed as follows:

$$\begin{aligned} Var(Y) &= Var(I_1) + Var(I_0) - 2Cov(I_1, I_0) \\ &= 2\sigma_I^2 - 2\,Cov(I_1, I_0) \\ Cov(Y, I_0) &= Cov(I_1 - I_0, I_0) \\ &= Cov(I_1, I_0) - \sigma_I^2 \end{aligned}$$

Substituting for σ^2_I and $Cov(I_1, I_0)$ from (3.17) and (3.18) gives (3.28) and (3.29).

The pair (I_0, Y) has a bivariate normal distribution with means $(K\text{-}n\bar{R}, 0)$ and covariance matrix:

(3.30) $$\Gamma = \sigma^2_R \begin{bmatrix} \dfrac{n^2}{2n\text{-}1} & -\dfrac{n}{2n\text{-}1} \\[2ex] -\dfrac{n}{2n\text{-}1} & \dfrac{2n}{2n\text{-}1} \end{bmatrix}$$

The correlation coefficent between Y and I_0 is

(3.31)
$$Corr(I_0, Y) = \frac{Cov(I_0, Y)}{\sqrt{Var(I_0)}\sqrt{Var(Y)}} = -\frac{1}{\sqrt{2n}}$$

The joint normal probability density function for (I_0, Y) is

(3.32)
$$f_{I_0, Y}(I, y) = \frac{1}{2\pi\sigma_R^2 \sqrt{\frac{n^2}{2n-1}} \sqrt{\frac{2n}{2n-1}} \sqrt{1-\frac{1}{2n}}}$$

$$\exp\left\{ -\frac{1}{2\left(1-\frac{1}{2n}\right)} \left[\left(\frac{I-\bar{I}}{\sigma_R\sqrt{\frac{n^2}{2n-1}}} \right)^2 + \left(\frac{y}{\sigma_R\sqrt{\frac{2n}{2n-1}}} \right)^2 \right. \right.$$

$$\left. \left. + \frac{2}{\sqrt{2n}} \frac{(I-\bar{I})y}{\sigma_R^2\sqrt{\frac{n^2}{2n-1}}\sqrt{\frac{2n}{2n-1}}} \right] \right\}$$

In order to use this joint density function to calculate $E(C)$, we transform the events A and B, defined above, into the (I_0, Y) sample space by substituting $I_1 = Y + I_0$:

$$A = \{I_0 < 0 < I_1\} = \{0 > I_0 > -Y\}, \quad \text{and}$$
$$B = \{I_0 > 0 > I_1\} = \{0 < I_0 < -Y\}.$$

Remembering that a crossing occurs if either A or B happens, we see that $Pr\{C(0,1] = 1\} = Pr\{A\} + Pr\{B\}$ and thus $E(C) = Pr\{A\} + Pr\{B\}$, since the random variable $C(0, 1]$ takes only values 0 or 1. Now $Pr\{A\}$ and $Pr\{B\}$ may be calculated by integrating (3.32) over the appropriate regions of the (I_0, Y) sample space. Since the bivariate normal distribution is symmetric around the origin and any line $I_0 = bY$ (b any scalar), $Pr\{A\} = Pr\{B\}$ so we need to integrate over one region only. Integrating over event A, we have

(3.33)
$$E(C) = 2Pr\{A\} = 2\int_0^{\infty}\int_{-y}^{0} f_{I_0, Y}(I, y)\, dI\, dy.$$

Now, if we insert the density function (3.32) into the integral, use the fact (from (3.16) and (3.19)) that $\bar{I} = \Psi^{-1}(S)\, \sigma_R \sqrt{n^2/(2n-1)}$, and make a change of variables by letting $x=\sigma_R I$ and $w=\sigma_R y$, we develop (3.25). This completes the proof of Theorem 2.

The integral in (3.25) gives the expected number of crossings in a unit interval (0, 1]. From the fact that $C(s, t]$ is a stationary point process, it follows that the expected number of crossings in any interval $(t, t+1]$ is $E(C)$. It also follows that the expected number of crossings in a larger interval $(t, t+h]$, is simply the interval size h multiplied by $E(C)$.

G. Evaluation of Expected Number of Crossings and Inferences About Run Lengths

We now use (3.25) to compute $E(C)$ and expected run lengths and to evaluate the impact of lead time choices on service.

First we remark on the surprising fact that σ_R does not appear in (3.25). Our intuitive expectation is that $E(C)$ would vary directly with σ_R. That it does not is due to the fact that σ_R has two opposing effects on $E(C)$: first σ_R affects σ_I, which moves directly with $E(C)$, and second σ_R affects $Cov(I_t, I_{t+1})$, which moves inversely with $E(C)$. Apparently the contributions of σ_R to these effects exactly offset one another. Since σ_R does cancel out, $E(C)$ depends only on n and S and not any parameters particular to a facility. Thus the following discussion and numerical results have general applicability to all manufacturing shops run in accordance with the TPM's control rule.

We computed values of $E(C)$ and expected run, success run and failure run lengths for $S=50\%$, 75%, 90% and 95% and for values of n from 1 to 10. Table 3.1 displays the results, Figure 3.3 is a graph of expected failure runs versus n, and Figure 3.4 is a graph of expected success runs versus n.

Table 3.1

Expected Run Lengths

Lead Time n	$S=50\%$ Expected Run Length	$S=50\%$ Expected Success Run	$S=50\%$ Expected Failure Run	$S=75\%$ Expected Run Length	$S=75\%$ Expected Success Run	$S=75\%$ Expected Failure Run
1	2.00	2.00	2.00	2.67	4.01	1.34
2	3.00	3.00	3.00	3.88	5.82	1.94
3	3.74	3.74	3.74	4.77	7.16	2.39
4	4.35	4.35	4.35	5.55	8.32	2.77
5	4.88	4.88	4.88	6.20	9.31	3.10
6	5.36	5.36	5.36	6.81	10.22	3.41
7	5.81	5.81	5.81	7.36	11.05	3.68
8	6.22	6.22	6.22	7.86	11.79	3.93
9	6.60	6.60	6.60	8.37	12.55	4.18
10	6.96	6.96	6.96	8.80	13.20	4.40

Lead Time n	$S=90\%$ Expected Run Length	$S=90\%$ Expected Success Run	$S=90\%$ Expected Failure Run	$S=95\%$ Expected Run Length	$S=95\%$ Expected Success Run	$S=95\%$ Expected Failure Run
1	5.52	9.94	1.10	10.52	19.98	1.05
2	7.38	13.28	1.48	13.00	24.71	1.30
3	8.96	16.13	1.79	15.55	29.55	1.56
4	10.21	18.39	2.04	17.67	33.57	1.77
5	11.40	20.52	2.28	19.65	37.33	1.96
6	12.42	22.36	2.48	21.41	40.69	2.14
7	13.48	24.26	2.70	23.15	43.98	2.31
8	14.37	25.86	2.87	24.57	46.68	2.46
9	15.17	27.31	3.03	25.97	49.35	2.60
10	16.00	28.80	3.20	27.47	52.20	2.75

As we anticipated, we see that increasing n leads to increased failure run lengths. Success run lengths increase also, and at higher S levels the increase in success runs is much more dramatic than the increase in failure run length. For example, at $S=90\%$, increasing n from 1 to 10 increases the average failure run by 2.20 from 1.10 to 3.20 periods, but the average success run is increased by 18.86 periods to 28.80. We also see that an increase in shipping level reduces failure run length and that the sensitivity of failure run length to changes in n is greater at lower levels of S.

Figure 3.3

Expected Failure Run Length Response to Lead Time

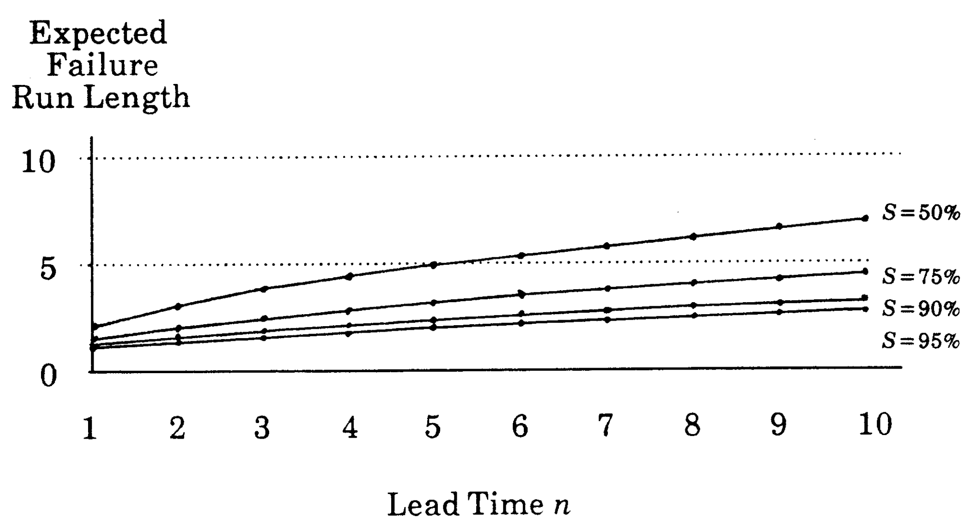
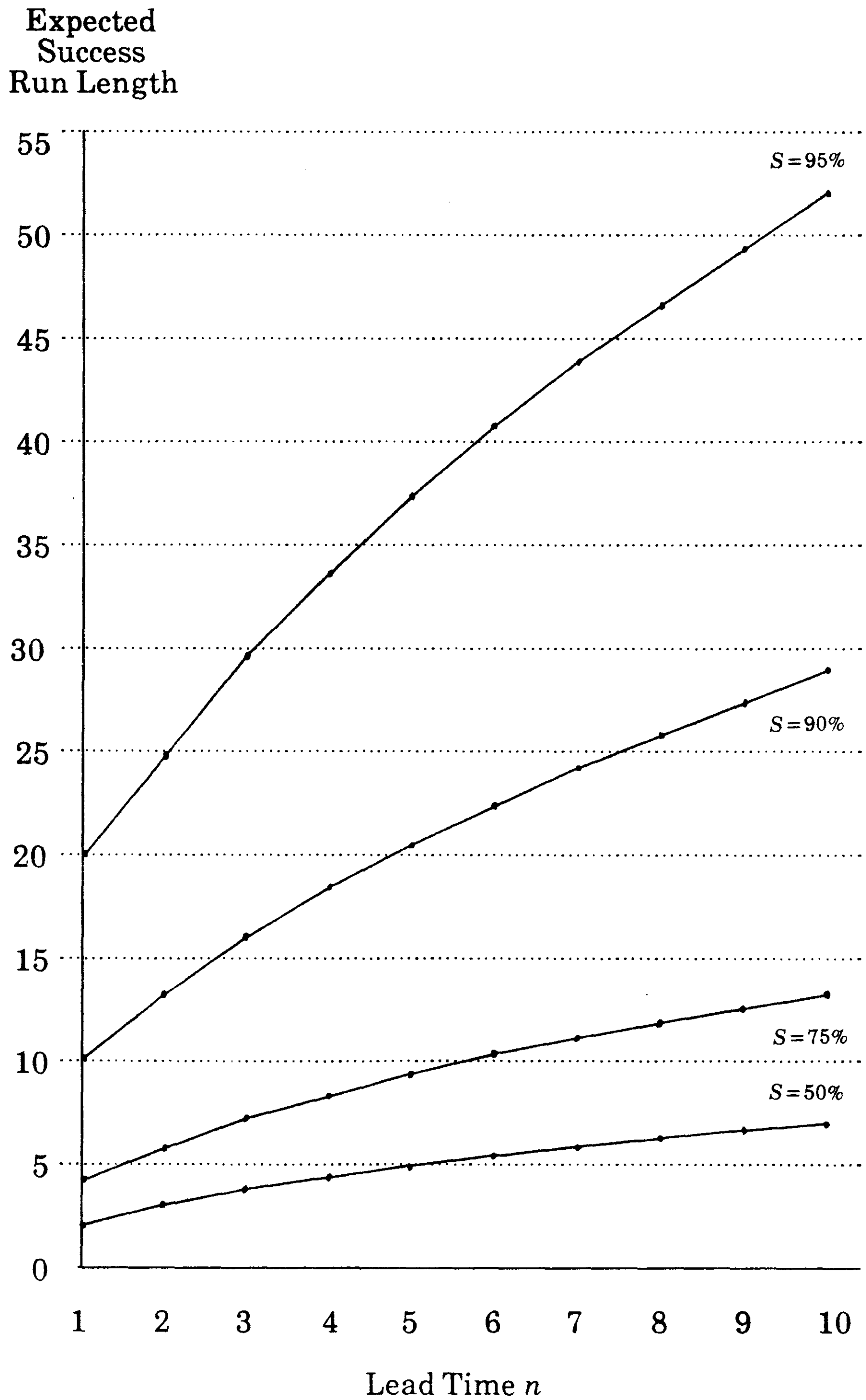


Figure 3.4

Expected Success Run Length Response to Lead Time



We can understand the interaction of n and S in determining failure run length by looking at Figures 3.5, 3.6, and 3.7, which depict a hypothetical randomly chosen sample path of the I_t process. First, we look at the effect of n in a system with no safety stock. This situation is depicted in Figure 3.5 and corresponds to the $S=50\%$ curve in Figures 3.3 and 3.4. Here the zero level corresponds with the mean of the I_t distribution, about which I_t is symmetric. Because of this symmetry, average run length is the same on either side of $I_t=0$. An increase in n will increase $Var\,(I_t)$, causing the extremes of the path on both sides of the mean to move farther out. The increase in n also increases $Cov(I_t, I_{t+h})$, which will cause a lengthening of average runs on both sides of the mean. Due to the symmetry, the increase in average run length will be the same on both sides, which is the effect we see in Table 3.1.

Next we look at the impact on failure run length of increasing S. The effect of increasing S by adding safety stock is seen by comparing Figure 3.5 to Figure 3.6, where we have now centered the I_t sample path around a positive level of safety stock, causing the zero level to be moved towards a tail of the distribution. As a result, the average length of failure runs will decrease while the average length of success runs increases, exactly the effect we see in Figures 3.3 and 3.4. So, increasing S while holding n constant causes improved service by two effects: first, directly through the increase of $S=Pr\{I_t \geqq 0\}$, and second by simultaneously reducing failure and increasing success run average lengths.

Finally, we consider a system with a positive level of safety stock and examine the effect of a change in n while holding the stock constant at A. Compare Figures 3.6 and 3.7. The resulting increase in variance and covariance increase the height of peaks and valleys and stretch out the path horizontally. Here we see the important relationship between S and the response of failure run lengths to n, which flattens out at higher levels of S in Figure 3.3. The higher the level of S, the farther the zero level is pushed into the tail of the I_t distribution. As the zero level moves farther out, failures become rarer, failure runs become shorter and the impact of n on failure runs becomes much smaller. Conversely, success runs are positively affected as n increases since more of the I_t distribution is in the success area. This accounts for the increasing sensitivity of average success run to n as S increases, as seen in Figure 3.4.

Figure 3.5

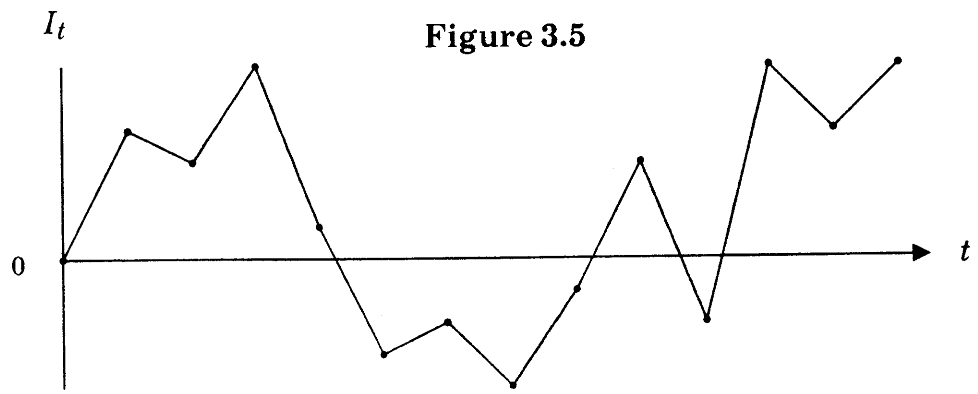


Figure 3.6

Effect of Increasing S

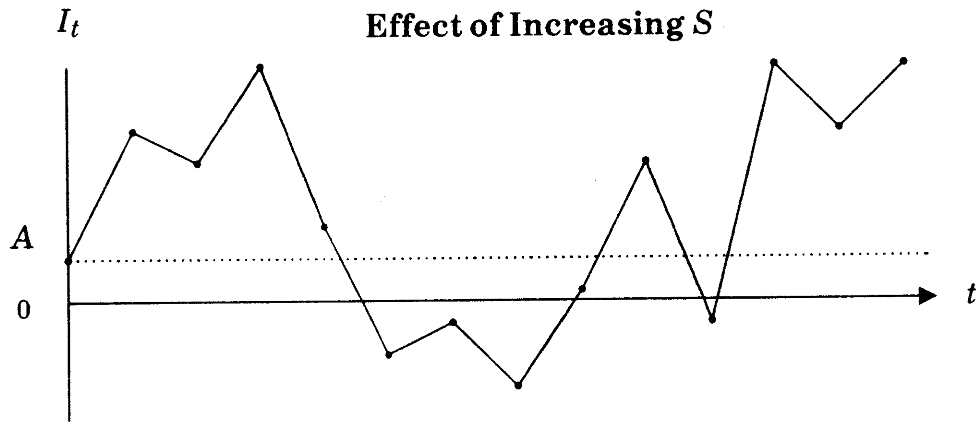
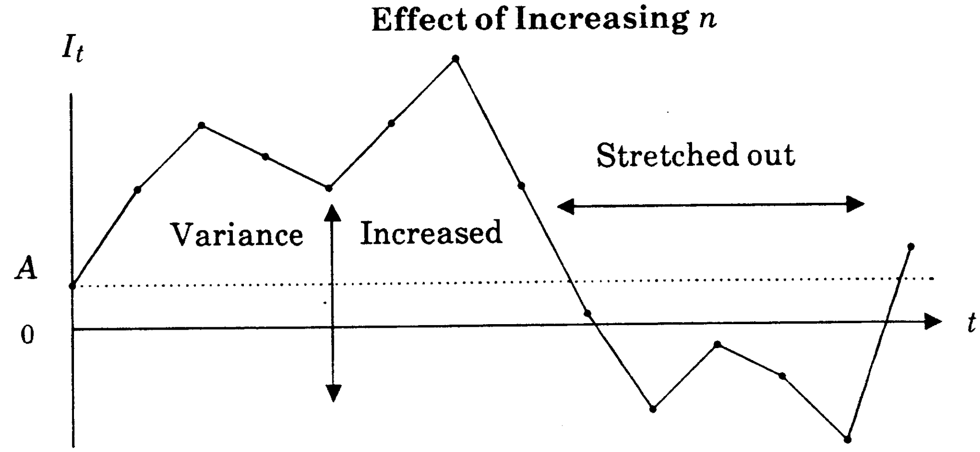


Figure 3.7

Effect of Increasing n



H. Management Implications: Achieving Balance Between Production Smoothing and Desired Service Levels

In Chapter 1, we saw that the TPM shows how to smooth production levels by increasing production lead times. Smoother production has the benefits of reduced required production capacity and reduced costs associated with managing the response to large fluctuations in production requirements. It was clear from the TPM results in Chapter 1 that these benefits are achieved at the expense of increasing the work-in-process stock.

Now, with the results of the previous section, we can assess the impact of lead time choices on the system service level as well. In this section we provide an integrated discussion of the impact of management decisions on all the performance measures of the system and provide some insights into understanding the trade-offs among competing measures.

We consider the performance measures to be expected production levels $\mathbf{E}(\mathbf{P}_t)$, variance of production levels $\mathrm{Var}(\mathrm{P}_t)$, the size of base stock K (consisting of in-process plus safety stock), expected failure run length $E(F)$, and the shipping level S. The management control parameters, upon which these performance measures depend, are the lead times n and the safety stock level, which we will denote by I^*. The performance measures also depend upon the mean and variance of the demand distribution, $\bar{R}$ and σ^2_R, which we will assume to be beyond management's control in our discussion. The following general equations indicate the dependence of each of the performance measures on the management control parameters (we have also included the demand distribution parameters for information only). The number or numbers following each equation is the reference to the full expressions elsewhere in this thesis.

(3.34) $$\mathbf{E}(\mathbf{P}_t) = f_1(\bar{R})$$ (1.9)

(3.35) $$\mathbf{Var}(\mathbf{P}_t) = f_2(n, \sigma_R^2)$$ (1.10)

(3.36) $$K = f_3(n, I^*, S, \bar{R}, \sigma_R^2)$$ (3.19)

(3.37) $$E(F) = f_4(n, I^*, S)$$ (3.24) and (3.25)

(3.38) $$S = f_5(n, I^*, \sigma_R^2)$$ (3.19)

Note that choices of n and I^* affect K and $E(F)$ both directly and indirectly via their effect on S, as we shall discuss below.

To illustrate the discussion, we present in Tables 3.2, 3.3, 3.4, and 3.5 in-process stock and safety stock for four hypothetical combinations of $\bar{R}$ and σ_R ($\bar{R} = 300$ and 150, $\sigma_R = 90$ and 45) for four different choices of S. In Figures 3.8 and 3.9, we present graphically the relationship between lead time and safety stock for the two different values of σ_R. In Figures 3.10 and 3.11 we present graphically the relationship between safety stock and expected failure run length for given levels of S and various choices of n. Figure 3.10 contains these relationships for $\sigma_R = 90$, and Figure 3.11 for $\sigma_R = 45$.

Now let us see what happens to the performance measures as management increases n, holding the safety stock I^* constant. Of course smoothing occurs as $\mathbf{Var}(\mathbf{P}_t)$ falls. Also the in-process stock $n\bar{R}$ increases, as noted above. From (3.38) and the equations underlying it, we see that the increase in n causes σ^2_R to rise, and thus S will fall. See also Figure 3.8 for an example: suppose $I^* = 200$, $n = 3$ and $S = 95\%$. If n is increased to 6 with no change in I^*, we move off the $S = 95\%$ curve and to another S curve where $S < 90\%$. Since I^* is held constant, the effect of the n and S changes on base stock K in (3.36) is restricted to the effect of n on in-process stock, as above. The combined changes of n and S will cause $E(F)$ in (3.37) to increase. This complex effect is most easily seen by examining Figure 3.10. Assuming again we start with $I^*=200$, $n=3$ and $S=95\%$, we increase n to 6 while holding I^* constant. We wind up moving to an S curve (not shown) somewhere to the right of the $S=90\%$ curve, and the expected failure run length has increased from 1.6 to more than 2.5 periods. In summary, as we increased n and held I^* constant, S fell and $E(F)$ increased.

Because of the complex interaction of n, I^*, S and $E(F)$ it is not possible in these numerical examples to vary both n and I^* to maintain the same level of both S and $E(F)$. If management desires to maintain a fixed S as n increases, I^* must be increased and a higher $E(F)$ must be accepted. If, as n increases, it is desired to maintain the same $E(F)$, I^* must be increased, which will also increase S. These results can be seen by working through several examples using Figures 3.10 and 3.11.

In conclusion, we see that increasing n to smooth production does have a negative impact on service as measured by both shipping level and expected failure run length. In order to prevent an increase in $E(F)$, a substantial increase in safety stock must be made (large enough that S actually increases). The existing level of S can be maintained with a smaller increase in I^* if management is willing to accept an increase in $E(F)$. The actual magnitude of these effects can be investigated for a specific facility by developing graphs like Figures 3.8 and 3.10, using the facility's actual $\overline{R}$ and σ^2_R.

Table 3.2

In-process and Safety Stock Response to

S and n for $\bar{R}=300$, $\sigma_R=90$

n	$n\bar{R}$	$S=50\%$		$S=75\%$		$S=90\%$		$S=95\%$	
		Safety Stock	Total Stock	Safety Stock	Total Stock	Safety Stock	Total Stock	Safety Stock	Total Stock
1	300	0	300	61	361	115	415	148	448
2	600	0	600	71	671	133	733	170	770
3	900	0	900	82	982	155	1055	198	1098
4	1200	0	1200	93	1293	174	1374	223	1423
5	1500	0	1500	102	1602	192	1692	246	1746
6	1800	0	1800	111	1911	208	2008	267	2067
7	2100	0	2100	119	2219	224	2324	287	2387
8	2400	0	2400	126	2526	238	2638	305	2705
9	2700	0	2700	134	2834	251	2951	322	3022
10	3000	0	3000	140	3140	264	3264	339	3339

Table 3.3

In-process and Safety Stock Response to

S and n for $\bar{R} = 150$, $\sigma_R = 90$

n	$n\bar{R}$	$S=50\%$		$S=75\%$		$S=90\%$		$S=95\%$	
		Safety Stock	Total Stock	Safety Stock	Total Stock	Safety Stock	Total Stock	Safety Stock	Total Stock
1	150	0	150	61	211	115	265	148	298
2	300	0	300	71	371	133	433	170	470
3	450	0	450	82	532	155	605	198	648
4	600	0	600	93	693	174	774	223	823
5	750	0	750	102	852	192	942	246	996
6	900	0	900	111	1011	208	1108	267	1167
7	1050	0	1050	119	1169	224	1274	287	1337
8	1200	0	1200	126	1326	238	1438	305	1505
9	1350	0	1350	134	1484	251	1601	322	1672
10	1500	0	1500	140	1640	264	1764	339	1839

Table 3.4

In-process and Safety Stock Response to

S and n for $\overline{R} = 300$, $\sigma_R = 45$

n	$n\overline{R}$	$S=50\%$		$S=75\%$		$S=90\%$		$S=95\%$	
		Safety Stock	Total Stock	Safety Stock	Total Stock	Safety Stock	Total Stock	Safety Stock	Total Stock
1	300	0	300	31	331	58	358	74	374
2	600	0	600	35	635	67	667	85	685
3	900	0	900	41	941	77	977	99	999
4	1200	0	1200	46	1246	87	1287	112	1312
5	1500	0	1500	51	1551	96	1596	123	1623
6	1800	0	1800	55	1855	104	1904	134	1934
7	2100	0	2100	59	2159	112	2212	143	2243
8	2400	0	2400	63	2463	119	2519	152	2552
9	2700	0	2700	67	2767	126	2826	161	2861
10	3000	0	3000	70	3070	132	3132	169	3169

Table 3.5

In-process and Safety Stock Response to

S and n for $\overline{R} = 150$, $\sigma_R = 45$

n	$n\overline{R}$	$S=50\%$		$S=75\%$		$S=90\%$		$S=95\%$	
		Safety Stock	Total Stock	Safety Stock	Total Stock	Safety Stock	Total Stock	Safety Stock	Total Stock
1	150	0	150	31	181	58	208	74	224
2	300	0	300	35	335	67	367	85	385
3	450	0	450	41	491	77	527	99	549
4	600	0	600	46	646	87	687	112	712
5	750	0	750	51	801	96	846	123	873
6	900	0	900	55	955	104	1004	134	1034
7	1050	0	1050	59	1109	112	1162	143	1193
8	1200	0	1200	63	1263	119	1319	152	1352
9	1350	0	1350	67	1417	126	1476	161	1511
10	1500	0	1500	70	1570	132	1632	169	1669

Figure 3.8

Safety Stock vs. Lead Time for the ECM Line

For $\sigma_R = 90$, from Table 3.4

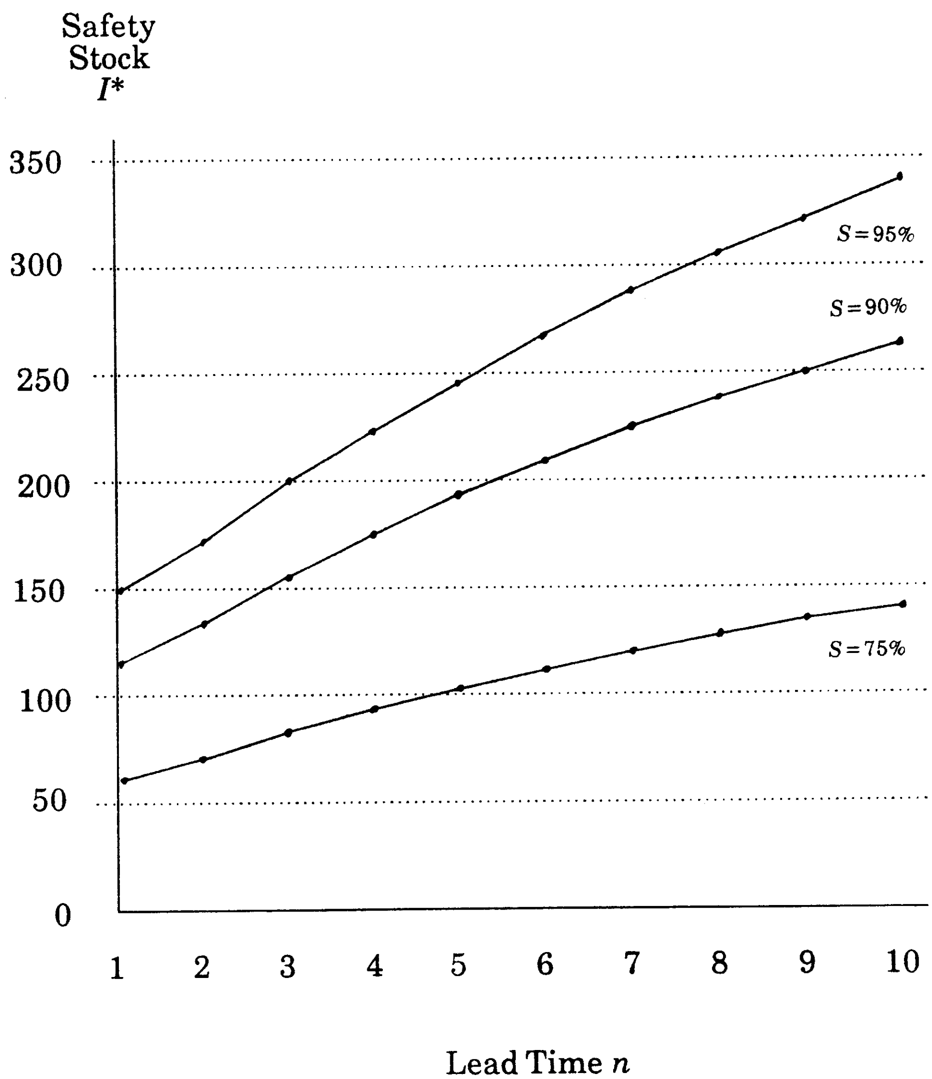


Figure 3.9

Safety Stock vs. Lead Time for the ECM Line

For $\sigma_R = 45$, from Table 3.4

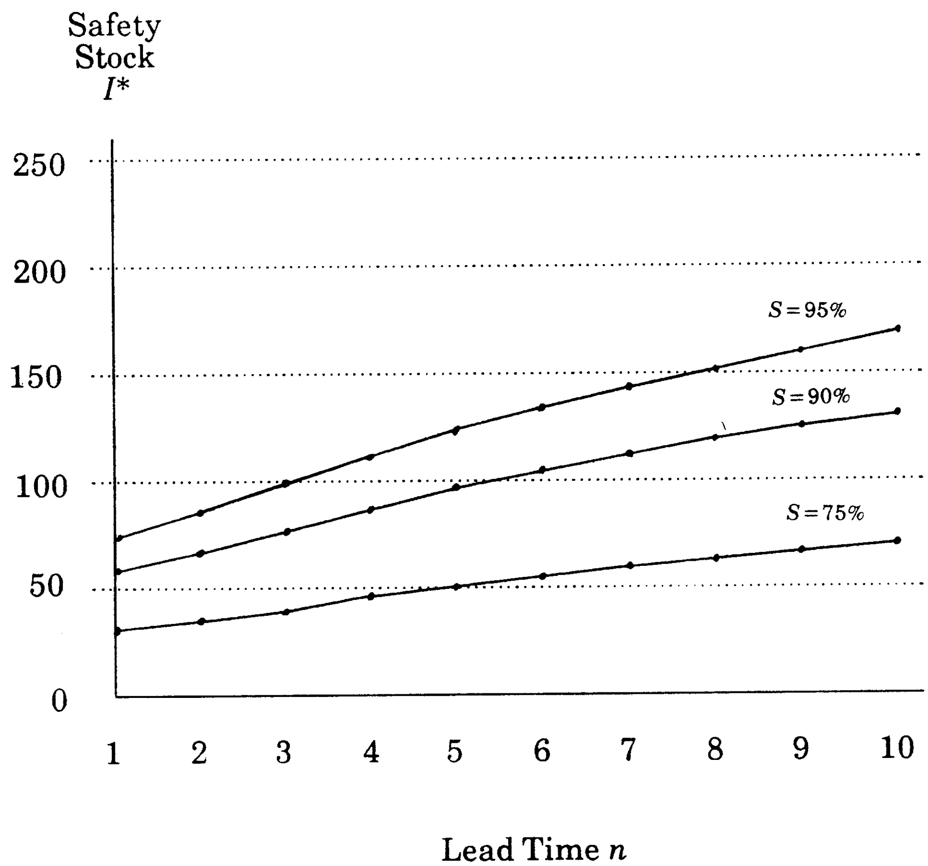


Figure 3.10

Safety Stock vs. Expected Failure Run Lengths

For $\sigma_R = 90$, from Tables 3.1 and 3.2

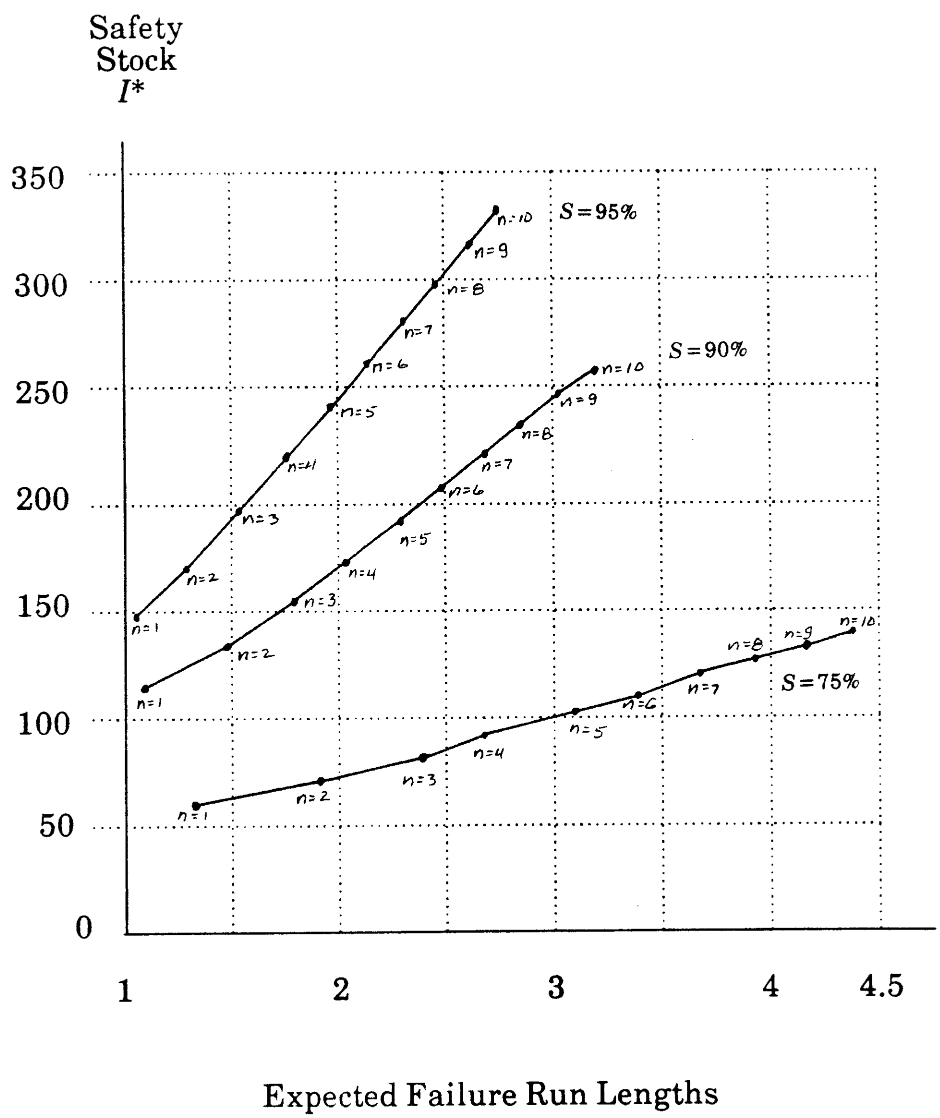
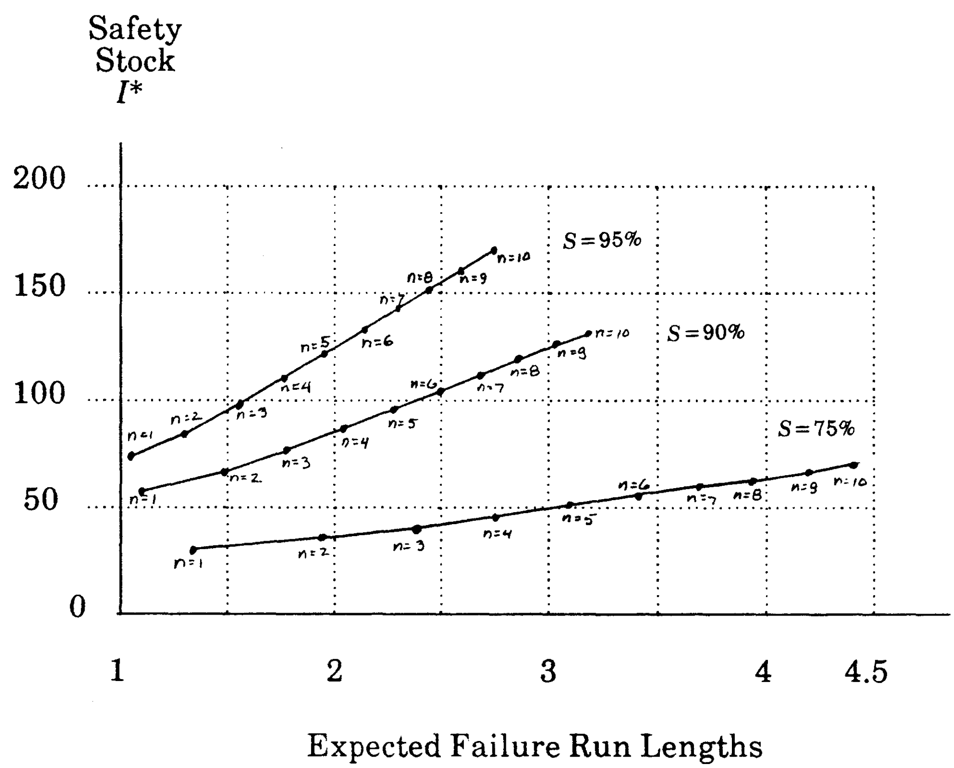


Figure 3.11

Safety Stock vs. Expected Failure Run Lengths

For $\sigma_R = 45$, from Tables 3.1 and 3.4



Chapter 4

Transient Behavior

A. The Need for Models of Transient Behavior

In "A Tactical Planning Model . . .", Graves solves for the equilibrium values of the production levels, $\mathbf{P}_t$, work-in-process inventory levels, $\mathbf{Q}_t$, their expectations, and their variance-covariance matrices, $\mathbf{Var}(\mathbf{P}_t)$ and $\mathbf{Var}(\mathbf{Q}_t)$ (equations 1.7 through 1.13 in Chapter 1). It is highly unusual, of course, for a production system to operate for long periods of time in an equilibrium environment. Most manufacturing systems are affected by environmental changes such as seasonal demand cycles, long term growth or decline of demand, and shorter term changes which affect the required production rate. If these changes are large or occur frequently, we might expect that the system would regularly be in transition from one equilibrium to another and would rarely have settled in at such a level. Thus the system's transient behavior may be of more relevance than the equilibrium level.

In this chapter we model the transient paths of $\mathbf{E}(\mathbf{P}_t)$ and $\mathbf{E}(\mathbf{Q}_t)$ to determine their characteristics and we focus attention on two aspects of transient behavior. The first is the speed of adjustment to a new equilibrium level. The second characteristic of interest is the shape of the paths $\mathbf{P}_t$ and $\mathbf{Q}_t$ take as the system adjusts to a new level. These paths might approach asymptotically, they might substantially overshoot the new levels before settling in, or they might oscillate in either a damped or explosive fashion around the new levels. The nature of the path behavior will, of course, have significant implications for management.

The equilibrium values are developed in the TPM by assuming that the terms of each time series of random variables $\{\varepsilon_{it}\}$, $(t = ..., -1, 0, 1, ...)$ are i.i.d. It may be recalled that the ε_{it} represent new arrivals to sector i from outside the system plus random fluctuations in the flow of work to sector i from other sectors. We will occasionally refer to these ε_{it} or the vector ε_t as "innovations". Equilibrium is imparted because, even though the value of ε_{it} will vary from period to period, the random variable is always drawn from the same distribution and thus has the same mean and variance in all periods. Equilibrium occurs when the ε_{it} have been drawn from their distributions for a long enough time that $\mathbf{E}(\mathbf{P}_t)$, as given by (1.9), is a good description of average system behavior. (Note that ε_{it} is i.i.d. for each sector i, but that it is perfectly acceptable to have the distribution of ε_{it} differ from that of ε_{jt}.)

We will define transient behavior as the behavior of $\mathbf{P}_t$, $\mathbf{Q}_t$ and their expectations in response to a change in the probability distributions from which the elements of the ε_t vector of equation (1.6) are drawn. Given the distributions, equilibrium expected production levels are a function of the vector of means of the distributions, μ, as in equation (1.9) If we change the distributions from which ε_{it} are drawn, then (1.9) can be used to identify the new equilibrium expected production levels as a function of the vector of means of the new distributions, μ^*. The system may take several periods to adjust to the new distributions of innovations, that is until $\mathbf{E}(\mathbf{P}_t)$ and $\mathbf{E}(\mathbf{Q}_t)$, both as functions of μ^*, are an adequate description of system performance. The behavior during this period of adjustment is the focus of our study of transients.

We model transient behavior by changing the distributions of innovations and then tracking system behavior. We will consider only changes to the means of the distributions and not to their variances, so we will focus only on changes in $\mathbf{E}(\mathbf{P}_t)$ and $\mathbf{E}(\mathbf{Q}_t)$. The equilibrium levels of $\mathbf{Var}(\mathbf{P}_t)$ and $\mathbf{Var}\ (\mathbf{Q}_t)$ will remain unchanged. We will consider four different types of changes to the distributions of innovations. The first is a one-period change with immediate return to drawing the ε_{it} from the original distribution. This corresponds to an unusual, one-time occurrence in the manufacturing process: for example, receipt of an unusually large order. The second type of change is an increase in the means to a new level, at which they remain. Acquisition of a new distributor and the consequent addition of a new market territory might cause this sort of change. The third change considered is one of linear growth, in which the mean innovations increases with time. The final change is to introduce cycles in which the mean will rise to a new level for a number of periods, return to the original level for another series of periods, and then repeat the pattern. This, of course, corresponds to seasonal cycles of orders.

We define the innovations at each sector as

(4.1) $$\Delta_{it} = \varepsilon_{it} + \delta_i(t),$$

where ε_{it} are the random variables defined previously and $\delta_i(t)$ is a nonnegative deterministic quantity, which may be different for each sector i. We continue to assume that each time series $\{ \varepsilon_{it} \}$ is i.i.d. with mean μ_i and variance $Var(\varepsilon_i)$. In vector form, we have the expectation vector and variance-covariance matrix as

(4.2) $$\mathbf{E}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\mu}$$

(4.3) $$\mathbf{Var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}$$

The different types of change can be implemented by defining in different ways the nature of the functional dependence of δ_i on t. The appropriate definition of $\delta_i(t)$, and thus the distribution of Δ_{it}, will be developed in each of the following sections.

In sections C, D, E and F, we develop expressions for the expected value of the vector $\mathbf{P}_t$ following the introduction of the change, the value of the new equilibrium level of $\mathbf{P}_t$, the vector ρ_1, and the difference at every value of t between the vector ρ_1 and $\mathbf{E}(\mathbf{P}_t)$, which will be called the "transient", $\mathrm{T}_t^{(i)}$, where the superscript i is an index which will be explained later. In sections G and H we will discuss the nature of the transients in all cases and in section I we will apply the results to the example of the ECM line. In the following section we establish some results which will be used in the subsequent development.

B. Preliminary Results

There are several algebraic results and points from matrix theory regarding the matrix $\mathbf{B} = (\mathbf{I} - \mathbf{D} + \mathbf{D\Phi})$ which will be used in the following sections. It is convenient to establish these in advance. The first four results will be used in sections C, D, E and F to derive expressions for the transients and the last two results will be used in sections G and H to evaluate the transient paths. First, we note that:

(4.4)
$$\sum_{s=0}^{\infty} \mathbf{B}^s = (\mathbf{I} - \mathbf{B})^{-1}.$$

For (4.4) to be valid, the spectral radius of **B** must be less than 1. Graves shows that this is true if and only if the spectral radius of $\mathbf{\Phi}$ is less than 1 and further that it is reasonable to expect this condition to be satisfied by a workflow matrix $\mathbf{\Phi}$. Therefore the sequence $\{B^s\}$ converges, and (4.4) is valid.

Second, it is true that:

(4.5)
$$\sum_{s=0}^{t-1} \mathbf{B}^s = (\mathbf{I} - \mathbf{B})^{-1} (\mathbf{I} - \mathbf{B}^t).$$

This result is developed by first writing

$$\sum_{s=0}^{t-1} \mathbf{B}^t = \mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \ldots + \mathbf{B}^{t-1}.$$

Then, pre-multiplying by B, gives:

$$\mathbf{B} \sum_{s=0}^{t-1} \mathbf{B}^t = \mathbf{B} + \mathbf{B}^2 + \ldots + \mathbf{B}^t.$$

Now subtract the second expression from the first, pre-multiply by $(\mathbf{I}-\mathbf{B})^{-1}$, and the result is (4.5).

Notice that the order of multiplication of the two terms on the right hand side of (4.5) can be reversed. This is true since we could have post-multiplied by $\mathbf{B}$ and then post-multiplied by $(\mathbf{I}-\mathbf{B})^{-1}$ in the derivation. As a result, we see that $(\mathbf{I}-\mathbf{B})^{-1}$ commutes with $\mathbf{B}^t$:

(4.6)
$$\mathbf{B}^t(\mathbf{I}-\mathbf{B})^{-1} \;=\; (\mathbf{I}-\mathbf{B})^{-1}\mathbf{B}^t, \qquad t = 0, 1, 2, \ldots\ .$$

Our fourth result is that

(4.7)
$$\sum_{s=0}^{t-1} s\,\mathbf{B}^s \;=\; (\mathbf{I}-\mathbf{B})^{-1}\,(\mathbf{I}-\mathbf{B}^t)\,(\mathbf{I}-\mathbf{B})^{-1} \;-\; (\mathbf{I}-\mathbf{B})^{-1}[\mathbf{I}+(t-1)\,\mathbf{B}^t].$$

This is derived by writing:

$$\sum_{s=0}^{t-1} s\,\mathbf{B}^s \;=\; \mathbf{B} + 2\mathbf{B}^2 + 3\mathbf{B}^3 + \ldots + (t-1)\,\mathbf{B}^{t-1}.$$

Now form a second series by pre-multiplying by $\mathbf{B}$, subtract the second series from the first, and then add $(t\text{-}1)\mathbf{B}^t$ to both sides, giving:

$$\left(\sum_{s=0}^{t-1} s\,\mathbf{B}^s \;-\; \mathbf{B}\sum_{s=0}^{t-1} s\,\mathbf{B}^s\right) + (t-1)\mathbf{B}^t \;=\; \mathbf{B} + \mathbf{B}^2 + \ldots + \mathbf{B}^{t-1}$$

$$= \sum_{s=0}^{t-1} B^s \;-\; \mathbf{I}.$$

From (4.5), we see that the right-hand side is $(\mathbf{I}-\mathbf{B}^t)\,(\mathbf{I}-\mathbf{B})^{-1} - \mathbf{I}$. Subtracting $(t\text{-}1)\,\mathbf{B}^t$ from both sides and then pre-multiplying by $(\mathbf{I}-\mathbf{B})^{-1}$, we have (4.7).

The next two results will be used in sections G and H to evaluate the transient paths. First, $(\mathbf{I}-\mathbf{B})^{-1}$ is a nonegative matrix. This follows from the fact that the spectral radius of $\mathbf{B}$ is less than 1 and that $\mathbf{B}$ is nonnegative. With these two features, a theorem in the theory of nonegative matrices [Lancaster and Tismenetsky, pg. 531, Theorem 2] guarantees that $(\mathbf{I}-\mathbf{B})^{-1} \geqq 0$. We can also state that $(\mathbf{I}-\mathbf{B})^{-1} \neq 0$ by virtue of the definition of $\mathbf{B}$.

Second, $\mathbf{B}$ has a real, positive eigenvalue equal to its spectral radius. This conclusion comes from a result in the Frobenius theory for reducible nonnegative matrices [Gantmacher, Vol. II, pg. 66, Theorem 3].

C. One-Period Change

Suppose the system is in an equilibrium state since the innovations ε_t have been drawn from the same distributions long enough for the effects of any prior innovation distributions to have dissipated. Then suppose that, at time $t=1$, ε_1 is drawn from the same distributions, and that a vector $\boldsymbol{\delta}$ is added to it. At time $t=2$ and all subsequent times, the innovation vectors are $\boldsymbol{\varepsilon}_t$ from the original distribution without the addition of $\boldsymbol{\delta}$.

We specify this experiment with the following equations. The inital state of the system is $\mathbf{P}_0$, which can be written as (from 1.7 and 1.8):

(4.8)
$$\mathbf{P}_0 = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D} \boldsymbol{\varepsilon}_{0-s}$$

(4.9)
$$\boldsymbol{\rho}_0 = \mathbf{E}(\mathbf{P}_0) = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D} \boldsymbol{\mu}$$

At time $t=1$, we replace the arrival vector ε_1 with

(4.10)
$$\boldsymbol{\Delta}_1 = \boldsymbol{\varepsilon}_1 + \boldsymbol{\delta},$$

where δ is deterministic. The mean of $\boldsymbol{\Delta}_1$ is $\boldsymbol{\mu} + \boldsymbol{\delta}$ and its variance-covariance matrix is Σ. At time $t=2$ and later the innovation vectors are ε_2, ε_3, . . ., all drawn from the original distributions.

By applying equation (1.6) successively, we have

$$\mathbf{P}_1 = \mathbf{B}\,\mathbf{P}_0 + \mathbf{D}\,\boldsymbol{\Delta}_1$$

$$\mathbf{P}_2 = \mathbf{B}^2\mathbf{P}_0 + \mathbf{B}\,\mathbf{D}\,\boldsymbol{\Delta}_1 + \mathbf{D}\,\boldsymbol{\varepsilon}_2$$

$$\mathbf{P}_3 = \mathbf{B}^3\mathbf{P}_0 + \mathbf{B}^2\mathbf{D}\,\boldsymbol{\Delta}_1 + \mathbf{B}\,\mathbf{D}\boldsymbol{\varepsilon}_2 + \mathbf{D}\boldsymbol{\varepsilon}_3$$

$$\vdots$$

Continuing the iteration, and using (4.10) to substitue for the $\boldsymbol{\Delta}_1$'s, we can write for any future time t,

$$\mathbf{P}_t = \mathbf{B}^t\,\mathbf{P}_0 + \sum_{s=0}^{t-1} \mathbf{B}^s\,\mathbf{D}\,\boldsymbol{\varepsilon}_{t-s} + \mathbf{B}^{t-1}\mathbf{D}\boldsymbol{\delta}.$$

By using the expression (4.8) for $\mathbf{P}_0$ in the first term, the first two terms can be combined, and the general expressions for $\mathbf{P}_t$ and its expectation become:

(4.11)
$$\mathbf{P}_t = \sum_{s=0}^{\infty} \mathbf{B}^s\,\mathbf{D}\,\boldsymbol{\varepsilon}_{t-s} + \mathbf{B}^{t-1}\,\mathbf{D}\boldsymbol{\delta}.$$

(4.12)
$$\mathbf{E}\,(\mathbf{P}_t) = \sum_{s=0}^{\infty} \mathbf{B}^s\,\mathbf{D}\boldsymbol{\mu} + \mathbf{B}^{t-1}\,\mathbf{D}\boldsymbol{\delta}.$$

The first term of (4.12) is simply the original equilibrium level of $\mathbf{E}(\mathbf{P}_t)$, $\boldsymbol{\rho}_0$, and the second term is the transient of the response:

(4.13)
$$\mathbf{T}_t^{(1)} = \mathbf{B}^{t-1}\mathbf{D}\boldsymbol{\delta}.$$

where the superscript 1 denotes the first case, the one-period change.

D. Step Change

Next we consider a step increase in which we let

(4.14)
$$\boldsymbol{\Delta}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\delta}.$$

where $\boldsymbol{\delta}$ is a constant vector, for time $t = 1$ and subsequently.

Starting from P_0, as defined in (4.8) and (4.9), we inject Δ_1 at $t=1$, Δ_2 at $t=2$, etc. Proceeding by iterative substitution, as we did in section C, we see that

$$P_1 = B P_0 + D\Delta_1$$

$$P_2 = B^2 P_0 + B D\Delta_1 + D\Delta_2$$

$$P_3 = B^3 P_0 + B^2 D\Delta_1 + B D\Delta_2 + D\Delta_3$$

.

.

.

$$P_t = B^t P_0 + \sum_{s=0}^{t-1} B^s D\Delta_{t-s} .$$

Since $\Delta_{t-s} = \varepsilon_{t-s} + \delta$, this can be rewritten as

$$P_t = B^t P_0 + \sum_{s=0}^{t-1} B^s D\varepsilon_{t-s} + \sum_{s=0}^{t-1} B^s D\delta .$$

As in the case of the one-period change, the first two terms can be combined into

$$\sum_{s=0}^{\infty} B^s D\, \varepsilon_{t-s} .$$

Using (4.5) to rewrite the third term we find that the general expression for P_t and its expectation is:

$$\text{(4.15)} \qquad P_t = \sum_{s=0}^{\infty} B^s D\varepsilon_{t-s} + (I-B^t)(I-B)^{-1} D\delta$$

$$\text{(4.16)} \qquad E(P_t) = \sum_{s=0}^{\infty} B^s D\mu + (I-B^t)(I-B)^{-1} D\delta .$$

P_t approaches a new equilibrium level, conditioned on the new innovations vector, which is given by

$$\text{(4.17)} \qquad \rho_1 = \sum_{s=0}^{\infty} B^s D(\mu+\delta) = (I-\Phi)^{-1} (\mu+\delta) .$$

We define the difference between the new equilibrium level and the expected value of P_t as the transient component of the response. Subtracting (4.16) from (4.17), and using (4.4) and (4.6), we find the expression for the transient is:

$$\text{(4.18)} \qquad \mathbf{T}_t^{(2)} = \boldsymbol{\rho}_1 - \mathbf{E}(\mathbf{P}_t) = \mathbf{B}^t(\mathbf{I}-\mathbf{B})^{-1}\mathbf{D}\boldsymbol{\delta}.$$

E. Linear Growth

Now we consider linear growth in the innovations vector and let

$$\boldsymbol{\Delta}_t = \boldsymbol{\varepsilon}_t + \mathbf{t}\,\boldsymbol{\delta}$$

for $t=1, 2, \ldots$, where $\boldsymbol{\delta}$ is a constant. Starting from $\mathbf{P}_0$ as defined in (4.8) and substituting iteratively as we did in sections C and D, we find

$$\mathbf{P}_t = \mathbf{B}^t\mathbf{P}_0 + \sum_{s=0}^{t-1}\mathbf{B}^s\mathbf{D}\,\boldsymbol{\varepsilon}_{t-s} + \sum_{s=0}^{t-1}(t-s)\,\mathbf{B}^s\mathbf{D}\boldsymbol{\delta}.$$

The first two terms combine to give

$$\sum_{s=0}^{\infty}\mathbf{B}^s\mathbf{D}\,\boldsymbol{\varepsilon}_{t-s}\ .$$

The first part of the third term is expanded by using (4.5):

$$\sum_{s=0}^{t-1} t\,\mathbf{B}^s\mathbf{D}\boldsymbol{\delta} = t(\mathbf{I}-\mathbf{B}^t)(\mathbf{I}-\mathbf{B})^{-1}\mathbf{D}\boldsymbol{\delta}.$$

Using (4.6), the second part is:

$$-\sum_{s=0}^{t-1} s\,\mathbf{B}^s\mathbf{D}\boldsymbol{\delta} = -\,[(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B}^t) - \mathbf{I} - (t-1)\mathbf{B}^t]\,(\mathbf{I}-\mathbf{B})^{-1}\mathbf{D}\boldsymbol{\delta}.$$

Combining these two parts we see that the third term is:

$$\sum_{s=0}^{t-1}(t-s)\,\mathbf{B}^s\mathbf{D}\boldsymbol{\delta} = \{(t+1)\mathbf{I} - (\mathbf{I}-\mathbf{B})^{-1} - \mathbf{B}^t[\mathbf{I} - (\mathbf{I}-\mathbf{B})^{-1}]\}(\mathbf{I}-\mathbf{B})^{-1}\mathbf{D}\boldsymbol{\delta}.$$

Putting all the parts together and taking the expectation, the entire expression for $\mathbf{E}(\mathbf{P}_t)$ is

$$\text{(4.20)} \qquad \mathbf{E}(\mathbf{P}_t) = \sum_{s=0}^{\infty}\mathbf{B}^s\mathbf{D}\boldsymbol{\mu} + \{(t+1)\mathbf{I} - (\mathbf{I}-\mathbf{B})^{-1} - \mathbf{B}^t[\mathbf{I} - (\mathbf{I}-\mathbf{B})^{-1}]\}\,(\mathbf{I}-\mathbf{B})^{-1}\mathbf{D}\boldsymbol{\delta}.$$

$\mathbf{E}(\mathbf{P}_t)$ will grow according to (4.20) and will eventually approach its long-run equilibrium growth path as the size of the term $-\mathbf{B}^t\,[\mathbf{I}-(\mathbf{I}-\mathbf{B})^{-1}]$ dies out. That this term does die out is guaranteed by the fact that the spectral radius of $\mathbf{B}$ is less than 1 (we will say more on this later). The long-run equilibrium growth path is thus (4.20) without that term:

$$\text{(4.21)} \qquad \mathbf{E}(\mathbf{P}_t)(\mathit{Long-run}) = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D}\boldsymbol{\mu} + [(t+1)\mathbf{I} - (\mathbf{I}-\mathbf{B})^{-1}](\mathbf{I}-\mathbf{B})^{-1}\mathbf{D}\boldsymbol{\delta}$$

The transient component in the growth path can be defined as the difference between the long-run equilibrium path, as given by (4.21) and the $\mathbf{E}(\mathbf{P}_t)$ at time t, as given by (4.20), and is:

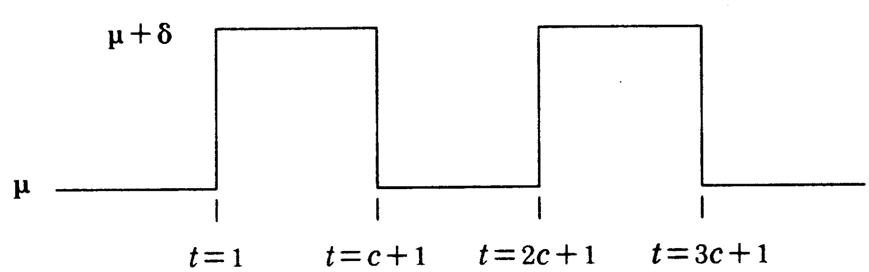
$$\text{(4.22)} \qquad \mathbf{T}_t^{(3)} = [\mathbf{I} - (\mathbf{I}-\mathbf{B})^{-1}]\, \mathbf{B}^t (\mathbf{I}-\mathbf{B})^{-1}\mathbf{D}\boldsymbol{\delta} \; .$$

F. Cyclical Pattern

Last, we consider a cyclical pattern of changes in which the vector of innovations takes on a new level for c periods, returns to the original equilibrium level for another c periods, and then repeats the cycle. This procedure generates a square wave cycle pattern of innovations, as depicted in Figure 4.1. The innovations vectors which generate the cycle are

Figure 4.1

Cyclical Pattern of Δ_t



$$\text{(4.23)} \qquad \boldsymbol{\Delta}_t = \boldsymbol{\varepsilon}_t, \qquad t < 1,\; t = (2i+1)c + 1, \ldots, (2i+1)c + c \;\; \mathit{for}\, i = 0, 1, 2, \ldots$$

$$\boldsymbol{\Delta}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\delta}, \qquad t = 2ic + 1, \ldots, 2ic + c \;\; \mathit{for}\, i = 0, 1, 2, \ldots$$

We will examine $\mathbf{P}_t$ and its transient components for the three periods $t=1$ to $t=c$, $t=c+1$ to $t=2c$, and $t=2c+1$ to $3c$, as the nature of all subsequent transients can be seen from these.

In the first period, $t=1$ to $t=c$, the system responds as to a step change, as described in section D. That is, $\mathrm{E}(\mathbf{P}_t)$ is given by (4.16), the equilibrium level ρ_1 by (4.17), and the transient extent to which ρ_1 has not been reached by (4.18). In particular, at $t=c$, just before the return to the lower arrival rate, the transient is

(4.24) $$\mathbf{T}_c^{(4)} = \mathbf{B}^c (\mathbf{I}-\mathbf{B})^{-1} \mathbf{D} \delta \; .$$

During the first down-cycle, $t = c+1$ to $t = 2c$, the system responds as follows. At $t=c$, $\mathbf{P}_c$ is, from (4.15)

$$\mathbf{P}_c = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D} \varepsilon_{c-s} + (\mathbf{I}-\mathbf{B}^c)(\mathbf{I}-\mathbf{B})^{-1} \mathbf{D} \delta \; .$$

Beginning at $t=c+1$, the innovations vector is $\Delta_t = \varepsilon_t$, so we have

$$\mathbf{P}_{c+1} = \mathbf{B} \mathbf{P}_c + \mathbf{D} \varepsilon_{c+1} \; ,$$

$$\mathbf{P}_{c+2} = \mathbf{B}^2 \mathbf{P}_c + \mathbf{B} \mathbf{D} \varepsilon_{c+1} + \varepsilon_{c+2}$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$\mathbf{P}_{c+j} = \mathbf{B}^j \mathbf{P}_c + \sum_{s=0}^{j-1} \mathbf{B}^s \mathbf{D} \varepsilon_{c+j-s} , \qquad where j \le c.$$

Substituting for $\mathbf{P}_c$, expanding the sum and simplifying gives

(4.25) $$\mathrm{E}(\mathbf{P}_t) = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D} \mu + \mathbf{B}^{t-c} (\mathbf{I}-\mathbf{B}^c)(\mathbf{I}-\mathbf{B})^{-1} \mathbf{D} \delta , \quad where c+1 \le t \le 2c$$

The equilibrium level is, of course, the old equilibrium level

$$\rho_0 = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D} \mu ,$$

so the transient component is:

(4.26) $$T_t^{(4)} = B^{t-c}(I-B^c)(I-B)^{-1}D\delta, \qquad where\, c+1 \le t \le 2c.$$

We start the next up-cycle from P_{2c} which is (from (4.25), after the expectation is undone):

$$P_{2c} = \sum_{s=0}^{\infty} B^s D \varepsilon_{2c-s} + B^c (I-B^c)(I-B)^{-1} D\,\delta$$

Injecting $\Delta_t = \varepsilon_t + \delta$, starting at $t=2c+1$, and following the same iterative process as previously, we derive

$$P_{2c+j} = B^j P_{2c} + \sum_{s=0}^{j-1} B^s D\, \varepsilon_{2c+j-s} + \sum_{s=0}^{j-1} B^s D\delta\,,$$

which simplifies to

(4.27) $$E(P_t) = \sum_{s=0}^{\infty} B^s D\mu + [B^{t-2c} B^c (I-B^c) + (I-B^{t-2c})]\,(I-B)^{-1} D\,\delta,$$

where $2c+1 \leqq t \leqq 3c$. The transient component is the difference between the equilibrium level ρ_1 from (4.17) and $E(P_t)$ in (4.27), and is

(4.28) $$T_t^{(4)} = [B^{t-2c} - B^{t-2c} B^c (I-B^c)]\,(I-B)^{-1} D\,\delta. \qquad where\, 2c+1 \le t \le 3c.$$

G. Transient Behavior of Production Levels

To evaluate the system's response during the period of adjustment to the new equilibrium, we use the expressions for the transients to investigate the length of the adjustment and the nature of the path. We first rewrite equations (4.13), (4.18), (4.22), (4.26) and (4.28), using the fact that $(I\text{-}B)^{-1} = (I\text{-}\Phi)^{-1} D^{-1}$, and thus $(I\text{-}B)^{-1}D\delta = (I\text{-}\Phi)^{-1}\delta$:

For the one-period increase:

(4.13') $$\mathbf{T}_t^{(1)} = \mathbf{B}^{t-1}\mathbf{D}\,\delta$$

For the step change:

(4.18') $$\mathbf{T}_t^{(2)} = \mathbf{B}^t\,(\mathbf{I}-\Phi)^{-1}\,\delta$$

For linear growth:

(4.22') $$\mathbf{T}_t^{(3)} = \mathbf{B}^t\,[\mathbf{I}-(\mathbf{I}-\mathbf{B})^{-1}]\,(\mathbf{I}-\Phi)^{-1}\,\delta$$

For cyclical changes:

(4.18') $$\mathbf{T}_t^{(4)} = \mathbf{B}^t(\mathbf{I}-\Phi)^{-1}\delta, \qquad for\; 1 \le t \le c,$$

(4.26') $$\mathbf{T}_t^{(4)} = \mathbf{B}^{t-c}(\mathbf{I}-\Phi)^{-1}\,\delta \;-\; \mathbf{B}^{t-c}\,\mathbf{B}^c(\mathbf{I}-\phi)^{-1}\delta, \; for\; c+1 \le t \le 2c$$

(4.28') $$\mathbf{T}_t^{(4)} = \mathbf{B}^{t-2c}(\mathbf{I}-\Phi)^{-1}\delta \;-\; \mathbf{B}^{t-2c}\,\mathbf{B}^c(\mathbf{I}-\mathbf{B}^c)(\mathbf{I}-\Phi)^{-1}\delta, \quad for\; 2c+1 \le t \le 3.$$

All of these have the same form in that they consist of a power of the matrix **B** multiplied by a constant vector. The exponent of **B** is a function of the number of periods elapsed since the change was injected into the sytem.

The rate and nature of decay of these transients is thus governed only by the powers of **B** through its eigenvalues. In section B above we pointed out that the maximal eigenvalue is real, positive and less than one. As a result, we can say conclusively that the transients decay asymptotically to zero, and that the expected production levels for each sector asymptotically approach their new equilibrium levels. If any other eigenvalues of **B** are negative or complex and have absolute value or modulus close to the maximal eigenvalue, then the adjustment paths may follow a damped oscillatory pattern. For a negative eigenvalue, the oscillation will be period-by-period, whereas the oscillation generated from a complex eigenvalue will be sinusoidal. We emphasize that the response paths will be determined primarily by the dominant eigenvalue, which is real and positive [Luenberger, pgs. 154-170].

The rapidity of decay depends, of course, on the magnitude of the dominant eigenvalue. For example, the number of periods required for the transient effect to decay to 1% of the original injected change for several possible eigenvalues are:

Eigenvalue λ	Periods n for λ^n to decay to 0.01λ
.50	8
.60	10
.70	14
.80	22
.90	45
.95	90

The transient response path is also affected by the signs of the elements in the constant vectors. We have assumed that $\boldsymbol{\delta} \geqq 0$. The following observations should be modified if it is desired to consider a $\boldsymbol{\delta} \leqq 0$. Recall from section B that $(\mathbf{I}-\mathbf{B})^{-1} \geqq 0$ and $(\mathbf{I}-\mathbf{B})^{-1} \neq 0$. As a result, $(\mathbf{I}-\boldsymbol{\Phi})^{-1}\boldsymbol{\delta} \geqq 0$ and $\neq 0$. By definition $\mathbf{D} \geqq 0$ and $\neq 0$.

Using this information we see that the transients for the one-period increase and the step change are straightforward as each consists of a constant term which is a nonnegative matrix and the exponentially decaying power of $\mathbf{B}$, so each decays to zero in a simple fashion. The transient for linear growth is more complicated since the matrix $[\mathbf{I}-(\mathbf{I}-\mathbf{B})^{-1}]$ will probably contain positive and negative terms. Clearly, however, the magnitude of the transient dies out according to the exponential multipler $\mathbf{B}^t$.

The most complex transients, of course, occur for the cyclical input since we now have lingering influences from previous cycles. During the first up-cycle, the transient, given by (4.18'), behaves like the step-change. At the beginning of the first down-cycle at time $t=c+1$, the remaining up-transient is $\mathbf{B}^c(\mathbf{I}-\boldsymbol{\Phi})^{-1}\boldsymbol{\delta}$. If the cycle is long enough, this will be essentially zero, but might be of significant size if the cycle is short. During the first down-cycle, equation (4.26') shows the lingering effects of this component. Equation (4.28') shows that the transient during the second up-cycle is composed of a component due to the second cycle, $\mathbf{B}^{t-2c}(\mathbf{I}-\mathbf{\Phi})^{-1}\boldsymbol{\delta}$, and whatever lingering effects are carried forward from the previous cycle, $-\mathbf{B}^{t-2c}\mathbf{B}^{c}(\mathbf{I}-\mathbf{B}^{c})(\mathbf{I}-\mathbf{\Phi})^{-1}\boldsymbol{\delta}$. If the cycles are long enough in relation to the maximal eigenvalue, this term will be essentially zero. For example, if the maximal eigenvalue of $\mathbf{B}$ is 0.80, the time unit is a day, and if we are examining an annual cycle ($2c=365$ days, $c=182$ days, or less if c is measured in work days), the largest entry in the matrix $\mathbf{B}$ will have shrunk to 0.01% of its original level during the first 40 days of the first up-cycle. By the time the first down-cycle starts at $t=183$ days, there will be no detectable effect remaining from the first cycle.

H. Transient Behavior of Work-In-Process Inventories

The transient behavior of the work-in-process inventories $\mathbf{Q}_t$ and their expectations $\mathrm{E}(\mathbf{Q}_t)$ is also described by the above discussion since $\mathbf{Q}_t=\mathbf{D}^{-1}\mathbf{P}_t$ and $\mathrm{E}(\mathbf{Q}_t)=\mathbf{D}^{-1}\mathrm{E}(\mathbf{P}_t)$, for all t. Thus the expected WIP levels and the WIP transients, T_{Q_t}, are:

For the one-period increase:

(4.29) $$\mathrm{E}(\mathbf{Q}_t) = \mathbf{D}^{-1}(\mathbf{I}-\mathbf{\Phi})^{-1}\boldsymbol{\mu} + \mathbf{D}^{-1}\mathbf{B}^{t-1}\mathbf{D}\boldsymbol{\delta}$$

(4.30) $$\mathrm{T}^{(1)}_{Q_t} = \mathbf{D}^{-1}\,\mathbf{B}^{t-1}\,\mathbf{D}\boldsymbol{\delta},$$

For the step increase:

(4.31) $$E(Q_t) = D^{-1}(I-\Phi)^{-1}\mu + D^{-1}(I-B^t)(I-\Phi)^{-1}\delta$$

(4.32) $$T^{(2)}_{Q_t} = D^{-1}B^t(I-\Phi)^{-1}\delta.$$

For linear growth:

(4.33) $$E(Q_t) = D^{-1}(I-\Phi)^{-1}\mu + D^{-1}[(t+1)I - B^t - (I-B^t)(I-B)^{-1}](I-\Phi)^{-11}\delta$$

(4.34) $$T^{(3)}_{Q_t} = D^{-1}[I - (I-B)^{-1}]B^t(I-\Phi)^{-1}\delta$$

For the cyclical pattern;

(4.35) $$E(Q_t) = D^{-1}(I-\Phi)^{-1}\mu + D^{-1}(I-B^t)(I-\Phi)^{-1}\delta$$

(4.36) $$T^{(4)}_{Q_t} = D^{-1}B^t(I-\Phi)^{-1}\delta \qquad \textit{for } 1 \le t \le c;$$

(4.37) $$E(Q_t) = D^{-1}(I-\Phi)^{-1}\mu + D^{-1}B^{t-c}(I-B^c)(I-\Phi)^{-1}\delta$$

(4.38) $$T^{(4)}_{Q_t} = D^{-1}B^{t-c}(I-B^c)(I-\Phi)^{-1}\delta, \qquad \textit{for } c+1 \le t \le 2c; \textit{ and}$$

(4.39) $$E(Q_t) = D^{-1}(I-\Phi)^{-1}\mu + D^{-1}B^{t-2c}B^c(I-B^c)(I-\Phi)^{-1}\delta + D^{-1}(I-B^{t-2c})(I-\Phi)^{-1}\delta,$$

(4.40) $$T^{(4)}_{Q_t} = D^{-1}B^{t-2c}(I-\Phi)^{-1}\delta - D^{-1}B^{t-2c}B^c(I-B^c)(I-\Phi)^{-1}\delta, \qquad \textit{for } 2c+1 \le t \le 3c.$$

I. Appplication to the ECM Line

In applying Grave's Tactical Planning Model to the ECM line, three different scenarios of lead times for the thirteen sectors were tried in order to generate what seemed to be adequate smoothing of production levels. These three different sets of lead times are shown in Table 4.1 along with the eigenvalues from the resulting **B** matrices. The vectors of lead times at the top of the table list the lead times in numerical order according to the sector numbers in Figure 1.1. The eigenvalues are listed in order of decreasing modulus or absolute value.

In Scenario 1, the spectral radius is 0.6696 and the decay is rather rapid since within 8 periods, the $\mathbf{B}^t$ matrix will have decayed to approximately 5% of its original value. The behavior will be complex and include oscillations since the first four eigenvalues are equal in modulus and include complex numbers and negative values.

In Scenario 2, the spectral radius is larger, reflecting the increase in lead times. Decay of the transient will still be fairly rapid since within 14 periods the $\mathbf{B}^t$ matrix will have decayed to approximately 5% of its original value. Non-oscillatory decay will dominate, since the largest two eigenvalues are positive. The next two complex values have a modulus of 0.5362, so there will be some minor sinusoidal oscillations for a few periods.

As some lead times are increased further to create Scenario 3, the spectral radius increases to 0.8333. This lengthens the decay period: it will take approximately 17 periods for $\mathbf{B}^t$ to decay to 5% of its original level.

In conclusion, then, we can see that the ECM Line is a system which responds rapidly to changes in innovations under choices of lead times which were considered relevant to acceptable system behavior. Increasing some of the sectoral lead times does cause the time of adjustment to become longer, but within the range of lead times used in Table 4.1 the adjustment times still seem reasonable.

Table 4.1

Lead Times and Eigenvalues for Three Scenarios in the ECM Line

Scenario 1 Lead Times (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)

Scenario 2 Lead Times (1, 2, 1, 2, 2, 1, 1, 1, 1, 2, 1, 2, 5)

Scenario 3 Lead Times (1, 3, 1, 3, 3, 1, 1, 2, 1, 2, 1, 2, 6)

Scenario 1 Eigenvalues	Scenario 2 Eigenvalues	Scenario 3 Eigenvalues
0.6696	0.8000	0.8333
-0.6696	0.7440	0.7440
$-0.0000 + 0.6696i$	$0.1129 + 0.5242i$	0.6667
$-0.0000 - 0.6696i$	$0.1129 - 0.5242i$	$0.6667 + 0.0000i$
Remaining eigenvalues are essentially zero	0.5000	$0.6667 - 0.0000i$
	0.5000	$0.1129 + 0.5242i$
	$0.5000 + 0.0000i$	$0.1129 - 0.5242i$
	$0.5000 - 0.0000i$	0.5000
	-0.4697	0.5000
	Remaining eigenvalues are essentially zero	-0.4697
		Remaining eigenvalues are essentially zero

Chapter 5

Summary and Suggestions for Additional Research

We would now like to briefly summarize what has been presented in the last three chapters and make two suggestions for additional research.

In Chapter 2 we extended the usefulness of the TPM by developing a model for determining releases from demand forecasts and the present state of a system's work-in-process inventories. As first developed by Graves, the TPM was appropriate only to model a job shop in which releases are determined by orders in hand. This extension permits it to be employed to model a facility which produces in response to demand forecasts and in which releases must be made based on the forecasts, in advance of actual demand.

In Chapter 3, we developed two measures of service provided by a facility and showed how these are affected by management's choice of lead times. We discussed how management might proceed in evaluating the trade offs among the four behavior characteristics affected by lead times: probability that demand is met, average failure run length, production smoothing, and size of inventories (in-process plus safety stock).

Finally, in Chapter 4 we developed a model of the dynamic behavior of the sytem as it adjusts from one equilibrium level to another in response to changes in demand.

We can suggest two areas in which this work might be extended to further enhance the usefulness of the TPM. First, it would be advantageous to have more information about the distribution of failure run lengths since we have determined only the mean. The literature on stationary processes contains expressions for the variance of run lengths for a continuous time process

[Cramer and Leadbetter, pgs. 202-215]. With some work, this could probably be adapted to the discrete time process contained in the TPM.

Second, the responsiveness of the various transient paths of Chapter 4 to changes in lead time has not been fully investigated in this thesis. A useful extension would be an evaluation of the way transient behavior would respond as lead times are varied. A full understanding of this would improve management decision making by clarifying the impact of lead time choices. Further research in this direction would require analysis of changes in eigenvalues of **B** in response to perturbations to elements of **B** created by lead time changes.

In closing, we express our hope that the extensions provided in this thesis will contribute to the usefulness of the TPM as a management planning tool.

References

Cox, D. R. and H. D. Miller, 1965.

The Theory of Stochastic Processeses, Chapman and Hall, New York.

Cramér, Harald and M. R. Leadbetter, 1967.

Stationary and Related Stochastic Processes, John Wiley and Sons, Inc., New York.

Gantmacher, F. R., 1959.

The Theory of Matrices, Vol. II. Chelsea Publishing Company, New York.

Graves, Stephen C., 1986.

"A Tactical Planning Model for A Job Shop." Operations Research 34, 522-533.

Karlin, Samuel, and Howard M. Taylor, 1975.

A First Course in Stochastic Processes, Second Edition. Academic Press, Inc., New York.

Lancaster, Peter and Miron Tismenetsky, 1985.

The Theory of Matrices, Second Edition. Academic Press, Inc., Orlando, Florida.

Luenberger, David G., 1979.

Introduction to Dynamic Systems: Theory, Models and Applications, John Wiley and Sons, Inc., New York.