CHAPTER 32
PRODUCTION PLANNING

Dennis B. Webster
Thomas G. Ray
Department of Industrial & Manufacturing Systems Engineering
Louisiana State University
Baton Rouge, Louisiana

32.1 INTRODUCTION

Changes that were unforeseen prior to the 1970s are now sweeping the field of manufacturing. Competition from outside the United States is driving the forces of change. There is a relentless push for improvement in total quality, which includes the quality of service to the customer. Service to the customer is related to two concepts: the delivered product quality and the timeliness of customer service.

The topics discussed in this section relate primarily to the second concept of customer service, timeliness. These topics relate either directly or indirectly to accomplishing the job in a timely manner. Forecasting, for example, provides the manufacturer with a basis for anticipating consumer demand so as to have adequate product on hand when it is demanded. Of course, the preferred approach would be to wait for a customer order and then produce and ship immediately when the order arrives. This approach is, for practical purposes, impossible for products with any significant lead time. What the manufacturer must do is to perform as well or better than his competition for the business area.

Job sequencing is an approach to reduce the completion times of the jobs to be performed. Materials requirements planning (MRP) is a technique for assuring that adequate inventory is available to complete the work required on products needed to meet a customer schedule. Inventory models are used in an effort to provide components for a manufacturing process in a timely manner at minimum cost when the demand for the item is constant.

Mechanical Engineers' Handbook, 2nd ed., Edited by Myer Kutz.
ISBN 0-471-13007-9

In a similar manner, each of the topics in this section relates to the subject of meeting customer demand, on time and at the lowest cost possible.

32.2 FORECASTING

32.2.1 General Concepts

The function of production planning and control is based upon establishing a plan, revising it as required, and adhering to it to accomplish desired objectives. Plans are based upon a forecast of future demand for the related products or services. Good forecasts are a requirement for a plan to be valid and functionally useful. Managers, when faced with a forecast, plan what actions must be taken to meet the requirements of the forecast. These actions prepare the organization to cope with the anticipated future state of nature that is predicated upon the forecast.

Forecasting methods are traditionally grouped into one of three categories: *qualitative techniques, time-series analysis,* or *causal methods.* Qualitative techniques are normally based on opinions or surveys. Time-series analysis is based on historical data and the study of its trends, cycles, and seasons. Causal methods try to find relationships between independent and dependent variables, determining which variables are predictive of the dependent variable of concern. The method selected for forecasting must relate to the type of information available for analysis.

Definitions

DESEASONALIZATION. The removal of seasonal effects from the data for the purpose of further study of the residual data.

ERROR ANALYSIS. The evaluation of errors in the historic forecasts, done as a part of forecasting model evaluation.

EXPONENTIAL SMOOTHING. An iterative procedure for the fitting of polynomials to data for use in forecasting.

FORECAST. Estimation of a future outcome.

HORIZON. A future time period or periods for which a forecast is required.

INDEX NUMBER. A statistical measure used to compare an outcome which is measured by a cardinal number with the same outcome in another period of time, geographic area, profession, etc.

MOVING AVERAGE. A forecasting method in which the forecast is an average of the data for the most recent *n* periods.

QUALITATIVE FORECAST. A forecast made without using a quantitative model.

QUANTITATIVE FORECAST. A forecast prepared by the use of a mathematical model.

REGRESSION ANALYSIS. A method of fitting a mathematical model to data by minimizing the sums of the squares of the data from a theoretical line.

SEASONAL DATA. Data that cycle over a known seasonal period, such as a year.

SMOOTHING. A process for eliminating unwanted fluctuations in data; normally accomplished by calculating a moving average or a weighted moving average.

TIME-SERIES ANALYSIS. A procedure for determining a mathematical model *for* data correlated with time.

TIME-SERIES FORECAST. Forecast prepared with a mathematical model *from* data correlated with time.

TREND. Underlying patterns of movement of historic data that become the basis for prediction of future forecasts.

32.2.2 Qualitative Forecasting

These forecasts are normally used for purposes other than production planning. Their validity is more in the area of policy-making or in dealing with generalities to be made from qualitative data. Among these techniques are the Delphi method, market research, consensus methods, and other techniques based upon opinion or historic relationships other than quantitative data.

The Delphi method is one of a number of nominal group techniques. It involves prediction with feedback to the group that gives the predictor's reasoning. Upon each prediction, the group is again polled to see if a consensus has been reached. If no common ground for agreement has occurred, the process continues moving from member to member until agreement is reached.

Surveys may be conducted of relevant groups and their results analyzed to develop the basis for a forecast. One group appropriate for analysis is customers. If a company has relatively few customers, this select number can be an effective basis for forecasting. Customers are surveyed and their responses combined to form a forecast.

Many other techniques are available for nonquantitative forecasting. An appropriate area to search if these methods seem relevant to a subjective problem at hand is the area of *nominal group techniques.*

32.2.3 Quantitative Forecasting

Quantitative forecasting involves working with numerical data to prepare a forecast. This area is further divided into two subgroups of techniques, according to the data type involved. If historic data are available and it is believed that the dependent variable to be forecast relates only to time, time-series analysis is used. If the data available suggest relationships of the dependent variable to be forecast to one or more independent variables, then the techniques used fall into the category of causal analysis. The most commonly used method in this group is regression analysis.

Methods of Analysis of Time Series

The following material will discuss in general several methods for analysis of time series. These methods provide ways of removing the various components of the series, isolating them, and providing information for their consideration should it be desired to reconstruct the time series from its components.

The movements of a time series are classified into four types: long-term or *trend* movements, *cyclical* movements, *seasonal* movements, and *irregular* movements. Each of these components can be isolated or analyzed separately. Various methods exist for the analysis of the time series. These methods decompose the time series into its components by assuming that the components are either multiplicative or additive. Assuming that the components are multiplicative, the following relationship holds:

$$Y = T \times C \times S \times I$$

Where Y is the outcome of the time series, T is the trend value of the time series, and C, S, and I are indices respectively for cyclical, seasonal, and irregular variations.

To process data for this type of analysis, it is best first to plot the raw data in order to observe their form. If the data are yearly, they need no deseasonalization. If the data are monthly or quarterly, they can be converted into yearly data by summing the data points that would add to a year before plotting. (Seasonal index numbers can be calculated to seasonalize the data later if required.) By plotting yearly data, the period of apparent data cycles can be determined or approximated. A centered moving average of appropriate order can be used to remove the cyclical effect in the data. Further, cyclical indices can be calculated when the order of the cycle has been determined. At this point, the data contain only the trend and irregular components of variation. Regression analysis can be used to estimate the trend component of the data, leaving only the irregular, which is essentially forecasting error.

Index Numbers. Index numbers are calculated by grouping data of the same season together, calculating the average over the season for which the index is to be prepared, and then calculating the overall average of the data over each of the seasons. Once the seasonal and overall averages are obtained, the seasonal index is determined by dividing the seasonal average by the overall average.

Example Problem 32.1

A business has been operational for 24 months. The sales data in thousands of dollars for each of the monthly periods are as shown in Table 32.1.

Table 32.1

	Year 1	Year 2
Jan.	20	24
Feb.	23	27
Mar.	28	30
Apr.	32	35
May	35	36
Jun.	26	28
Jul.	25	27
Aug.	23	23
Sep.	19	17
Oct.	21	22
Nov.	18	19
Dec.	12	14

The overall average is 584 divided by 24, or 24.333. The index for January would be

$$I_{\text{Jan}} = \frac{(20 + 24)}{24.333} = .904$$

The index for March would be

$$I_{\text{Mar}} = \frac{(28 + 30)}{24.333} = 1.192$$

To use the index, a trend value for the year's sales would be calculated, the average monthly sales would be obtained, and then this figure would be multiplied by the index for the appropriate month to give the month's forecast.

It should be noted that a season can be defined as any period for which the data is available for appropriate analysis. If there are seasons within a month, i.e., four weeks in which the sales vary considerably according to a pattern, a forecast could be indexed within the monthly pattern also. This would be a second indexing within the overall forecast. Further, seasons could be chosen as quarters rather than months or weeks. This choice of the period for the analysis is dependent upon the requirements for the forecast.

Data given on a seasonal basis can be deseasonalized by dividing them by the appropriate seasonal index. Once this has been done, they are labeled *deseasonalized data.* They still contain the trend, cyclical, and irregular components after this adjustment.

Moving Average. A moving average can normally be used to remove the seasonal or cyclical components of variation. This removal is dependent upon the choice of a moving average that contains sufficient data points to bridge the season or cycle. For example, a seven-period-centered moving average should be sufficient to remove seasonal variation from monthly data. A disadvantage to the use of moving averages is the loss of data points due to the inclusion of multiple points into the calculation of a single point. For the monthly data related to the calculation of index numbers given in the previous section, only the months of April of Year 1 through September of Year 2 would be available for analysis when a seven-month-centered moving average is used. Six data points are not available for calculation due to the requirements of the method.

Example Problem 32.2

See Table 32.2. Note that in this case the five-year moving average lost four data points, two on each end of the data series. Observation of the moving average indicated a steady downward trend in the data. The raw data had fluctuations that might tend to confuse an observer, initially due to the apparent positive changes from time to time.

Weighted Moving Average. A major disadvantage of the moving average method, the effect of extreme data points, can be overcome by using a weighted moving average. In this average, the effect of the extreme data points may be decreased by weighing them less than the data points at the center

Table 32.2

Year	Data	5-Year Moving Total	5-Year Moving Average
1	60		
2	56.5		
3	53.0	275.3	55.1
4	54.6	269.2	53.8
5	51.2	261.1	52.2
6	53.9	257.2	51.4
7	48.4	250.9	50.2
8	49.1	242.1	48.4
9	48.3	232.8	46.6
10	42.4		
11	44.6		

Table 32.3

Year	Data	5-Year Moving Total	5-Year Total Less Center Value	Weighted Average (.5 Col 2/4 + .5 Col 4)
1	60			
2	56.5			
3	53.0	275.3	222.3	54.3
4	54.6	269.2	214.6	54.1
5	51.2	261.1	209.9	51.8
6	53.9	257.2	203.3	52.4
7	48.4	250.9	202.5	49.5
8	49.1	242.1	193.0	48.7
9	48.3	232.8	184.5	47.2
10	42.4			
11	44.6			

of the group. There are many ways to do this. One method would be to weight the center point of a five-period average as 50% of the total, with the remaining points weighted for the remaining 50%. For the example in the previous section, the yield would be as shown in Table 32.3.

Example Problem 32.3

See Table 32.3.

Table 32.4 displays the two forecasts. The results are very comparable, with the weighted average forecast distinguishing a slight upswing from period 5 to 6 that was ignored by the moving average method.

Exponential Smoothing. This method determines the forecast (F) for the next period as the weighted average of the last forecast and the current demand (D). The current demand is weighted by a constant α and the last forecast is weighted by the quantity $1 - \alpha$ ($0 \leq \alpha \leq 1.0$).

new forecast = α (demand for current period) + $(1 - \alpha)$ (forecast for current period)

This can be expressed symbolically as

$$F_t = \alpha D_{t-1} + (1 - \alpha) F_{t-1}$$

Normally the forecast for the first period is taken to be the actual demand for that period (i.e., forecast and demand are the same for the initial data point). The smoothing constant is chosen as a result of analysis of error by a method such as mean absolute deviation coupled with the judgment of the analyst. A high value of α makes the forecast very responsive to the occurrence in the last period. Similarly, a small value would lead to a lack of significant response to the current demand. Evaluations must be made in light of the cost effects of the errors to determine what value of α is best for a given situation. The following example problem shows the relationship between actual data and forecasts for various values of α.

Table 32.4

Period	Moving Average Forecast	Weighted Average Forecast
3	55.1	54.3
4	53.8	54.1
5	52.2	51.8
6	51.4	52.4
7	50.2	49.5
8	48.4	48.7
9	46.6	47.2

Example Problem 32.4

See Table 32.5.

Causal Methods

These methods assume that there are certain factors that have a cause–effect relationship with the outcome of the quantity to be forecast and that a knowledge of these factors will allow a more accurate prediction of the dependent quantity. The statistical models of regression analysis fall within this category of forecasting.

Basic Regression Analysis. The simplest model for regression analysis is the linear model. The basic approach involves the determination of a theoretical line that passes through a group of data points that appear to follow a linear relationship. The desire of the modeler is to determine the equation for the line that would minimize the sums of the squares of the deviations of the actual points from the corresponding theoretical points. The values for the theoretical points are obtained by substituting the values of the independent variable x_i into the functional relationship

$$\hat{Y}_i = a + bx_i$$

The difference between the data and the forecasted value of point i is

$$Y_i - \hat{Y}_i$$

Squaring this value and summing the relationship over the N related points yields

$$L = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$

Substituting the functional relationship for the forecasted value of Y gives

$$L = \sum_{i=1}^{N} (Y_i - \hat{a} + \hat{b}x_i)^2$$

By using this relationship, taking the partial derivatives of L with respect to a and b and solving the resulting equations simultaneously, the normal equations for least squares for the linear regression case are obtained. These are

$$\Sigma\, Y = aN + b\,\Sigma\, X$$
$$\Sigma\, XY = a\,\Sigma\, X + b\,\Sigma\, X^2$$

Solving these equations yields values for a and b. These values are given by

Table 32.5

		Forecasts for Various α Values		
Period	Demand	$\alpha = .1$	$\alpha = .2$	$\alpha = .3$
1	85	85	85	85
2	102	85	85	85
3	110	86.7	88.4	90.1
4	90	89.0	92.7	96.1
5	105	89.1	92.2	94.3
6	95	90.7	94.8	97.5
7	115	91.1	94.8	96.8
8	120	93.5	98.8	102.3
9	80	96.2	103.0	107.6
10	95	94.6	98.4	99.3

$$a = \frac{\Sigma X^2 \Sigma Y - \Sigma X \Sigma XY}{n \Sigma X^2 - (\Sigma X)^2}$$

and

$$b = \frac{n \Sigma XY - \Sigma X \Sigma Y}{N \Sigma X^2 - (\Sigma X)^2}$$

The regression equation is then $Y_i = a + bx_i$ and the correlation coefficient r, which gives the relative importance of the relationship between x and y as

$$r = \frac{n \Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N \Sigma X^2 - (\Sigma X)^2][N \Sigma Y^2 - (\Sigma Y)^2]}}$$

This value of r can range from $+1$ to -1. The plus sign would indicate a positive correlation (i.e., large values of x are associated with large values of y; a negative correlation implies that large values of x are associated with small values of y) and the negative sign negative correlation.

Quadratic Regression. This regression model is used when the data appear to follow a simple curvilinear trend and the fit of a linear model is not adequate. The procedure for deriving the normal equations for quadratic regression is very similar to that for linear regression. The quadratic model has three parameters that must be estimated, however. These are the constant term a, the coefficient of the linear term b, and the coefficient of the square term c. The model is

$$Y_i = a + bx_i + cx_i^2$$

Its normal equations are

$$\begin{aligned} \Sigma Y &= N a + b \Sigma X + c \Sigma X^2 \\ \Sigma XY &= a \Sigma X + b \Sigma X^2 + c \Sigma X^3 \\ \Sigma X^2Y &= a \Sigma X^2 + b \Sigma X^3 + c \Sigma X^4 \end{aligned}$$

The normal equations for least squares for a cubic curve, quartic curve, and so on can be generalized from the expressions for the linear and quadratic models.

32.2.4 Forecasting Error Analysis

One common method of evaluation of forecast accuracy is termed *mean absolute deviation* (MAD) from the procedure used in its calculation. For each available data point, a comparison of the forecasted value is made to the actual value. The absolute value of the differences is calculated. This absolute difference is then summed over all values and its average calculated to give the evaluation.

$$\begin{aligned} \text{MAD} &= \frac{\text{Sum of the absolute deviations}}{\text{number of deviations}} \\ &= \Sigma \frac{|(Y_i - \hat{Y}_i)|}{N} \end{aligned}$$

Alternative forecasts can be analyzed to determine the value of MAD and a comparison can be made using this quantity as an evaluation criterion. Other criteria can also be calculated. Among these are the mean square of error (MSE) and the standard error of the forecast (s_{yx}). These evaluation criteria are calculated as shown below:

$$MSE = \sum_{i=1}^{N} \frac{(Y_i - \hat{Y}_i)^2}{N}$$

and

$$s_{yx} = \frac{\sqrt{\Sigma Y^2 - a \Sigma Y - b \Sigma XY}}{N - 2}$$

In general, these techniques are used to evaluate the forecast and then the results of the various evaluations, together with the data and forecasts, are studied. Conclusions may then be drawn as to which method is preferred or the results of the various methods compared to determine what they in effect distinguish.

32.2.5 Conclusions on Forecasting

A number of factors should be considered in choosing a method of forecasting. One of the most important factors is *cost*. The problem of valuing an accurate forecast is presented. If the question "How will the forecast help and how will it save money?" can be answered, a decision can be made regarding the allocation of a percent of the savings to the cost of the forecasting process. Further, concern must be directed to the required *accuracy* of a forecast in order to achieve desired cost reductions. Analysis of past data and the testing of the proposed model using this historic data provide a possible scenario for hypothetical testing of the effects of cost of variations of actual occurrences from the plan value (forecast).

In many cases, an inadequate data base will prohibit significant analysis. In others the data base may not be sufficient for the desired projection into the future.

The answers to each of these questions are affected by the type of product or service for which the forecast is to be made, as well as the value of the forecast to the planning process.

32.3 INVENTORY MODELS

32.3.1 General Discussion

Normally items waiting to be purchased or sold are considered to be in inventory. One of the most pressing problems in the manufacturing and sale of goods is the control of this inventory. Many companies experience financial difficulties each year due to a lack of an adequate control in this area. Whether the items in question be raw material used to manufacture a product or products waiting to be sold, problems arise when too many or too few items are available. The greatest number of problems arise when too many items are held in inventory.

The primary factor in the reduction of inventory costs is deciding when to order, how much to order, and if back ordering is permissible. Inventory control involves the making of decisions by management as to the source from which the inventory is to be procured and the quantity to be procured at the time. This source could be from another division of the company handled as an intrafirm transfer, outside purchase from any of a number of possible vendors, or manufacture of the product in-house.

The basic decisions to be made once a source has been determined are how much to order and when to order. Inherent in this previous analysis is the concept of demand. Demand can be known or unknown, probabilistic or deterministic, constant or lumpy. Each of these characteristics affects the method of approaching the inventory problem.

For the *unknown demand* case, a decision must be made as to how much the firm is willing to risk. Normally, the decision would be to produce some "k" units for sale and then determine, after some period of time, to produce more or to discontinue production due to insufficient demand. This amounts to the reduction of the unknown demand situation to one of a lumpy demand case after the decision has been made to produce a batch of finite size. Similarly, if a decision is made to begin production at a rate of n per day until further notice, the unknown demand situation has been changed to a constant known demand case.

Lumpy demand, or demand that occurs periodically with quantities varying, is frequently encountered in manufacturing and distribution operations. It is distinguished from the known demand case. This second case is that of a product which has historic data from which forecasts of demand can be prepared. Factor of concern in these situations are the lead time and the unit requirement on a periodic basis. The following are the major factors to be considered in the modeling of the inventory situation.

Demand

Demand is the primary stimulus on the procurement and inventory system; it is, in fact, the justification for its existence. Specifically, the system may exist to meet the demand of customers, the spare parts demand of an operational weapons system, the demand of the next step in a manufacturing process, and so on. The characteristic of demand, although independent of the source chosen to replenish inventories, will depend upon the nature of the environment giving rise to the demand.

The simplest demand pattern may be classified as deterministic. In this special case, the future demand for an item may be predicted with certainty. Demand considered in this restricted sense is only an approximation of reality. In the general case, demand may be described as a random variable that takes on values in accordance with a specific probability distribution.

Procurement Quantity

Procurement quantity is the order quantity, which in effect determines the frequency of ordering and is related directly to the maximum inventory level.

Maximum Shortage

The maximum shortage quantity is also related to the inventory level.

Item Cost

Item cost is the basic purchase cost of a unit delivered to the location of use. In some cases, delivery cost will not be included if that cost is insignificant in relation to the unit cost. In these cases, the delivery cost will be added to overhead and not treated as a part of direct material costs.

Holding Cost

Inventory holding costs are incurred as a function of the quantity on hand and the time duration involved. Included in these costs are the real out-of-pocket costs, such as insurance, taxes, obsolescence, and warehouse rental and other space charges, and operating costs, such as light, heat, maintenance, and security. In addition, capital investment in inventories is unavailable for investment elsewhere. The rate of return forgone represents a cost of carrying inventory.

The inventory holding cost per unit of time may be thought of as the sum of several cost components. Some of these may depend upon the maximum inventory level incurred. Others may depend upon the average inventory level. Still others, like the cost of capital invested, will depend on the value of the inventory during the time period. The determination of holding cost per unit for a specified time period depends on a detailed analysis of each cost component.

Ordering Cost

Ordering cost is the cost incurred when an order is placed. It is composed of the cost of time and materials, and any expense of communication in placing an order. In the case of a manufacturing model it is replaced by *setup cost,* which is the cost incurred when a machine's tooling or jigs and fixtures must be changed to accommodate the production of a different part or product.

Shortage Cost

Shortage cost is the penalty incurred for being unable to meet a demand when it occurs. This cost does not depend upon the source chosen to replenish the stock, but is a function of the number of units short and the time duration involved.

The specific dollar penalty incurred when a shortage exists depends on the nature of the demand. For instance, if the demand is that of customers of a retail establishment, the shortage cost will include the loss of goodwill. In this case, the shortage cost will be small relative to the cost of the item. If, however, the demand is that of the next step of a manufacturing process, the cost of the shortage may be high relative to the cost of the item. Being unable to meet the requirements for a raw material or a component part may result in lost production or even closing of the plant. Therefore, in establishing shortage cost, the seriousness of the shortage condition and its time duration must be considered.

32.3.2 Types of Inventory Models

Deterministic

Deterministic models assume that quantities used in the determination of relationships for the model are all known. These quantities include demand per unit of time, lead time for product arrival, and costs associated with such occurrences as a product shortage, the cost of holding the product in inventory, and the cost associated with placing an order for a product.

Constant Demand

Constant demand is one case that can be analyzed within the category of deterministic models. It represents very effectively the case for some components or parts in an inventory that are used in multiple parents, these multiple parent components having a composite demand that is fairly constant over time.

Lumpy Demand

Lumpy demand is varying demand that occurs at irregular points in time. This type of demand is normally a dependent demand that is driven by an irregular production schedule affected by actual customer requirements. Although the same assumptions are made regarding the knowledge of related quantities as in the constant demand case, this type of situation is analyzed separately under the topic of materials requirements planning (MRP). This separation of methodology is due to the different inputs to the modeling process in that the knowledge about demand is approached by different methods in the two cases.

Probabilistic

Probabilistic models consider the same quantities as do the deterministic models, but treat the quantities that are not cost-related as random variables. Hence, demand and lead time have their associated probability distributions. The added complexity of the probabilistic values requires that these models be analyzed by radically different methods.

Definitions

The following terms are defined in order to clarify their usage in sections of the material related to inventory that follow. Where appropriate, a literal symbol is assigned to represent the term.

INVENTORY (I). Stock held for the purpose of meeting a demand either internal or external to the organization.
LEAD TIME (T). The time required to replenish an item of inventory by either purchasing from a vendor or manufacturing the item in-house.
DEMAND (D). The number of units of an inventory item required per unit of time.
RE-ORDER POINT. The point at which an order must be placed in order for the procured quantity to arrive at the proper time or, for the manufacturing case, the finished product to begin flowing into inventory at the proper time.
RE-ORDER QUANTITY (Q). The quantity for which an order is placed when the re-order point is reached.
DEMAND DURING LEAD TIME. This quantity is the product of lead time and demand. It represents the number of units that will be required to fulfill demand during the time that it takes to receive an order that has been placed with a vendor.
REPLENISHMENT RATE (R). This quantity is the rate at which replenishment occurs when an order has been placed. For a purchase situation it is infinite (when an order arrives, in an instant the stock level rises from 0 to Q). For the manufacturing situation it is finite.
SHORTAGE. The units of unsatisfied demand that occur when there is an out-of-stock situation.
BACK ORDER. One method of treating demand in a shortage situation when it is acceptable to the customer. (A notice is sent to the customer saying that the item is out of stock and will be shipped as soon as it becomes available.)
LUMPY DEMAND. Demand that occurs in an aperiodic manner for quantities whose volume may or may not be known in advance. Constant demand models should normally never be used in a lumpy demand situation. The exception would be a component that is used for products that experience lumpy demand, but that itself experiences constant demand. The area of MRP (materials requirements planning) was developed to deal with the lumpy demand situations.

32.3.3 The Modeling Approach

Modeling in operations research involves the representation of reality by the construction of a model in one of several alternative ways. These models may be iconic, symbolic, or mathematical. For inventory models, the mathematical model is normally the selection of choice. The model is developed to represent a concept whose relationships are to be studied. As much detail can be included in a particular model as is required to represent the situation effectively. The detail omitted must be of little significance as to its effect on the model. The model's fidelity is the extent to which it accurately represents the situation for which it is constructed.

Inventory modeling involves building mathematical models to represent the interactions of the variables of the inventory situation to give results adequate for the application at hand. In this section, treatment will be limited to deterministic models for inventory control. Probabilistic or stochastic models may be required for some analysis. References 1–3 may be consulted if more sophisticated models are required.

General

Using the terminology defined above, a basic logic model of the general case inventory situation will be developed. The objective of inventory management will normally be to determine an operating policy that will provide a means to reduce inventory costs. To reduce costs, a determination must first be made as to what costs are present. The general model is as follows:

$$\text{total cost} = \text{cost of items} + \text{cost of ordering} + \text{cost of holding items in stock} + \text{cost of shortage}$$

This cost is stated without a base period specified. Normally it will be stated as a per-period cost, with the period being the same period as the demand rate (D) period.

Models of Inventory Situations

Purchase Model with Shortage Prohibited. This model is also known as an infinite replenishment rate model with infinite shortage costs because of the slope of the replenishment rate line (it is vertical) when the order arrives. The quantity on hand instantaneously changes from zero to "Q". The shortage condition is preempted by the assignment of an infinite value to shortage cost. See Fig. 32.1. For this case, the item cost per period is symbolically

$$C_iD$$

The ordering cost is

$$(C_pD)/Q$$

The shortage cost is zero since shortage is prohibited and the inventory holding cost is

$$(C_hQ)/2$$

The equation for total cost is then

$$TC(Q) = C_iD + (C_pD)/Q + (C_hQ)/2$$

Analysis of this model reveals that the first component of cost, the cost of items, does not vary with Q. (Here we are assuming a constant unit cost; purchase discounts models will be covered later.) The second component of cost, the cost of ordering, will vary on a per-period basis with the size of the order (Q). For larger values of Q, the cost will be smaller since fewer orders will be required to receive the fixed demand for the period. The third component of cost, cost of holding items in stock, will increase with increasing order size Q and conversely decrease with smaller order sizes. The fourth component of cost, cost of shortage, is affected by the re-order point. It is not affected by the order size and for this case shortage is not permitted.

It should be noted that this equation is essentially obtained by determining the cost of each of the component costs on a per cycle basis and then dividing that expression by the number of periods per cycle (Q/D). To obtain the extreme point(s) of the function, it is necessary to take the derivative of $TC(Q)$ with respect to Q, equate this quantity to zero, and solve for the corresponding value(s) of Q. This yields

$$0 = 0 + (C_pD)/Q^2 + C_h/2$$

or

$$\hat{Q} = (2C_pD)/C_h$$
$$L = DT$$

Inspection of the sign of the second derivative of this function reveals that the extreme point is a

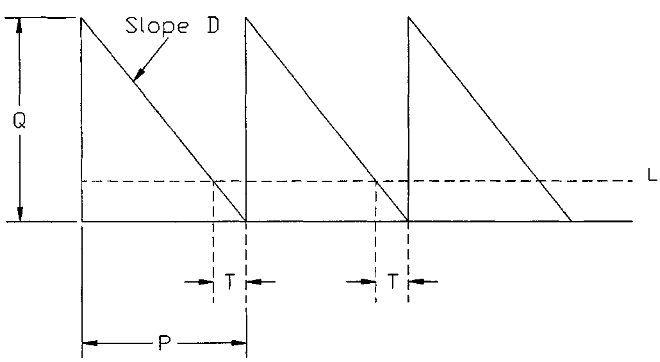

Fig. 32.1 Purchase model with shortage prohibited.

minimum. This fits the objective of the model formulation. The quantity to be ordered at any point in time is then $\hat{Q}$ and the time to place the order will be when the inventory level drops to L (the units consumed during the lead time for receiving the order).

Purchase Model with Shortage Permitted. This model is also known as an infinite replenishment rate model with finite shortage costs. See Fig. 32.2. For this model, the product cost and the ordering cost are the same as for the previous model

$$C_iD + (C_pD)/Q$$

The holding cost is different, however. It is given by

$$C_h[Q - (DT - L)]^2/2Q$$

This represents the unit periods of holding per cycle times the holding cost per unit period. The unit periods of holding is obtained from the area of the triangle whose altitude is $Q - (DT - L)$ and whose base is the same quantity divided by the slope of the hypotenuse. The unit periods of shortage is calculated in the same manner. For that case, the altitude is $(DT - L)$ and the base is $(DT - L)$ divided by D. The shortage cost component is then

$$C_s(DT - L)^2/2Q$$

The total cost per period is given by

$$TC(Q,DT - L) = C_iD + \frac{(C_pD)}{Q} + \frac{C_h[Q - (DT - L)]^2}{2Q} + \frac{C_s[Q - (DT - L)]^2}{2Q}$$

Note that the quantity $(DT - L)$ is used as a variable. This is done for the purposes of simplifying the equations that result when the partial derivatives are taken for the function. Taking these derivatives and solving the resulting equations simultaneously for the values of Q and $(DT - L)$ yields the following relationships:

$$\hat{Q} = \sqrt{\frac{(2C_pD)}{C_h} + \frac{C_pD}{C_s}}$$

$$\hat{L} = DT - \sqrt{\frac{2C_nC_PD}{C_s(C_h + C_s)}}$$

Manufacturing Model with Shortage Prohibited. This model is also known as a finite replenishment rate model with infinite shortage costs. Figure 32.3 illustrates the situation.

$$\hat{Q} = \sqrt{\frac{2C_pD}{C_h(1 - D/R)}}$$

$$L = DT$$

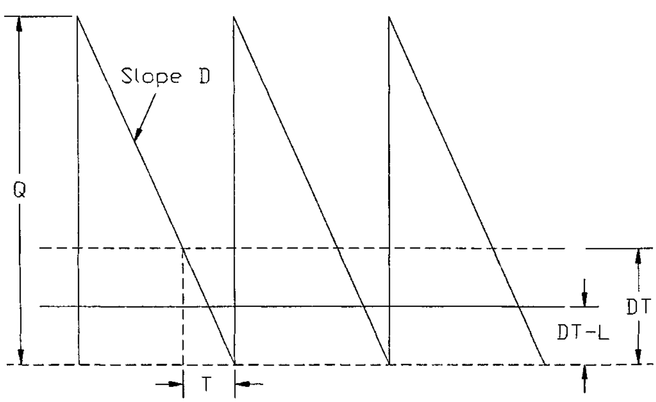

Fig. 32.2 Purchase model with shortage permitted.

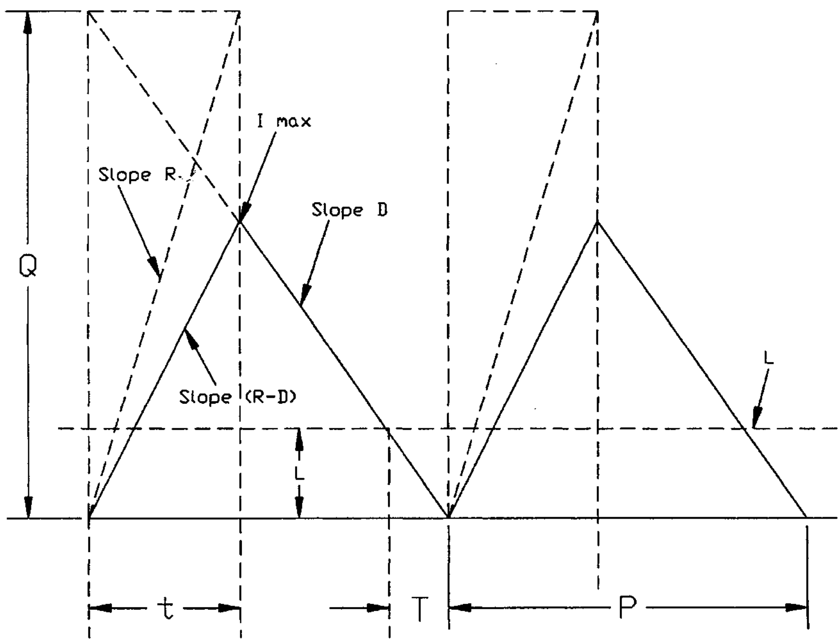

Fig. 32.3 Manufacturing model with shortage prohibited.

Manufacturing Model with Shortage Permitted. This model is also known as a finite replenishment rate model with finite shortage costs. It is the most complex of the models treated here, as it is the general case model. All of the other models can be obtained from it by properly defining the replenishment rate and shortage cost. For example, the purchase model with shortage prohibited is obtained by defining the manufacturing rate and the shortage cost as infinite. When this is done, the equations reduce to those appropriate for the stated situation. See Fig. 32.4.

For this model, the expressions for Q and L are as shown below:

$$\hat{Q} = \sqrt{\frac{1}{1 - D/R}} \quad \sqrt{\frac{2C_pD}{C_h} + \frac{2C_pD}{C_s}}$$

$$\hat{L} = DT\sqrt{1 - D/R} \quad \sqrt{\frac{2C_pD}{C_s\left(1 + \frac{C_s}{C_h}\right)}}$$

Models for Purchase Discounts

Purchase Discount Model with Fixed Holding Cost. In this situation, the hold cost (C_w) is assumed to be fixed, not a function of units costs.

A supplier offers a discount for ordering a larger quantity. The normal situation is as shown in Table 32.6.

The decision-maker must apply the appropriate EOQ purchase model, either finite or infinite shortage cost. Upon choice of the appropriate model, the following procedure will apply.

1. Evaluate $\hat{Q}$ and calculate $TC(\hat{Q})$.
2. Evaluate $TC(q_{k+1})_1$ where q_{k+1} is the smallest quantity in the price break interval above that interval where $\hat{q}$ lies.
3. If $TC(\hat{q}) < TC(qk + 1)$, the ordering quantity will be $\hat{q}$. If not, go to step 4.
4. Since the total cost of the minimum quantity in the next interval above that interval containing $\hat{q}$ is a basic amount, an evaluation must be made successively of total costs of the minimum quantities in the succeeding procurement intervals until one reflects an increase in cost or the last choice is found to be the minimum. For example, in a situation where shortage is not

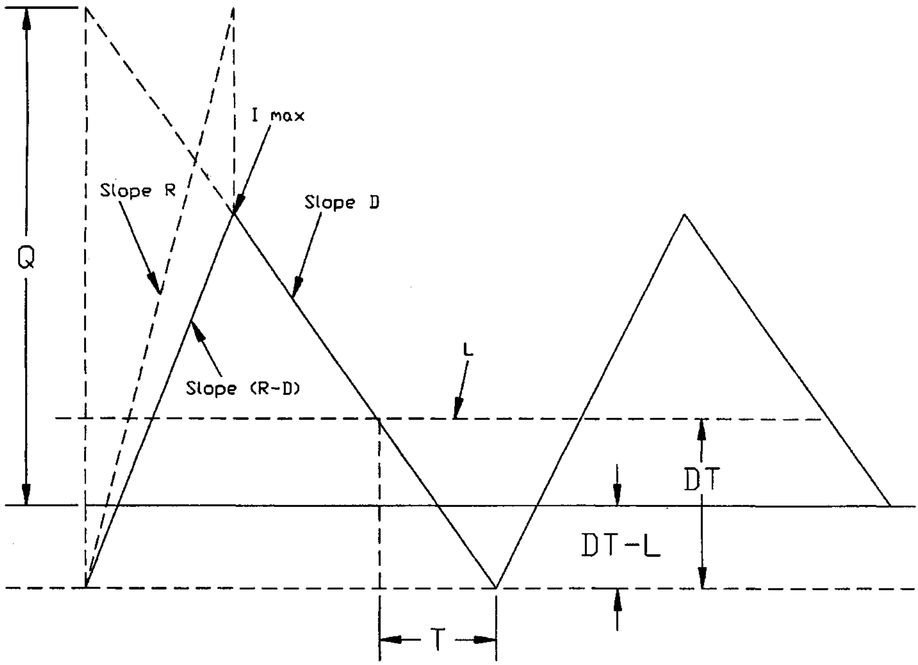

Fig. 32.4 Manufacturing model with shortage permitted.

permitted, the ordering cost is \$50, the holding cost is \$1 per unit year, and the demand is 10,000 units per year.

$$\hat{Q} = \sqrt{\frac{2C_pD}{C_w}}$$

$$= \sqrt{\frac{2(\$50)10{,}000}{1}}$$

$$= 1000$$

$$TC(Q) = C_iD + \frac{C_nQ}{2} + \frac{D}{Q}C_p$$

$$TC(\hat{Q}) = \$20(10{,}000) + \$1\left(\frac{1000}{2}\right) + \$50\left(\frac{\$810{,}000}{1{,}000}\right)$$

$$= \$201{,}000$$

Table 32.6

Range of Quantity Purchased	Price
$1–q_1$	P_1
$q_1 + 1–q_2$	P_2
$q_2 + 1–q_3$	P_3
. .	.
. .	.
. .	.
$q_{m-1} + 1–q_m$	P_m

Table 32.7

Q	P
0–500	22.00
501–1199	20.00
1200–1799	18.00
1800 +	16.50

The question is then whether the smallest quantity in the next discount interval (1200–1799) would give a lower total cost. See Table 32.7.

$$TC(1200) = \$18(10{,}000) + \$1\left(\frac{1200}{2}\right) + \$50\left(\frac{10{,}000}{1200}\right)$$
$$= 181{,}033$$

Since this a lower cost, an evaluation must be made of the smallest quantity in the next interval, 1800.

$$TC(1800) = 16.50(10{,}000) + \$1\left(\frac{1800}{2}\right) + \$50\left(\frac{10{,}000}{1800}\right)$$
$$TC(1800) = 166{,}175$$

Since there are no further intervals for analysis, this is the lowest total cost and its associated q, 1800, should be chosen as the optimal $\hat{Q}$.

The total cost function for this model is shown in Fig. 32.5.

Quantity Discount Model with Variable Holding Cost. In this case, the holding cost is variable with unit cost, i.e., $C_w = KC_i$. Again, the appropriate model must be chosen for shortage conditions. For the infinite shortage case,

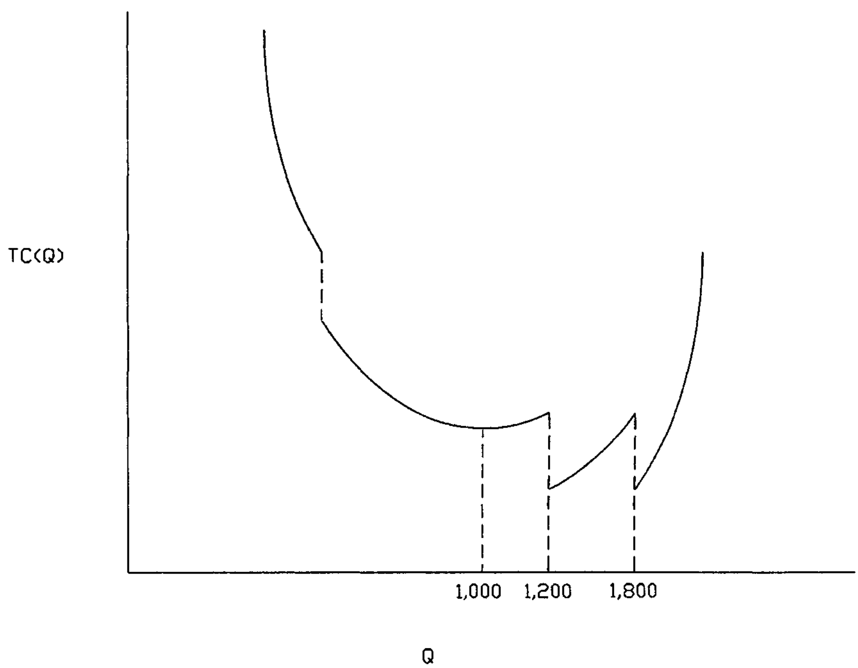

Fig. 32.5 Total cost function for a quantity discount model with fixed percentage holding cost.

$$\hat{Q} = \sqrt{\frac{2C_pD}{KC_i}}$$

To obtain the optimal value of Q in this situation, the following procedure must be followed:

1. Evaluate $\hat{Q}$ using the expression above and the item cost for the first interval.
   a. If the value of $\hat{Q}$ falls within the interval for C_i, use this $TC(\hat{Q})$ for the smallest cost in the interval.
   b. If $\hat{Q}$ is greater than the maximum quantity in the interval, use Q_{max} where Q_{max} is the greatest quantity in the interval and evaluate $TC(Q_{max})$ as the lowest cost point in the interval.
   c. If $\hat{Q}$ is less than the smallest Q in the interval, use Q_{min} where Q_{min} is the smallest quantity in the interval as the best quantity and evaluate $TC(Q_{min})$.
2. For each cost interval, follow the steps of part 1 to evaluate best values in the interval.
3. Choose the minimum total cost from the applications of steps 1 and 2.

Example Problem 32.5

Using the same data as in the previous example, assume $C_h = .05C_I$

$$\hat{Q} = \sqrt{\frac{2(50)10{,}000}{(.05)(22)}}$$
$$\cong 990$$

Since 990 > 500, the smallest cost in the interval will be at

$$TC(500) = 22.00(10{,}000) + \$1\left(\frac{500}{2}\right) + \$50\left(\frac{10{,}000}{500}\right)$$
$$= 220{,}000 + 250 + 1000$$
$$= \$221{,}250$$

Using the second interval,

$$\hat{Q} = \sqrt{\frac{2(50)10{,}000}{.05(20)}} = 1000$$

Since this value falls within the interval, $TC(1{,}000)$ is calculated

$$TC(1000) = 20(10{,}000) + \$1\left(\frac{1000}{2}\right) + 50\left(\frac{10{,}000}{1000}\right)$$
$$= 201{,}000$$

for the next interval (1201–1799)$C_{\pm}$ = 18.00 and

$$\hat{Q} = \sqrt{\frac{2(50)10000}{.05(18.00)}}$$
$$\hat{Q} \cong 1050 \text{ (which falls outside the interval to the left)}$$

Hence, the smallest quantity in the interval will be used.

$$TC(Q) = 18(10{,}000) + \$1\left(\frac{1200}{2}\right) + \$50\left(\frac{10{,}000}{1200}\right)$$
$$\cong \$181{,}016$$

This is lower than either interval previously evaluated. For the next interval

$$\hat{Q} = \sqrt{\frac{2(50)10{,}000}{.05(16.50)}}$$

$$\hat{Q} \cong 1101 \text{ (evaluate on calculator)}$$

Hence, the smallest quantity in the interval must be used.

$$\begin{aligned} TQ(1800) &= 16.50(10{,}000) + \$1\left(\frac{1800}{2}\right) + \$50\left(\frac{10{,}000}{1800}\right) \\ &= 165{,}000 + 900 + 272 \\ &\cong 166{,}172 \end{aligned}$$

This is chosen because it is the last interval for evaluation and yields the lowest total cost. Where the previous total cost function for fixed holding cost was a segmented curve with offsets, this is a combination of different curves, each valid over a specific range. In the case of the fixed holding cost, it had only one minimum point, yet the offsets in the total cost due to changes in applicable unit cost in an interval could change the overall minimum. In this situation, there are different values of $\hat{Q}$ for each unit price. The question becomes whether the value of $\hat{Q}$ falls within the price domain. If it does, the total cost function is evaluated at that point; if not, a determination must be made as to whether the value of $\hat{Q}$ lies to the left or right of the range. If it is to the left, the smallest value in the range is used to determine the minimum cost in the range. If it is to the right, the maximum value in the range is used.

Conclusions Regarding Inventory Models

It should be noted that the material discussed here has covered only a very small percentage of the class of deterministic inventory models, albeit these models represent a large percentage of applications. Should the models discussed here not adequately represent the situation under study, further research should be directed at finding a model with improved fidelity for the situation. Other models are covered in Refs. 1, 4, and 5.

32.4 AGGREGATE PLANNING—MASTER SCHEDULING

Aggregate planning is the process of determining overall production, inventory, and workforce levels required to meet forecasted demand over a finite time horizon while trying to minimize the associated production, inventory, and workforce costs. Inputs to the aggregate planning process are forecasted demand for the products (either aggregated or individual); outputs from aggregate planning after disaggregation into the individual products are the scheduled end products for the master production schedule. The time horizon for aggregate planning normally ranges from 6–18 months, with a 12-month average.

The difficulties associated with aggregate planning are numerous. Product demand forecasts vary widely in their accuracy; the process of developing a suitable aggregate measure to use for measuring the value or quantity of production in a multiple product environment is not always possible; actual production does not always meet scheduled production; and unexpected events occur, including material shortages, equipment breakdowns, and employee illness.

Nevertheless, some form of aggregate planning is often required because seldom is there a match between the timing and quantity for product demand versus product manufacture. How the organization should staff and produce to meet this imbalance between production and fluctuating demand is what aggregate planning is about.

32.4.1 Alternative Strategies to Meet Demand Fluctuations

Manufacturing managers use numerous approaches to meet changes in demand patterns for both short and intermediate time horizons. Among the more common are the following.

1. Produce at a constant production rate with a constant workforce, allowing inventories to build during periods of low demand and supplying demand from inventories during periods of high demand. This approach is used by firms with tight labor markets, but customer service may be adversely affected and levels of inventories may widely fluctuate between being excessively high to being out of stock.
2. Maintain a constant workforce, but vary production within defined limits by using overtime, scheduled idle time, and, potentially, subcontracting of production requirements. This strategy allows for rapid reaction to small or modest changes in production when faced with similar demand changes. It is the approach generally favored by many firms, if overall costs can be kept within reasonable limits.

3. Produce to demand, letting the workforce fluctuate by hiring and firing, while trying to minimize inventory carrying costs. This approach is used by firms that typically use low skilled labor where the availability of labor is not an issue. Employee morale and loyalty, however, will always be degraded if this strategy is followed.

32.4.2 Aggregate Planning Costs

Aggregate planning costs can be grouped into one or more of the following categories:

1. *Production costs* include all of those items that are directly related to or necessary in the production of the product, such as labor and material costs. Supplies, equipment, tooling, utilities, and other indirect costs are also included, generally through the addition of an overhead term. Production costs are usually divided into fixed and variable costs, depending upon whether the cost is directly related to production volume.
2. *Inventory costs* include the same ordering, carrying, and shortage costs discussed in Section 32.3.
3. *Costs associated with workforce and production rate changes* are in addition to the regular production costs and include the additional costs incurred when new employees are hired and exiting employees are fired or paid overtime premiums. They may also include costs when employees are temporarily laid off or given alternative work that underutilizes the employees' skills, or when production is subcontracted to an outside vendor.

32.4.3 Approaches to Aggregate Planning

Researchers and practitioners alike have been intrigued by aggregate planning problems, and numerous approaches have been developed over the decades. Although difficult to categorize, most approaches can be grouped as in Table 32.8.

Optimal formulations take many forms. Linear programming models are popular formulations and range from the very basic,[7,8] which assume deterministic demand, a fixed workforce, and no shortages, to complicated models, which use piecewise linear approximations to quadratic cost functions, variable demand, and shortages.[9,10] The linear decision rule (LDR) technique, developed in an extensive project,[11–13] is one of the few instances where the approach was implemented. Nevertheless, due to the very extensive data-collection, updating, and processing requirements to develop and maintain the rules, no other implementation has been reported. Lot-size models usually are either of the capacitated (fixed capacity)[14] or uncapacitated (variable capacity)[15] variety. Although a number of lot-size models have been developed and refined, including some limited implementation,[16,17] computational complexity constrains consideration to relatively small problems. Goal-programming models are attempts at developing more realistic formulations by including multiple goals and objectives. Essentially these models possess the same advantages and disadvantages of LP models, with the additional benefit of allowing tradeoffs among multiple objectives.[18,19] Other optional approaches have modeled the aggregate planning problem using queueing,[20] dynamic programming,[21–23] and Lagrangian techniques.[24,25]

Nonoptional approaches have included the use of search techniques (ST), simulation models (SM), production switching heuristics (PSH), and management coefficient models (MCM). STs involve first the development of a simulation model that describes the system under study to develop the system's response under various operating conditions. A standard search technique is then used to find the parameter settings that maximize or minimize the desired response.[26,27] SMs also develop a model describing the firm and are usually run using a restricted set of schedules to see which performs best. SMs allow the development of very complex systems, but computationally may be so large as to disallow exhaustive testing.[28] PSHs were developed to avoid frequent rescheduling of workforce sizes and production rates.[29] For example, the production rate P_t in period t is determined by[30]

Table 32.8 Classification of Aggregate Planning Approaches[a]

Optimal	Nonoptimal
A. Linear programming	A. Search techniques
B. Linear decision rules	B. Simulation models
C. Lot-size models	C. Production switching heuristics
D. Goal programming	D. Management coefficient models
E. Other analytical approaches	

[a]Modified from Ref. 6.

$$P_t = \begin{cases} L \text{ if } F_t - I_{t-1} < N - C \\ H \text{ if } F_t - I_{t-1} > N - A \\ N \text{ otherwise} \end{cases}$$

where F_t = demand forecast for period t
I_{t-1} = net inventory level (inventory on-hand minus backorder) at the beginning of period t
L = low-level production rate
N = normal-level production rate
H = high-level production rate
A = minimum acceptable target inventory level
C = maximum acceptable target inventory level

Although this example shows three levels of production, fewer or more levels could be specified. The fewer the levels, the less rescheduling and vice versa. However, with more levels, the technique should perform better, because of its ability to better track fluctuations in demand and inventory levels. MCMs were developed by attempting to model and duplicate management's past behavior.[31] However, consistency in past performance is required before valid models can be developed, and it has been argued that if consistency is present, the model is not required.[32]

32.4.4 Levels of Aggregation and Disaggregation

It should be obvious from the previous discussion that different levels of aggregation and disaggregation can be derived from use of the various models. For example, many of the linear programming formulations assume aggregate measures for multiple production and demand units such as production-hours, and provide output in terms of the number of production-hours that must be generated per planning period. For the multiple-product situation, therefore, a scheduler at the plant level would have to disaggregate this output into the various products by planning periods to generate the master production schedule. However, if data were available to support it, a similar, albeit more complex, model could be developed that considered the individual products in the original formulation, doing away with the necessity of disaggregation. This is not often done because of the increased complexity of the resultant model, the increased data requirements, and the increased time and difficulty in solving the formulation. Also, it should be noted that aggregate forecasts that are used as input to the planning process are generally more accurate than forecasts for individual products.

A major task facing the planner, therefore, is determining the level of aggregation and disaggregation required. Normally this is determined by the following:

1. The decision requirements and the level of detail required. Aggregate planning at a corporate level is usually more gross than that done at a division level.
2. The amount, form, and quality of data available to support the aggregate planning process. The better the data, the better the likelihood that more complex models can be supported. Complex aggregate models may also require less disaggregation.
3. The timing, frequency, and resources available to the planner. Generally, the more repetitive the planning, the simpler the approach becomes. Data and analysis requirements as well as analyst's capabilities significantly increase as the complexity of the approach increases.

32.4.5 Aggregate Planning Dilemma

Although aggregate planning models have been available since 1955 and many variations have been developed in the ensuing decades, few implementations of these models have been reported. Aggregate planning is still an important production-planning process, but many managers are unimpressed by the modeling approach. Why? One answer is that aggregate planning does occur throughout the organizational structure, but is done by different individuals at different levels in the organization for different purposes. For example, a major aggregate planning decision is that of plant capacity, which is a constraint on all lower-level aggregate planning decisions. Determining if and when new plant facilities are to be added is generally a corporate decision. However, input for the decision comes from both division and plant levels. Division-level decision-makers may then choose between competing plant facilities in their aggregate planning process in determining which plants will produce which quantity of which products within certain time frames, with input from the individual plant facilities. Plant level managers may aggregate plan their production facilities for capacity decisions, but then must disaggregate these into a master production schedule for their facility. This schedule is constrained by corporate and division decisions.

Most models developed to date do not explicitly recognize that aggregate planning is a hierarchical decision-making process performed on different levels by different people. Therefore, AP is not performed by one individual in the organization, as implicitly assumed by many modeling approaches, but by many people with different objectives in mind.

Other reasons that have been given for the lack of general adoption for AP models include:

1. The AP modeling approach is viewed as a top-down process, whereas many organizations operate AP as a bottom-up process.
2. The assumption used in many of the models, such as linear cost structures, the aggregation of all production into a common measure, or that all workers are equal, are too simplistic or unrealistic.
3. Data requirements are too extensive or costly to obtain and maintain.
4. Decision-makers are intimidated or unwilling to deal with the complexity of the models' formulations and required analyses.

Given this, therefore, it is not surprising that few modeling approaches have been adopted in industrial settings. Although research continues on AP, there is little to indicate any significant modeling breakthrough in the near future that will dramatically change this situation.

One direction, however, is to recognize the hierarchical decision-making structure of AP and to design modeling approaches that utilize it. These systems may be different for different organizations and will be difficult to design, but currently appear to be one approach for dealing with the complexity necessary in the aggregate planning process if a modeling approach is to be followed. For a comprehensive discussion of hierarchical planning systems, see Ref. 33.

32.5 MATERIALS REQUIREMENTS PLANNING

Materials requirements planning (MRP) is a procedure for converting the output of the aggregate planning process, the master production schedule, into a meaningful schedule for releasing orders for component inventory items to vendors or to the production department as required to meet the delivery requirements of the master production schedule.

Materials requirements planning is used in situations where the demand for a product is irregular and highly varying as to the quantity required at a given time. In these situations, the normal inventory models for quantities manufactured or purchased do not apply. Recall that those models assume a constant demand and are inappropriate for the situation where demand is unknown and highly variable. The basic difference between the independent and dependent demand systems is the manner in which the product demand is assumed to occur. For the constant demand case, it is assumed that the daily demand is the same. For dependent demand, a forecast of required units over a planning horizon is used. Treating the dependent demand situation differently allows the business to maintain a much lower inventory level in general than would be required for the same situation under an assumed constant demand. This is so because the average inventory level will be much less in the case where MRP is applied. With MRP, the business will procure inventory to meet high demand just in advance of the requirement and at other times maintain a much lower level of average inventory.

Definitions

Available units. Units of stock that are in inventory and are not in the category of buffer or safety stock and are not otherwise committed.

Gross requirements. The quantity of material required at a particular time that does not consider any available units.

Inventory unit. A unit of any product that is maintained in inventory.

Lead time. The time requirement for the conversion of inventory units into required subassemblies or the time required to order and receive an inventory unit.

MRP. Materials Requirements Planning: a method for converting the end item schedule for a finished product into schedules for the components that make up the final product.

MRP-II. Manufacturing Resources Planning: a procedural approach to the planning of all resource requirements for the manufacturing firm.

Net requirements. The units of a requirement that must be satisfied by either purchasing or manufacturing.

Product structure tree. A diagram representing the hierarchical structure of the product. The trunk of the tree would represent the final product as assembled from the subassemblies and inventory units that are represented by level one, which come from sub-subassemblies, and inventory units that come from the second level, and so on ad infinitum.

Scheduled receipts. Material that is scheduled to be delivered in a given time bucket of the planning horizon.

Time bucket. The smallest distinguishable time period of the planning horizon for which activities are coordinated.

32.5.1 Procedures and Required Inputs

The *master production schedule* is devised to meet the production requirements for a product during a given planning horizon. It is normally prepared from fixed orders in the short run and product