In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

3/17/65
b

A STUDY OF MANAGEMENT CONTROL SYSTEMS WITH AN APPLICATION TO SEASONAL GOODS INVENTORY PROBLEMS

A THESIS

Presented to

The Faculty of the Graduate Division

by

Sang Hoon Chang

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

In the School of Industrial Engineering

Georgia Institute of Technology

February, 1967

A STUDY OF MANAGEMENT CONTROL SYSTEMS WITH AN

APPLICATION TO SEASONAL GOODS INVENTORY PROBLEMS

Approved:

Chairman

Date approved by Chairman: 4-27-67

ACKNOWLEDGMENTS

I would like to express my gratitude to the members of the thesis advisory committee: Dr. David E. Fyffe (Chairman), Dr. Harrison M. Wadsworth, and Dr. Joseph L. Hammond, Jr. The work from which this thesis evolved could not have been done without their advice and encouragement. Special credit is also due to Dr. Albert F. Hanken for his helpful criticisms and suggestions on the manuscript.

I am appreciative of the aid given by Professor Frank F. Groseclose (retired) and Dr. Robert N. Lehrer of the School of Industrial Engineering during the years 1961-66. I am also thankful for the aid given by the Systems Engineering Committee and Dr. Joseph L. Hammond, Jr., of the School of Electrical Engineering during the years 1964-67.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

Various forms of management control problems arise in Industrial Engineering; for example, production control, inventory control, and budget control. A common characteristic of these control problems is the control scheme; that is, the process of making decisions on the basis of *a priori* information so as to improve future performance of a system. In this sense, the functional scheme of a control process may be conceptualized by a feedback control analogue of physical systems. Although the techniques of control theory may be advantageously applied to study such management control problems, a preliminary consideration is needed before such an application can be made. Since the control theory techniques have been mainly developed for use in physical systems, one needs to carefully define the boundaries of management problems so as to fit the techniques to a given situation. In view of the complexities associated with management problems, it is desirable to have a procedure which can be used as a basis for modeling such management control problems.

The general objectives which underly this research are two-fold; (1) to analyze the common characteristics of those problems which are peculiar and important to the concepts of management control, and the procedures involved in management decision processes for planning and control; (2) to develop a theoretical frame of reference for use in modeling management control systems so that the study of management control problems may be consistently carried out with respect to the

overall problem situation, and a feedback control scheme to be used for determining analytical solutions for such management control systems.

The general objectives of this research are pursued by way of two specific tasks of investigation. The first task is concerned with management control problems in general, and the second is with an application of the general concepts to a specific problem of modeling for a seasonal goods inventory situation of retail firms.

According to the existing knowledge in the field, it appears that system theory offers the most helpful and logical basis for modeling complex situations. By making new interpretations of existing concepts in system theory, a concise and unified body of theory is formulated in this thesis which may be particularly useful in modeling management control problems. Two categories of modeling problems are recognized; namely, the problems of modeling the spatial boundaries and the dynamic boundaries of a situation.

Given a situation for management control, the first step in the modeling procedure is to define the spatial boundaries of a problem so that the problem can be structured as a system. Such a system may be modeled by considering the topics of hierarchical system structure, attributes, and system goals. In particular, the results of analysis may be used: (1) to model a heirarchical structure with respect to nontransferable attributes so that separable boundaries for the system, components, and environment can be identified; (2) to recognize the difference in modeling system equations with respect to energy attributes and information attributes; and (3) to formulate a single measure of performance for the system when the system is characterized with a multiplicity of goals.

Once the spatial boundaries of a system are defined, the subsequent step in the modeling procedure is to define the dynamic boundaries of the system process within the frame of the already defined spatial system boundaries. For this purpose, it is first necessary to identify the controllable and uncontrollable subsystems of a given problem situation so that the system state, control input, environmental input, and system objectives can be recognized. The multi-stage control process can then be described in terms of a state equation and an objective equation. The individual stage of the multi-stage control process can be further described in terms of the feedback control sequence which consists of measurement, estimation, computation, optimization, decision, and actuation.

The general modeling procedure is illustrated with an inventory situation of retail firms. First, the spatial boundaries of an inventory situation are defined so that the inventory problem can be recognized as a relatively isolated system within the overall organizational structure. Subsequently, the inventory system is modeled within the framework of multi-stage control processes.

In the formulation of a multi-stage control process, a method is required to estimate the statistical characteristics of a random process which underlies the system stage. In the case of the inventory control processes, this situation often pertains to the problem of obtaining the demand forecasts. The nature of the forecasting techniques used in inventory control may vary depending upon particular circumstances in a given situation. This research has developed a statistical method which may be used to estimate the demand probabilities for seasonal goods inventory items.

The most commonly used procedure in the literature is that which assumes the probabilities of demand are estimated before the beginning of a season. Such *a priori* estimates of demand probabilities are referred to as initial estimates. The procedure proposed in this thesis also accepts such initial estimates; however, a filtering procedure[1] is applied so that the initial estimation errors can be corrected as more data become available after the season begins. The filtering procedure is designed to re-estimate the seasonal demand; however, the re-estimated results can be also used to predict the subperiod demand for the season. Within this framework, the filtering problem of estimating the seasonal demand coincides with the prediction problem of estimating the subperiod demands. The proposed filtering procedure is very sensitive to the parameter values used in the model.

The general procedure for modeling and forecasting is subsequently applied to formulate a seasonal goods inventory control model of retail situations. The seasonal goods inventory problems have been solved in the literature for the case where re-estimates of demand probabilities are not allowed in the model. In practice, however, a seasonal period is often divided by a finite number of time points such that the estimates of demand as well as the determination of order quantities are allowed to take place at each of these time points.

---

[1]Shaw, L. G., "Optimum Stochastic Control," in J. Peschon (ed.), *Disciplines and Techniques of System Control*, Blaisdell, 1965.

In a recent publication, Murray *et al.*[2] considered a seasonal goods inventory model which allows the re-estimation of demand probabilities. However, their model is applicable only when the size of demand population is exactly known, since they assumed that the demand pattern follows the beta binomial probability function. However, the size of demand population is often unknown in the real situation of seasonal goods inventory problems. The linear feedback filter procedure does not require knowledge of the size of the demand population. On the other hand, the linear feedback filter procedure assumes that the historical data are available for the purposes of estimation. It seems that this assumption is reasonable and logical in view of a case study reported by Cyert *et al.*[3] and Hertz *et al.*[4] The formulation of the model has resulted to an adaptive optimization problem.

A specific inventory situation of retail firms is used in this study to provide a background for the theoretical analysis and development. The general outcome of the study may be applied to other situations in management control problems with appropriate modifications to meet specific characteristics of individual problems; for example, some additional research may be suggested for the following situations:

---

[2]Murray, G. R., Jr., and E. A. Silver, "A Bayesian Analysis of the Style-Goods Inventory Problems," *Management Science*, 1966, pp. 785-797.

[3]Cyert, R. M., and J. G. March, *A Behavioral Theory of the Firm*, Prentice-Hall, 1963.

[4]Hertz, D. B., and K. H. Schaffir, "A Forecasting Method for Management of Seasonal Style-Goods Inventories," *Operations Research*, 1960, pp. 45-52.

(1) when the level of acceptable performance is specified for the system; (2) when the time lag between the activities of individual stages of a multi-stage control process is significant; (3) when the demand generating subsystem can be regarded as a controllable subsystem.

CHAPTER I

INTRODUCTION

Background

Various forms of control problems arise in industrial engineering; for example, production planning, inventory control, and budget control. A common characteristic of these control problems is the control scheme; that is, the process of making decisions on the basis of *a priori* information so as to improve future performance of the system. In this sense, the functional scheme of a control process may be conceptualized by a feedback control analogue of physical systems.

Control theory was originally developed for automatic control of electrical and mechanical systems. Since the Second World War, the importance of control theory has received much attention, not only with respect to physical systems, but also biological and business systems.[1] Application of control theory to management problems was considered by many.[2] Some of the more important contributions were made by Holt *et al.*,[3] and Forrester.[4] Forrester's method of Industrial Dynamics has been widely accepted as an effective tool in the simulation approach to busi-

---

[1]Wiener (40).

[2]For a literature review, see Chang (5).

[3]Holt, *et al.* (20).

[4]Forrester (9).

ness problems. Although the simulation approach depends largely upon computer utilization, the importance of control theory as a theoretical foundation of Industrial Dynamics was well emphasized by Forrester.[5]

As it is currently known in the literature, the study by Holt *et al.* has made, perhaps, the most use of "classical" control theory concepts in the analytical development of production systems. They derived the certainty equivalence theorem,[6] which was used in connection with the quadratic cost function to determine optimum solutions for their problem. Although the quadratic performance indexes are commonly used in control theory,[7] their use in management problems is limited to restrictive cases. They are applicable only when the error cost of performance is proportional to the square of errors, which implies that both positive and negative errors are equally undesirable.

Although Holt *et al.* and many others have made use of the "classical" control theory concepts, the recent developments in "modern" control theory[8] have not yet been fully applied to the study of management control problems. In fact, the state-space approach and the optimization techniques of modern control theory are very well suited for studying management problems. The reason for this is that their use permits the formulation of a wide range of problems. This applies to

---

[5] *Ibid.*, pp. 355-356.

[6] Holt *et al.* (20) Chapter 6; also see Hadley (14), pp. 448-454, and Whittle (39), p. 137.

[7] Kalman *et al.* (22).

[8] Bellman (3); Tou (37).

both maximization problems and minimization problems which can be either linear or nonlinear as well as deterministic or stochastic, or even adaptive.

Since management systems are typically stochastic or adaptive, any attempt to use a control model involves the problem of obtaining *a priori* information on the underlying stochastic processes. One of the most commonly known methods in forecasting is Brown's exponential smoothing.[9] The other is the method of regression analysis.[10] Although these methods are well known in management and economics literature, they are not well suited to use within the frame of control theory. On the other hand, the spectral analysis, which is commonly used for prediction in control theory, is often not applicable to management control problems. In this sense, forecasting is often a critical problem in developing a control model for management problems.

Although the techniques of control theory may be advantageously applied to studying management control problems, there are preliminary considerations that must be dealt with before such an application can be made. Since the control theory techniques were mainly developed for use in physical systems, one needs to carefully define the boundaries of management problems so as to fit the techniques to the given situations.[11] The concepts of system theory may prove to be quite helpful

---

[9]Brown *et al.* (4); also see Winter (42) and Cohen (7).

[10]For example, see Johnston (21).

[11]See Mesoravic *et al.* (28).

for use in defining complex situations of management problems of the real world. Although system theory has received much interest in various publications, it has not been fully introduced in the area of management control problems.[12]

Study Objectives

Industrial engineers are often faced with various forms of management control problems. The general objectives which underly this research are two-fold:

1. To analyze: (a) the common characteristics of those problems which are peculiar and important to the concepts of management control, and (b) the procedures involved in management decision processes for planning and control.

2. To develop: (a) a theoretical frame of reference for use in modeling management systems so that the study of management control problems may be consistently carried out with respect to the overall problem situation, and (b) a feedback control scheme to be used for determining analytical solutions for such management systems.

The general objectives of this research are pursued by way of two specific investigations. The first is concerned with management control problems in general, and the second with an application of the general concept to a specific problem situation in management control. The latter investigation is largely devoted to the development of a seasonal goods inventory model which gives a realistic representation

---

[12]See Hall (16); also Goode *et al.* (11).

of the inventory situation in practice. The method of forecasting the seasonal goods demand is recognized as a critical problem in the development of such a model.

Study Procedure

This research begins with a consideration of system theory for modeling management control problems, then proceeds to the development of a seasonal goods inventory model. As illustrated in Figure 1, this thesis consists of four main chapters which are concerned with the following specific problem areas:

1. Chapter II gives an interpretation of the existing concepts in system theory which may be particularly helpful in modeling management control problems. Since management problems are predominantly influenced by factors which arise from socio-economic considerations, an attempt is made to analyze the relationship between a purely physical system and a management system. A knowledge of such relationship may be useful in applying the control theory concepts to modeling management control problems.

2. Chapter III is primarily concerned with the description of a dynamic model for management control problems. The system theory of Chapter II is applied, first, to define the spatial boundaries of an inventory system so that the system can be modeled as a relatively isolated subsystem of a larger problem. The department store is used to provide a prototype example of the retail situation. After having defined the spatial boundaries of a system, Chapter III considers the problems involved in defining the dynamic characteristics of a multi-

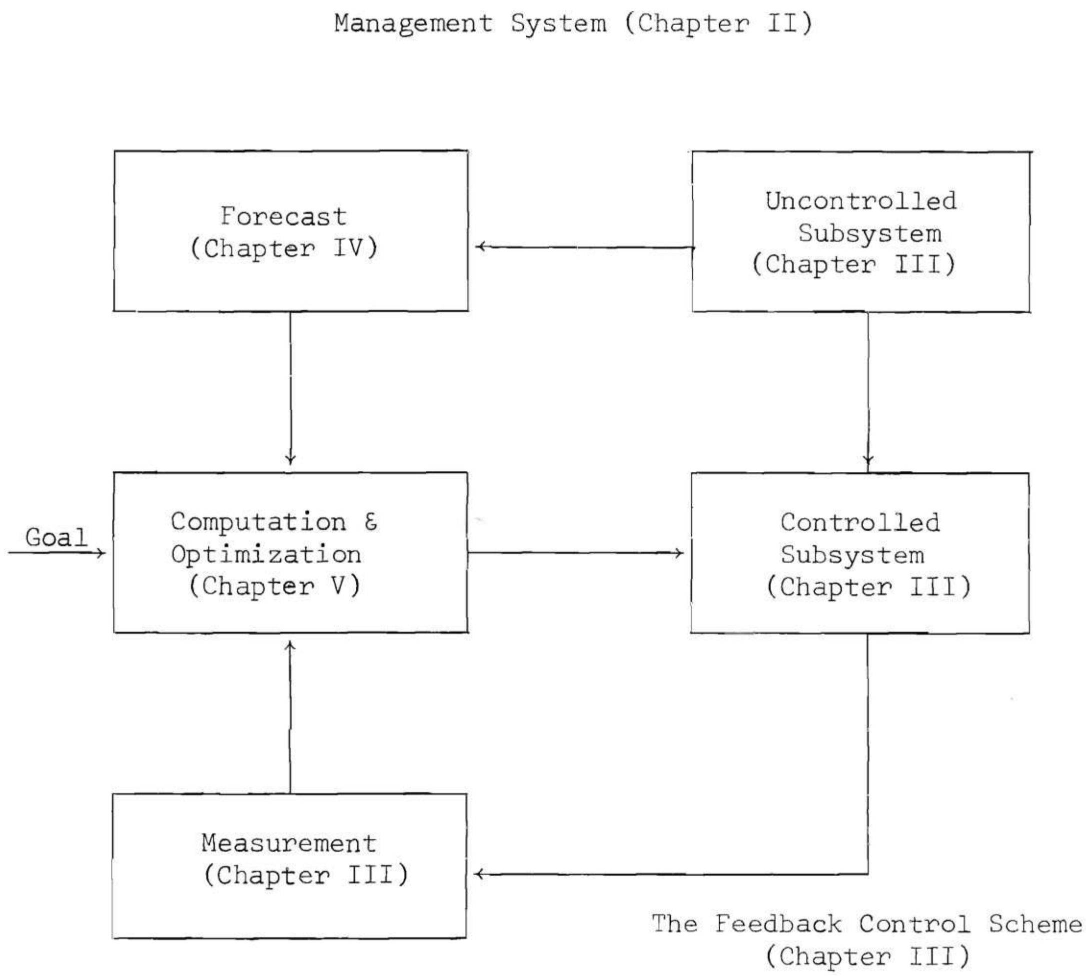

Figure 1. A Feedback Control Scheme. (The Numbers in the Parentheses Refer to the Appropriate Chapters Where the Indicated Topics are Mainly Discussed.)

stage decision process. By applying the system theory of Chapter II as well as the existing knowledge in control theory, a procedure for constructing dynamic models for management control problems is described. The description of the general procedure is subsequently applied in Chapter V to model the seasonal goods inventory problem.

3. The statistical considerations needed for forecasting are first analyzed in Chapter IV, and subsequently a method of forecasting which can be conveniently applied to the seasonal goods inventory problem is developed in the chapter.

4. The general dynamic model of Chapter III and the forecasting method of Chapter IV are applied to the development of a seasonal goods inventory model in Chapter V. The model recognizes forecasting as an integral part of the multi-stage control process, so that at each time point re-estimates of demand are allowed within the control scheme.

CHAPTER II

MODELING MANAGEMENT PROBLEMS FROM A GENERAL SYSTEMS VIEWPOINT

General

Because of the complexity of real world problems, it is often necessary to carefully consider the problem of modeling for given problem situations. It is commonly recognized that system theory provides a useful basis for modeling complex situations. The objective of this chapter is to review and interpret the known concepts in system theory in order to formulate a framework of system theory which may be particularly useful to define the spatial boundaries for modeling management control problems. This objective is pursued by considering the topics of system structure, attributes, and goals.

Management Systems

In recent years, the importance of system theory has received much attention in various publications. Since, in such publications, the word "system" is frequently used to represent many possible systems, it is desirable to make a specific definition of the term "management system" which can be consistently used to discuss management control systems. For the purposes of this study, a system may be considered as belonging to one of two categories: namely, the naturally existing systems and the man-made systems. For example, the solar system is a naturally existing system; and an inventory system may be regarded as a man-made system. A management system shall be regarded as a man-made system which exists for the purpose of satisfying certain specific needs of man.

Let S be a system which exists for the purpose of satisfying a specific need N^*. In this situation, it is appropriate to describe the system S with reference to the need N^*. Assume that S is an "open" system; that is, it has both input and output. Let ε and ζ denote the input and output of S, respectively, as shown in Figure 2.

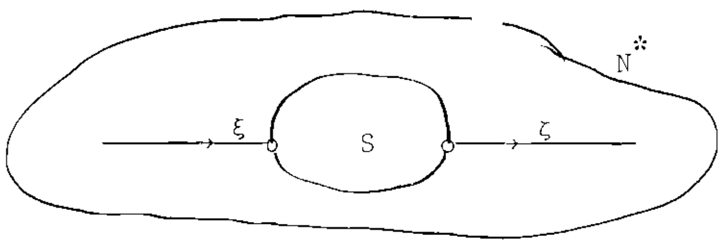

Figure 2. A Schematic Description of System S.

According to Gosling,[1] it is convenient to think of a system as being enclosed within an imaginary boundary which separates the system from its surrounding environment. Suppose there are two imaginary terminals on the system boundary such that one of them serves as an input terminal and the other as an output terminal. The two-terminal system of Figure 2 is said to be unilateral in the sense that its input and output do not reverse their direction of flow. For a unilateral

[1]Gosling (12), p. 12.

system, the input-output relation may be expressed in the form:

$$\zeta = S(\xi, N^*) . \tag{2-1}$$

In the expression above, the input ξ represents some valuable resources which are expended in the system, and the output ζ represents some useful product which the system is required to produce as specified by the given need N^*. The problems of properly identifying the need N^*, analyzing the output ζ, and determining the input ξ are the fundamental considerations involved in management systems.

The discussion above concerns a system that is viewed as a single entity. However, a system is usually composed of two or more parts which are interconnected in such a way that the overall function of the system is the interrelated product of those parts within the system. Such parts are sometimes referred to as subsystems, components, or elements of the system. Sometimes, the structure of a system is such that many smaller parts can be recognized within a part of the system. Such a system is said to have a hierarchical structure.

In a study of organization theory, Simon[2] has observed that most real organizations have hierarchical structures. The degree of hierarchy and the efficiency of system information are often considered in connection with the problem of centralization and decentralization.[3]

---

[2]Simon (35), p. 41.

[3]See Zannetos (43).

In a study of adaptive behavior of living organisms, Ashby[4] has observed that the behavioral pattern of animate beings can be explained by the efficiency of the hierarchical structure of their internal parts. Such observations may be extended and applied to model a hierarchical structure for management systems.

In systems literature,[5] the structure of a system is sometimes described in terms of the universe, environment, system, subsystem, components, and elements A unique definition of these terms, which may be conveniently used to model management system problems, will be made in the following:

Given a "problem," let the *universe* U^* be the problem itself. On this universe, suppose it is possible to define a *system* S and its environment $\bar{S}$ such that $\bar{S}$ is the complementary set of S; i.e.

$$S \bigcup \bar{S} = U^* . \tag{2-2}$$

where $\bigcup$ denotes the union. The internal structure of a system S is then defined by a finite number of k components, k 1, which are disjoint to one another. If C_i denotes component i of the system S, i = 1, 2,..., k, then it follows that:

$$\bigcup_{i=1}^{k} C_i = S \tag{2-3}$$

---

[4]Ashby (2), pp. 148-153.

[5]Hall (16); Hammond *et al.* (17).

$$\bigcap_{i=1}^{k} C_i = \phi \,, \tag{2-4}$$

where $\bigcap$ denotes the intersection, and ϕ denotes an empty set. A *subsystem* of S is defined as any subset of the system S. According to the order of system hierarchy, subsequently smaller parts may be recognized within a component. Such smaller parts can be defined as *elements* of the system S.

The components of S defined above are idealized subsystems of S which are finite in number and disjoint to one another. Because of the interacting forces acting among the parts of a system, such isolated components of a system may not practically exist. The definition may be justified, however, if the system is defined with respect to the "non-transferable attributes." This topic will be further discussed in the following section.

System Attributes

In the previous section, the management system S is characterized as having both input and output. The input and output of a system may be referred to as system attributes. For the purpose of modeling a system problem, the system attributes may be categorized into two classes: namely, the *transferable* and *nontransferable* attributes.[6] The transferable attributes are those quantities which can be described, for

[6]This classification was originally made by J. L. Hammond. See Hammond *et al.* (17); also see Gosling (12), p. 11, for a discussion of "transfer properties."

example, by movement, flow, or force; hence, they are usually expressed in terms of vector quantities which have both magnitude and direction. On the other hand, the nontransferable attributes may be regarded as those properties which can be described over a fixed time interval; hence, they are usually expressed in terms of scalar quantities.

Depending upon one's point of interest, a system equation can be modeled with respect to either the transferable attributes or the nontransferable attributes. The system equation is often formulated in the form of a differential equation with respect to the transferable attributes. In this case, the focus of analysis is usually placed upon the dynamic characteristics of a system. On the other hand, the performance measure of a system, such as the quadratic criterion, or the objective function of linear and dynamic programming, is usually expressed in terms of nontransferable attributes.

In the preceding section of this chapter, a system S was defined as consisting of a finite number of disjoint components. Such disjoint, separable components can be defined when the focus of analysis is placed upon the nontransferable attributes of a system. There is no loss of generality caused by this restriction, sin e management problems in the final analysis are always concerned with the evaluation of system performance, and all the relevant system attributes can be considered in terms of nontransferable attributes.

Once the system structure is modeled in terms of nontransferable attributes, the next step is to analyze the system behavior in terms of transferable attributes. For the purpose of modeling system equations, the transferable attributes of the system may be classified into two categories: namely, the *energy attributes* and the *information attributes*.[7] These two attributes can be distinguished by the fact that the information attributes contain little or no energy. It appears that the consideration of energy attributes is needed when the focus of analysis is placed upon the work aspect of a system, and the consideration of information attributes is needed when the focus of analysis is placed upon the control aspect of a system.

At this point, it is appropriate to introduce the four-port and two-port representation of a system. According to Koenig *et al.*,[8] a system can be modeled as having four-port terminals with respect to the energy attributes, and as having two-port terminals with respect to the information attributes. Figure 3 shows a schematic diagram of a four-port terminal model and a two-port terminal model.

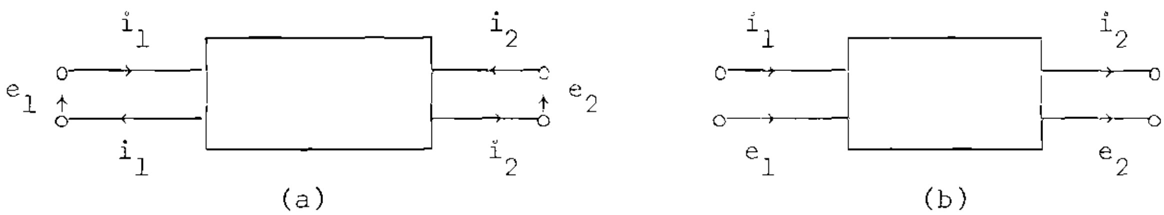

Figure 3. Modeling System Attributes with:
(a) Four-port Terminals and
(b) Two-port Terminals.

Four terminals are used for modeling the system energy attributes, since two terminals are needed to describe the input, and another two

---

[7]See Hammond *et al.* (17).

[8]Koenig *et al.* (23); also see Hammond *et al.* (17).

terminals are needed to describe the output. Two terminals are needed to describe the input or the output of energy attributes: namely, one terminal for the "level" of the energy and the other terminal for the "flow rate" of the energy. In other words, the energy attributes can be described only jointly by means of the level and the flow rate. For example, both the voltage (level) and the current (flow rate) are needed to describe the energy attributes of an electric circuit. It is also possible to describe a management system model in terms of energy attributes. For example, suppose the symbols of Figure 3 can be interpreted as follows:

e_1: the level of total investment of a firm.

i_1: the rate of investment return of the firm.

e_2: the level of inventory investment of the firm.

i_2: the rate of inventory investment return of the firm.

Since the level of the investment and the rate of the investment return are dependent on one another, such attributes of the firm may be represented in terms of energy attributes. The four-port terminal representation, therefore, places in evidence the interrelations among all four attributes: i.e., e_1, i_1, e_2, and i_2. A system equation for the four-port model may be expressed as:

$$f(e_1, e_2, i_1, i_2) = 0 \tag{2-5}$$

Although the four-port terminal model gives a logical representation of system energy attributes, it is generally used when the internal structure of a system is exactly known. When the internal structure is not exactly known, it is difficult or impossible to construct a system equation of the form of Equation (2-5). In the case of some simple physical systems, such as an electric circuit, the interdependency of system energy attributes can be often determined deterministically. However, in the case of complex systems, such as a management system, it is usually impossible to give a deterministic description of system attributes in the form of Equation (2-5).

Under certain assumptions, the system energy attributes may be modeled with a two-port representation. For example, suppose it is possible to assume an independency between the attributes e_1 and i_1 as well as between e_2 and i_2 for the system of Figure 3(a). Under this assumption, the system attributes may be represented by the two-port model of Figure 3(b). A set of system equations for the two-port model may be expressed as:

$$\begin{aligned} e_2 &= f_1(e_1, i_1) \\ i_2 &= f_2(e_1, i_1) \; . \end{aligned} \tag{2-6}$$

A comparison of Equations (2-5) and (2-6) indicates that Equation (2-6) is restricted by the assumption that e_2 and i_2 are independent; whereas, such restriction is not needed in Equation (2-5).

Since information attributes are free of energy considerations, a system equation with respect to information attributes can be always modeled with a two-port terminal representation.

System Goals

The system structure and the system attributes were discussed in the preceding sections. This section is concerned with the topic of system goals. In considering management control problems, one often presupposes that there exists a single goal or a single measure of performance which serves as a basis for evaluating the system behavior. For a complex system with many components or individual groups, there can be many possible component goals or individual goals within the system. When this is the case, it is of interest to analyze the relationship between the system goal and the component goals.

The problems associated with multiplicity of goals is a subject of much interest in organization theory.[9] In a recent publication, Mesarovic *et al.*[10] introduced the concept of a multi-level-multi-goal system. This work may be briefly summarized as follows. When a system is structured in a hierarchical order, it is appropriate to recognize a multiplicity of levels of goals as well as a multiplicity of goals. The *level* of goals is defined so that a higher level goal dominates its lower level goals. In other words, the lower level can be regarded as a subset of the higher level goal. In the terminology of Mesarovic *et al.*, for example, a single-level-single-goal system can be a system with many goals, but none of the goals dominates any other goal of the system. The simplest system of this type is a single-goal system.

---

[9] See, for example, March *et al.* (26) and Cyert *et al.* (8).

[10] Mesarovic *et al.* (28).

According to Mesarovic *et al.*, a single-goal system can be regarded as consisting of two subsystems: namely, the *causal unit* and the *goal-seeking unit*. A single-goal system with causal unit P and goal-seeking unit G is schematically described in Figure 4. The causal unit is often called a plant in control theory, from which the symbol P is derived. The control input to P from G is denoted by ξ_G, and the plant input is denoted by ξ_P. The letter ζ_P denotes the plant output, and η denotes the plant performance observed by G. The small letter g denotes the system goal.

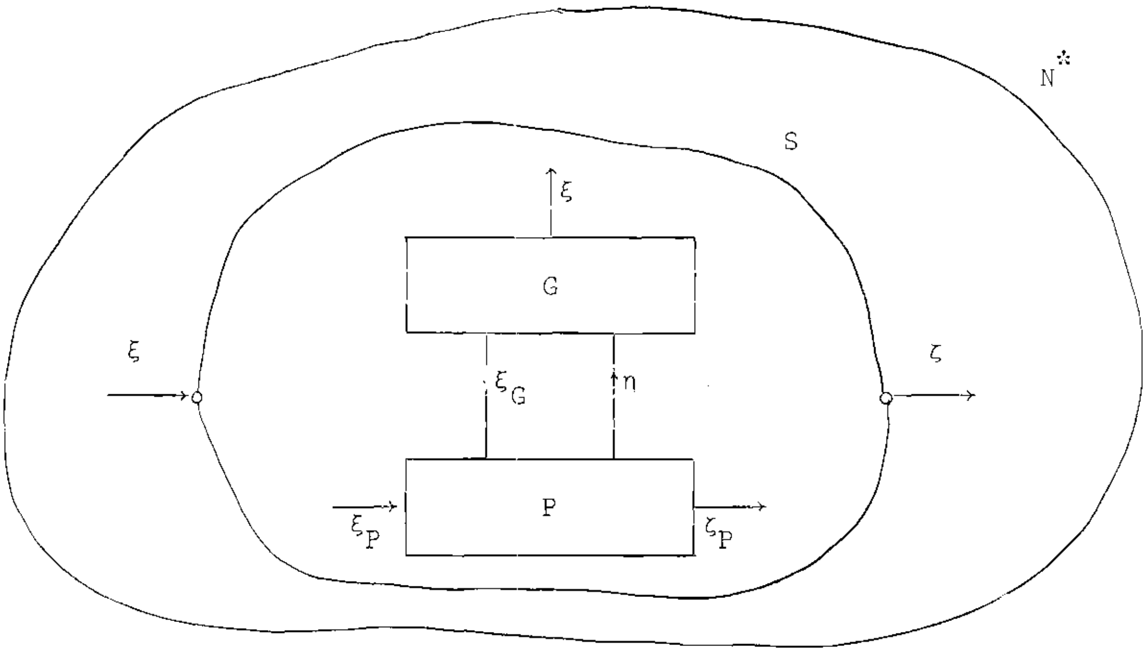

Figure 4. A Single-goal System with Causal
Unit P and Goal-seeking Unit G.

Making use of the notation introduced above, a system goal g may be expressed in a functional form:

$$g = g(\eta, \xi_G; N^*) , \tag{2-7}$$

where g is expressed as a function which depends on the need N^*, the system performance η, and the control input ξ_G. When the system S is a "pure" control system, such as a servo-mechanism, then g is usually expressed in the form of a reference input.[11] In this case, the reference input is regarded as a signal, and the task of the goal-seeking unit is to properly identify the signal. On the other hand, there are many control systems which have no reference input. For example, it may happen that the goal of a management system can not be regarded as a reference input, but rather is established by the goal-seeking unit G. Once the system goal is established, the subsequent task of goal-seeking unit G is to determine the control input ξ_G so as to have the plant produce some desired output ζ_P.

In modeling control problems, one often presupposes the existence of a single measure of performance, or a single-goal for the system. It is possible to have a single measure of performance for a purely physical system, when there are no interacting goals within the system. In modeling management control problems, however, it may be necessary to carefully examine the multiplicity of goals of the system before making

---

[11] In organization theory, "reference input" is equivalently referred to as "aspiration level." See March *et al.* (26).

such an assumption. This consideration is of particular importance when human elements are included in a system. Since a single measure of performance is needed in order to analyze an overall system problem, it may be sometimes necessary to reduce the multiplicity of goals of the system to an appropriate single goal. The following procedure may be used when it is necessary to reduce the complexity of a single-level-multi-goal system to a single-goal system.

Suppose there are k "component" goals in a single-level-multi-goal system S. Let $G_1, G_2, \ldots, G_i, \ldots, G_k$ denote the component goals and G_0 denote a system goal which represents a set containing all component goals of the system. As shown in Appendix 1, the system goal G_0 can be either an unordered set or an ordered set. An unordered system goal may be described as an unordered set:

$$G_0 = \{G_1, G_2, \cdots, G_i, \ldots, G_k\} \ . \tag{2-8}$$

When it is possible to rank the order of preference among component goals, then the system goal may be expressed as an ordered set:

$$G_0 = (G_{(1)}, G_{(2)}, \cdots, G_{(k)}) \ , \tag{2-9}$$

where the parenthesized subscripts refer to the order of ranking among the component goals. When all component goals are identical to one another, then the system can be regarded as a single-goal system.

Summary

This chapter has reviewed the known concepts in system theory, and analyzed the problem of modeling management control systems with respect to the three main topics: system structure, system attributes, and system goals. The results of study may be used for the following:

1. To define a specific management system so that its hierarchical structure is modeled with respect to nontransferable attributes.

2. To recognize the differences in modeling system equations with respect to energy attributes and information attributes.

3. To formulate a single measure of performance for a system when the system is characterized with a multiplicity of goals.

Once the spatial boundaries of a system are identified, the subsequent problem in modeling is to define the dynamic boundaries of the system. This topic is discussed in the following chapter.

CHAPTER III

MODELING FOR AN INVENTORY CONTROL SYSTEM

General

An inventory system may be regarded as a subsystem when it is viewed from an overall organizational standpoint. In order to make a systems approach to modeling an inventory problem, two main considerations must be dealt with. First, it is necessary to define the spatial boundaries of the system in order to place in evidence the effects of organizational constraints which act upon the given inventory situation. Second, it is necessary to define the dynamic boundaries of the system process in order to analyze and evaluate its time dependent behavior. The general concepts for modeling management systems were discussed in Chapter II. These concepts are applied in this chapter to define the spatial boundaries of an inventory problem of retail firms. Subsequently, an inventory process is described as a multi-stage control process.

Modeling a Retail Inventory Situation

Industrial firms may be categorized as being either retail or manufacturing firms. The main business of retail firms is characterized by the activities of buying goods from producers or wholesalers and selling those goods to customers. Although retail firms often manufacture or process some of their goods, such activities are only incidental or subordinate to the main activities of buying-to-sell. In this manner, the buying and selling activities are closely integrated in retail firms. This contrasts significantly with manufacturing firms where production and marketing activities are usually separated within the organization. For this reason, the inventory control situations of retail firms and manufacturing firms have somewhat different characteristics. For example, while the objective function of an inventory system of retail firms may be expressed in terms of maximizing the net revenue, the objective function of an inventory system of manufacturing firms may be expressed in terms of minimizing the relevant inventory cost. In view of such differences, an inventory system of a typical retail firm is considered in this chapter. For this purpose, the department store will be regarded as a typical retail firm.

The essential characteristics of a department store can be described as follows. The market structure of a department store can be regarded as an oligopolistic competition if the store is relatively small.[1] The store sales closely reflect the state of economy in the form of disposable personal income. As a matter of fact, the store sales seldom exceed a certain fraction of the disposable personal income of a given consumer population.[2] For this reason, expanding and maintaining a market share is one of the most important goals of a department store.

---

[1]Cf. Holdren (19).

[2]According to Snyder (36), this fraction is approximately 7 per cent.

The market share may be regarded as a measure which represents a store's utility to the consumer public. The utility may be attributed to the three factors:[3] quality, availability, and accessibility of the consumer goods which the store offers to the public. The quality may depend upon the price, reliability, and the degree of customer satisfaction of the items sold by the store. Availability refers to the variety and quantity of commodities, and the range of choices offered to consumers. An inventory problem can be regarded as a subproblem of the general problem concerned with such availability. Accessibility depends on considerations such as: multi-departmental effects, advertising, credit policies, store location, parking facilities, etc. These three utility factors jointly influence a consumer's concept of the store's reputation as well as the store's market share.

In order to consider the problem of modeling an "inventory control system" for the retail situation described above, the terms: state, control input, environmental input, controllability, and observability are needed for the discussion. According to MacFarlane,[4] *state* is defined as:

> A state of a physical object is a quantitative measure of a physical condition of the object which remains unchanged with lapse of time if the object is suitably isolated.

MacFarlane's definition of state can be conveniently used to describe the physical conditions of a system in terms of its state at a certain time. Previously in Chapter II, the system attributes were discussed

---

[3]Cf. Regan (33).

[4]MacFarlane (25), p. 12.

in terms of input and output. However, in order to place in evidence the dynamic characteristics of a system, it is more convenient to represent system attributes by input, state, and output. For example, the physical conditions of a system may be described by many variables. Depending upon one's point of interest, a specific variable or a set of variables can be selected among many possible system variables to define a state or a state vector for the system. Once a system state has been defined, it can be used as an intermediate variable to relate the effect of the input upon the output. The state is changed by the input, and the output is an observation of the state.

The input which acts upon the system state may be recognized either as a *control input* or as an *environmental input*. The control input is deliberately exerted upon the system in order to transform its state into a more desirable one. On the other hand, the environmental input is an exogenous force which affects the system state, but is not subject to a control.

At this time, it is appropriate to consider the concept of controllability and observability. According to Gilbert,[5] a system can be partitioned into four possible subsystems: namely, S_A, S_B, S_C, and S_D which are designated as:

S_A: the controllable and observable subsystem.

S_B: the uncontrollable but observable subsystem.

S_C: the controllable but unobservable subsystem.

---

[5]Gilbert (10). Originally, Gilbert used these terms to discuss a linear deterministic system.

S_D: the uncontrollable and unobservable subsystem.

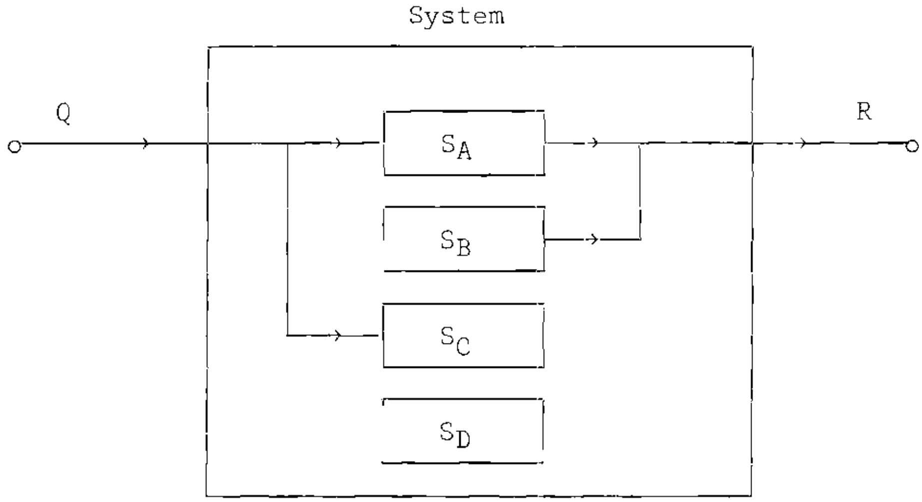

Figure 5. A System Partitioned into Four Subsystems.

Figure 5 shows a two-part representation of a system partitioned into four such subsystems. In the figure, Q denotes the control input, and R denotes the output. The control input is shown in connection with the controllable subsystems S_A and S_C, and the output is shown in connection with the observable subsystems S_A and S_B.

The concepts described above may be used to define the boundaries of the inventory control system of a department store. Given a department store situation, the store activities may be partitioned into four possible classes of activities which can be referred to as subsystems. Among these subsystems, the inventory control system may be defined as the controllable and observable subsystem S_A of the overall system. For such an inventory system, the system state can be designated as the levels of inventory at a given point in time. The system is controllable by means of inventory replenishment, and the inventory levels are observable. In other words, the system states, i.e., the inventory levels are both controllable and observable. Once the inventory system has been modeled and shown to be controllable and observable, then all other activities of the store can be categorized into subsystems which are uncontrollable and/or unobservable. For example, the "demand" factor can be regarded as a subsystem S_B which is observable but uncontrollable: hence, demand plays the role of an environmental input for the inventory control system. As another example, the cash level may be controlled partially by restricting the amount of inventory replenishment. When the cash level is not considered as a part of the control system, then it may be regarded as a subsystem S_C. All other activities of the store which are irrelevant to the inventory control problem may be relegated to the uncontrollable and unobservable subsystem S_D.

Modeling for a Multi-Stage Control Process

Having defined the spatial boundaries of a control system, one can proceed to model the control process within the defined spatial boundaries. In modeling a control process, it is necessary to link the present state of the system with the past and future states of the system. A general model of a multi-stage control process, which is subsequently used to develop a seasonal goods inventory control model in Chapter V, is considered in the following discussion.

Suppose a control process is considered over a planning horizon which covers a time interval (t_o, t_n). This time period may be divided into a finite number, say n of subperiods such that the system state at each given point in the subperiods can be bserved and controlled. For subperiod i, i=1, 2,..., n, the system state, input, and output are schematically represented in Figure 6.

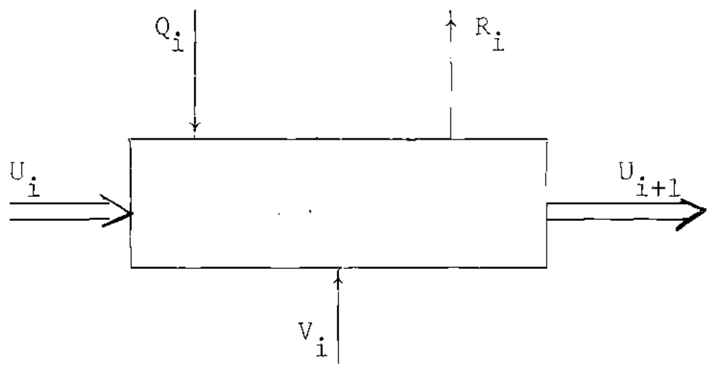

Figure 6. A System Representation for Subperiod i.

The symbols used in the figure are interpreted as follows. The transition of the system state from one subperiod to the next is indicated by double lines with direction arrows. The present state of the system is designated as U_i. The system inherits the present state from the previous (i-1)-th subperiod. The solid single lines with directional arrows indicate the system input. The control input is designated as Q_i, and the environmental input is designated as V_i. The dotted line with a directional arrow indicates the system output which is designated as R_i. The dotted line is used to denote the fact that the system output R_i is a scalar variable which represents the system utility for subperiod i.

In general, the output R_i can be expressed as a single-valued function of state and input; that is,

$$R_i = R_i(U_i, Q_i, V_i) . \tag{3-1}$$

The expression above is sometimes referred to as the objective function of the system. The transitional relation between states U_i and U_{i+1} may be expressed in the form:

$$U_{i+1} = T_i(U_i, Q_i, V_i) . \tag{3-2}$$

The expression above is commonly referred to as the state equation of the system.

In Chapter II the transferable and nontransferable attributes of a system were discussed. The input, state, and output are system attributes. Among those attributes shown in Figure 6, the state and inputs may be regarded as transferable attributes, since they have transferable or transitional effects upon one another. On the other hand, the output R_i may be regarded as a nontransferable attribute, since it has no direct effect upon the other attributes of the system.

When a system is modeled on the basis of nontransferable attributes, as was discussed in Chapter II, then the system can be structured so as to consist of a finite number of disjoint components. Such an analysis can be applied to structure the system model for a multi-stage process. For instance, suppose a control process is considered over a planning horizon (t_o, t_n). When the planning horizon consists of n subperiods, then it may be possible to model the control process as a system S which consists of n components. Suppose the component attribute is described in terms of its nontransferable attribute R_i, then it follows from Equations (2-3) and (2-4) of Chapter II that:

$$\bigcup_{i=1}^{n} R_i = R \tag{3-3}$$

$$\bigcap_{i=1}^{n} R_i = \phi , \tag{3-4}$$

where R denotes the total outcome of the system over the entire planning horizon. In general, R can be expressed as:

$$R = R(R_1, R_2, \ldots, R_i, \ldots, R_n) . \tag{3-5}$$

In particular, when R_i is represented by a nontransferable attribute, then the disjointness or separability condition can be applied, so that R can be expressed by a more convenient form of Equation (3-3).

In order to make a dynamic programming formulation of a multi-stage process, let G_i denote the partial sum of the total output which is defined as:

$$G_i = G_i(R_i, R_{i+1}, \ldots, R_n) \tag{3-6}$$

$$= \sum_{j=1}^{n} R_j .$$

When the separability condition is applied to the expression above, it follows that:

$$G_i = G_i[R_i, G_{i+1}(R_{i+1}, R_{i+2}, \ldots, R_n)] = R_i + G_{i+1} . \tag{3-7}$$

Now, let $f_i(U_i)$ denote the optimum output that can be expected from the system over subperiod j, j-i, i+1, i_2,..., n, provided the optimum control inputs Q_j^* are used for all j subperiods. Then $f_i(U_i)$ can be written as:

$$f_i(U_i) = \underset{Q_i, Q_{i+1}, \ldots, Q_n}{\text{Max}} \{G_i[R_i(U_i, Q_i, V_i), \ldots, R_n(U_n, Q_n, V_n)]\} \tag{3-8}$$

$$= \underset{Q_i, Q_{i+1}, \ldots, Q_n}{\text{Max}} \{G_i[R(U_i, Q_i, C_i), G_{i+1}]\} .$$

If G_i is a monotonically nondecreasing function of G_{i+1} for every R_i, then:[6]

$$f_i(U_i) = \underset{Q_i}{\text{Max.}}[R_i(U_i, Q_i, V_i), \underset{Q_i+1, Q_{i+2}, \ldots, Q_n}{\text{Max}} (G_{i+1})] \tag{3-9}$$

$$= \underset{Q_i}{\text{Max}}[R_i(U_i, Q_i, V_i), f_{i+1}(U_{i+1})]$$

---

[6]Nemhauser (31), p. 35.

$$= \underset{Q_i}{\mathrm{Max}}[R_i(U_i,Q_i,V_i) + f_{i+1}(U_{i+1})] .$$

For the n-th and last subperiod of the process:

$$f_n(U_n) = \underset{Q_n}{\mathrm{Max}}[R_n(U_n,Q_n,V_n)] . \tag{3-10}$$

The solution to the problem above may be obtained subject to the constraint:

$$Q_i^* \varepsilon \Omega(Q_i), \; i=1,2,\ldots,n . \tag{3-11}$$

where $\Omega(Q_i)$ denotes the allowable region of control inputs Q_i.

<u>Feedback Sequences in Control Processes</u>

The control process pattern which repeats itself in every subperiod of the multistage process may be described in terms of a feedback sequence. For subperiod i, consider a sequence of time points designated as i_a, i_b, i_c, and i_d at which the following activities may take place:

1. Measurement and estimation at time point i_a.
2. Optimization at time point i_b.
3. Decision making at time point i_c.
4. Actuation at time point i_d.

Making use of the symbols shown in Figure 6, the characteristics of these activities are described as follows.

1. Measurement and Estimation. At time point i_a, the state U_i of the system, inherited from subperiod (i-1), is observed. The state U_i is related to the previous state U_{i-1} by the state equation:

$$U_i \quad T_{i-1}(U_{i-1}, Q_{i-1}, V_{i-1}), \tag{3-12}$$

which is derived from Equation (3-2) by making appropriate adjustments on the subscripts.

In order to control the process for the i-th subperiod, knowledge of the characteristics of the environmental input V_i must be obtained so that an appropriate control input can be determined to optimize the i-th stage of the process. In order to optimize the control process for the entire planning horizon, however, a knowledge of the estimates for V_i is needed so that a set of optimum control inputs Q_j^* can be determined, for j=1, i+1,..., n.

The level of complexity involved in measurement and estimation often depends on the characteristics of the environmental input V_j. Consider the following three cases:

a. V_j is deterministic.

b. V_j is stochastic with known probability density functions.

c. V_j is stochastic with unknown probability density functions.

The simplest of these three is the deterministic case. For the second case, a knowledge of the probability density function p_j is assumed to be available for all j, where:

$$p_j = p(V_j, a_j), \quad j=1, i+1,..., n. \tag{3-13}$$

In the expression of the probability density function p_j shown above, V_j denotes the underlying random variable, and a_j denotes the parameters of the density function. A commonly made assumption in this case is that p_i is independent for all i, i-1, 2,..., n. When p_i is not independent, then a knowledge of the joint probability density function:

$$p(V_1, V_2, \ldots, V_n; \; a_1, a_2, \ldots, a_n) \tag{3-14}$$

is needed for estimation.

The most complex situation arises when the environmental input originates from a stochastic process with unknown statistical characteristics. Suppose the form of the probability density function p_j is known but the parameter values are unknown. In this case, past observations on the random variable V_i can be used to generate statistical estimates for the unknown parameters. Let X_{i-1}^N denote a set of N observations:

$$X_{i-1}^N = (V_{i-N}, V_{i-N+1}, \ldots, V_{i-1}) \; . \tag{3-15}$$

Suppose X_{i-1}^N is available at time i_a. On the basis of X_{i-1}^N, the estimate $\hat{a}_i$ may be obtained as:

$$\hat{a}_i = f_{i-1,j}(X_{i-1}^N), \; j=1, \; i+1, \ldots, n. \tag{3-16}$$

A specific problem of estimating demand for a seasonal goods inventory situation is considered in Chapter IV.

2. Optimization. At time point i_b, the stage of measurement and estimation is completed. The information obtained from this stage is now used to analyze the effect of each possible control vector Q_j, $j=i, i+1, \ldots, n$, on the future state of the process. The criterion of optimality, such as the objective function of Equation (3-9), may be used to determine the optimum control input Q_i^* for the i-th subperiod, $i=1, 2, \ldots, n$.

3. Decision. At time point i_c, the analysis is completed, and a decision is made to apply the optimum control input Q_i^* upon the system state.

4. Actuation. At time point i_d, the decision is implemented, and the control input Q_i^{**} takes actual effect upon the system state. When the implementation process is subject to errors, then the actual control input Q_i^{**} may not be identical with the optimum control input Q_i^*. At this point, the control sequence has completed its cycle for the i-th subperiod.

A schematic diagram of the control sequence is shown in Figure 7. The dashed lines denote the information loop which connects all the stages in the control sequence. Figure 8 gives an illustration of the time spacing between the stages of the control sequence. Figure 8(a) shows that all time points, i_a, i_b, i_c, and i_d, are closely located at the beginning of subperiod i. In Figure 8(b), they are shown as being widely scattered within the subperiod. In Figure 8(c), the point i_d is shown as being located in subperiod (i+1). In this case, the control is not actuated until after the i-th subperiod. Such a situation would arise in inventory control when the order replenishment or production

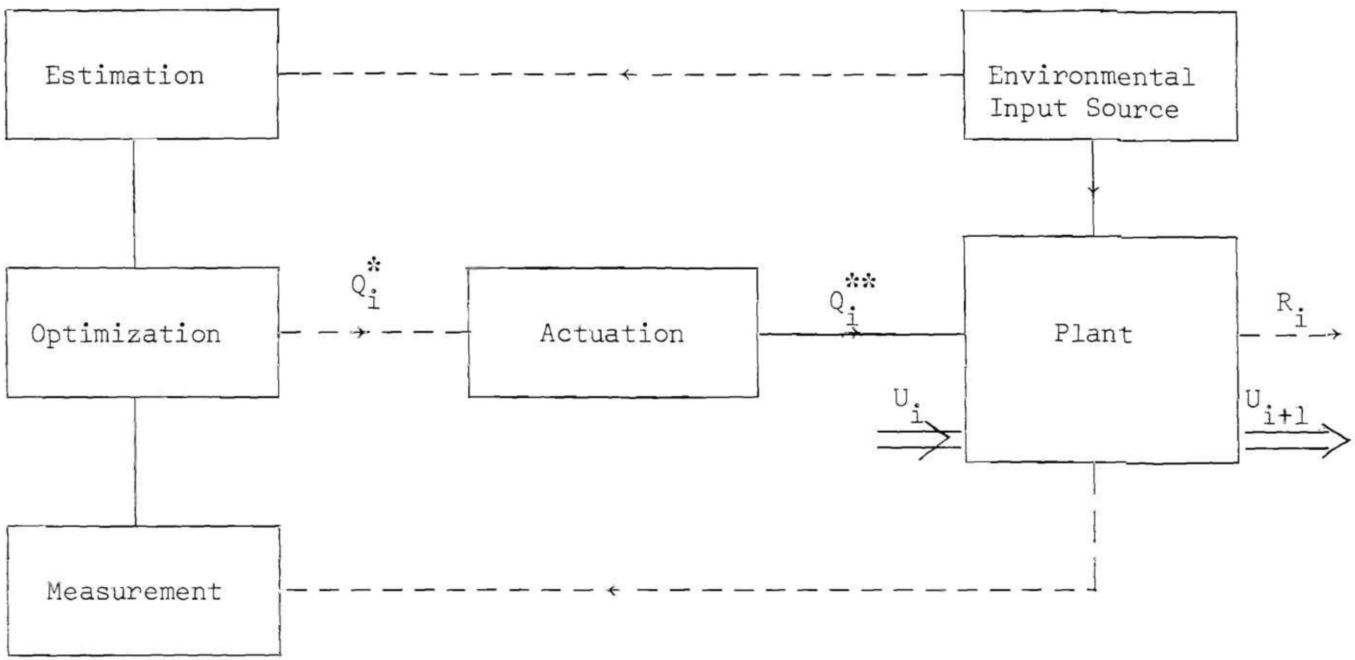

Figure 7. A Schematic Diagram of the Control Sequence.

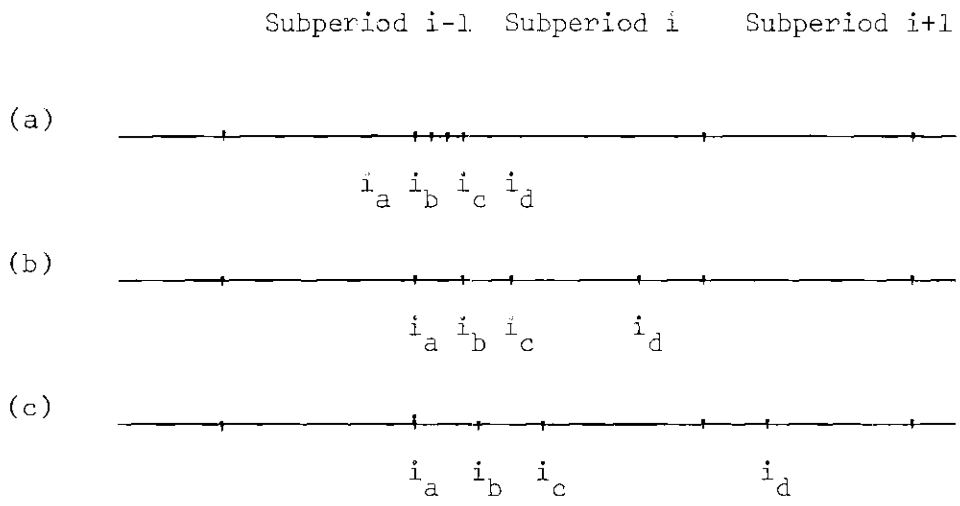

Figure 8. Time Intervals in Control Sequence.

lead time is longer than the time unit of the subperiod. This situation may be avoided if a subperiod is conveniently chosen to cover a sufficiently long time interval so that i_d can be located within the subperiod. For the development of a seasonal goods inventory model in Chapter V, it is assumed that the time points, i_a, i_b, i_c, and i_d, can be spotted at the beginning of every subperiod, as shown in Figure 8(a).

The general discussion presented above of the problem of modeling a multi-stage control process and the feedback sequence of control can be applied to model an inventory control system as follows. For the inventory system, the level of inventory observed at time point i_a is designated as state U_i. When Q_i and V_i denote the inventory replenishment and demand for subperiod i, respectively, then the system state equation can be written as:

$$U_{i+1} = U_i + Q_i - V_i \ . \tag{3-17}$$

At time point i_a, the estimates on future demand V_j, $j=i, i+1, \ldots, n$, may also be obtained. At time point i_b, the optimum replenishment Q_i^* is determined. In this case, the criterion of optimality may be expressed in the form of maximizing the expected return over all j subperiods, $j=1, i+1, \ldots, n$. At time point i_c the order is placed. At time point i_d the replenishment is received. This completes the control cycle for subperiod i.

Summary

The system theory concepts of Chapter II are applied to the modeling of a retail inventory control problem. It is shown that the model may be regarded as a relatively isolated subsystem. Subsequently, the known concepts in control theory are applied to analyze a procedure for modeling a multi-stage control process. The results of the study may be used for the following:

1. To define controllable and uncontrollable subsystems for a given inventory situation.

2. To formulate a multi-stage control system by recognizing system state, control input, environmental input, and system objectives.

3. To recognize the time dependent feedback sequence for individual stages in control processes.

The generality of the discussion facilitates the formulation of dynamic models for similar problems in management control.

CHAPTER IV

DEMAND FORECASTING FOR SEASONAL GOODS ITEMS

General

A procedure for modeling a multi-stage control process was discussed in Chapter III. In the formulation of such a model, a method is needed to estimate the statistical characteristics of a random process if the system state is under the influence of the random process. In the case of an inventory control process, this situation applies to the problem of demand forecasting. The nature of the forecasting techniques used in inventory control may vary depending upon particular circumstances in a given situation. It may involve only the use of historical data on the system state, or may involve predicting some economic indices and correlating the resulting prediction to a demand variable under consideration.

This chapter is concerned with an investigation of the procedure used to forecast demand for seasonal goods inventory items. In the development of stochastic models for inventory problems, it is usually assumed in the literature that the probability of demand is known. Such an assumption is also commonly made in the literature with respect to the demand probabilities of seasonal goods inventory items. Suppose a seasonal period covers a time interval (t_o, t_n). The assumption mentioned above implies that the demand probabilities are determined before the time t_o. The assumption is justified if the seasonal period is a very brief time interval so that the actual demand cannot be observed until the end of the season.

Suppose (n-1) time points t_i, i=1, 2,..., (n-1), can be identified between the time points t_o and t_n, and the situation allows to make observations on the demand at these time points t_i. Then, it may be possible to use the observed data to make re-estimates of demand probabilities. An application of feedback filter theory[1] is made in this chapter in order to consider such a re-estimation problem.

The Best Linear Estimate

Consider two random variables X and Z which are related by some rule; for example, the relation may be expressed in the form of a joint probability density function:

$$P(X, Z) \ . \tag{4-1}$$

Suppose it is possible to directly observe X, but Z cannot be directly observed. In this situation, the values of Z may be estimated on the basis of given observations on X. In order to develop an estimation procedure, a criterion is needed to identify the best among all possible estimates.

Let $\hat{Z}$ denote the best estimate of Z which is defined over the ensemble of all possible combinations of X and Z. The estimation loss is denoted by a loss function $L(\hat{Z})$:

---

[1]Shaw (34); Papoulis (47), Chapter II.

$$L(\hat{Z}) = E(Z - \hat{Z})^2 , \tag{4-2}$$

where E is an operator denoting the expectation. The estimate which minimizes the loss function $L(\hat{Z})$ is commonly referred to as the least mean square estimate. The least mean square estimate has many desirable statistical characteristics--which are discussed later in this chapter--and is frequently considered as the *best* estimate.

Let the small letters x and z be particular values of the random variables X and Z, respectively. When the estimate $\hat{Z}$ is obtained on the basis of observations on X, it can be expressed as a function of X, or $\hat{Z}(X)$. If the conditional density function of Z given X, i.e., $p_{Z|X}(z|x)$, is known, the loss function of Equation (4-2) can be expressed as:

$$L(\hat{Z}) = \int_{-\infty}^{\infty} [z - \hat{z}(x)]^2 \, p_{Z|X}(z|x) \, dz , \tag{4-3}$$

where $\hat{z}(x)$ denotes the estimate of Z for a particular observation x of X. The best estimate $\hat{z}(x)$ in the sense of least mean squares is that which minimizes the loss function of Equation (4-3). This is well known to be $E(Z|X)$ or the mean of the conditional density function $p_{Z|X}(z|x)$; i.e.,

$$\hat{z}(x) = E(Z|X) . \tag{4-4}$$

When the joint distribution of Z and X is normal, it is also known that:

$$E(Z|X) = E(Z) + \frac{\sigma_Z}{\sigma_X}\rho[x - E(X)] \ , \qquad (4\text{-}5)$$

where E(Z) and E(X) are expected values of Z and X, respectively, $\sigma_Z{}^2$ and $\sigma_X{}^2$ are the variances of Z and X; and ρ is the correlation coefficient. In summary, when the conditional density function of the random variables is known, the best estimate in the sense of least mean squares can be obtained as the conditional expectation of Z given the observation of X.

The best estimate in the form of conditional mean estimates, however, is often difficult to obtain, since it requires a knowledge of the conditional density function. For a special case with a single observation x, the best estimate may be easily obtained; for instance, as that shown by Equation (4-5). When observations are made from a large number of different sample populations, however, the conditional density of the desired variable Z and the observable variable X may become quite complex. The linear mean square estimate requires less prior information about the random behavior of the desired variable and the observation variable than would be the case for the conditional mean estimates. Furthermore, the linear mean square estimates have many desirable properties which are described as follows.

As estimate $\hat{z}(x)$ of a random variable Z based on an observation vector x is defined as linear if it satisfies the condition:

$$\hat{z}(a_1x_1 + a_2x_2) = a_1\hat{z}(x_1) + a_2\hat{z}(x_2) \ , \qquad (4\text{-}6)$$

where a_1 and a_2 are any constants. Suppose a finite number of observations $x_1, x_2, \ldots, x_N$ are available on the random variable X. Let $\chi = (x_1, x_2, \ldots, x_N)$ be a set of such observations. The estimate $\hat{z}(\chi)$ is linear, by definition, if it is a linear combination of N observations; namely,

$$\hat{z}(\chi) = \sum_{i=1}^{N} a_i x_i \ , \tag{4-7}$$

where a_i are constants whose values need to be specified. Suppose it is needed to determine the values of a_i for all i such that the estimate $\hat{z}(\chi)$ is the best estimate in the sense of least mean squares. A method for determining the best estimate $\hat{z}(\chi)$, which is well known in the literature,[2] is briefly reviewed as follows.

Since z and x_i are the particular values of the random variables Z and X, respectively, Equation (4-7) can be equivalently written as:

$$\hat{Z} = \sum_{i=1}^{N} a_i X_i \ . \tag{4-8}$$

When this expression is substituted into Equation (4-2), it results in:

$$L(\hat{Z}) = E(Z - \sum_{i=1}^{N} a_i X_i)^2 \ . \tag{4-9}$$

---

[2] *Ibid.*

The optimum values of a_i in the sense of least mean squares are those which minimize the right-hand side of Equation (4-9). This can be determined by differentiating $L(\hat{Z})$ with respect to a_i, setting the partial derivatives equal to zero, and solving the resulting N simultaneous equations. Namely,

$$\frac{\partial L(\hat{Z})}{\partial a_i} = - E[2(Z - \sum_{i=1}^{N} a_i X_i) X_j] \tag{4-10}$$

$$j \quad 1,2,\ldots,N \ .$$

Setting the partial derivatives equal to zero will yield N simultaneous equations:[3]

$$E(ZX_j) = \sum_{i=1}^{N} a_i E(X_i X_j) \tag{4-11}$$

$$j \quad 1,2,\ldots,N \ .$$

The second partial derivative with respect to a_i is always positive so that the values of a_i determined from Equation (4-11) are minimizing values. The expression $E(ZX_j)$ is commonly called the cross-correlation function between the random variables Z and X_j, and the expression

---

[3]Sometimes, these equations are referred to as Wiener-Hopf equations. See Wiener (41).

$E(X_i X_j)$ is the auto-correlation function of the random process X with respect to the random variables X_i and X_j at a time interval (j-i).

When a knowledge of the correlation functions $E(ZX_j)$ and $E(X_i X_j)$ is available, the N simultaneous Equations (4-11) may be solved to obtain the best linear mean square estimate $\hat{Z}$. It is to be noted that, in this case, a knowledge of the conditional density function is not needed.

At this point, it is appropriate to comment on the orthogonality of linear estimates.[4] The linear mean square estimate $\hat{Z}$ has the interesting property that it is orthogonal with its residuals. A residual is the error resulting from the use of an estimate, and is denoted by $\tilde{Z}$, i.e., $\tilde{Z} = Z - \hat{Z}$. The orthogonality of the linear mean square estimate with its residual can be shown as follows. By use of Equation (4-8), the cross product moment of an estimate $\hat{Z}$ and its residual $\tilde{Z}$ is:

$$E[\hat{Z}(Z - \hat{Z})] = E\left[\left(\sum_{i=1}^{N} a_i X_i\right)\left(Z - \sum_{j=1}^{N} a_j X_j\right)\right] \qquad (4\text{-}12)$$

$$= \sum_{i=1}^{N} a_i \left[E(X_i Z) - \sum_{j=1}^{N} a_j E(X_i X_j)\right]$$

$$= 0$$

---

[4]Two random variables are said to be orthogonal if their cross product moments are zero.

The last step in the expression above follows from the optimum condition of Equation (4-11) for the linear mean square estimates. This orthogonality property would, in some cases, permit a simplification of the estimation procedure. It is for this reason that this property is subsequently applied in this chapter to develop a linear feedback prediction procedure.

Since the normal density function plays an important role in linear estimation theory as well as in describing the probability laws of various random phenomena, it is briefly reviewed here to serve as a basis for a subsequent discussion. A random process is said to be Gaussian if all the probability density functions (i.e., first, second, third, etc.) describing the statistical properties of the process are of normal form. The general form of the n-th order normal density function of a random process X is expressed as:

$$p(x_1,x_2,\ldots,x_n) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp[-\tfrac{1}{2}(X-\mu)\ V^{-1}(X-\mu)] \ , \qquad (4\text{-}13)$$

where:

$$X \qquad \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} , \text{ an } (n \times 1) \text{ vector,}$$

$$X' = (x_1, x_2, \ldots, x_n) ,$$

$$\mu = E(X) ,$$

$$V = E[(X-\mu)(X-\mu)'], \text{ an } (n \times n) \text{ matrix} ,$$

$$|V| = \text{determinant of } V ,$$

$$V^{-1} = \text{inverse of } V .$$

For the special case of zero-mean random variables, i.e., $E(X) = 0$, the joint normal density function of Equation (4-13) reduces to:

$$p(x_1,x_2,\ldots,x_n) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left(-\frac{1}{2} X'V^{-1}X\right) , \qquad (4\text{-}14)$$

where:

$$V = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} ,$$

$$\sigma_{ij} = E(x_i x_j) .$$

It can be noted from Equation (4-14) that the probability density function of the zero-mean normal random variable can be specified by its auto-correlation functions $E(x_i x_j)$.

The conditional density function of a variable X_1 given all of the other x's; i.e., $x_2, x_3, \ldots, x_n$, of a Gaussian random process with the joint density function of Equation (4-14) can be expressed as:[5]

$$p(X_1 | x_2, x_3, \ldots, x_n) = \frac{1}{(2\pi)^{1/2} M} \exp[- \tfrac{1}{2}(X_1 - V_{12}V_{22}^{-1}X_2)' M^{-1} (X_1 - V_{12}V_{22}^{-1}X_2)] \tag{4-15}$$

where:

$X_2' = (x_2, x_3, \ldots, x_n)$, a [1 x (n-1)] vector,

$V_{12} = (\sigma_{12}, \sigma_{13}, \ldots, \sigma_{1n})$, a [1 x (n-1)] vector,

$$V_{22} = \begin{bmatrix} \sigma_{22}, & \sigma_{23}, & \cdots & \sigma_{2n} \\ \sigma_{32}, & \sigma_{33} & \cdots & \sigma_{3n} \\ & \cdots & \cdots & \\ \sigma_{n2}, & \sigma_{n3} & & \sigma_{nn} \end{bmatrix},$$ an [(n-1) x (n-1)] matrix,

$M = \sigma_{11} - V_{12}V_{22}^{-1}V_{21}$, a scalar.

The conditional mean of X_1 given particular values of observations of

---

[5] Graybill (13), p. 63.

X_2 is:

$$E(X_1|x_2,x_3,...,x_n) = V_{12}V_{22}^{-1}X_2 . \tag{4-16}$$

It can be noted from Equation (4-16) that the conditional mean of X_1 given observations of X_2 is equal to the linear combination $V_{12}V_{22}^{-1}X_2$ when the random process X is Gaussian. In other words, $V_{12}V_{22}^{-1}X_2$ gives the best linear mean square estimate of X_1 for a Gaussian random process X, and may be expressed as:

$$V_{12}V_{22}^{-1}X_2 = \sum_{i=2}^{n} a_i x_i . \tag{4-17}$$

This section has considered the problem of obtaining the best linear estimate of a random variable Z given observations of a random variable X. The best estimate of Z can be expressed as the conditional mean of Z given observations of X, if the conditional probability is known. In particular, when the random variables are Gaussian, then the best estimate of Z can be expressed as a linear combination of observations of X, where a knowledge of the conditional probability is not needed for the estimation.

Feedback Filter Procedure for Re-estimation

Linear mean square estimation will be further considered in this section with respect to the problem of re-estimating parameter values of random variables, where initial estimates of the parameter values are assumed to be known.

Preliminary Considerations

Consider a sequence of discrete time points:

$$T_1, T_2, \ldots, T_k, \ldots, T_s, T_{(s+1)} \; ,$$

where the time points T_k are regarded as consisting of their own time intervals $(t_{k,0}, t_{k,n})$ as shown in Figure 9.

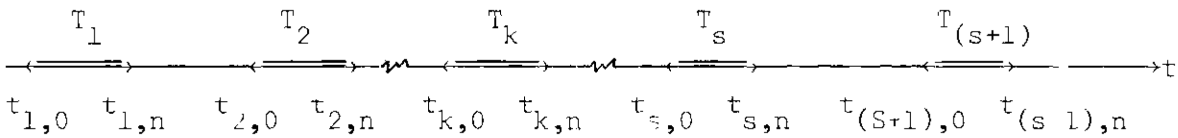

Figure 9. The Time Points T_k and their Intervals

Let the time point $T_{(s+1)}$ be defined as occurring in the future and the other time points $T_k, k = 1,2,\ldots,s$, in the past.

Now, consider a stochastic process:

$$Z_1, Z_2, \ldots, Z_k, \ldots, Z_s, Z_{(s+1)} \; ,$$

where the random variables Z_k are defined for the time points T_k. For example, Z_k may be the sum, or the average of the sum, of the number of random occurrences of certain events which take place in the intervals $(t_{k,0}, t_{k,n})$. The parameter values of the random variables Z_k may be time-variant with respect to the time points T_k, $k=1,2,\ldots,s,(s+1)$; however, it is assumed that the parameter values of one of the Z_k are time-invariant over the interval $(t_{k,0}, t_{k,n})$. For example, the parameter values of Z_1 and Z_2 may be time-variant; however, the parameter values of Z_1 are time-invariant over the interval $(t_{1,0}, t_{1,n})$.

Suppose the outcomes of Z_k can be observed only at times $t_{k,n}$, and one wishes to obtain *a priori* estimates of the unknown constant values of Z_k at times $t_{k,0}$. For instance, consider the problem of estimating the unknown constant value of $Z_{(s+1)}$. Such an estimate may be computed in terms of certain past values of Z_k, $k=1,2,\ldots,s$, or by means of regression analysis with respect to some other time series. Let I_0 be a collection of some available data which are used to compute the estimate of $Z_{(s+1)}$, and $\hat{Z}_{(s+1)}$ be designated as the estimate of $Z_{(s+1)}$ computed on the basis of I_0. Then, the least mean square estimate $\hat{Z}_{(s+1)}$ can be expressed as the conditional mean given I_0; namely,

$$\hat{Z}_{(s+1)} = E(Z_{(s+1)} | I_0) . \tag{4-18}$$

At the same time the variance of the estimation error, which is designated as $\sigma^2_{\hat{Z}_{(s+1)}}$ may be computed as the conditional mean:

$$\sigma^2_{\hat{Z}_{(s+1)}} = E[(Z_{(s+1)} - \hat{Z}_{(s+1)})^2 | I_0] . \tag{4-19}$$

In addition to the description of the situation given above, suppose (n-1) time points can be identified over the intervals $(t_{k,0}, t_{k,n})$, as shown in Figure 10.

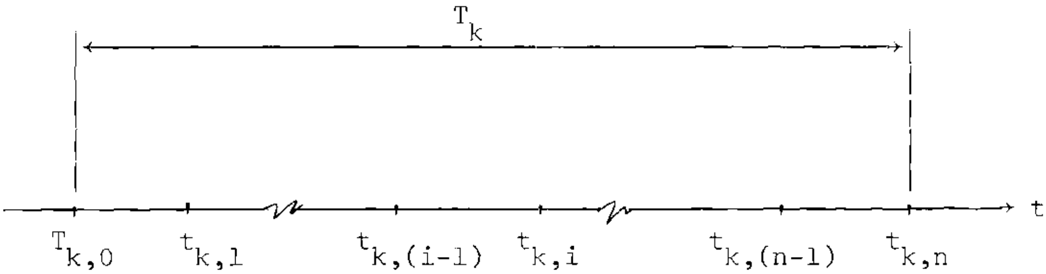

Figure 10. Subintervals of the Interval $(t_{k,0}, t_{k,n})$

Although the *a posteriori* value of Z_k is unobtainable until the time $t_{k,n}$, suppose some observations on Z_k can be made at each of $t_{k,i}$ for $i = 1,2,\ldots,n$. Such observations will be designated as $x_{k,i}$.

Similarly, the time points $t_{(s+1),i}$ and the observations $x_{(s+1),i}$ can be described for the period $T_{(s+1)}$. Given the initial estimate $\hat{Z}_{(s+1)}$ at the time $t_{(s+1),0}$, the observations $x_{(s+1),i}$ may be used at times $t_{(s+1),i}$ to improve the initial estimate. Such a re-estimation procedure will be considered in the following subsection.

The Filtering Problem[6,7]

The problem of obtaining the re-estimation of $Z_{(s+1)}$ was briefly described in the preceding discussion. In order to simplify the sub-

---

[6]It is commonly known that the idea of recursive filtering is originally due to Kalman (46). However, to the best of this writer's knowledge, this filtering problem is first considered by Shaw (34). The procedure used here is almost identical to that given by Shaw except the definitions of m_i and n_i in Equation (4-22) and the subsequent consequences. Shaw assumed that the values of m_i are the same for all i and n_i is the white noise with a common variance. It is hoped that the notation used here is less likely to be misleading than that used by Shaw.

[7]A similar problem is also discussed in Papoulis (47), pp. 419; 425.

script notation, the first subscript (s+1) will be eliminated in the following presentation. In particular, the previous symbols of $Z_{(s+1)}$, $\hat{Z}_{(s+1)}$, $t_{(s+1),i}$ and $x_{(s+1),i}$ will be replaced by the simplified forms of Z, $\hat{Z}_o$, t_i, and x_i, respectively. The simplified symbols will be used to rewrite Equations (4-18) and (4-19) as :

$$\hat{Z}_o = E(Z|I_o) . \tag{4-20}$$

$$\sigma^2_{\hat{Z}_o} = E[(Z - \hat{Z}_o)^2|I_o] . \tag{4-21}$$

The problem is now stated as follows:

1. The initial estimate $\hat{Z}_o$ of an unknown constant Z is made available at time t_o.

2. The estimated variance $\sigma^2_{\hat{Z}_o}$ of the initial estimation error is also made available at time t_o.

3. At times t_i, $i = 1,2,\ldots,n$, observed data x_i are made available, where x_i are related to Z by some rule.

4. It is required to have a procedure to compute the re-estimates of Z at times t_i.

The relation between the unknown constant Z and the observed data x_i is postulated as follows. For the purpose of estimation, the unknown constant Z can be regarded, *a priori*, as a random variable. The best possible point estimate which can be made at any time on the random variable Z is the expected value of Z. Let μ_Z denote the expectation of Z, and assume that the true value of μ_Z is also unknown. Suppose some random variables X_i can be defined by the following relation:

$$X_i = m_i Z + n_i \ , \tag{4-22}$$

where m_i are known constants; n_i are the Gaussian *noise* with zero means and known variances $\sigma^2_{n_i}$, and assume that n_i are independent and orthogonal to Z; namely:

$$E(n_i) = 0 \ , \tag{4-23}$$

$$E(n_i n_j) = 0 \quad \text{if} \quad i \neq j \tag{4-24}$$

$$= \sigma^2_{n_i} \quad \text{if} \quad i = j \ ,$$

$$E(Z \ n_i) = 0 \ . \tag{4-25}$$

Furthermore, let x_i be the observations obtained on the random variables X_i and write:

$$x_i = m_i \mu_Z + e_i \ , \tag{4-26}$$

where e_i are the amount of noise in the observed values of x_i. The numerical values of x_i can be observed at times t_i; however, the values of μ_Z and e_i are not observable, but can be only estimated in terms of statistics. It can be noted that the problem of estimating the unknown constant Z is equivalent to the problem of estimating its expected value μ_Z.

Let χ_i denote sets of data which are used at times t_i for the purpose of the estimation, where $i = 0,1,2,\ldots,n$. Then χ_o, $\chi_1,\ldots$, χ_n are expressed as:

$$
\begin{aligned}
\chi_o &= (\hat{Z}_o, \sigma_{\hat{Z}_o}{}^2) \\
\chi_1 &= (\chi_o, x_i) \\
\chi_2 &= (\chi_o, x_1, x_2) = (\chi_1, x_2) \\
&\;\;\vdots \\
\chi_n &= (\chi_{n-1}, x_n) \; .
\end{aligned}
\tag{4-27}
$$

Let $\hat{Z}_i$ be designated as the best estimates of Z which can be computed at times t_i. Then, Z_i can be expressed as the conditional means:

$$\hat{Z}_i = E[Z|\chi_i] \; . \tag{4-28}$$

Also, let $\hat{X}_{i,i-1}$ be designated as the best *a priori* estimates of the random variables X_i which are computed at times t_{i-1}. Then $\hat{X}_{i,i-1}$ can be also expressed as the conditional means:

$$\hat{X}_{i,i-1} = E(\chi_i|\chi_{i-1})$$

$$= E[(m_i Z + n_i)|\chi_{i-1}]$$

$$= m_i E(Z|\chi_{i-1}) + E(n_i|\chi_{i-1}) \tag{4-29}$$

$$= m_i \hat{Z}_{i-1}$$

Let $\tilde{Z}_i$ and $\tilde{X}_{i,i-1}$ be defined as residuals which result from using estimates $\hat{Z}_i$ and $\hat{X}_{i,i-1}$, respectively; that is:

$$\tilde{Z}_i = Z - \hat{Z}_i \tag{4-30}$$

$$\tilde{X}_{i,i-1} = X_i - \hat{X}_{i,i-1} \; . \tag{4-31}$$

The best estimates of $\tilde{Z}_{i-1}$ which are computed at times t_i will be denoted by $\hat{\tilde{Z}}_{i-1,i}$. Then $\hat{\tilde{Z}}_{i-1,i}$ can be expressed as the conditional means:

$$\hat{\tilde{Z}}_{i-1,i} = E[\tilde{Z}_{i-1}|\chi_i]$$

$$= E[(Z - \hat{Z}_{i-1})|\chi_i]$$

$$= E[Z|\chi_i] - E[\hat{Z}_{i-1}|\chi_i] \tag{4-32}$$

$$= \hat{Z}_i - \hat{Z}_{i-1} \; .$$

It can be observed from Equation (4-32) that $\hat{Z}_i$ may be computed at times t_i as sums of the previous estimates $\hat{Z}_{i-1}$ and the estimates of their residuals, i.e.:

$$\hat{Z}_i = \hat{Z}_{i-1} + \hat{\hat{Z}}_{i-1,i} \; . \tag{4-33}$$

The values of $\hat{\hat{Z}}_{i-1,i}$, however, cannot be determined by use of Equation (4-32) unless the conditional probabilities of $\hat{Z}_{i-1}$ given χ_i are known.

The estimation procedure will be radically simplified if only linear operations are allowed on the data--instead of using the nonlinear method through conditional means. For instance, suppose there exist constants a_i such that Z_i can be expressed as linear combinations of observations x_i and the previous estimates $\hat{Z}_{i-1}$. Namely,

$$\begin{aligned}
\hat{Z}_1 &= \hat{Z}_o + a_i x_i \\
\hat{Z}_2 &= \hat{Z}_1 + a_2 x_2 = \hat{Z}_o + \sum_{j=1}^{2} a_j x_j \\
&\vdots \\
\hat{Z}_i &= \hat{Z}_{i-1} + a_i x_i = \hat{Z}_o + \sum_{j=1}^{i} a_j x_j \; .
\end{aligned} \tag{4-34}$$

Suppose:

$$Z_o = 0 \; . \tag{4-35}$$

Then, Equations (4-34) can be simply written as:

$$\hat{Z}_1 = a_1 x_1$$

$$\hat{Z}_2 = \hat{Z}_1 + a_1 x_1 = \sum_{i=1}^{2} a_j x_j \qquad (4\text{-}36)$$

$$\vdots$$

$$\hat{Z}_j = \hat{Z}_{i-1} + a_i x_i = \sum_{j=1}^{i} a_j x_j \ .$$

Let $\tilde{x}_{i,i-1}$ be defined as the residuals of the observations x_i and the estimates $\hat{X}_{i,i-1}$; namely,

$$\tilde{x}_{i,i-1} = x_i - \hat{X}_{i,i-1} \ . \qquad (4\text{-}37)$$

Then, for any a_j, suitable constants α_j can be found so that Equations (4-36) are expressed in the form:[8]

$$\hat{Z}_1 = \alpha_1 \tilde{x}_{1,0}$$

$$\hat{Z}_2 = \hat{Z}_1 + \alpha_2 \tilde{x}_{2,1} \qquad (4\text{-}38)$$

$$\vdots$$

$$\hat{Z}_i = \hat{Z}_{i-1} + \alpha_i \tilde{x}_{i,i-1} \ .$$

---

[8]Shaw (34), p. 155.

A comparison of the last terms on the right-hand sides of Equations (4-33) and (4-38) suggests that

$$\hat{\tilde{Z}}_{i-1,i} = \alpha_i \tilde{x}_{i,i-1} \quad (4\text{-}39)$$

In order that $\alpha_i \tilde{x}_{i,i-1}$ be the best estimate of $\tilde{Z}_{i-1,i}$ in the sense of the least mean squares, the values of α_i can be determined by solving the following:

$$\frac{\partial}{\partial \alpha_i} E[\tilde{Z}_{i-1} - \alpha_i \tilde{X}_{i,i-1}]^2 = 0 \quad (4\text{-}40)$$

Then,

$$E[\tilde{Z}_{i-1}\tilde{X}_{i,i-1}] = \alpha_i E(\tilde{X}_{i,i-1}^2) ,$$

and

$$\alpha_i = \frac{E[\tilde{Z}_{i-1}\tilde{X}_{i,i-1}]}{E(\tilde{X}_{i,i-1}^2)} . \quad (4\text{-}41)$$

The numerator on the right-hand side of Equation (4-41) is:

$$E[\tilde{Z}_{i-1}\tilde{X}_{i,i-1}] = E[\tilde{Z}_{i-1}(X_i - \hat{X}_{i,i-1})] \quad (4\text{-}42)$$

$$= E[\tilde{Z}_{i-1}(m_i \tilde{Z}_{i-1} + n_i)]$$

$$= m_i E(\tilde{Z}_{i-1}^2) .$$

The denominator on the right-hand side of Equation (4-41) is:

$$E[\hat{X}_{i,i-1}^{\,2}] = E[(m_i\tilde{Z}_{i-1} + n_i)^2]$$

$$= m_i^{\,2}E(\tilde{Z}_{i-1}^{\,2}) + E(n_i^{\,2}) \qquad (4\text{-}43)$$

$$= m_i^{\,2}E(\tilde{Z}_{i-1}^{\,2}) + \sigma_{n_i}^{\,2} \;.$$

By use of Equations (4-42) and (4-43), α_i of Equation (4-41) can be written as:[9]

$$\alpha_i = \frac{m_i E(\tilde{Z}_{i-1}^{\,2})}{m_i^{\,2}E(\tilde{Z}_{i-1}^{\,2}) + \sigma_{n_i}^{\,2}} \;. \qquad (4\text{-}44)$$

This value of α_i can be used in Equations (4-38) to compute $\hat{Z}_i$. However, the values of $E(\tilde{Z}_{i-1}^{\,2})$ are still unknown. A recursive relation can be used to compute $E(\tilde{Z}_i^{\,2})$ as shown below.[10]

$$E(Z_i^{\,2}) = E[(Z - \hat{Z}_i)^2]$$

$$= E\{[Z - (\hat{Z}_{i-1} + \alpha_i\hat{X}_{i,i-1})]^2\}$$

$$= E\{[\tilde{Z}_{i-1} - \sigma_i(m_i\tilde{Z}_{i-1} + n_i)]^2\}$$

---

[9] Cf. *Ibid.*, p. 155.

[10] Cf. *Ibid.*, p. 156.

$$= E\,[(1 - \alpha_i m_i)\tilde{Z}_{i-1} - \alpha_i n_i]^2$$

$$= (1 - \alpha_i m_i)^2\, E(\tilde{Z}_{i-1}^{\,2}) + \alpha_i^{\,2} E(n_i)^2 .$$

When α_i of Equation (4-44) is substituted into the expression above, it results in:

$$E(\tilde{Z}_i^{\,2}) = \frac{\sigma_{n_i}^{\,2}}{m_i^{\,2}[E(\tilde{Z}_{i-1}^{\,2})] + \sigma_{n_i}^{\,2}}\, E(\tilde{Z}_{i-1}^{\,2}) \qquad (4\text{-}45)$$

$$= (1 - \alpha_i m_i)\, E(\tilde{Z}_{i-1}^{\,2}) .$$

A schematic diagram of the linear feedback filter model is shown in Figure 11. The part of the diagram, which is shown within the dotted outline and designated as the source of information, represents the model of Equation (4-26). The other part of the diagram, which is shown within another dotted outline and designated as the iterative scheme, represents the procedure for computing $\hat{Z}_i$.

In summary, the iterative scheme consists of the following main steps.

1. Given the initial estimate $\sigma_{\hat{Z}_o}^{\,2}$, and the values of m_i and $\sigma_{n_i}^{\,2}$, then compute α_i by use of Equations (4-44) and (4-45).

2. Use the computed values of α_i and the observations x_i to obtain:

$$\hat{Z}_i = \hat{Z}_{i-1} + \alpha_i \tilde{x}_{i,i-1} . \qquad (4\text{-}46)$$

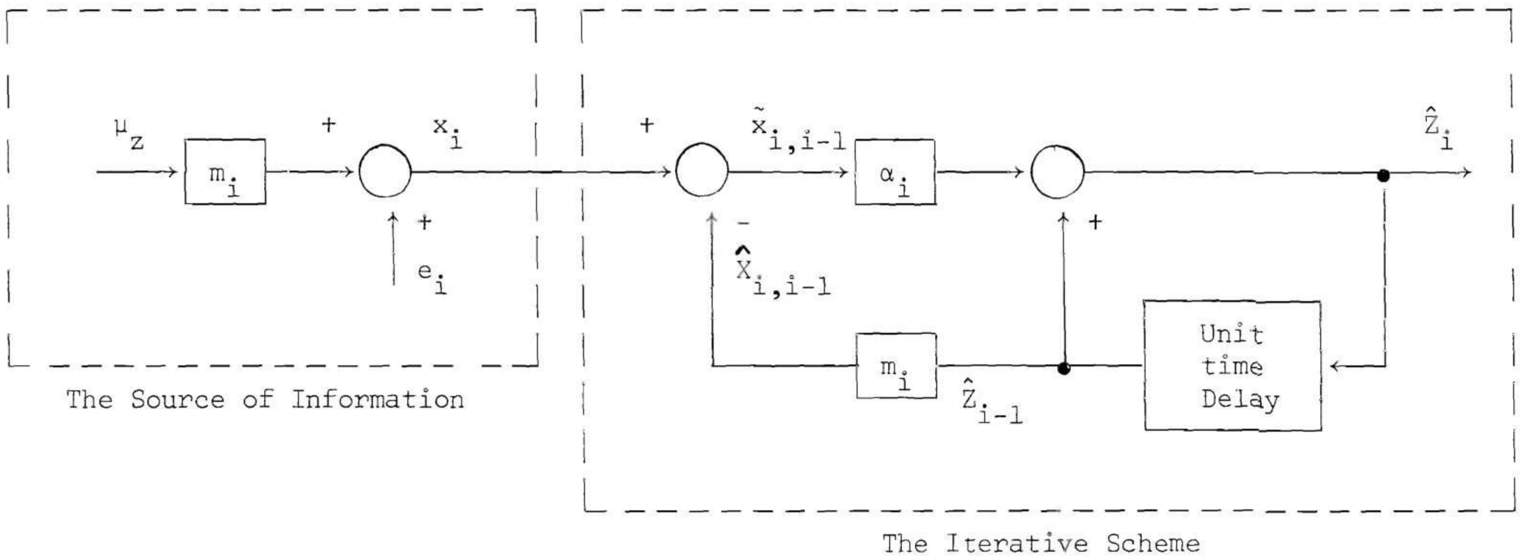

Figure 11. The Linear Feedback Filter Model

This scheme is computationally convenient, because the new estimate can be determined as a sum of the previous estimate and the correction term.

<u>The Reduction of Estimation Errors</u>

The use of the re-estimation procedure should result in the reduction of estimation errors. It can be observed from Equation (4-45) that the mean square errors of the new estimate $\hat{Z}_i$ are proportional to the mean square errors of the previous estimates $\hat{Z}_{i-1}$ by the factor of $(1 - \alpha_i m_i)$. In order to determine the lower and upper bounds of $(1 - \alpha_i m_i)$, the assumptions of the model will be more precisely stated with respect to the values of $\sigma_{\hat{Z}_o}^2$, $\sigma_{n_i}^2$, and m_i.

1. $\sigma_{\hat{Z}_o}^2$ is not zero. If $\sigma_{\hat{Z}_o}^2$ is zero, then $\hat{Z}_o$ has no error of estimation, and the re-estimation would not be required.

2. $\sigma_{\hat{Z}_o}^2$ has a finite value; i.e., $\sigma_{\hat{Z}_o}^2 < \infty$. If $\sigma_{\hat{Z}_o}^2$ is infinitely large, then it implies that the estimate $\hat{Z}_o$ is unknown.

3. $\sigma_{n_i}^2$ are not zero for all i, $i=1,2,\ldots,n$. If any one of $\sigma_{n_i}^2$ is zero, then the corresponding value of x_i can be used to compute without error the constant value of Z.

4. $\sigma_{n_i}^2$ have finite values; i.e., $\sigma_{n_i}^2 < \infty$ for all i, $i=1,2,\ldots,n$. If $\sigma_{n_i}^2$ are infinitely large, then the random variables X_i can take any real number, and it is impossible to have any reasonable means of estimation.

5. $0 \quad m_i < 1$ for all i, and $\sum_i m_i = 1$, for $i = 1,2,\ldots,n$.

Under these conditions, the following inequalities will hold:

$$0 < (1 - \alpha_i m_i) < 1 \, . \tag{4-47}$$

When these inequalities are considered together with Equation (4-45), it is observed that:

$$E(\tilde{Z}_{i+1}^{2}) < E(\tilde{Z}_{i}^{2}) . \tag{4-48}$$

In conclusion, the reductions in the mean square error of estimation can be made in the filtering procedure, and the new estimates $\hat{Z}_i$ are, on the average, improved estimates of the previous estimates $\hat{Z}_{i-1}$.

Special Cases of the Filtering Problem When the Initial Estimates are not Available but $\sigma_{n_i}^{2}$ are Known

As a special case, it is interesting to analyze the consequences of large values of $\sigma_{\hat{Z}_o}^{2}$. Suppose $\sigma_{\hat{Z}_o}^{2} = M$, where M is an arbitrary number, and let $M \to \infty$. Then, from Equation (4-44):

$$\lim_{M \to \infty} \alpha_1 = \lim_{M \to \infty} \frac{m_1 M}{m_1^{2} M + \sigma_{n_1}^{2}} = \frac{1}{m_1} , \tag{4-49}$$

where $\sigma_{n_1}^{2}$ is assumed to be a relatively small number; i.e., $\sigma_{n_1}^{2} < M$. In this case,

$$\lim_{M \to \infty} \hat{Z}_1 = \frac{1}{m_1} x_1 , \tag{4-50}$$

and the estimate $\hat{Z}_1$ of Equation (4-50) becomes identical to that which can be computed by means of the simple average without knowledge of the initial estimate. In fact, when $M \to \infty$, then the random number Z can be any real number, and the best initial estimate could be $\hat{Z}_o = 0$.

Now, consider the problem of computing $E(\tilde{Z}_1{}^2)$ by use of Equation (4-45); namely:

$$E(\tilde{Z}_1{}^2) = (1 - \alpha_1 m_1)M \; . \tag{4-51}$$

However, if $M \to \infty$, then $(1 - \alpha_1 m_1) \to 0$, since $\alpha_1 \;\; \frac{1}{m_1}$. In this case, the right-hand side of Equation (4-51) results in an indeterminate form. Making the appropriate substitution for α_1, one can rewrite the right-hand side of Equation (4-51) as:

$$E(\tilde{Z}_1{}^2) = \left[1 - \frac{m_1{}^2 M}{m_1{}^2 M + \sigma_{n_1}{}^2}\right] M \tag{4-52}$$

$$= \frac{\sigma_{n_1}{}^2 M}{m_1{}^2 M + \sigma_{n_1}{}^2} \; ,$$

and if $\sigma_{n_1}{}^2 << M$, then:

$$\lim_{M \to \infty} E(\tilde{Z}_i{}^2) = \frac{\sigma_{n_1}{}^2}{m_1{}^2} \; . \tag{4-53}$$

This is an interesting result which indicates that, although the error of the initial estimate can be very large, the mean square error of the first estimate at time t_1 is a bounded value, provided $\sigma_{n_1}{}^2 < \infty$.

This result is interpreted as follows: when the initial estimates $\hat{Z}_o$ and $\sigma_{\hat{Z}_o}{}^2$ are unknown, but $\sigma_{n_i}{}^2$ are known for all i and $0 < \sigma_{n_i}{}^2 < \infty$, the filtering method can be applied to compute the estimates $\hat{Z}_i$ for i=1,2,...,n by use of Equations (4-44), (4-45), and (4-46).

When the Initial Estimates are not Available and the Noises Have a Common Variance

As in the preceding case, suppose $\sigma_{\hat{Z}_o}^2 = M$ and let $M \to \infty$. In addition, let

$$\sigma_{n_1}^2 = \sigma_{n_2}^2 = \ldots = \sigma_{n_n}^2 = \sigma_n^2 .$$

In this case, the values of α_i and $E(\tilde{Z}_i^2)$ can be computed for i = 1, 2, and 3 as shown below:

$$\alpha_1 = \frac{1}{m_1} ,$$

$$\alpha_2 = \frac{m_2}{m_1^2 + m_2^2} ,$$

$$\alpha_3 = \frac{m_3}{m_1^2 + m_2^2 + m_3^2} ,$$

and,

$$E(\tilde{Z}_1^2) = \frac{\sigma_n^2}{m_1^2} ,$$

$$E(\tilde{Z}_2^2) = \frac{\sigma_n^2}{m_1^2 + m_2^3} ,$$

$$E(\tilde{Z}_3^2) = \frac{\sigma_n^2}{m_1^2 + m_2^2 + m_3^2} .$$

By a generalization, it can be shown that, for $i = 1,2,\ldots,n$:

$$\alpha_i = \frac{m_i}{\sum_{j=1}^{i} m_j^2} , \tag{4-54}$$

$$E(\tilde{Z}_i^2) = \frac{\sigma_n^2}{\sum_{j=1}^{i} m_j^2} . \tag{4-55}$$

When the values of α_i of Equation (4-54) are used for the values of α_i in Equation (4-46), the estimates $\hat{Z}_i$ can be computed as:

$$\hat{Z}_i = \hat{Z}_{i-1} + \alpha_i(x_i - m_i\hat{Z}_{i-1})$$

$$= \hat{Z}_{i-1} + \frac{m_i}{\sum_{j=1}^{i} m_j^2}(x_i - m_i\hat{Z}_{i-1}) . \tag{4-56}$$

This result is interpreted as follows. When the initial estimates are not available and the noises have a common variance, then the estimates $\hat{Z}_i$ can be computed by Equation (4-56). It is very interesting to observe that, as shown in Equation (4-56), the iterative procedure is completely independent of the noise. In other words, the magnitude of the noise variance does not affect the estimating iteration (although it affects the outcomes in x_i).

Comparison with the Simple Moving Averages

Let $\hat{\hat{Z}}_i$ denote the estimate of Z computed at times t_i, $i=1,2,\ldots,n$, by the method of the simple moving averages; that is:

$$\hat{\hat{Z}}_i = \frac{1}{i} \sum_{j=1}^{i} \frac{1}{m_j} x_j \ . \tag{4-57}$$

It can be easily shown that the expression above can be written in the following form:

$$\hat{Z}_i = \hat{Z}_{i-1} + \beta_1 (x_i - m_i \hat{Z}_{i-1}) = \hat{Z}_{i-1} + \frac{1}{im_i} (x_i - m_i \hat{Z}_{i-1}) \ , \tag{4-58}$$

where

$$\beta_i = \frac{1}{im_i} \ .$$

It is interesting to compare β_i of Equation (4-58) with α_i of Equation (4-54).

(a) If $m_j < m_i$ for $j=1,2,\ldots,(i-1)$, then:

$$\beta_i < \alpha_i \ .$$

The inequality above holds true, since:

$$\alpha_i = \frac{1}{m_i \sum_{j=1}^{i} \left[\frac{m_j}{m_i}\right]^2} \ ,$$

and if $m_j < m_i$ for $j=1,2,\ldots,(i-1)$, then:

$$\frac{1}{im_i} < \frac{1}{m_i \sum_{j=1}^{i} \left[\frac{m_j}{m_i}\right]^2}$$

(b) If $m_i < m_j$ for all $j=1,2,\ldots,n$, then:

$$\alpha_i < \beta_i \ .$$

(c) If $m_1 = m_2 = \ldots = m_n = m$, then

$$\alpha_i = \beta_i = \frac{1}{im} \ . \tag{4-59}$$

These results are given the following interpretations. When the initial estimates are not available, and the noises have a common variance, then the weighting factors α_i of the filtering method and β_i of the simple moving averages are different if $m_i \neq m_j$ for all $j=1,2,\ldots,i$ and $i=1,2,\ldots,n$, but are identical if $m_i = m_j$ for all $j=1,2,\ldots,i$ and $i=1,2,\ldots,n$.

In other words, if:

(i) the initial estimates are not available,

(ii) the time points t_i can be assigned in such a way that $m_1 = m_2 = \ldots = m_n$, and

(iii) the magnitude of the noise variances is bounded and the same for all i, $i=1,2,\ldots,n$,

then:

(i) the simple moving averages give identical estimates of

Z_i as can be obtained by the filtering method, and

(ii) it is not necessary to know the value of the noise variance.

Numerical Examples

The Data and the Situation

Suppose a seasonal period (t_o, t_n) is identified with five sub-periods; i.e., n=5. In order to generate the data x_i, i=1,2,3,4, and 5, two sets of five random normal numbers with zero means and unit variances are selected from a random number table,[11] and shown in Table 1.

Table 1. Two Sets of Five Random Normal Numbers with Zero Means and Unit Variances

i	Set A	Set B
1	0.91	-0.51
2	1.18	-0.99
3	-1.50	0.97
4	-0.69	0.98
5	1.37	-1.10

The random number table contains 56 sets of 5 random normal numbers. Among these sets, a set was randomly chosen, and is used as

[11] Churchman, *et al.*, (44), p. 181.

the data in Set A of Table 1. On the other hand, the selection of the data in Set B of Table 1 was not made on a random basis. All numbers except one in Set B have smaller deviations from the means than those in Set A. In fact, the numbers in Set B have the smallest overall deviations from the means among the 56 sets contained in the random number table. The random numbers in Set A are used to generate the experimental data for Examples 1, 2 and 3, and the random numbers in Set B for Examples 4 and 5.

It is assumed in all examples that the unknown constant μ_Z is equal to 5. For the purpose of illustration, Z is regarded as a random variable with EZ = μ_Z = 5. The value of μ_Z, which is equal to 5, is, of course, unknown to the estimator. At time t_o, the initial estimate $\hat{Z}_o$ is given as zero; i.e., $\hat{Z}_o = 0$. Various values of $\sigma^2_{\hat{Z}_o}$ are used in the examples to study their effects on the subsequent estimation. Different values of $\sigma^2_{n_i}$ and m_i are considered in the examples to illustrate their effects on the estimation errors.

Example 1

The particular situation for this example is specified by the following:

(a) $m_1 = m_2 = \ldots = m_5 = 0.2$

(b) $\sigma^2_{n_1} = \sigma^2_{n_2} = \ldots = \sigma^2_{n_5} = 1$

(c) Use the random numbers in Set A of Table 1 to generate the data x_i.

Since the value of μ_Z is given as 5, it follows that the random variables X_i have the means $EX_i = m_i\mu_Z = 1$ and the variances $\sigma^2_{n_i} = 1$ for all i, i=1,2,...,5. The simulated data x_i are shown in Table 2.

Table 2. Simulated Data x_i for Example 1

i	x_i	EX_i	$\sigma_{n_i}^2$
1	1.91	1	1
2	2.18	1	1
3	-0.50	1	1
4	0.31	1	1
5	2.37	1	1

For the purpose of analysis, six different values of $\sigma_{\hat{Z}_o}^2$ are considered in this example; namely,

$$\sigma_{\hat{Z}_o}^2 = 0,\ 5,\ 10,\ 20,\ 40,\ \text{and}\ 200\ .$$

The values of α_i, $E(\hat{Z}_i^2)$, and $\hat{Z}_i$ are computed by use of Equations (4-44), (4-45), and (4-46), respectively. The computational results are tabulated in Table 3, and also shown in the graph of Figure 12.

The estimates by the simple moving averages are also computed by use of Equation (4-57), and tabulated in Table 3 as well as graphed in Figure 12.

The sum of squares of estimation errors, defined by:

$$\sum_{i=1}^{n} (\mu_Z - \hat{Z}_i)^2 = \sum_{i=1}^{5} (5 - \hat{Z}_i)^2 , \tag{4-60}$$

Table 3. Computational Results for Example 1

$$m_1 = m_2 = \dots = m_5 = 0.2$$

$$\sigma_{n_1} = \sigma_{n_2} = \dots = \sigma_{n_5} = 1$$

	$\sigma_{\hat{Z}_o}^2$	i	α_i	$E(\tilde{Z}_i^2)$	$\hat{Z}_i$	$\sum_i(\mu_Z-\hat{Z}_i)^2$
The Linear Filtering Method with $\hat{Z}_o = 0$	0	1	0	0	0	125.0
		2	0	0	0	
		3	0	0	0	
		4	0	0	0	
		5	0	0	0	
	5	1	0.83	4.15	1.6	33.9
		2	0.76	3.51	3.0	
		3	0.61	3.07	2.3	
		4	0.55	2.73	2.2	
		5	0.50	2.46	3.2	
	10	1	1.43	7.20	2.7	12.4
		2	1.12	5.59	4.6	
		3	0.92	4.57	3.3	
		4	0.78	3.89	3.1	
		5	0.72	3.33	4.3	
	20	1	2.22	11.20	4.3	4.0
		2	1.55	7.73	6.3	
		3	1.11	6.02	4.4	
		4	0.97	4.80	3.8	
		5	0.78	4.03	5.1	
	40	1	3.08	15.36	5.9	9.7
		2	1.90	9.52	7.8	
		3	1.38	6.89	4.9	
		4	1.08	5.40	4.2	
		5	0.89	4.44	5.6	
	200	1	4.45	22.00	8.5	35.3
		2	2.34	11.65	9.6	
		3	1.59	7.95	5.8	
		4	1.20	6.04	4.8	
		5	0.97	4.89	6.1	
The Simple Moving Averages		1			9.6	50.9
		2			10.2	
		3			6.0	
		4			4.9	
		5			6.3	

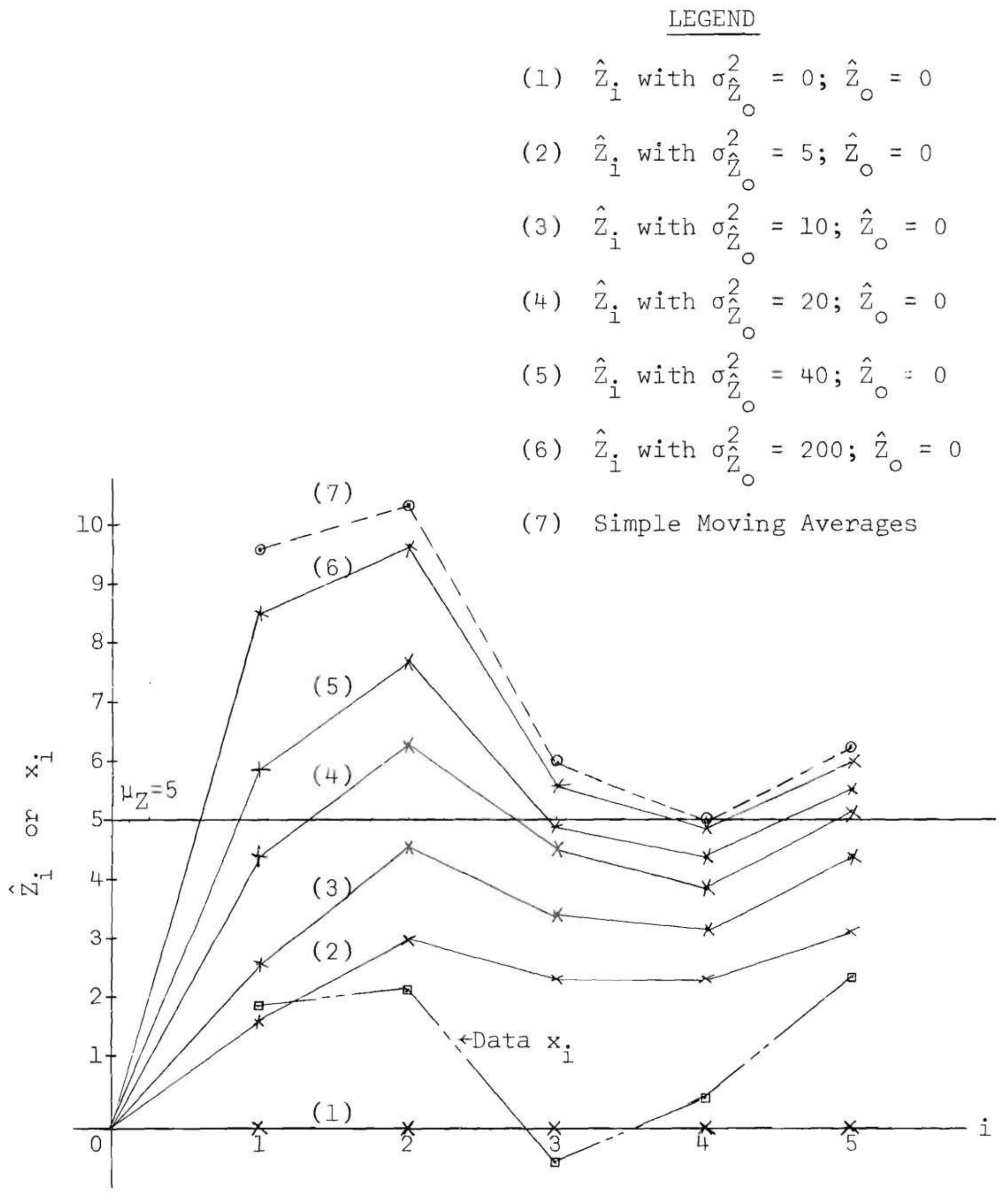

Figure 12. Graph of the Data, the Estimates, and the Moving Averages for Example 1

is computed as shown in the last column of Table 3. The sum of squares of estimation errors, which will be simply denoted by S.S.E., can be used as a measure to evaluate the accuracies of estimation in various outcomes.

The numerical results of this example are summarized as follows.

1. The case when only the initial estimate is used. If the initial values are given by the pair, $\hat{Z}_o = 0$ and $\sigma_{\hat{Z}_o}^2 = 0$, then the situation implies that the re-estimation is not required. In this case, the S.S.E. (the sum of squares of estimation errors) results in a large number; i.e., 125.

2. The case when the initial estimates are used with the data to obtain the re-estimates. Five different values of $\sigma_{\hat{Z}_o}^2$ are considered for this case; i.e., $\sigma_{\hat{Z}_o}^2 = 5, 10, 20, 40$, and 200. In this case, the values of S.S.E. are much smaller than the case without the re-estimation.

3. The case when only the data are used. If the initial estimates are unknown, then the simple moving averages can be used in this case with the assumption that $\sigma_{\hat{Z}_o}^2 \to \infty$. (Also note the common values of m_i as well as of $\sigma_{n_i}^2$ in this example.) This phenomenon can be readily observed in Figure 12; namely, as the values of $\sigma_{\hat{Z}_o}^2$ increase, the estimated values $\hat{Z}_i$ approach the simple moving averages.

Example 2

The particular situation for this example is specified by the following:

(a) $m_1 = m_2 = \dots = m_5 = 0.2$

(b) $\sigma_{n_1}^2 = \sigma_{n_2}^2 = \ldots = \sigma_{n_5}^2 = 4$

(c) Use the random numbers in Set A of Table 1 to generate the data x_i.

The only difference in the situations of this example and the preceding is the value of the common variance, $\sigma_{n_i}^2$. The simulated data are shown in Table 4.

Table 4. Simulated Data for Example 2

i	x_i	EX_i	$\sigma_{n_i}^2$
1	2.82	1	4
2	3.36	1	4
3	-2.00	1	4
4	-0.38	1	4
5	3.74	1	4

The computational results for this example are tabulated in Table 5, and also shown in the graph of Figure 13. The results indicate that when the deviation of the data from their mean are large, then the simple moving averages result in large errors of estimation.

Example 3

The particular situation for this example is specified by the following:

Table 5. Computational Results for Example 2

$$m_1 = m_2 = \cdots = m_5 = 0.2$$

$$\sigma_{n_1} = \sigma_{n_2} = \cdots = \sigma_{n_5} = 2$$

	$\sigma^2_{\hat{Z}_o}$	i	α_i	$E(\tilde{Z}_i^{\,2})$	$\hat{Z}_i$	$\sum_i (\mu_Z-\hat{Z}_i)^2$
The Linear Filtering Method with $\hat{Z}_o = 0$	0	1	0	0	0	125.0
		2	0	0	0	
		3	0	0	0	
		4	0	0	0	
		5	0	0	0	
	20	1	0.91	16.36	2.6	19.7
		2	0.70	14.07	4.6	
		3	0.61	12.35	2.8	
		4	0.55	10.99	2.3	
		5	0.50	9.90	3.9	
	40	1	1.43	28.56	4.0	10.2
		2	1.11	22.22	6.9	
		3	0.91	18.18	3.8	
		4	0.77	15.38	2.9	
		5	0.67	13.32	5.0	
	200	1	3.33	66.00	9.4	78.4
		2	1.99	39.73	12.3	
		3	1.42	28.45	6.0	
		4	1.05	22.48	4.3	
		5	0.92	18.34	7.0	
The Simple Moving Averages		1			14.1	199.0
		2			15.5	
		3			7.0	
		4			4.8	
		5			7.6	

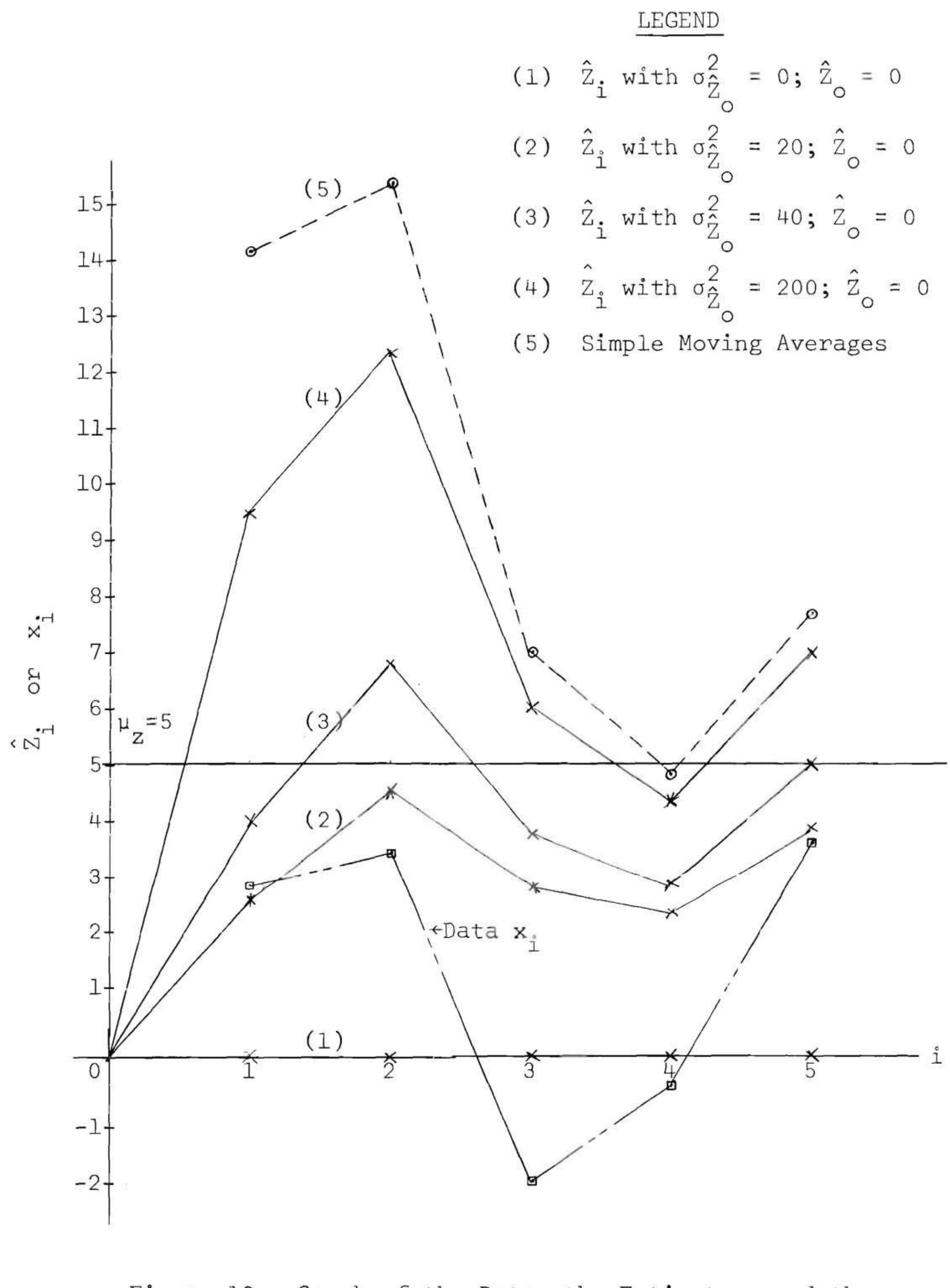

Figure 13. Graph of the Data, the Estimates, and the Moving Averages for Example 2

(a) $m_1 = 0.1,\ m_2 = m_3 = 0.3,\ m_4 = 0.2,\ m_5 = 0.1$

(b) $\sigma_{n_1}^2 = 1,\ \sigma_{n_2}^2 = \sigma_{n_3}^2 = 4,\ \sigma_{n_4}^2 = \sigma_{n_5}^2 = 1$.

(c) Use the random numbers in Set A of Table 1 to generate the data x_i.

The differences in the situations of this example and the two preceding are in the different values used for m_i and $\sigma_{n_i}^2$. The simulated data are shown in Table 6.

Table 6. Simulated Data for Example 3

i	x_i	EX_i	$\sigma_{n_i}^2$
1	1.41	0.5	1
2	3.86	1.5	4
3	-1.50	1.5	4
4	0.31	1.0	1
5	1.87	0.5	1

The computational results for this example are tabulated in Table 7, and also shown in the graph of Figure 14. An interesting phenomenon to be observed in the graph is that, as the values of $\sigma_{\hat{Z}_o}^2$ increase, the estimated values $\hat{Z}_i$ do not approach the simple moving averages in this case. (Note the different values of m_i and of $\sigma_{n_i}^2$.)

Table 7. Computational Results for Example 3

$m_1 = 0.1,\ m_2 = m_3 = 0.3,\ m_4 = 0.2,\ m_5 = 0.1$

$\sigma_{n_1} = 1,\ \sigma_{n_2} = \sigma_{n_3} = 2,\ \sigma_{n_4}\quad \sigma_{n_5} = 1$

	$\sigma_{\hat{Z}_o}^{2}$	i	α_i	$E(\tilde{Z}_i^{\,2})$	$\hat{Z}_i$	$\sum_i (\mu_Z-\hat{Z}_i)^2$
The Linear Filtering Method with $\hat{Z}_o = 0$	0	1	0	0	0	125.0
		2	0	0	0	
		3	0	0	0	
		4	0	0	0	
		5	0	0	0	
	20	1	1.67	16.66	2.4	14.3
		2	0.91	12.11	5.4	
		3	0.71	9.44	3.6	
		4	1.37	6.85	3.0	
		5	0.64	6.41	3.9	
	40	1	2.86	28.56	4.0	11.5
		2	1.30	17.42	8.0	
		3	0.94	12.51	5.1	
		4	1.67	8.33	3.8	
		5	0.77	7.69	4.7	
	200	1	6.67	66.60	9.4	100.9
		2	2.00	26.64	13.4	
		3	1.25	16.65	8.2	
		4	2.01	10.00	5.5	
		5	0.91	9.09	6.2	
The Simple Moving Averages		1			14.1	172.0
		2			13.5	
		3			7.3	
		4			5.9	
		5			8.4	

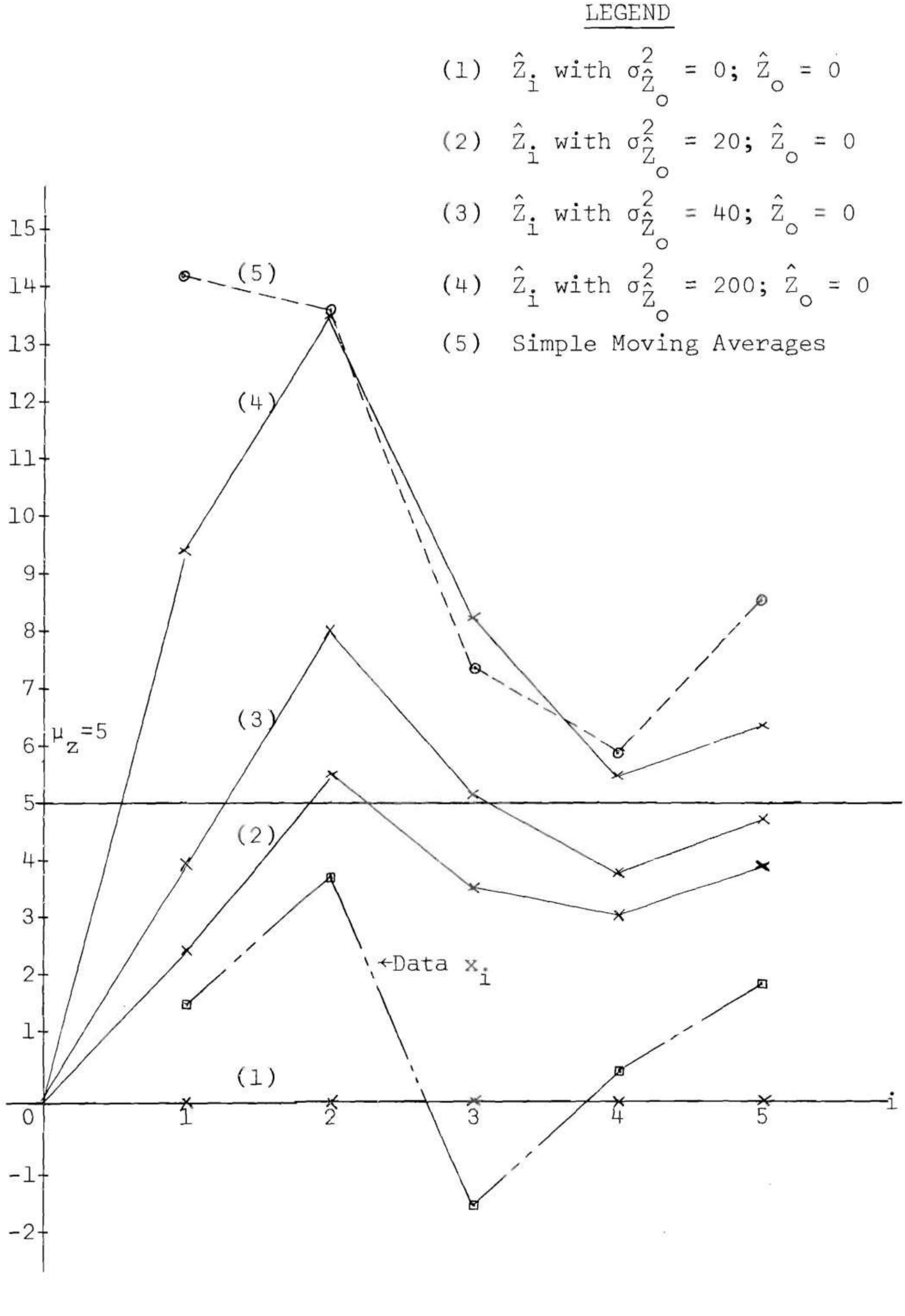

Figure 14. Graph of the Data, the Estimates, and the Moving Averages for Example 3

Example 4

The particular situation for this example is specified by the following:

(a) $m_1 = m_2 = \ldots = m_5 = 0.2$

(b) $\sigma_{n_1}^2 = \sigma_{n_2}^2 = \ldots = \sigma_{n_5}^2 = 1$

(c) Use the random numbers in Set B of Table 1 to generate the data x_i.

The only difference in this example and Example 1 lies in the different sets of random numbers used to generate the data. The simulated data are shown in Table 8. The computational results for this example are tabulated in Table 9, and also shown in the graph of Figure 15.

Table 8. Simulated Data for Example 4

i	x_i	EX_i	$\sigma_{n_i}^2$
1	0.49	1	1
2	0.01	1	1
3	1.97	1	1
4	1.98	1	1
5	-0.10	1	1

Table 9. Computational Results for Example 4

$$m_1 = m_2 = \cdots = m_5 = 0.2$$

$$\sigma_{n_1} = \sigma_{n_2} = \cdots = \sigma_{n_5} = 1$$

	$\sigma_{\hat{Z}_o}^2$	i	α_i	$E(\tilde{Z}_i^2)$	$\hat{Z}_i$	$\sum_i (\mu_Z - \hat{Z}_i)^2$
The Linear Filtering Method with $\hat{Z}_o = 0$	0	1	0	0	0	125.0
		2	0	0	0	
		3	0	0	0	
		4	0	0	0	
		5	0	0	0	
	20	1	2.22	11.20	1.1	41.3
		2	1.55	7.73	0.8	
		3	1.11	6.02	2.8	
		4	0.97	4.80	4.2	
		5	0.78	4.03	3.5	
	40	1	3.08	15.36	1.5	32.3
		2	1.90	9.52	1.0	
		3	1.38	6.89	3.4	
		4	1.08	5.40	4.8	
		5	0.89	4.44	3.9	
	200	1	4.45	22.00	2.2	24.5
		2	2.34	11.65	1.2	
		3	1.59	7.95	3.9	
		4	1.20	6.04	5.4	
		5	0.97	4.89	4.2	
The Simple Moving Averages		1			2.5	21.9
		2			1.3	
		3			4.1	
		4			5.6	
		5			4.4	

LEGEND

(1) $\hat{Z}_i$ with $\sigma^2_{\hat{Z}_o} = 0$; $\hat{Z}_o = 0$

(2) $\hat{Z}_i$ with $\sigma^2_{\hat{Z}_o} = 20$; $\hat{Z}_o = 0$

(3) $\hat{Z}_i$ with $\sigma^2_{\hat{Z}_o} = 40$; $\hat{Z}_o = 0$

(4) $\hat{Z}_i$ with $\sigma^2_{\hat{Z}_o} = 200$; $\hat{Z}_o = 0$

(5) Simple Moving Averages

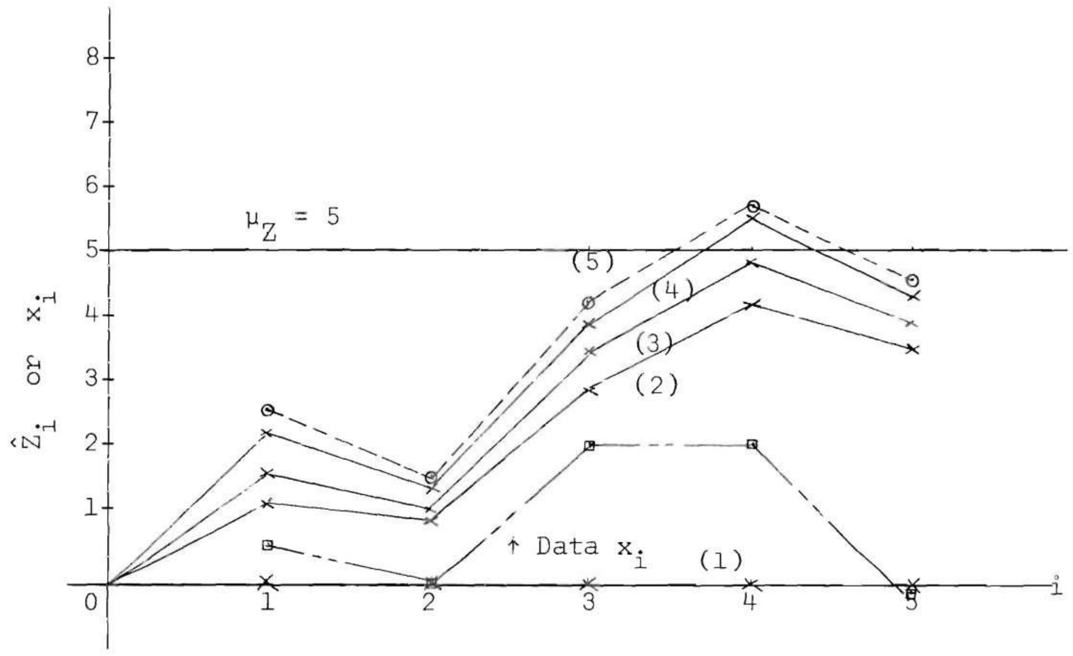

Figure 15. Graph of the Data, the Estimates, and the Moving Averages for Example 4

The numerical results indicate that, in this particular situation, the simple moving averages give better estimates than the others. It should be recalled, however, that the random numbers in Set B of Table 1 are such that their deviations from the mean are very small.

Example 5

The particular situation for this example is specified by the following:

(a) $m_1 = m_2 = \ldots = m_5 = 0.2$

(b) $\sigma_{n_1}^2 = \sigma_{n_2}^2 = \ldots = \sigma_{n_5}^2 = 4$

(c) Use the random numbers in Set B of Table 1 to generate the data x_i.

The only difference in this example and the preceding lies in the values of $\sigma_{n_i}^2$. The simulated data are shown in Table 10.

Table 10. Simulated Data for Example 5

i	x_i	EX_i	$\sigma_{n_i}^2$
1	0.02	1	4
2	-0.98	1	4
3	2.94	1	4
4	2.96	1	4
5	-1.20	1	4

The computational results for this example are tabulated in Table 11, and also shown in the graph of Figure 16. The results indicate that, when the deviations of data are relatively large from their means, then the simple moving averages result in the large value of S.S.E.

Summary of Results in Examples

The numerical results obtained in the preceding five examples are summarized as follows:

1. If the re-estimation is not made, then the bias error in the initial estimate cannot be corrected. In this case, a high accuracy in the initial estimate would be required to eliminate the chance of probable bias errors.

2. If only the data x_i are used without the initial estimate, then the estimates are highly sensitive to the large deviations in the data, which results in large errors of estimation.

3. If the filtering method of re-estimation is used, then the bias error in the initial estimate can be eventually corrected; i.e., the magnitude of $E(\tilde{Z}_i^2)$ smoothly decreases, and the estimation is not too sensitive to large deviations in the data. In all examples (except Example 4), the filtering methods resulted in the smallest estimation errors. As explained earlier, the situation in Example 4 was the least likely case.

Application of the Feedback Filter Procedure to Forecast Demand of Seasonal Goods Inventory Items

This section is concerned with the application of the feedback

Table 11. Computational Results for Example 5

$$m_1 = m_2 = \cdots = m_5 = 0.2$$

$$\sigma_{n_1} = \sigma_{n_2} = \cdots = \sigma_{n_5} = 2$$

	$\sigma^2_{\hat{Z}_o}$	i	α_i	$E(\tilde{Z}_i^{\,2})$	$\hat{Z}_i$	$\sum_i (\mu_Z-\hat{Z}_i)^2$
The Linear Filtering Method with $\hat{Z}_o = 0$	0	1	0	0	0	125.0
		2	0	0	0	
		3	0	0	0	
		4	0	0	0	
		5	0	0	0	
	20	1	0.91	16.36	0	87.4
		2	0.70	14.07	-0.7	
		3	0.61	12.35	1.2	
		4	0.55	10.99	2.7	
		5	0.50	9.90	1.8	
	40	1	1.43	28.56	0	78.3
		2	1.11	22.22	-1.0	
		3	0.91	18.18	1.9	
		4	0.77	15.38	3.8	
		5	0.67	13.32	2.5	
	200	1	3.33	66.00	0.1	78.1
		2	1.99	39.73	-1.9	
		3	1.42	28.45	2.8	
		4	1.05	22.48	5.3	
		5	0.92	18.34	3.4	
The Simple Moving Averages		1			0.1	84.8
		2			-2.4	
		3			3.3	
		4			6.2	
		5			3.7	

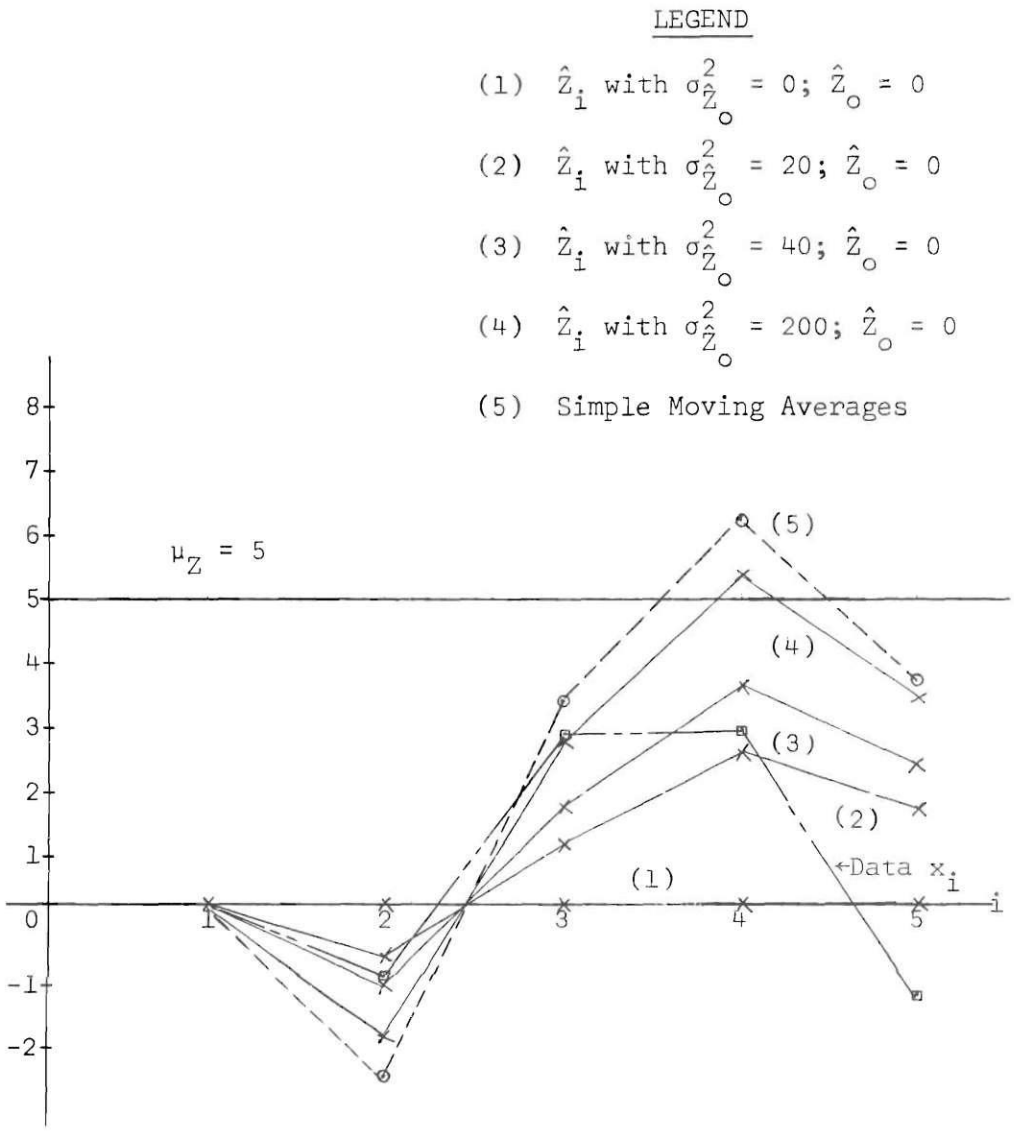

Figure 16. Graph of the Data, the Estimates, and the Moving Averages for Example 5

filter procedure to the problem of demand forecasting of seasonal goods inventory items. Consider a seasonal period which is defined by the time interval (t_o, t_n). Let D be designated as the number of the seasonal items in demand for the season. When the constant value of D is unknown, it can be regarded *a priori* as a random variable D. The problem of computing a point estimate of the random variable D is equivalent to that of estimating the expected value of D.

Let $\hat{D}_o$ be designated as the initial estimate of D which is made available at time t_o; $\sigma^2_{\hat{D}_o}$ as the variance of the initial estimate error; and $\hat{D}_i$ as the estimate of D which is made at time t_i, where $i=1,2,\ldots,n$. Furthermore, the symbol Z, which was used in the preceding sections, is given the following definition:

$$Z = D - \hat{D}_o \, . \tag{4-61}$$

In other words, Z is defined as the residual of the initial estimate $\hat{D}_o$. When Z is given such a definition, the symbols $\hat{Z}_i$ and $\tilde{Z}_i$ can be expressed as follows:

$$\hat{Z}_i = \hat{D}_i - \hat{D}_o \, . \tag{4-62}$$

$$\tilde{Z}_i = Z - \hat{Z}_i \tag{4-63}$$

$$D - \hat{D}_i \, .$$

Once the value of $\hat{Z}_i$ is known, then the value of $\dot{D}_i$ can be determined by the relation of Equation (4-62).

The i-th subperiod of the season (t_o, t_n) was defined in the preceding section as the subinterval (t_{i-1}, t_i). Let V_i be designated as the number of the seasonal items in demand for the i-th subperiod; $\hat{V}_{i,j}$ as the estimate of V_i which is computed at time t_j, where $j < i$; and v_i as the actual demand for the i-th subperiod which can be observed at time t_i.

Similar to the postulate stated by Equation (4-22), suppose the following relation holds for V_i:

$$V_i = m_i D + n_i, \tag{4-64}$$

where m_i are known constants whose values satisfy the following conditions:

$$\sum_{i=1}^{n} m_i = 1 \quad \text{and} \quad 0 < m_i < 1. \tag{4-65}$$

The Gaussian noise n_i is the same as defined in Equations (4-22), (4-23), and (4-24). Since D is a constant quantity, D is independent of (and orthogonal to) the noise n_i. In real situations, there can be many factors which contribute to the noise; for example, the customer's buying habit and the weather conditions could be such factors which explain variations in n_i.

Suppose the quantities $\hat{V}_{i,j}$ are computed by the following rule $j < i$:

$$\hat{V}_{i,j} = m_i \hat{D}_j, \quad j < i. \tag{4-66}$$

Then, the variables X_i and their associated quantities $\hat{X}_{i,i-1}$ and $\tilde{X}_{i,i-1}$ can be expressed as follows:

$$\begin{aligned} X_i &= m_i Z + n_i \\ &= m_i(D - \hat{D}_o) + n_i \\ &= (m_i D + n_i) - m_i \hat{D}_o \\ &= V_i - \hat{V}_{i,o} \end{aligned} \tag{4-67}$$

$$\begin{aligned} \hat{X}_{i,i-1} &= m_i \hat{Z}_{i-1} \\ &= m_i(\hat{D}_{i-1} - \hat{D}_o) \\ &= \hat{V}_{i,i-1} - \hat{V}_{i,o}. \end{aligned} \tag{4-68}$$

$$\begin{aligned} \tilde{X}_{i,i-1} &= X_i - \hat{X}_{i,i-1} \\ &= (V_i - \hat{V}_{i,o}) - (\hat{V}_{i,i-1} - \hat{V}_{i,o}) \\ &= V_i - \hat{V}_{i,i-1}. \end{aligned} \tag{4-69}$$

Furthermore, define x_i by the following:

$$x_i = v_i - \hat{V}_{i,o} \tag{4-70}$$

Since the values of v_i and $\hat{V}_{i,o}$ are made known at time t_i, the value of x_i can be determined by the definition above. Then, it follows that:

$$\begin{aligned} \tilde{x}_{i,i-1} &= x_i - \hat{X}_{i,i-1} \\ &= (v_i - \hat{V}_{i,o}) - (\hat{V}_{i,i-1} - \hat{V}_{i,o}) \\ &= v_i - \hat{V}_{i,i-1}. \end{aligned} \tag{4-71}$$

As shown in Equation (4-62), the problem of computing $\hat{D}_i$, given the initial estimate $\hat{D}_o$, is equivalent to that of computing $\hat{Z}_i$. The estimate $\hat{Z}_i$ can be computed by use of Equation (4-46). The values of α_i, which are needed in Equation (4-46), can be determined by use of Equations (4-44) and (4-45). In order to use Equation (4-45), the value of the initial estimate $\sigma^2_{\hat{Z}_o}$ is needed. From Equations (4-61) and (4-62):

$$\begin{aligned} \sigma^2_{\hat{Z}_o} &= E[(Z - \hat{Z}_o)^2] \\ &= E[(Z)^2] \\ &= E[(D - \hat{D}_o)^2] \\ &= \sigma^2_{\hat{D}_o} \end{aligned} \tag{4-72}$$

The problem of computing the values of $\hat{D}_o$ and $\sigma^2_{\hat{D}_o}$ is considered in the following section.

On the Assumptions of the Model

The Seasonal Period (t_o, t_n)

Suppose a seasonal period of a seasonal goods item is defined over the time interval (t_o, t_n), which is referred to as a season. The time points t_o and t_n are called the opening time and the closing time of the season, respectively. In real situations, the opening and closing times of a season are subject to random variations, and these time points are often determined arbitrarily. In some cases[12], the season is defined as being open at time t_o when demand to date reaches 5 per cent of the seasonal total demand, and as being closed at time t_n when demand to date reaches 95 per cent of the seasonal total demand. In such cases, the determination of the time points would be based on a long run history of past seasons

The Initial Estimates: $\hat{D}_o$ and $\sigma^2_{\hat{D}_o}$

The initial estimates of a seasonal demand may be obtained subjectively or objectively. It is difficult to say, generally, whether the subjective or the objective method of estimation is preferable over the other. In the case of department stores, the estimates are often made by a person or persons who are responsible for estimating the demand, obtaining the budget, buying the stock, and selling the items. In such cases, the subjective estimates of demand are often made on the

---

[12]Hertz *et al.* (18).

low side.[13] This is due to a psychological reason: if the estimates are made on the higher side and the sales fall short of the estimated target, then the sales performance may be judged unfavorably by management; on the other hand, if the estimates are low and the sales exceed the estimated target, then the sales performance may be judged favorably by management. This is an illustrative case of the multi-level-multi-goal system.

When a firm has a long-run history over past seasons, it may be possible to make an objective estimate of the seasonal demand. Suppose the firm has data over s past seasons. Let season k be one of the s seasons, where k = 1,2,...,s. The seasonal period of season k is defined by a time interval $(t_{k,o}, t_{k,n})$. The following symbols are defined as:

D_k : the random variable representing the seasonal demand for season k.

$\hat{D}_{k,o}$: the *a priori* estimate of D_k which is computed at time $t_{k,o}$.

$\sigma^2_{\hat{D}_{k,o}}$: the variance of the estimation error.

d_k : the actual demand for season k which can be observed at time $t_{k,n}$.

Suppose the (s + 1)-st season lies in the future, and consider the problem of obtaining $\hat{D}_{(s+1),o}$ and $\sigma^2_{\hat{D}_{(s+1),o}}$. Two methods will be illustrated.

---

[13] Cyert *et al.* (8), Chapter 6.

Method 1. Suppose a collection of data d_k and estimates $\hat{D}_{k,o}$ are available over s seasons; i.e.,

Seasons:	1	2	...	k	...	s
Data:	d_1	d_2	...	d_k	...	d_s
Estimates:	$\hat{D}_{1,o}$	$\hat{D}_{2,o}$	...	$\hat{D}_{k,o}$	...	$\hat{D}_{s,o}$

In this case, $\hat{D}_{(s+1),o}$ and $\sigma^2_{\hat{D}_{(s+1),o}}$ may be computed by the following rule:

$$\hat{D}_{(s+1),o} = \frac{1}{s}\sum_{k=1}^{s} d_k, \tag{4-73}$$

$$\sigma^2_{\hat{D}_{(s+1),o}} = \frac{\sum_{k=1}^{s} (d_k - \hat{D}_{k,o})^2}{s - 1} \tag{4-74}$$

Method 2. Suppose the seasonal demand D_k can be explained by some observable variable W_k; for example, the following relation may be postulated.

$$D_k = \alpha + \rho W_k + \varepsilon_k, \tag{4-75}$$

$$k = 1,2,\ldots,s,s+1,$$

where α and ρ are constants; and ε_k are[14].

$$E(\varepsilon_k) = 0 \qquad \text{for all } k, \tag{4-76}$$

$$\begin{aligned} E(\varepsilon_k \varepsilon_j) &= 0 \qquad \text{for } k \neq j, \\ &= \sigma_\varepsilon^2 \qquad \text{for } k = j. \end{aligned} \tag{4-77}$$

Suppose the following data are available at time $t_{(s+1),0}$.

Seasons	1	2	...	k	...	s, (s+1)
Data on D_k	d_1	d_2	...	d_k	...	d_s
Data on W_k	W_1	W_2	...	W_k	...	W_s, $W_{(s+1)}$.

In this case, $\hat{D}_{(s+1),o}$ and $\sigma^2_{\hat{D}_{(s+1),o}}$ can be computed by the following rule:[15]

$$\hat{D}_{(s+1),o} = \hat{\alpha} + \hat{\beta}\, W_{(s+1)}, \tag{4-78}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the familiar least squares estimates of α and β, and:

---

[14] For the case of autocorrelated disturbances, see Johnston (21) p. 178 and p. 195.

[15] Johnston (21), p. 36.

$$\sigma^2_{\hat{D}_{(s+1),o}} = \sigma^2_\epsilon \left[\frac{1}{s} + \frac{\left(W_{s+1} - \frac{\sum_{k=1}^{s} W_k}{s} \right)^2}{\sum_{k=1}^{s} W_k^2} \right] \tag{4-79}$$

The Values of m_i and $\sigma^2_{n_i}$

It has been assumed in the preceding discussions that the values of m_i and $\sigma^2_{n_i}$ are known and given in the problem. In a case study of seasonal goods inventory problems, Hertz *et al.*[16] suggested that such parameter values may be computed on the basis of historical data.

Suppose a firm has demand histories for individual seasonal items or groups of similar seasonal items. The group of items which have similar demand characteristics is sometimes referred to as a line of items. The items may be grouped in a line which are sold in a single distribution channel, at a same price range, and for a same functional use. For example, men's overcoat selling in the price range of $100 and $150 per unit through the channel of a men's wear department may be grouped in a line. Another group of men's overcoats selling in the price range of $50 and $75 per unit through the channel of a basement store may be grouped as another line.

It is assumed that sufficient historical data are available over s seasons on the basis of either individual items or lines. Let d_k denote the k-th season demand, and $v_{k,i}$ denote the i-th subperiod demand

---

[16]Hertz *et al.* (18).

within the k-th season. Assuming that m_i is a fixed constant for the i-th subperiod over all seasons, the following relation is postulated.

$$v_{k,i} = m_i d_k + e_{k,i}, \qquad (4\text{-}80)$$

$$k = 1,2,\ldots,s,$$

where $v_{k,i}$ and d_k are the given data, m_i is the constant, and $e_{k,i}$ is the disturbance. At first, it appears that the value of m_i may be computed by the familiar least squares estimate:[17]

$$\hat{m}_i = \frac{\sum_{k=1}^{s} v_{k,i} d_k}{\sum_{k=1}^{s} d_k^2} \qquad (4\text{-}81)$$

However, the values of m_i computed by Equation (4-81) may not satisfy the condition specified by Equation (4-65); i.e.,

$$\sum_{i=1}^{n} m_i = 1, \qquad o < m_i < 1.$$

A method which does work is to approximate $\hat{m}_i$ by m_i':

---

[17]Johnston (21), p. 18.

$$\hat{m}_i \doteq m_i' = \frac{\sum_{k=1}^{s} v_{k,i}}{\sum_{k=1}^{s} d_k} . \qquad (4\text{-}82)$$

If the disturbance $e_{k,i}$ is independent, then the value of $\sigma^2_{n_i}$ can be simply estimated by:

$$\hat{\sigma}^2_{n_i} = \frac{\sum_{k=1}^{s} (v_{k,i} - m_i' d_k)^2}{s - 1} \qquad (4\text{-}83)$$

If it is suspected that the disturbances are serially correlated, then the significance of autocorrelation may be tested by means of the Durbin-Watson statistic[18] or by some other methods.[19] The method of the Durbin-Watson statistic is briefly outlined as follows.

Let $u_{k,i}$ denote the autocorrelated disturbance, and write:

$$v_{k,i} = m_i d_k + u_{k,i}, \qquad (4\text{-}84)$$

where $u_{k,i}$ is assumed to follow the first order autocorrelation scheme:

$$u_{k,i} = \rho_i u_{k-1,i} + e_{k,i}. \qquad (4\text{-}85)$$

In the expression above, ρ_i is a constant; and $e_{k,i}$ is an independent

---

[18] Durbin *et al.* (45); also Johnston (21), p. 192.

[19] Theil *et al.* (48).

disturbance.

Suppose m_i' of Equation (4-82) is used to approximate m_i. Then, $u_{k,i}$ can be approximated by:

$$u'_{k,i} = v_{k,i} - m_i' d_k , \tag{4-86}$$

The Durbin-Watson statistic is computed as:

$$d = \frac{\sum_{k=2}^{s} (u'_{k,i} - u'_{(k-1),i})}{\sum_{k=2}^{s} (u'_{k,i})^2} . \tag{4-87}$$

If the value of d exceeds the limit given in the Durbin-Watson table, then it can be concluded that the autocorrelation is significant. If this is the case, the value of ρ_i is estimated by the familiar least squares estimate:

$$\rho'_i = \frac{\sum_{k=2}^{s} (u'_{k,i})(u'_{(k-1),i})}{\sum_{k=2}^{s} (u'_{k,i})^2} . \tag{4-88}$$

By use of ρ'_i, the data d_k and $v_{k,i}$ are transformed into:

$$d'_k = d_k - \rho'_i d_{(k-1)} \tag{4-89}$$

$$v'_{k,i} = v_{k,i} - \rho'_i v_{k,(i-1)} \tag{4-90}$$

The transformed data d'_k and $v'_{k,i}$ will then be used to compute new estimates m'_i and $(\hat{\sigma}'_{n_i})^2$ of m_i and $\sigma^2_{n_i}$, respectively, as:

$$m'_i = \frac{\sum_{k=2}^{s} v'_{k,i}}{\sum_{k=2}^{s} d'_k} \qquad (4\text{-}91)$$

$$(\hat{\sigma}'_{n_i})^2 = \frac{\sum_{k=2}^{s} (v'_{k,i} - m'_i d'_k)^2}{s - 2} \qquad (4\text{-}92)$$

Summary

This chapter has investigated the statistical procedures which can be used to forecast demand for seasonal goods inventory items. The procedure which is most frequently considered in the literature is that which assumes the probabilities of demand are estimated once for all before the beginning of a season. Such *a priori* estimates of the demand probabilities are referred to as the initial estimates. The procedure proposed in this chapter also accepts the initial estimates; however, the focus of analysis is placed upon the problem of correcting the initial estimation errors as more data becomes available after the season begins.

The methods of least mean square estimation and filtering theory are used as the theoretical basis for the development of the statistical procedure. The best estimate of a random variable in terms of the least mean squares can be given by the conditional mean based on observations. For the case of a Gaussian random variable, the conditional mean can be expressed as a linear combination of observations. When the estimation errors are regarded as the Gaussian random variables, the linear filter theory may be applied to consider the problem of estimating the initial estimation errors.

The basic models for the proposed procedure are given by equations (4-22) and (4-61). The formulas which can be used to compute the initial estimation errors are given by Equations (4-44) through (4-46). The development of these formulas is largely based on Shaw's linear filter model.[20] Once the initial estimation errors are estimated, then Equation (4-62) can be used recursively to re-estimate the seasonal demand. As the seasonal demand is re-estimated, the re-estimated result can be used to predict the subperiod demand. Within this framework, the filtering problem of estimating the seasonal demand will coincide with the predicting problem of estimating the subperiod demand.

If the estimated variance of the initial estimation error is very small, then the filtering method is very insensitive to correct the bias in the initial estimation. On the other hand, if the estimated variance of the initial estimation error is very large, then the filtering method becomes quite sensitive to the fluctuations in the data. In other words, it is important to have a reliable means of estimating the variance of the initial estimation error.

The accuracy of the filtering estimation depends also on the accuracy of the estimated parameter values of m_i and the noise vari-

---

[20] Shaw (34).

ances. As an extreme case, if the values of m_i are the same for all subperiods and the noise variances are also the same for all subperiods, then the re-estimates computed by the filtering method will approach the simple moving averages as the variance of the initial estimation error approaches an infinitely large number.

The methods for determining the values of the initial estimates as well as the parameter values of the model are also outlined in the later part of the chapter.

CHAPTER V

INVENTORY CONTROL FOR SEASONAL GOODS ITEMS

General

The procedures used in practice to control inventories of seasonal goods items are often such that the inventory control situation may be modeled as a multi-stage control process. The problems associated with defining the spatial boundaries of an inventory control system in retail situations, as well as defining the dynamic boundaries of such a control process, were discussed in some detail in Chapter III. When modeled as a multi-stage control process, the problem of forecasting demand becomes an integral part of the control process in such a way that, at each control point in time, the system is allowed to estimate demand as well as to determine control input.

An approach to modeling the seasonal goods inventory problem as a multi-stage control process was considered by Murray *et al.* in a recent publication.[1] The Bayesian approach to forecasting demand was made in their model on the assumption that the demand pattern follows the beta binomial probability function. Under such an assumption, their model is applicable only when the size of a demand population is exactly known. However, the size of demand population is often unknown in real situations of seasonal goods inventory control problems.

---

[1]Murray *et al.* (30).

The linear feedback filtering procedure presented in Chapter IV does not require knowledge of the size of the demand population. On the other hand, the procedure assumes that Equations (4-22) and (4-64) can be defined for the inventory situation. It seems that this assumption is reasonable and logical in view of the case studies reported by Cyert *et al.*[2] and Hertz *et al.*[3]

The general characteristics of the seasonal goods inventory problems are first considered in this chapter, and the filtering procedure is applied to formulate a seasonal goods inventory control model in the form of a multi-stage control process. The analysis is illustrated by numerical examples.

The Seasonal Goods Inventory Problem

Inventory stock items may be classified according to whether they are seasonal or nonseasonal. For a retail department store, for example, the majority of hardware items may be regarded as nonseasonal and the majority of clothing items as seasonal. The essential characteristics of seasonal goods inventories as opposed to nonseasonal inventories can be listed as follows:[4]

1. Seasonal goods inventory items have a finite demand period with well-marked opening and closing times for the season.

2. The demand rate of the items usually varies within the

---

[2]Cyert *et al.* (8).

[3]Hertz *et al.* (18).

[4]Murray *et al.* (30).

seasonal period.

3. There are only a limited number of opportunities to purchase or produce the items at varying costs which depend on the time at which decisions are made to obtain them.

4. The price of the item can be changed within the season. At the close of the season, unsold units result to high cost of obsolescence.

An examination of these characteristics will suggest that the seasonal goods inventory problem is a class of the newsboy problem or the slow-moving item inventory problem. An extensive study of the latter problem has been reported by Hadley.[5] The present problem, however, differs from Hadley's model in two aspects. First, instead of a single procurement opportunity as in Hadley's model, more than one opportunity is allowed for procurement in the present problem. Second, instead of a single estimate of the seasonal demand, a limited number of opportunities are allowed to re-estimate demand in the present problem. In the case of retail situations, the present model is a more realistic representation than Hadley's single period model; particularly for the case of department store operations.

Cyert *et al.*[6] has reported a case study of inventory control practices in department stores. According to their study, the firm divides replenishment orders into two categories of orders; namely, advance orders and reorders. Advance orders are placed early enough to

---

[5] Hadley (15), Chapter 6.

[6] Cyert *et al.* (8).

allow the firm and its suppliers to avoid uncertainties by providing contractual commitments; hence, advance orders may be obtained at a lower cost than reorders. Reorders are placed after the season begins, and are used for the purpose of controlling the uncertainty in demand as well as other uncertainties in purchase costs and selling prices.

In a case study reported by Cyert *et al.* , the amount of advance orders constitute approximately 50 per cent to 75 per cent of total seasonal orders; in a particular season, the amount of advance orders placed for Easter-season was 50 per cent, for Summer-season was 60 per cent, for Fall-season was 75 per cent, and for holiday-season was 65 per cent. Since advance orders seldom meet the total seasonal demand, the remainder of demand is filled by reorders.

A schematic diagram of the inventory ordering process is shown in Figure 17. As shown in the diagram, there are three factors which influence the amount of reorders; namely, the current inventory level, the amount of advance orders already placed, and the sales re-estimate which is made after the season begins.

According to Cyert *et al.*,[7] the re-estimate of demand may be determined by the following simple rule:

$$S_{(T-\tau)} = \frac{S'_{(T-\tau)}}{S'_{\tau}} S_{\tau} \ . \tag{5-1}$$

where

[7] *Ibid.*, p. 136.

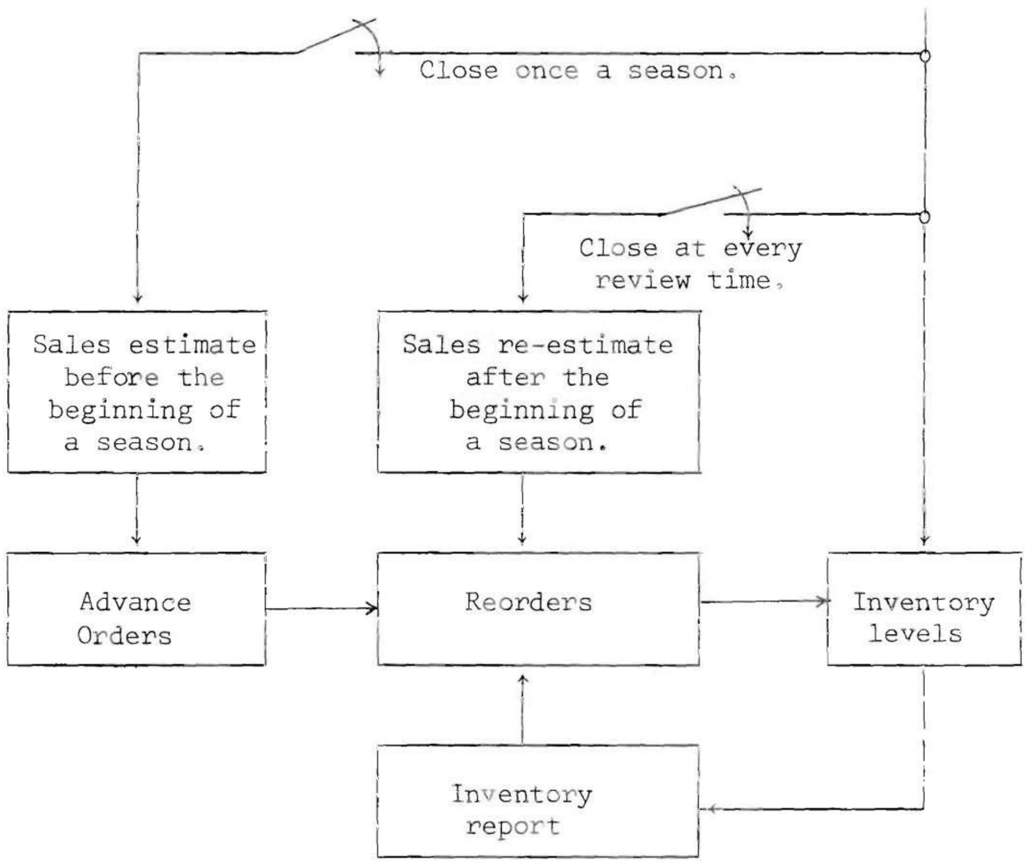

Figure 17. Advance Orders and Reorders in Inventories

S_τ = actual sales up to time τ from the beginning of a season.

$S_{(T-\tau)}$ = estimate of sales for the remainder of the season.

S'_τ = the amount of last year's S_τ.

$S'_{(T-\tau)}$ = the amount of last year's $S_{(T-\tau)}$.

If S_T denotes the total sales of a season and S'_T denotes the amount of last year's S_T, then the rule given by Cyert *et al.* in Equation (5-1) may be applied to obtain:

$$S_\tau = \frac{S'_\tau}{S'_T} S_T \ . \tag{5-2}$$

When the symbols of S_τ, $\frac{S'_\tau}{S'_T}$, and S_T are replaced by V_i, m_i, and D, respectively, then the deterministic relation of Equation (5-2) can be used as a basis to model the stochastic relation in the form of Equation (4-64):

$$V_i = m_i D + n_i \ ,$$

where n_i denotes the random disturbance. Once it is possible to model the relation expressed above, then the filtering procedure of Chapter IV may be used to obtain the re-estimates of demand probabilities.

The Seasonal Goods Inventory Model

Consider a seasonal goods inventory process for which the planning horizon is defined as the seasonal period (t_o, t_n). As discussed in the preceding chapters, (n-1) time points t_i, $i = 1,2,\ldots,(n-1)$, may be defined between t_o and t_n so that the seasonal period is divided into n subperiods. Let the time interval (t_{i-1}, t_i) be the i-th subperiod of the season.

The state, input, and output variables for the inventory control process are defined as follows. For the i-th subperiod, $i = 1,2,\ldots,n$:

y_i : the observable state variable which represents the inventory level at the beginning of the i-th subperiod. The inventory level is measured at time t_{i-1} before the replenishment q_i has arrived.

q_i : the control input which represents the replenishment. The replenishment is instantaneously made at time t_{i-1}.

v_i : the environmental input which represents the subperiod total demand. A demand may occur at any time during the subperiod; however, the subperiod total demand is observable only at the end of the subperiod.

R_i : the output which represents the return in revenue for the subperiod.

The state equation of the process can be expressed in the familiar form:

$$y_{i+1} = y_i + q_i - v_i. \tag{5-3}$$

The assumption which underlies the relation shown above is that the feedback sequence of control can take place at the beginning of the subperiods. In other words, as shown in Figure 8(a), the activities of measurement, computation, decision, and actuation can take place at the beginning of the subperiods.

Figure 18 is a schematic diagram showing the inventory control process over the planning horizon consisting of n subperiods. The initial state of the inventory process is denoted by y_1 and the post-season inventory is denoted by y_{n+1}. The horizontal flows indicated by solid lines represent the flow of material units. The vertical dotted lines represent the flow of information concerning the subperiod return R_i. The letter G_i denotes the goal-seeking unit for the i-th subperiod, which seeks to optimize the subperiod return. The letter G denotes the overall system goal which seeks to optimize the total seasonal return R.

Suppose G_i is an operator which assigns values or costs to resources utilized by the inventory process. The resources are material units which are expressed in terms of sales v_i, inventory level y_i, and replenishment q_i. Suppose G_i assigns values to these variables to give a relation with R_i which can be expressed as:

$$R_i = G_i(v_i,\ q_i,\ y_i). \tag{5-4}$$

The system goal G is also considered as an operator which relates the subperiod revenues R_i to the total revenue R; namely,

$$R = G(R_1, R_2, \ldots, R_n, R_{n+1}). \tag{5-5}$$

where R_{n+1} denotes the post season salvage return with respect to y_{n+1}.

For the inventory problem, suppose R_i can be expressed as consisting of three separable components:

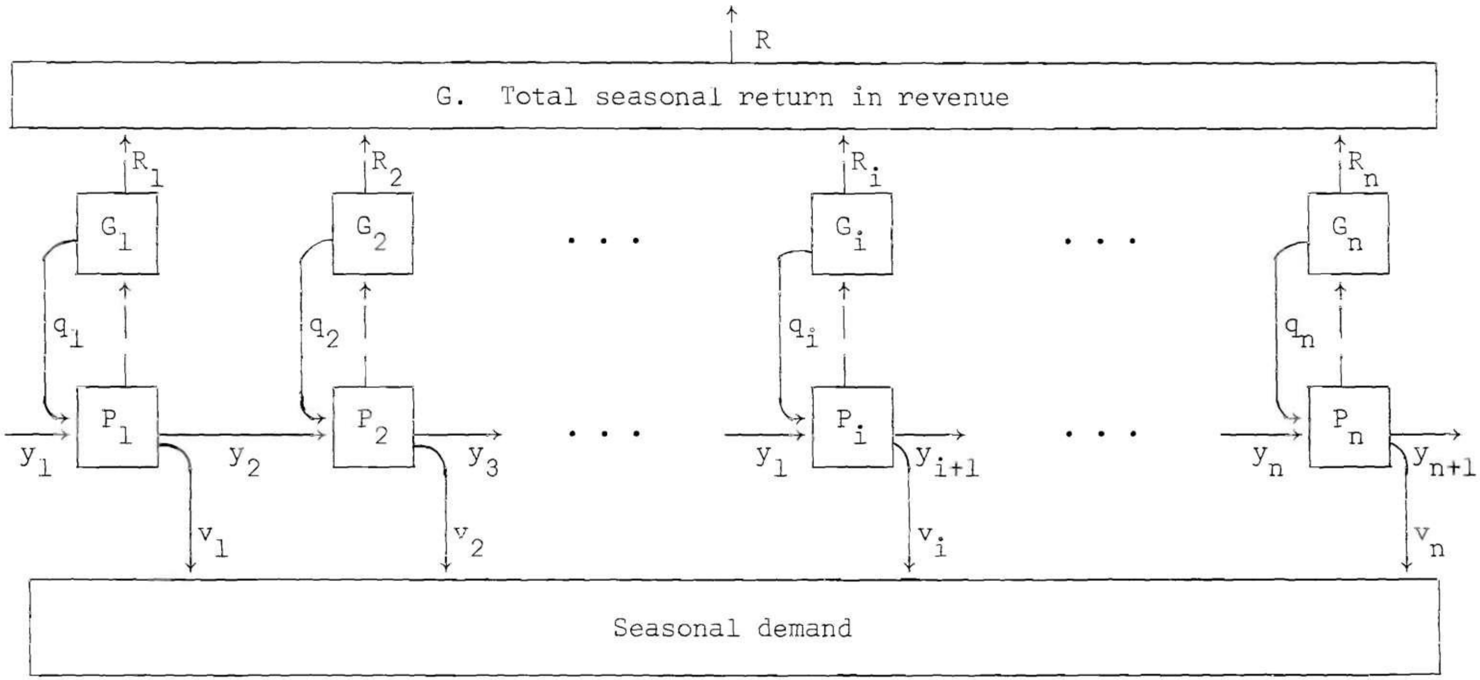

Figure 18. The n-subperiods Inventory Processes over a Season

$$R_i = G_{v,i} - G_{q,i} - G_{y,i}, \tag{5-6}$$

where,

$G_{v,i}$ = the value is units sold,

$G_{q,i}$ = the cost of units replenished,

$G_{y,i}$ = the cost of inventory holding.

Each of these components is modeled as follows. Suppose the demand can be described in terms of a random variable V_i with the probability function $p(V_i)$.

1. The expected value of units sold:

$$G_{v,i} = r_i \left[\sum_{V_i=0}^{y_i+q_i-1} V_i p(V_i) + \sum_{V_i=y_i+q_i}^{\infty} (y_i+q_i) p(V_i) \right], \tag{5-7}$$

where r_i denotes the unit selling price minus selling expenses per unit.

2. The cost of replenishment: Suppose that the unit replenishment cost, denoted by $s_i(q_i)$, is a deterministic function which depends on the volume of replenishment; for instance, this includes the situation where the volume-discount is considered. Then, $G_{q,i}$ may be expressed as:

$$G_{q,i} = [s_i(q_i)]q_i. \tag{5-8}$$

3. The expected cost of holding inventory for unsold units at the end of the i-th subperiod:

$$G_{y,i} = C_i \sum_{V_i=0}^{y_i+q_i-1} (y_i + q_i - V_i)p(V_i), \tag{5-9}$$

where C_i denotes the unit inventory holding cost.

The expected return R_i for the i-th subperiod can be expressed by use of Equations (5-7) through (5-9) as:

$$R_i = r_i \left[\sum_{V_i=0}^{y_i+q_i-1} V_i p(V_i) + \sum_{V_i=y_i+q_i}^{\infty} (y_i+q_i)p(V_i) \right] - [s_i(q_i)]q_i \tag{5-10}$$

$$- C_i \sum_{V_i=0}^{y_i+q_i-1} (y_i+q_i-V_i)p(V_i).$$

Let $g_i(y,V)$ denote the sum of expected revenues for the time interval (t_i, t_n), provided the optimum replenishment policies are employed at times $t_{i+1}, t_{i+2}, \ldots, t_n$; i.e.,

$$g_i(y,V) = \underset{q_{i+1}, q_{i+2}, \ldots, q_n}{\text{Max.}} \sum_{j=i}^{n+1} R_j. \tag{5-11}$$

Also let $f_i(y,V)$ denote the maximum revenue expected from subperiod i to the remainder of the season provided the optimum replenishment policies are employed at all the time points: $t_i, t_{i+1}, t_{i+2}, \ldots, t_n$; i.e.,

$$f_i(y,V) = \underset{q_i}{\text{max.}} \{g_i(y,V)\}. \tag{5-12}$$

The expressions $g_i(y,V)$ and $f_i(y,V)$ are shown as functions of two variables: The inventory level y which is to be controlled, and the random demand V which is to be estimated.

Combining Equations (5-10), (5-11), and (5-12), the following dynamic programming formulation can be obtained:[8]

$$f_i(y,V) = \underset{q_i}{\text{Max.}} \left\{ \sum_{V_i=0}^{y_i+q_i-1} [r_i V_i - C_i(y_i+q_i-V_i) + f_{i+1}(y_i+q_i-V_i,V)]p(V_i) + \sum_{V_i=y_i+q_i}^{\infty} [r_i(y_i+q_i) + f_{i+1}(0,V)]p(V_i) - [s_i(q_i)]q_i \right\}, \tag{5-13}$$

subject to:

$$y_{i+1} = y_i + q_i - V_i, \quad \text{and} \quad q_i \geq 0 \text{ for all } i.$$

For the post season at $i = n+1$, it is assumed that $V_{n+1} = 0$ and $q_{n+1} = 0$; hence,

$$f_{n+1}(y,0) = r_{n+1}y_{n+1}. \tag{5-14}$$

---

[8] Cf. Murray *et al.* (30).

Equation (5-13) can be rearranged to give a computationally more convenient form:

$$f_i(y,V) = \underset{q_i}{\text{Max.}} \{r_i y_i + [r_i - s_i(q_i)]q_i \tag{5-15}$$

$$- \sum_{V_i=0}^{y_i+q_i-1} [(r_i+C_i)(y_i+q_i-V_i) - f_{i+1}(y_i+q_i-V_i,V)]p(V_i)$$

$$+ \sum_{V_i=y_i+q_i}^{\infty} f_{i+1}(0,V)p(V_i)\} .$$

The optimum replenishment policy q_i^* can be determined in a straightforward manner by solving Equations (5-13) or (5-15), provided the estimates of demand are available. Obviously, the simplest approach to this problem can be found when the probability function $p(V_i)$ of the i-th subperiod demand is known for all subperiods.

Hadley[9] considered a case where $p(V_i)$ is the Poisson density function with mean λ_i:

$$f_p(V_i|\lambda_i) = \frac{(\lambda_i)^{V_i} e^{-\lambda_i}}{V_i!} . \tag{5-16}$$

$$V_i = 0,1,2,\ldots .$$

$$i = 1,2,\ldots,n .$$

---

[9]Hadley (15), p. 310.

If the random behavior of demand can be described by the Poisson density function for each subperiod i, then it may be possible to obtain the estimates of the mean λ_i, and subsequently compute the optimum solutions for ordering quantities. The key to this problem is, however, the procedure used for estimating the unknown means for future subperiods. Hadley assumed that the means are either exactly known or determinable from a functional relationship. He did not consider the situation where the re-estimates of the seasonal demand are obtained on the basis of sales observations made within the season.

In a recent publication, Murray *et al.*[10] reported a study where the re-estimates of future demand are obtained from the sales performance in the earlier part of the season. They made a Bayesian approach to forecast demand probabilities by assuming that the random behavior of demand can be described by the beta binomial probability density function. Let N be the number of total potential customers, V be the number of actual customers, and p be the fraction of N that generates the actual demand. They assumed that the fraction p is distributed as the beta normalized density function:[11]

$$f_\beta(p|V,N) = \frac{1}{B(V,N-v)} p^{V-1} (1-p)^{N-V-1} , \qquad (5\text{-}17)$$

$$0 \le p \le 1 ,$$

$$N > V > 0 ,$$

[10]Murray *et al.* (30).

[11]Raiffa *et al.* (32), p. 218.

where $B(V,N-V)$ is the complete beta function:

$$B(V,N-V) = \frac{(V-1)!\,(N-V-1)!}{(N-1)!} . \qquad (5\text{-}18)$$

Now suppose the seasonal period can be divided into subperiods, and let N_i be the number of potential customers for subperiod i. Suppose N_i is known exactly for all subperiods, but the fraction of N_i who will purchase the seasonal item is unknown. Let V_i be the number of actual customers who will purchase the item during subperiod i, $\bar{v}_i$ be the cumulative number of customers who have purchased the item prior to subperiod i, and $\bar{N}_i$ be the cumulative number of potential customers prior to subperiod i for the season. Under these assumptions, the probability that the N_i potential customers in subperiod i will generate demand for V_i units given observations on $\bar{v}_i$ and $\bar{N}_i$ can be expressed as the beta probability function:[12]

$$f_{\beta b}(V_i|\bar{v}_i,\bar{N}_i,N_i) = \int_0^1 f_b(V_i|p_i,N_i)f_\beta(p_i|\bar{v}_i,\bar{N}_i)dp_i \ , \qquad (5\text{-}19)$$

where $f_b(V_i|p_i,N_i)$ is the binomial function:

$$f_b(V_i|p_i,N_i) = \frac{N_i!}{V_i!(N_i-V_i)!} p_i^{V_i}(1 - p_i)^{N_i-V_i} \ , \qquad (5\text{-}20)$$

$$0 \le p_i \le 1 \ ,$$

---

[12]Raiffa (32), p. 237.

and $f_\beta(p_i|\bar{v}_i,\bar{N}_i)$ is the beta normalized density function of Equation (5-17). The beta binomial probability function of Equation (5-19) can be expressed in the computational form:

$$f_{\beta b}(V_i|\bar{v},\bar{N}_i,N_i) = \frac{(V_i+\bar{v}_i-1)!(N_i+\bar{N}_i-V_i-\bar{v}_i-1)!N_i!(\bar{N}_i-1)!}{(\bar{v}_i-1)!(N_i-V_i)!(\bar{N}_i-\bar{v}_i-1)!(N_i+\bar{N}_i-1)!V_i!} \qquad (5\text{-}21)$$

If demand follows the beta binomial probability law, then one can use the Bayesian approach to compute the future demand estimates, provided the number of potential customers is known with certainty. In the case of seasonal goods inventory situations, it is often unrealistic to make such an assumption that the number of potential customers is known; except, perhaps, for some special cases.[13] When the size of the demand population is not exactly known, then the Bayesian approach using the beta binomial probability function to estimate the future demand is not applicable.

The feedback filtering procedure presented in Chapter IV does not require the *a priori* knowledge of the size of the customer population. On the other hand, the application of the filtering procedure requires that the assumptions underlying Equation (4-64) are satisfied in the given situation. In reference to the case studies reported by Hertz *et al.* and Cyert *et al.*, this requirement seems to be a reasonable one. Some numerical examples will illustrate the application of the filtering procedure to solve the inventory problem.

---

[13] Murray *et al.* (30) mentioned that the mail order situation is one of such special cases.

Numerical Examples

To consider the simplest possible situation for illustration, suppose a seasonal period can be divided into two subperiods; i.e., $n = 2$. For this situation, the dynamic programming formulation of Equations (5-14) and (5-15) can be expressed as follows:

$$f_3(y,0) = n_3 y_3 \ , \tag{5-22}$$

$$\begin{aligned} f_2(y,V) &= \underset{q_2}{\text{Max.}} \ \{g_2(y,V)\} \\ &= \underset{q_2}{\text{Max.}} \ \{r_2 y_2 + [r_2 - s_2(q_2)]q_2 \\ &\qquad - \sum_{V_2=0}^{y_2+q_2-1} [(r_2+C_2)(y_2+q_2-V_2) - r_3 y_3]p(V_2)\} \\ &= \underset{q_2}{\text{Max.}} \ \{r_2 y_2 + [r_2 - s_2(q_2)]q_2 \\ &\qquad - (r_2+C_2-r_3) \sum_{V_2=0}^{y_2+q_2-1} (y_2+q_2-V_2)p(V_2)\} \ , \end{aligned} \tag{5-23}$$

$$\begin{aligned} f_1(y,V) &= \underset{q_1}{\text{Max.}} \ \{g_1(y,V)\} \\ &= \underset{q_1}{\text{Max.}} \ \{r_1 y_1 + [r_1 - s_1(q_1)]q_1 \\ &\qquad - \sum_{V_1=0}^{y_1+q_1-1} [(r_1+C_1)(y_1+q_1-V_1) - f_2(y,V)]p(V_1) \end{aligned} \tag{5-24}$$

$$+ \sum_{V_1=y_1+q_1}^{\infty} f_2(0,V)p(V_1)\} .$$

For the purpose of illustration, the following hypothetical data will be used in this example:[14]

$$r_1 = 12, \quad r_2 = 12, \quad r_3 = 3,$$

$$s_1(q_1) = 5, \quad s_2(q_2) = 8,$$

$$C_1 = C_2 = 0,$$

$$y_1 = 0.$$

In other words, the price remains constant at 12 per unit within the season, but its post season salvage price is only 3 per unit. The purchase costs are independent of the volume, but depend on the time of purchase; namely, 5 per unit at time t_1 and 8 per unit at time t_2. The inventory holding cost is assumed to be negligible. The initial inventory level y_1 is assumed to be zero. When these numerical values are substituted into Equations (5-22), (5-23), and (5-24), then:

$$f_3(y,0) = 3y_3 , \tag{5-25}$$

$$f_2(y,V) = \underset{q_2}{\text{Max.}} \{12y_2 + 4q_2 - 9 \sum_{V_2=0}^{y_2+q_2-1} (y_2+q_2-V_2)p(V_2)\} , \tag{5-26}$$

---

[14]These are the same data used in Murray *et al.* (30).

$$f_1(0,V) = \max_{q_1} \{7q_1 - \sum_{V_1=0}^{q_1-1} [12(q_1-V_1) - f_2(y,V)]p(V_1) \quad (5\text{-}27)$$

$$+ \sum_{V_1=q_1}^{\infty} f_2(0,V)p(V_1)\} .$$

With respect to the demand probabilities, Murray *et al.* assumed that the number of potential customers is exactly known to be 3 for the first subperiod and 5 for the second subperiod. The *a priori* estimate of the fraction of customers who will generate the demand is assumed to be 0.5. In summary, the forecasting procedure of Murray *et al.* is based on the following main assumptions:

1. The probability of demand can be described by the beta binomial probability function.

2. The size of the demand population is exactly known; e.g., $N_1 = 3$ and $N_2 = 5$.

3. The *a priori* estimate of the probability that any member of the demand population will generate a demand is available before the season begins; e.g., $p_1 = 0.5$.

Under these assumptions, the probabilities of demand for subperiods 1 and 2 can be expressed as:

$$p(V_1) = f_{\beta b}(V_1|\bar{v}_1,\bar{N}_1,N_1), \qquad 0 \le V_1 \le 3, \qquad (5\text{-}28)$$

$$p(V_2) = f_{\beta b}(V_2|\bar{v}_2,\bar{N}_2,N_2) \qquad (5\text{-}29)$$

$$= f_{\beta b}(V_2|\bar{v}_1+V_1,\ \bar{N}_1+N_1,\ N_2), \quad 0 \leq V_2 \leq 5. \qquad (5\text{-}29)$$

The values of $p(V_1)$ and $p(V_2)$ can be used to solve equation (5-27) for the optimum ordering quantities.

The method for using the feedback filtering procedure of Chapter IV is illustrated as follows. Suppose it is possible to relate the sub-period demand V_i to the seasonal total demand D by a linear relation of equation (4-64); i.e.,

$$V_i = m_i D + n_i \ ,$$

where m_i is a fractional number, $0 \leq m_i < 1$, and n_i is a zero mean normal random variable with variance $\sigma^2_{n_i}$.

The following values are assumed for the illustration:

$$\hat{D}_o = 4 \ , \qquad \sigma^2_{\hat{D}_o} = 4;$$

$$m_1 = (3/8), \quad m_2 = (5/8);$$

$$\hat{\sigma}^2_{n_1} = 1.5, \qquad \hat{\sigma}^2_{n_2} = 2.5.$$

On the basis of the given data, the following quantities can be computed at time t_o. By use of Equation (4-66):

$$\hat{V}_{1,o} = m_1\hat{D}_o = (3/8)(4) = 1.5. \qquad (5\text{-}30)$$

By use of Equation (4-44):

$$\alpha_1 = \frac{m_1 E(\tilde{Z}_o^2)}{(m_1)^2 \; E(\tilde{Z}_o^2) + \sigma_{n_1}^2} = \frac{(3/8)(4)}{(3/8)^2(4) + 1.5} = 0.73 \quad (5\text{-}31)$$

When the new data v_1 becomes available at time t_1, the following quantities can be computed. By use of Equations (4-62), (4-46), (4-71) and (5-30):

$$\hat{D}_1 = \hat{D}_o + \alpha_1(v_1 - \hat{V}_{1,o}) \quad (5\text{-}32)$$

$$= 4 + (0.73)(v_1 - 1.5)$$

$$= 2.90 + 0.73(v_1) \; .$$

Once the values of $\hat{D}_1$ is calculated, then it can be used to compute the *a priori* estimate $\hat{V}_{2,1}$ as well as the *a posteriori* estimate $\hat{V}_{1,1}$; i.e., by use of Equation (4-66):

$$\hat{V}_{2,1} = m_2\hat{D}_1 = (5/8)\;\hat{D}_1 \; , \quad (5\text{-}33)$$

$$\hat{V}_{1,1} = m_1\hat{D}_1 = (3/8)\;\hat{D}_1 \; . \quad (5\text{-}34)$$

The values of $\hat{V}_{2,1}$ and $\hat{V}_{1,1}$ are used at time t_1 to be the expected values of the random variables V_2 and V_1, respectively.

When the noise n_i in Equation (4-64) is a Gaussian random variable, then V_i is also Gaussian. For example, $p(V_1)$ is the normal density function whose mean and variance are estimated at time t_o to be the values of $\hat{V}_{1,o}$ and $\hat{\sigma}^2_{n_1}$, respectively. The parameter values of $p(V_1)$ are then re-estimated at time t_1 to be the values of $\hat{V}_{1,1}$ and $\hat{\sigma}^2_{n_1}$, respectively, provided the magnitude of $\sigma^2_{n_1}$ is not affected by the estimated values of V_1. In some cases,[15] the magnitude of V_i affects the magnitude of $\sigma^2_{n_i}$, and it is necessary to re-estimate $\sigma^2_{n_i}$.

For the purposes of computational conveniences, the Poisson approximation to the normal distribution will be made in this example. Although this assumption is a very restrictive one, it facilitates the amount of computations required for the example under consideration. The Poisson density function with mean λ_i is shown in Equation (5-16). Under the assumption stated above, the means λ_1 and λ_2 are estimated at time t_o to be the values of $\hat{V}_{1,o}$ and $\hat{V}_{2,o}$, respectively; and subsequently re-estimated at time t_1 to be the values of $\hat{V}_{1,1}$ and $\hat{V}_{2,1}$, respectively. Such computed values are shown in Table 12.

Table 12. The Values of $\hat{V}_{2,1}$ and $\hat{V}_{1,1}$ Computed at Time t_1 Given the Values of Data v_1

v_1	0	1	2	3	4	5	6	7
$\hat{V}_{2,1}$	1.8	2.3	2.7	3.2	3.6	4.1	4.5	5.0
$\hat{V}_{1,1}$	1.1	1.4	1.6	1.9	2.2	2.5	2.7	3.0

[15] Johnston (21), pp. 207-211.

The range of v_1 in Table 12 is shown as $0 \leq v_1 \leq 7$, since $f_p(V_1/\lambda_1) \doteq 0$ for $V_1 \quad 7$.

The probability functions $p(V_1)$ and $p(V_2)$ can be described at time t_1 as follows:

$$p(V_1) = f_p(V_1/\lambda_1 = \hat{V}_{1,1};\ v_1) \tag{5-35}$$

$$p(V_2) = f_p(V_2/\lambda_2 = \hat{V}_{2,1};\ v_1). \tag{5-36}$$

The probability values of $p(V_1)$ and $p(V_2)$ can be obtained from a Poisson probability table[16] corresponding to each of the estimated means shown in Table 12. The probability values are subsequently used to solve Equations (5-26) and (5-27) for the optimum ordering quantities q_1^* and q_2^*. The computational scheme for the dynamic programming problem is relatively straightforward for the present example; however, the computational requirements would have been very great if the example was not made as simple as the present one. The computational results are summarized in Table 13.

The optimum solutions can be found from Table 13 to be:

$$q_1^* = 4, \quad \text{and} \quad f_1(0,V) = 19.77.$$

The solution for the second period depends upon the actual outcome v_1 as shown in the tabulation on page 128.

---

[16] Molina (29).

Table 13. Computational Results for Example: When the Filtering Procedure is Used for Estimation

SECOND SUBPERIOD

v_1	0		1		2		3		4		5		6		7	
$\hat{V}_{2,1}$	1.8		2.3		2.7		3.2		3.6		4.1		4.5		5.0	
y_2	q_2^*	$f_2(y,V)$	q_2^*	$f_2(y,V)$	q_2^*	$f_2(y,V)$	q_2^*	$f_2(y,V)$	q_2^*	$f_2(y,V)$	q_2^*	$f_2(y,V)$	q_2^*	$f_2(y,V)$	q_2^*	$f_2(y,V)$
0	1	2.51	2	4.12	2	5.16	2	6.08	2	6.63	3	9.08	3	9.81	4	12.09
1	0	10.51	1	12.12	1	13.16	1	14.08	1	14.63	2	17.08	2	17.81	3	20.09
2	0	18.34	0	20.12	0	21.16	0	22.08	0	22.63	1	25.08	1	25.81	2	28.09
3	0.	23.81	0	26.77	0	28.73	0	30.66	0	31.91	0	33.08	0	33.81	1	36.09
4	0	27.73	0	31.57	0	34.28	0	37.25	0	39.27	0	41.35	0	42.70	0	44.09
5	0	31.07	0.	35.31	0	38.54	0	42.22	0	44.90	0	47.87	0	49.91	0	52.08
6	0	34.20	0	38.57	0	42.05	0	46.21	0	49.28	0	52.91	0	55.59	0	58.57
7	0	37.20	0	41.70	0	45.23	0	49.58	0	52.99	0	57.03	0	60.13	0	63.67

FIRST SUBPERIOD

q_1	0	1	2	3	4	5	6	7
$g_1(0,V)$	6.57	12.24	16.4[illegible]	18.8[illegible]	19.77	19.01	17.25	14.62

v_1	0	1	2	3	4	5	6	7
q_2^*	0	0	0	1	2	3	3	4

For purposes of comparison, consider a situation where the parameters of the demand variable are estimated only once at the beginning of the season, and no re-estimates are allowed after the season begins. Making use of the data given in the example, suppose $p(V_1)$ and $p(V_2)$ can be represented by the Poisson density functions:

$$p(V_1) = f_p(v_1/\lambda_1 = 1.5) \ , \qquad (5\text{-}37)$$

$$p(V_2) = f_p(v_2/\lambda_2 = 2.5) \ , \qquad (5\text{-}38)$$

where the means are estimated on the basis of the initial estimate $\hat{D}_o$; namely:

$$\lambda_1 = m_1\hat{D}_o = (3/8)(4) = 1.5,$$

$$\lambda_2 = m_2\hat{D}_o = (5/8)(4) = 2.5.$$

The seasonal goods inventory problem with this type of demand probabilities is well known in the literature.[17] When Equations (5-26) and (5-27) are solved with $p(V_1)$ and $p(V_2)$ shown in Equations (5-37) and (5-38), the results can be obtained as tabulated in Table 14.

---

[17]For example, see Hadley (15), Chapter 6.

Table 14. Computational Results for Example: The Case Without Re-estimation

SECOND SUBPERIOD			FIRST SUBPERIOD	
y_2	q_2^*	$f_2(y,V)$	q_1	$g_1(0,V)$
0	2	4.66	0	4.66
1	1	12.66	1	10.77
2	0	20.66	2	15.54
3	0	27.77	3	19.10
4	0	32.98	4	21.45
5	0	36.93	5	22.46
6	0	40.31	6	22.27
7	0	43.44	7	21.23

The optimum solution can be found from Table 14 to be:

$$q_1^* = 5, \qquad \text{and} \qquad f_1(0,V) = 22.46$$

The solution for the second subperiod depends upon the actual outcome v_1 as shown in the tabulation below:

v_1	0	1	2	3	4	5	6	7
q_2^*	0	0	0	0	1	2	2	2

A comparison of this solution (when the re-estimation of demand is not allowed in the model) with the preceding (when the re-estimation of demand is made by means of the filtering procedure) indicates that the advance order quantities are not the same; namely, the advance order is smaller when the re-estimation is allowed in the model. The reorder quantities are also not the same in two cases; namely, for $v_1 > 3$, the reorder is greater when the re-estimation is allowed in the model. It is further noted that the optimum expected return is higher when the demand is not re-estimated; i.e., $f_1(0,V) = 22.46$, and is lower when the demand is re-estimated; i.e., $f_1(0,V) = 19.77$.

These results are interpreted as follows: The re-estimation scheme allows a reduction in the initial investment (i.e., the advance order quantities); however, it allows a greater flexibility in the second investment (i.e., the reorders). The difference in the values of $f_1(0,V)$ is interpreted as follows: If the re-estimation is not allowed, it is equivalent to assuming that the variance of the initial estimation error is very small. On the other hand, if the re-estimation is allowed, it is equivalent to assuming that the variance of the initial estimation error is not small. In other words, the re-estimation would be required if the uncertainty in the initial estimates is greater. It then follows that the expected return would be smaller when the uncertainties in the future events are greater.

Summary

The general model of the multi-stage control process, which was discussed in Chapter III, is used in this chapter as a basis for formulating a seasonal goods inventory model. A seasonal period is divided by a finite number of time points so that the estimation of demand as well as the determination of order quantities are allowed to take place at each of these time points.

This problem was recently considered by Murray *et al.*[18] The main difference between their model and the present model lies in the procedure used for estimating demand. The forecasting procedure of Murray *et al.* is based on the following main assumptions.

1. The probability of demand can be described by the beta binomial probability function.

2. The size of demand population is exactly known; e.g., $N_1 = 3$ and $N_2 = 5$.

3. The *a priori* estimate of the probability that any member of the demand population will generate a demand is available before the season begins; e.g., $p_i = 0.5$.

The forecasting procedure of the present model is based on the following main assumptions.

1. The linear feedback filter procedure can be used to estimate the trend in demand probabilities.

2. The model of Equation (4-22) and Equation (4-64) can be defined for the inventory situation.

3. The *a priori* estimates of the seasonal demand are available at the beginning of the season.

In reference to the case studies reported by Cyert *et al.*[19] and Hertz *et al.*,[20] it seems that the assumptions of the present model are

---

[18] Murray *et al.* (30).

[19] Cyert *et al.* (8).

[20] Hertz *et al.* (18).

a more reasonable and logical representation of the inventory practice than the assumptions given by Murray *et al.*

The use of the proposed model is illustrated by numerical examples. The results of the examples indicate that, when the re-estimation is made, the amount of advance orders is smaller than the case without re-estimation.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The complexities associated with management control problems often make it necessary to carefully examine the procedure used for modeling the real world situation. This research is directed toward two main objectives: (1) to develop a theoretical frame of reference which can be conveniently used to model management control problems in general; and (2) to develop a seasonal goods inventory model which gives a realistic representation of the inventory situation in practice. The results and conclusions evolved from this research are summarized as follows:

1. According to the existing knowledge in the field, it appears that system theory offers the most helpful and logical basis for modeling complex situations. By making new interpretations of existing concepts in system theory, a concise and unified body of theory is formulated and discussed in Chapter II which may be particularly useful in modeling management control problems. Given a situation for management control, the first step in the modeling procedure is recognized as the definition of the spatial boundaries of a problem so that the problem can be structured as a system. Such a system may be modeled by considering the topics of hierarchical system structure, system attributes, and system goals. In particular, a management system may be

structured with respect to the nontransferable attributes; the system behavior may be analyzed with respect to the information attributes; and the system goal may be identified with a single-level-single-goal system.

2. Once the spatial boundaries of a system problem is defined, the subsequent step in the modeling procedure is to define the dynamic boundaries of the system process. For this purpose, the topics of the multi-stage control processes and the feedback control sequences are considered in Chapter III. The general procedure is illustrated with an inventory situation of retail firms. First, the spatial boundaries of the inventory situation are defined so that the inventory problem can be recognized as a relatively isolated system within the overall organizational structure. Subsequently, the inventory system is modeled within the framework of the multi-stage control processes.

3. In the formulation of a multi-stage control process, a method is required to estimate the statistical characteristics of a random process which underlies the system state. In the case of the inventory control process, this situation applies to the problem of demand forecasting. A method which can be used to forecast demand for seasonal goods inventory items is developed in Chapter IV. The procedure which is most frequently considered in the literature is that which assumes the probabilities of demand are estimated once for all before the beginning of a season. The proposed procedure also accepts such initial estimates; however, a filtering procedure[1] is

[1] Shaw (34).

applied so that the initial estimation errors can be corrected as more data become available after the season begins. The filtering procedure is primarily used to re-estimate the seasonal demand; however, the re-estimated results can be also used to predict the subperiod demand for the season. Within this framework, the filtering problem of estimating the seasonal demand coincides with the prediction problem of estimating the subperiod demands.

The proposed filtering procedure is very sensitive to the parameter values used in the model. If the estimated variance of the initial estimation error is very small, then the procedure is very slow in the correction of the large bias errors in the initial estimation. On the other hand, if the estimated variance of the initial estimation error is very large, then the filtering procedure becomes very sensitive to the fluctuations in the data. If the variance of the initial estimation error approaches an infinitely large number, then the re-estimated values computed by the filtering method will approach the simple averages in a special case considered in the study.

4. The general procedure for modeling and forecasting is subsequently applied to model a seasonal goods inventory control situation of retail firms. The seasonal goods inventory problems have been solved in the literature for the case where re-estimates of demand probabilities are not allowed in the model. In practice, however, a seasonal period is often divided by a finite number of time points such that the estimation of demand as well as the determination of order quantities are allowed to take place at each of these time points.

In a recent publication, Murray *et al.*[2] considered a seasonal goods inventory model which allows re-estimates of demand probabilities. However, their model is applicable only when *a priori* knowledge of the size of the demand population is available. Such a knowledge is not needed in the present model which makes use of the linear feedback filtering procedure. In reference to the case studies reported by Cyert *et al.*[3] and Hertz *et al.*,[4] the present model appears to be a logical representation of the seasonal goods inventory situation in practice.

Recommendations

A specific inventory situation of a retail firm is used in this study to provide a background for the theoretical analysis and development. The general outcome of the study may be applied to other situations in management control problems with appropriate modifications to meet specific characteristics of individual problems. Some possible topics for additional research may be suggested as follows.

1. The objective function is expressed in the form of a maximization problem in the present study. This is based on the assumption that the goal of the system is to maximize the net return in revenue as specified by the objective function. According to the Simon-March hypothesis,[5] the system goals are often concerned with the discovery

---

[2]Murray *et al.* (30).

[3]Cyert *et al.* (8).

[4]Hertz *et al.* (18).

[5]March (26).

and selection of acceptable alternatives rather than optimal alternatives. If the system goal is to meet an acceptable level of performance, then the objective function may be expressed in the form of minimizing a quadratic cost function.

If the acceptable level of performance is known to the system, then a straight-forward application of control theory of physical systems can be made to study the situation.[6] On the other hand, if the acceptable level of performance is not exactly known, then the problem becomes relatively difficult and complicated.[7]

2. The individual stage of a multi-stage control process can be described in terms of a feedback control sequence which consists of measurement, estimation, computation, optimization, decision, and actuation. The present study assumed that the time lag between these activities in sequence is not significant enough to affect the outcome of a solution. In many cases, however, the time lags cause serious problems; for example, the replenishment lead time. For such a situation, the actuation aspect of sequence may be analyzed in detail.

3. In this study, the seasonal goods inventory problem is formulated with only one decision variable representing the order quantities. In the retail situations, the level of promotional efforts may be regarded as another decision variable. In such a case, the demand generating subsystem is no longer uncontrollable, but can be regarded as a controllable subsystem. In order to analyze this situa-

---

[6] This was the case in the study of Holt *et al.* (20).

[7] See Aseltine (1); Charnes (6).

tion, a system equation is needed which describes a relationship between the levels of consumer response and promotional efforts; for example, the level of consumer response may be defined as another state variable. When such a knowledge is available, then the multi-stage control processes can be modeled with two state variables and two decision variables.

APPENDICES

APPENDIX 1

GOALS OF A SINGLE-LEVEL-MULTI-GOAL SYSTEM

The concept of multi-goal-multi-level systems was introduced by Mesarovic *et al.* in a recent publication.[1] A single-level-multi-goal system is a special case of such multi-level-multi-goal systems. In an earlier paper, the problem of ranking multiple goals was considered by Marschak.[2] These concepts are jointly applied in this Appendix in order to develop a procedure for recognizing the *unordered* system goals as well as the *ordered* system goals.

Suppose a system S consists of k components: $C_1, C_2, \ldots, C_i, \ldots, C_k$. It is assumed that each component may have its own goal, and let G_i denote the goal of component C_i. Furthermore, let G_0 denote the system goal. For purposes of the present discussion, suppose there are three alternatives; say x, y, and z, over which the goals of system components can establish their own preferences. Making use of the notation introduced by Marschak, let xG_iy be interpreted as: from the viewpoint of G_i, the alternative x is as good as the alternative y. In other words, for G_i, the alternative x is preferable or equivalent to y. Suppose the goals are *rational* in the sense that the following conditions are satisfied:

---

[1]Mesarovic *et al.* (28).

[2]Marschak (27)

1. Transitive condition: i.e., xG_iy and yG_iz implies xG_iz.

2. Irreflexive condition: i.e., xG_iy or yG_ix can hold true, but both cannot hold true at the same time unless x and y are identical.

When the component goals are rational, it may be possible to consider a system goal. On the other hand, when component goals are not rational, then it would be meaningless to consider a system goal. When the component goals are rational, the system goal, or the group goal, G_0, can be regarded as being *unordered* if: when xG_iy does not hold for $i = 1,2,\ldots,k$, then xG_0y does not hold; stated equivalently, xG_0y holds only if xG_iy holds for all i, $i = 1,2,\ldots,k$. The unordered system goal may be represented by an unordered set:

$$G_0 = \{G_1,G_2,\ldots,G_i,\ldots,G_k\} \ . \tag{A1-1}$$

The system goal G_0 can be regarded as being *ordered* if: xG_0y holds, even when xG_iy may not hold for some i, $i = 1,2,\ldots,k$. In this case, the ordered system goal can be expressed by an ordered set:

$$G_0 = (G_{(1)},G_{(2)},\ldots,G_{(k)}) \ . \tag{A1-2}$$

In the expression above, the parenthesized subscripts refer to the order of preference of component goals. Finally, the system goal G_0 can be regarded as a single-goal, if all component goals are identical to the system goal.

For an illustration,[3] consider an industrial firm as a system S which consists of two components. Let component C_1 be the management of the firm, and component C_2 be the labor. Also let G_0, G_1, and G_2 denote the goals of S, C_1, and C_2, respectively. Suppose x, y, and z be the three alternatives over which the goals of system components can establish their own preferences. In the event that a labor dispute takes place, the system goal can be described by an unordered set with two component goals. When it is possible to have a negotiation over the labor dispute, then, during the time of negotiation, the system goal can be described by an ordered set of two component goals. Finally, when an agreement is made between labor and management, then the system goal can be regarded as a single-goal.

---

[3] This problem was originally given by Marschak (27).

REFERENCES

REFERENCES

1. Aseltine, J. A., A. R. Mancini, and C. W. Sarture, "A Survey of Adaptive Control Systems," *IRE Transactions*, The Institute of Radio Engineers, December, 1958, pp. 102-108.

2. Ashby, W. R., *Design for a Brain*, Wiley, 1960.

3. Bellman, R., *Adaptive Control Processes: A Guided Tour*, Princeton University Press, 1961.

4. Brown, R. G., and R. Meyer, "The Fundamental Theorem of Exponential Smoothing," *Operations Research*, Vol. 9, 1961, pp. 673-685.

5. Chang, S. H., "Application of Control Theory to the Design of Production Systems," 1964, unpublished.

6. Charnes, A. and A. C. Stedry, "Search-Theoretic Models of Organization Control by Budgeted Multiple Goals," *Management Science*, Vol. 12, No. 5, 1966, pp. 448-456.

7. Cohen, G. D., "A Note on Exponential Smoothing and Autocorrelated Inputs," *Operations Research*, 1963, pp. 361-367.

8. Cyert, R. M., and J. G. March, *A Behavioral Theory of the Firm*, Prentice-Hall, 1963.

9. Forrester, J. W., *Industrial Dynamics*, M.I.T. Press, 1961.

10. Gilbert, E. G., "Controllability and Observability in Multivariate Control Systems," *Journal of the Society for Industrial and Applied Mathematics*, Control Series A, Vol. 2, No. 1, 1963, pp. 128-151.

11. Goode, H. H., and R. E. Machol, *Systems Engineering*, McGraw-Hill, 1957.

12. Gosling, W., *The Design of Engineering Systems*, Wiley, 1962.

13. Graybill, F. A., *An Introduction to Linear Stochastic Models*, Vol. I, McGraw-Hill, 1961.

14. Hadley, G., *Nonlinear and Dynamic Programming*, Addison-Wesley, 1964.

15. Hadley, G., and T. M. Whitin, *Analysis of Inventory Systems*, Prentice-Hall, 1963.

16. Hall, A. D., *A Methodology for Systems Engineering*, Van Nostrand, 1962.

17. Hammond, J. L., Jr., and S. H. Chang, "Systems Engineering I," Class notes used at Georgia Institute of Technology, unpublished, 1966.

18. Hertz, D. B., and K. H. Schaffir, "A Forecasting Method for Management of Seasonal Style-Goods Inventories," *Operations Research*, Vol. 8, 1960, pp. 45-52.

19. Holdren, R. B., *The Structure of a Retail Market and the Market Behavior of Retail Units*, Prentice-Hall, 1960.

20. Holt, C. C., F. Modigliani, J. F. Muth, and H. A. Simon, *Planning Production, Inventories, and Work Forces*, Prentice-Hall, 1960.

21. Johnston, J., *Econometric Methods*, McGraw-Hill, 1963.

22. Kalman, R. E., and R. W. Koepeke, "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes," *ASME Transactions*, American Society of Mechanical Engineers, Vol. 80, 1958, pp. 1820-1826.

23. Koenig, H. E., and W. A. Blackwell, *Electro-Mechanical System Theory*, McGraw-Hill, 1961.

24. Lehrer, R. N., *Work Simplification*, Prentice-Hall, 1957.

25. MacFarlane, A. G. J., *Engineering Systems Analysis*, Addison-Wesley, 1964.

26. March, J. G., and H. A. Simon, *Organizations*, Wiley, 1958.

27. Marschak, J., "Toward an Economic Theory of Organization and Information," in R. M. Thrall (ed.), *Decision Processes*, Wiley, 1954.

28. Mesarovic, M. D., J. L. Sanders, and C. F. Sprague, "An Axiomatic Approach to Organizations from a General Systems Viewpoint," in W. W. Cooper (ed.), *New Perspectives in Organization Research*, Wiley, 1964.

29. Molina, E. C., *Poisson's Exponential Binomial Limit*, Van Nostrand, 1942.

30. Murray, G. R., Jr., and E. A. Silver, "A Bayesian Analysis of the Style Goods Inventory Problems," *Management Science*, Vol. 12, No. 11, 1966, pp. 785-797.

31. Nemhauser, G. L., *Introduction to Dynamic Programming*, Wiley, 1966.

32. Raiffa, H., and R. Schlaifer, *Applied Statistical Decision Theory*, Harvard University Press, 1961.

33. Regan, W. J., "The Stages of Retail Development," in R. Cox (ed.), *Theory in Marketing*, Irvin, 1964.

34. Shaw, L. G., "Optimum Stochastic Control," in J. Peschol (ed.), *Disciplines and Techniques of Systems Control*, Blaisdell, 1965.

35. Simon, H. A., *The New Science of Management Decisions*, Harper, 1960.

36. Snyder, R. M., *Measuring Business Cycles*, Wiley, 1955.

37. Tou, J., *Modern Control Theory*, McGraw-Hill, 1964.

38. Von Bertalanffy, L., "The Theory of Open Systems in Physics and Biology," *Science*, Vol. 111, 1950, pp. 23-29.

39. Whittle, P., *Prediction and Regulation*, Van Nostrand, 1963.

40. Wiener, N., *Cybernetics*, Wiley, 1961.

41. Wiener, N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, M.I.T. Press, 1949.

42. Winter, P. R., "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, 1960, pp. 324-342.

43. Zannetos, Z. S., "On the Theory of Divisional Structures: Some Aspects of Centralization and Decentralization of Control and Decision Making," *Management Science*, Vol. 12, No. 14, 1965, pp. B49-B68.

Other References

44. Churchman, C. W., R. L. Ackoff, and E. L. Arnoff, *Introduction to Operations Research*, Wiley, 1957.

45. Durbin, J., and G. S. Watson, "Testing for Serial Correlation in Least-squares Regression," Pts. I and II, *Biometrika*, 1950 and 1951.

46. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *ASME Transactions*, American Society of Mechanical Engineers, vol. 82D, March, 1960.

47. Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1965.

48. Theil, H., and A. L. Nager, "Testing the Independence of Regression Disturbances," *Journal of American Statistical Association*, Vol. 56, 1961, pp. 793-806.

VITA

Sang Hoon Chang was born in Chong-ju, Korea, on May 7, 1929. He is the son of Ung Doo and Do Ra Chang. In April, 1962, he was married to Jung Wha Park of Sang-ju, Korea. He attended public schools in Korea until the breakout of the Korean War in 1950. During the War, he was a civilian employee of the U. S. Army in Korea. In 1954, he came to the United States to begin a study program with a full scholarship awarded by the Atlanta Rotary Educational Foundation.

He was graduated from the Georgia Institute of Technology with the degrees: Bachelor of Electrical Engineering (1958) and Master of Science in Industrial Engineering (1963). The topic of the Master's thesis was: *Determination of Optimum Reject Allowances in Manufacturing*, and the thesis was written under the direction of Dr. Harrison M. Wadsworth. The result of the research was published in the *Journal of Industrial engineering* (1964, pp. 127-132). In September, 1963, he was enrolled in the Ph.D. program of the School of Industrial Engineering at the Georgia Institute of Technology.

In March, 1967, he accepted a position as a Senior Research Engineer in Systems Research with the Goodyear Tire & Rubber Company.