

### 1. MPS

MPS stands for Master Production Schedule. It is a company’s plan of how many and when items will be delivered to the customer. MPS includes deliveries that are fixed by the customer and promised by the company. It can include deliveries that are not yet promised but that are still negotiated. It may also include a forecast of customer demand. MPS covers a planning horizon of several months into the future. Initially a bigger portion of MPS is constituted by the firm customer orders and promised deliveries. But towards the end of the planning horizon, a larger portion of MPS is driven by demand forecasts.

Since firm orders cannot be altered by the customer and the company tends to hold delivery promises, initial months of MPS is fairly stable. Because of this, the portion of MPS associated with those months is said to be *frozen*. Outside the frozen horizon, orders can be modified relatively easily so MPS is *flexible*; see Figure 1.

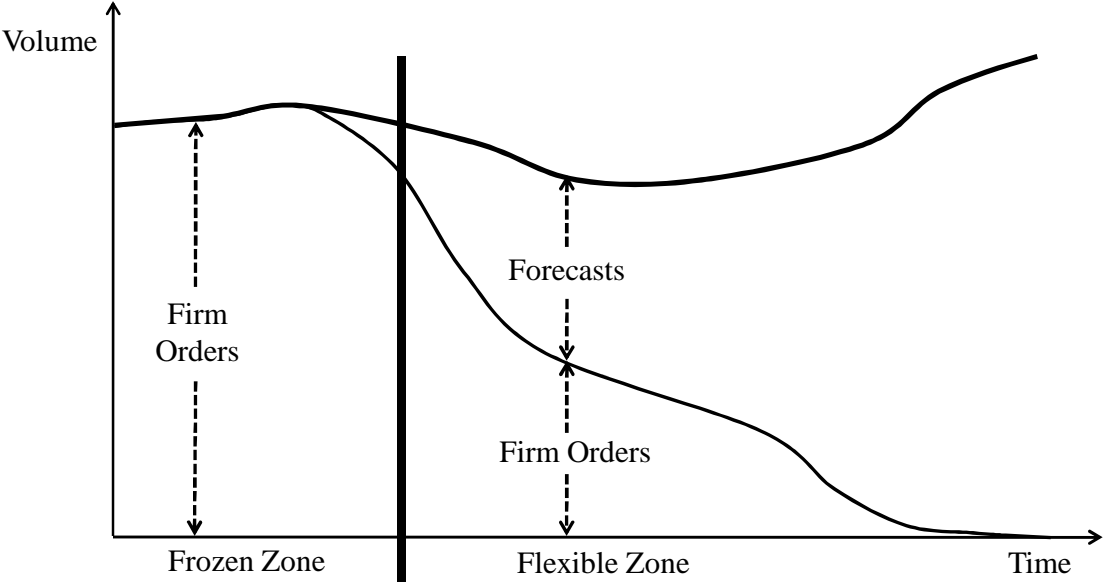


Figure 1: Frozen and flexible parts of MPS.

Having fixed the demand during the frozen horizon, companies use linear and integer programming formulations to make production/distribution plans. On the other hand, demands are random during the flexible zone. There are ways of incorporating randomness into linear and integer programs or ways to create models from scratch. What is important is to recognize the random demand during the flexible horizon. Failing to do so, i.e., assuming pseudocertainty, will result in plans that are too adamant to adjust for various demand scenarios that can materialize.

Demand forecasting is naturally an important ingredient in MPS construction, especially during the flexible horizon. We will not discuss forecasting methods. We want to emphasize that there are non-

traditional forecasting models (Guerrero and Elizondo [3], and Bodily and Freeland [2]) where demand is assumed to reveal itself in steps over time. Partial and earlier observations of demand is used to forecast the future demand later on. For example, in one of the models it is assumed that the ratio of orders received up to a certain time to the whole demand is approximately constant. After the proportionality constant is estimated, forecasts are readily generated from the partially observed demands.

We also note that forecasts for immediate future are more accurate (in terms of less randomness or lower variance) than those for the far future. Thus, planners should handle forecasts of future demands with some suspicion. If possible, it is a smart strategy to wait for forecasts to become more accurate before committing to meet demands. An example of this strategy is the postponement of product differentiation. Demand forecasts are the drivers of the SC operations, once they are ready we can plan the operations as discussed in the next section.

## 2. Aggregate Planning Strategies

1. Chase (the demand) strategy: produce at the instantaneous demand rate. Example: fast food restaurants.
2. Level strategy: produce at the rate of long run average demand. Example: swim wear production.
3. Time flexibility strategy: high levels of workforce or capacity that suffices to meet any realistic amount of demand. Example: Machining shops, army.
4. Deliver late strategy: convince the customer to wait for the delivery. Example: Spare parts for your Jaguar.

These strategies are extreme although we often resort to compromises among the extremes. To achieve such compromises, we need to detail the strategies further, which can be done through a more quantitative approach.

## 3. Aggregate Planning with Linear and Integer Programming

We start with an aggregate formulation example. Suppose a production manager is responsible for scheduling the monthly production levels of a certain product for a planning horizon of twelve months. For planning purposes, the manager was given the following information:

- The total demand for the product in month  $j$  is  $d_j$ , for  $j = 1, 2, \dots, 12$ . These could either be targeted values or be based on forecasts.
- The cost of producing each unit of the product in month  $j$  is  $c_j$  (dollars), for  $j = 1, 2, \dots, 12$ . There is no setup/fixed cost for production.
- The inventory holding cost per unit for month  $j$  is  $h_j$  (dollars), for  $j = 1, 2, \dots, 12$ . These are incurred at the end of each month.
- The production capacity for month  $j$  is  $m_j$ , for  $j = 1, 2, \dots, 12$ .

The manager's task is to generate a production schedule that minimizes the total production and inventory-holding costs over this twelve-month planning horizon.

To facilitate the formulation of a linear program, the manager decides to make the following simplifying assumptions for now:

1. There is no initial inventory at the beginning of the first month.
2. Units scheduled for production in month  $j$  are immediately available for delivery at the beginning of that month. This means in effect that the production rate is infinite.
3. Shortage of the product is not allowed at the end of any month.

To understand things better, let us consider the first month. Suppose, for that month, the planned production level equals 100 units and the demand,  $d_1$ , equals 60 units. Then, since the initial inventory is 0 (Assumption 1), the ending inventory level for the first month would be  $0+100-60=40$  units. Note that all 100 units are immediately available for delivery (Assumption 2); and that given  $d_1 = 60$ , one must produce no less than 60 units in the first month, to avoid shortage (Assumption 3). Suppose further that  $c_1 = 15$  and  $h_1 = 3$ . Then, the total cost for the first month can be computed as:  $15 \cdot 100 + 3 \cdot 40 = 1380$  dollars.

At the start of the second month, there would be 40 units of the product in inventory, and the corresponding ending inventory can be computed similarly, based on the initial inventory, the scheduled production level, and the total demand for that month. The same scheme is then repeated until the end of the entire planning horizon.

### 3.1 The Decision Variables

The manager's task is to set a production level for each month. Therefore, we have twelve decision variables:

- $x_j =$  the production level for month  $j$ ,  $j = 1, 2, \dots, 12$ .

### 3.2 The Objective Function

Consider the first month again. From the discussion above, we have:

- The production cost equals  $c_1x_1$ .
- The inventory-holding cost equals  $h_1(x_1 - d_1)$ , provided that the ending inventory level,  $x_1 - d_1$ , is nonnegative.

Therefore, the total cost for the first month equals  $c_1x_1 + h_1(x_1 - d_1)$ .

For the second month, we have:

- The production cost equals  $c_2x_2$ .
- The inventory-holding cost equals  $h_2(x_1 - d_1 + x_2 - d_2)$ , provided that the ending inventory level,  $x_1 - d_1 + x_2 - d_2$ , is nonnegative. This follows from the fact that the starting inventory level for this month is  $x_1 - d_1$ , the production level for this month is  $x_2$ , and the demand for this month is  $d_2$ .

Therefore, the total cost for the second month equals  $c_2x_2 + h_2(x_1 - d_1 + x_2 - d_2)$ .

Continuation of this argument yields that:

- The total production cost for the entire planning horizon equals

$$\sum_{j=1}^{12} c_j x_j \equiv c_1 x_1 + c_2 x_2 + \dots + c_{12} x_{12} ,$$

where we have introduced the standard summation notation (" $\equiv$ " means by definition).

- The total inventory-holding cost for the entire planning horizon equals

$$\begin{aligned}
\sum_{j=1}^{12} h_j \left[ \sum_{k=1}^j (x_k - d_k) \right] &\equiv h_1 \left[ \sum_{k=1}^1 (x_k - d_k) \right] + h_2 \left[ \sum_{k=1}^2 (x_k - d_k) \right] + \dots \\
&\quad + h_{12} \left[ \sum_{k=1}^{12} (x_k - d_k) \right] \\
&= h_1 [x_1 - d_1] + h_2 [(x_1 - d_1) + (x_2 - d_2)] + \dots \\
&\quad + h_{12} [(x_1 - d_1) + (x_2 - d_2) + \dots + (x_{12} - d_{12})].
\end{aligned}$$

Since our goal is to minimize the total production and inventory-holding costs, the objective function can now be stated as

$$\text{Min } \sum_{j=1}^{12} c_j x_j + \sum_{j=1}^{12} h_j \left[ \sum_{k=1}^j (x_k - d_k) \right].$$

### 3.3 The Constraints

Since the production capacity for month  $j$  is  $m_j$ , we require

$$x_j \leq m_j$$

for  $j = 1, 2, \dots, 12$ ; and since shortage is not allowed (Assumption 3), we require

$$\sum_{k=1}^j (x_k - d_k) \geq 0$$

for  $j = 1, 2, \dots, 12$ . This results in a set of 24 functional constraints. Of course, being production levels, the  $x_j$ 's should be nonnegative.

### 3.4 LP Formulation

In summary, we have arrived at the following formulation:

$$\begin{aligned}
\text{Min } &\sum_{j=1}^{12} c_j x_j + \sum_{j=1}^{12} h_j \left[ \sum_{k=1}^j (x_k - d_k) \right] \\
\text{Subject to : } & \\
&x_j \leq m_j \quad \text{for } j = 1, 2, \dots, 12 \\
&\sum_{k=1}^j (x_k - d_k) \geq 0 \quad \text{for } j = 1, 2, \dots, 12 \\
&x_j \geq 0 \quad \text{for } j = 1, 2, \dots, 12.
\end{aligned}$$

This is a linear program with 12 decision variables, 24 functional constraints, and 12 nonnegativity constraints. In an actual implementation, we need to replace the  $c_j$ 's, the  $h_j$ 's, the  $d_j$ 's, and the  $m_j$ 's with explicit numerical values.

### 3.5 An Alternative Formulation for Production Planning

In the above formulation, the expression for the total inventory-holding cost in the objective function involves a nested sum, which is rather complicated. Notice that for any given  $j$ , the inner sum in that expression,  $\sum_{k=1}^j (x_k - d_k)$ , is simply the ending inventory level for month  $j$ . This motivates the introduction of an additional set of decision variables to represent the ending inventory levels. Specifically, let

- $y_j =$  the ending inventory level for month  $j, j = 1, 2, \dots, 12$ ;

then, the objective function can be rewritten in the following simpler-looking form:

$$\text{Min} \quad \sum_{j=1}^{12} c_j x_j + \sum_{j=1}^{12} h_j y_j .$$

With these new variables, the no-shortage constraints also simplify to  $y_j \geq 0$  for  $j = 1, 2, \dots, 12$ . However, we now need to introduce a new set of constraints to “link” the  $x_j$ 's and the  $y_j$ 's together.

Consider the first month again. Denote the initial inventory level as  $y_0$ ; then, by assumption, we have  $y_0 = 0$ . Since the production level is  $x_1$  and the demand is  $d_1$  for this month, we have  $y_1 = y_0 + x_1 - d_1$ . Continuation of this argument shows that for  $j = 1, 2, \dots, 12$ ,

$$y_j = y_{j-1} + x_j - d_j ;$$

and these relations should appear as constraints to ensure that the  $y_j$ 's indeed represent ending inventory levels. We have, therefore, arrived at the following new formulation:

$$\begin{array}{ll} \text{Min} & \sum_{j=1}^{12} c_j x_j + \sum_{j=1}^{12} h_j y_j \\ \text{Subject to :} & \\ & x_j \leq m_j \quad \text{for } j = 1, 2, \dots, 12 \\ & y_j = y_{j-1} + x_j - d_j \quad \text{for } j = 1, 2, \dots, 12 \\ & x_j \geq 0 \quad \text{for } j = 1, 2, \dots, 12 \\ & y_j \geq 0 \quad \text{for } j = 1, 2, \dots, 12. \end{array}$$

which is a linear program with 24 decision variables, 24 functional constraints, and 24 nonnegative variables.

Although there are twice as many decision variables in the new formulation, both formulations have the same number of functional constraints. We will show in a later section that the total amount of effort necessary to arrive at an optimal solution to a linear program depends primarily on the number of functional constraints. In general, it is not uncommon to have several equivalent formulations of the same problem.

### 3.6 Remarks

- If Assumption 1 is relaxed, so that the initial inventory level is not necessarily zero, we can simply set  $y_0$  to whatever given value.
- In our formulation, we assumed that there is no production delay (Assumption 2). This assumption can be easily relaxed. Suppose instead there is a production delay of one month; that is, the scheduled production for month  $j, x_j$ , is available only after a delay of one month, i.e., in month  $j + 1$ . Then,

in the alternative formulation, we can simply replace the constraint  $y_j = y_{j-1} + x_j - d_j$  by  $y_j = y_{j-1} + x_{j-1} - d_j$  (with  $x_0 \equiv 0$ ), for  $j = 1, 2, \dots, 12$ . Of course, for the first month, the given value of  $y_0$  must be no less than  $d_1$ ; otherwise, the resulting LP will not have any solution.

- Assumption 3 can also be relaxed. If shortages are allowed, we can let  $y_t$  be negative as well. Then we define positive part of  $y_j$  as on-hand inventory  $I_j$  and negative part as backorder  $S_j$ :

$$I_j = \max\{0, y_j\}, \quad S_j = \max\{0, -y_j\}$$

so that

$$y_j = I_j - S_j$$

while  $I_j, S_j \geq 0$ . We also need to introduce a backorder penalty cost of, say,  $p_j$  per unit of shortage at the end of month  $j$ .

### 3.7 Solved Exercise with Integer Variables

PlaToy Company produces toys in Plano and Richardson. It has developed three new barbie dolls — Betsy, Vicky and Wendy — for possible inclusion in its product line for Xmas season. Setting up the production facilities to begin production would cost the same amount at both Plano and Richardson factories, see set up costs in dollars below. For administrative reasons, the same factory would be used for all new doll production: All dolls are produced at either Plano or Richardson Factories.

Regardless where they are produced, the contribution to margin for all the dolls are the same as given below. Also production rates at factories (units per hour) are also given below. Plano and Richardson Factories, respectively, have 300 hours and 200 hours of production time available before Xmas.

	Set up cost	Contribution to margin	Plano Factory Production rate	Richardson Factory Production rate
Betsy	50000	5	50	40
Vicky	60000	6	80	60
Wendy	40000	7	60	70

It is known that practically all barbies produced until Xmas can be sold. But these dolls will not be produced after Xmas. The problem is to determine how many units (if any) of each new toy should be produced before Xmas and at which factory to maximize the total profit. Formulate an MILP (Mixed Integer-Linear Program).

**Solution:** Let  $y_{BP} = 1$  if Betsy is produced at Plano, 0 otherwise. Define binary variables  $y_{BR}, y_{VP}, y_{VR}, y_{WP}, y_{WR}$  similarly. Let  $x_{BP}$  be the number Betsy produced at Plano. Define variables  $x_{BR}, x_{VP}, x_{VR}, x_{WP}, x_{WR}$  similarly.

Maximize  $5(x_{BP} + x_{BR}) + 6(x_{VP} + x_{VR}) + 7(x_{WP} + x_{WR}) - 50000(y_{BP} + y_{BR}) - 60000(y_{VP} + x_{VR}) - 40000(y_{WP} + y_{WR})$

ST:

$$y_{BP} + y_{BR} \leq 1; y_{BP} + y_{VR} \leq 1; y_{BP} + y_{WR} \leq 1.$$

$$y_{VP} + y_{VR} \leq 1; y_{VP} + y_{BR} \leq 1; y_{VP} + y_{WR} \leq 1.$$

$$y_{WP} + y_{WR} \leq 1; y_{WP} + y_{BR} \leq 1; y_{WP} + y_{VR} \leq 1.$$

These constraints say “the same factory would be used for all new doll production”:

$$y_{BP} = 1 \Rightarrow y_{BR} = 0, y_{VR} = 0, y_{WR} = 0; y_{BR} = 1 \Rightarrow y_{BP} = 0, y_{VP} = 0, y_{WP} = 0;$$

$$y_{VR} = 1 \Rightarrow y_{BP} = 0, y_{VP} = 0, y_{WP} = 0; y_{WR} = 1 \Rightarrow y_{BP} = 0, y_{VP} = 0, y_{WP} = 0;$$

$$y_{VP} = 1 \Rightarrow y_{VR} = 0, y_{BR} = 0, y_{WR} = 0; y_{WP} = 1 \Rightarrow y_{WR} = 0, y_{BR} = 0, y_{VR} = 0.$$

$$x_{BP}/50 + x_{VP}/80 + x_{WP}/60 \leq 300; x_{BR}/40 + x_{VR}/60 + x_{WR}/70 \leq 200.$$

These constraints say that no more than available hours used for production.

$$x_{BP} \leq 300(50)y_{BP}; x_{VP} \leq 300(80)y_{VP}; x_{WP} \leq 300(60)y_{WP};$$

$$x_{BR} \leq 200(40)y_{BR}; x_{VR} \leq 200(60)y_{VR}; x_{WR} \leq 200(70)y_{WR};$$

These constraints say that production can take place if the factory is set up.

## 4. Deterministic Capacity Planning

Until now it is always assumed that resource capacities are fixed and the aggregate plan is constructed taking those given capacities. When planning horizons are long enough, there is enough time to build up capacities within the planning horizon. In that case capacities must be considered as (decision) variables rather than given parameters. Capacity planning decisions are very hard to reverse. Once they are made and implemented, they have long lasting consequences. Therefore, capacity planning is critical for financial success. At a broader perspective, SCM can be thought as matching demand with capacity. Intel CEO, Andy Grove put this as:

- Think of every enterprise ... involved in adjusted capacity-demand pricing. ... If this can be done ... in real time, you'll see another power of 10 increase in the efficiency of the work in economic system. Now how do we get there?

Capacity planning models attempt to answer this question by telling the size of capacity increments and the time between the increments.

It is hard to forecast demand with long planning horizons. Remember the forecasting discussion and the fact that forecast become more variable as they are made for further periods in to the future. Traditionally this variability is studied by treating demands as random variables. Then the stochastic optimization techniques become the venue of choice. Because of the complexity of these techniques, we will restrict our scope to deterministic capacity planning. Even within the deterministic capacity planning stream, there are many different approaches. We next look at what criteria could be useful to classify deterministic capacity planning approaches:

1. Single vs. Multiple facilities. When the capacity of a facility is planned independent of other existing facilities such that facilities cannot transfer products among themselves, the problem is said to be single facility type.
2. Single vs. Multiple resource capacities. Almost all operations require multiple resources (various machines, workforce, etc.), but one can adapt an OPT-like approach and focus only on the single resource. Then the capacity of the bottleneck resource is planned and the remaining resources are adjusted according to the bottleneck resource capacity. This is typically how single resource capacity models are used. In the multiple resource case, each resource's capacity is studied explicitly and optimized jointly. In this case, the sequence of capacity increments (increment first resource A, second resource B, third resource A, fourth resource C, ...) needs to be decided on as well as the size of increments and the time between increments.
3. Single vs. Multiple demands. In aggregate planning, demands of several products are lumped together to obtain a single demand stream. This is of course an approximation. The exact way of handling multiple demands is representing them as a multidimensional vector. Then the capacity is planned to match demand in every dimension. Clearly handling multiple demand streams explicitly is a big challenge and there are very few models in this category.

4. Expansion vs. Contraction. Many models consider only capacity expansions. Capacity contractions are not financially important, especially when machine capacities are concerned. This is because, machines tend to fill their economic life (amortized value of 0 dollars) before contraction, so companies do not care to deal with the issue of getting rid of these no-dollar-value machines. The situation is different for labor, because as long as the labor is employed there is an associated cost. Therefore, contraction needs to be studied more carefully when studying labor capacity.
5. Discrete vs. Continuous expansion times. The choice here is made for modelling convenience. Depending on the structure, sometimes discrete models are easier to handle, and other times continuous ones. The quality of solution whether the expansion times are discrete or continuous does not change much. Sometimes vendors accept/deliver orders on certain times, say every monday, such cases give rise to discrete times. But these cases can be studied almost equivalently well with continuous times.
6. Discrete vs. Continuous capacity increments. Almost all capacities can be incremented only in discrete sizes. We can not buy half a machine or tenth of a truck. Thus, it will not be wrong to say that capacity comes in quantum. However, sometimes the scale of increment is so large that capacity can be treated as continuous. This is generally the case in workforce planning. The difference between say 15 workers and 15.5 is not much. The justification of treating capacity as a continuous variable is the same as the justification of using continuous variables for intrinsically-discrete activities in linear programming.
7. Capacity costs. In most of the models, resources are priced according to their type and the increment size. Naturally 1 lathe costs different than 1 press. The more interesting concept is the economies of scale: 2 lathes cost less than twice the price of 1 lathe. When there is economies of scale, costs are represented as increasing and concave costs. There can also be fixed costs associated with capacity expansions. For example, when a building is expanded a big portion of the cost is independent of the size of the expansion and it is fixed.
8. Penalty for demand-capacity mismatch. Many models plans capacity such that demand is always met. However sometimes it is more profitable to delay an expansion (to save from the opportunity cost of the expansion) while falling short of the demand. In order to evaluate the cost of this option properly, we must charge ourselves when shortages happen. Shortage costs are generally proportional to the magnitude of shortages. Penalty costs for not meeting demands are similar to those of inventory.
9. Single vs. Multiple decision makers. In competitive markets, companies pay attention to each other's capacity expansions to avoid too much slack capacity. Then, expansion decisions become interdependent across companies. There are models that use game-theoretic techniques to study multi-player markets with or without information symmetries.

**A Model of Constant Capacity Increments:** We now study a simple capacity expansion model where capacity increments are all constant and equal to  $x$ . We keep the deterministic demand assumption and let demand be  $D(t) = \mu + \delta t$ . Here  $\mu$  is the demand at the present time  $t = 0$  and  $\delta$  is the rate of demand increase. Note that such an increasing trend in the demand cannot happen forever but for long enough periods in specific industries. Examples can include global PC demand in the 1990s, global cell phone demand in the 2000s, electricity demand in developing countries in the 2010s.

We assume that capacity is determined in such a way that the demand is always met. This assumption applies to industries where storage is not possible/economical, such as power (electricity generation) industry. Since capacity increments of size  $x$  are depleted in  $x/\delta$  time units, there will be a new capacity increment every  $x/\delta$  time units. Thus,  $x/\delta$  is the time between successive expansions. See Figure 2.



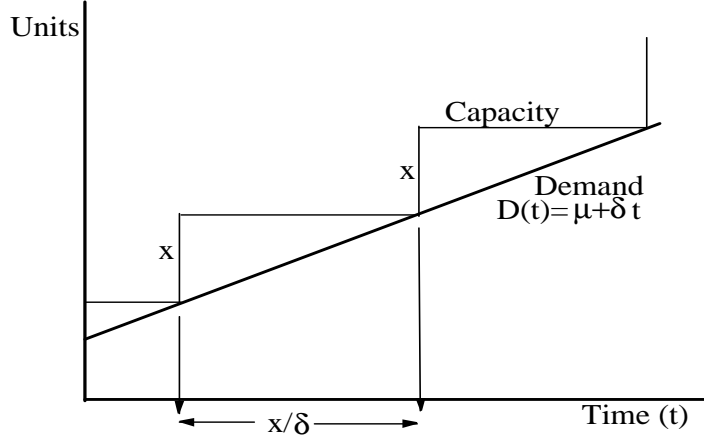


Figure 2: Constantly growing demand vs. capacity with fixed increments of  $x$ .

Let  $f(x)$  be the cost of an expansion of size  $x$ . In general, expansion costs have economies of scale and  $f(x)$  is an increasing but a concave curve. Let  $r$  be the interest rate (say, about 5%), i.e., discount rate of money. Then the present value of the cost of an expansion of size  $x$  made  $t$  time units later is  $\exp(-rt)f(x)$ . There is a brief discussion of continuous discounting and "exp" in the Appendix.

Let  $C(x)$  be the discounted cost of expansions of size  $x$  over an infinite horizon:

$$\begin{aligned}
 C(x) &= f(x) + \exp(-rx/\delta)f(x) + \exp(-2rx/\delta)f(x) + \cdots + \exp(-kx/\delta)f(x) + \dots \\
 &= \exp(-0rx/\delta)f(x) + \exp(-1rx/\delta)f(x) + \exp(-2rx/\delta)f(x) + \cdots + \exp(-kx/\delta)f(x) + \dots \\
 &= f(x) (\exp(-0rx/\delta) + \exp(-1rx/\delta) + \exp(-2rx/\delta) + \cdots + \exp(-kx/\delta) + \dots) \\
 &= f(x) \sum_{k=0}^{\infty} \exp(-kx/\delta),
 \end{aligned}$$

where term  $\exp(-kx/\delta)f(x)$  is the present value of the  $k$ th expansion. We can write the present value of the total cost as

$$C(x) = f(x) \sum_{k=0}^{\infty} (\exp(-rx/\delta))^k = f(x) \frac{1}{1 - \exp(-rx/\delta)},$$

where we have used the identity

$$\sum_{k=0}^{\infty} a^k = \frac{1}{1 - a} \text{ for } a < 1.$$

For illustration we choose  $f(x) = x^{0.5}$  dollars,  $r=5\%$ ,  $\delta=1$  unit per week, and draw  $C(x)$  in Figure 3. The minimum cost is achieved around the optimal expansion size  $x^* = 30$  units. That is every expansion must be of the size equivalent to 30 weeks demand. The total discounted cost of implementing  $x^*$  is about 7 dollars. In general, we suggest that  $C(x)$  graphed and  $x^*$  is found from the graph. This will be a safe way to find  $x^*$  especially when  $C(x)$  is not convex. The details of this simple capacity expansion model can be found in [5].

An important assumption in the previous model was 100% demand satisfaction. Capacity is often expensive and companies may choose to pay for shortage costs rather than install capacity. Once shortage

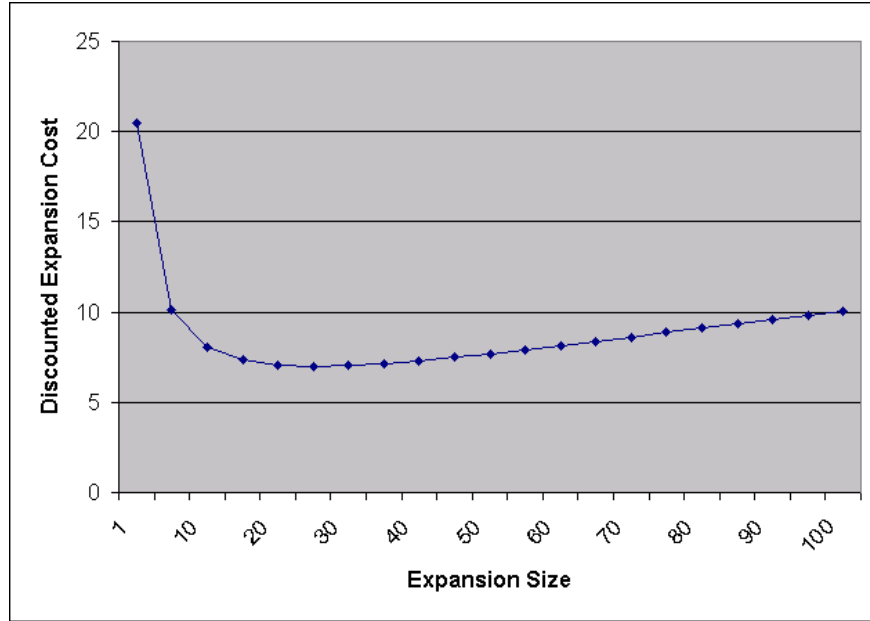


Figure 3: Discounted expansion cost  $C(x)$  as  $x$  varies.

becomes substantial, capacity is installed. Right before capacity installation, companies may use subcontracting to boost up their capacity. Inventories can be used to regulate capacity and demand mismatch. Inventories can be built up when there is surplus capacity. These inventories are later depleted when demand grows above capacity. We summarize these observations in Figure 4.

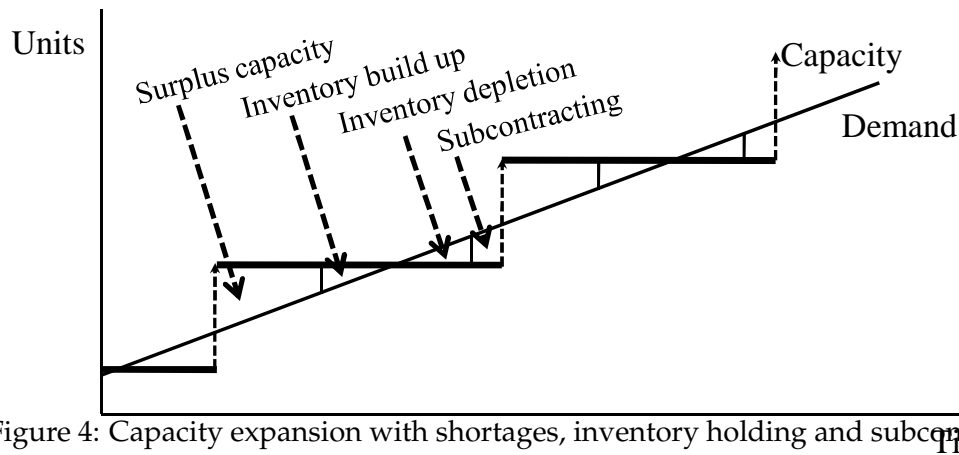


Figure 4: Capacity expansion with shortages, inventory holding and subcontracting.

Clearly, when the options of accumulating inventory and subcontracting are allowed, capacity planning problem becomes more complex. On the other hand, a good capacity plan must be able to match demand so closely that there should not be a need for inventories or subcontracting. Then the motivation for capacity planning is to avoid inventories. Partly because of this motivation and partly because of its complexity, many capacity planning models do not emphasize/study inventory holding or subcontracting.

## 5. Stochastic Capacity Planning: The case of flexible capacity

Many manufacturing companies own several plants where they produce various products. But not all products can be produced at all plants. In order to produce a product at a plant, special tooling may need to be installed at the plant. A plant that can produce many products is called flexible. To be accurate, it is said to have product variety flexibility (compare this concept to product mix and volume flexibility). Flexible plants provide a richer set of options when allocating products to plants.

First consider a small example with 3 plants and 2 products A and B. Suppose that Plant 1 and 2 produces only A, Plant 3 produces only B. Neither of these plants are flexible. Suppose further that we retool Plant 2 so that it can produce B as well. With this added flexibility, we may produce all A or all B's at Plant 2. However, most often the production mixture is such that some As are produced along side with some Bs at Plant 2. The more interesting issue is that the actual mixture of production at Plant 2 can be determined after observing the actual demand. For example when demand for A is high, Plant 2 produces more As and fewer Bs, and vice versa. Note that in both cases, the total production at Plant 2 is stable although individual product volumes vary. When individual products are substitutes for each other (i.e. they have negatively correlated demand), variations in A and B demand will cancel each other, keeping the total demand constant. Therefore, the resulting system is more manageable.

In order to appreciate the flexibility of Plant 2, we must be explicitly model the correlation between product A and B demand. Note that correlations are ignored when decisions are based on expected revenues / profits. As a result, flexibility will have limited value if only discounted expected cash flows are studied after ignoring demand correlations. The flexibility will be most valuable when the correlation is strong but negative. By now, it must be clear that demands must be treated as random variables.

Define the following parameters

- Use  $i$  to denote plants and  $j$  to denote products.
- $D_j$ : Random demand for product  $j$ .
- $c_{ij}$ : Tooling cost to configure plant  $i$  to produce product  $j$ .
- $m_j$ : Contribution to margin of producing a unit of  $j$  during the regular time.
- $r_i$ : Regular time capacity in units available at plant  $i$ .

As for the decision variables, we have the plant to product assignment binary variable  $y_{ij}$  and production levels  $x_{ij}$ . Then the objective function immediately follows

$$\max - \sum_{i,j} c_{ij} y_{ij} + \sum_{i,j} m_j x_{ij}$$

There are also capacity constraints

$$\sum_j x_{ij} \leq r_i \quad \text{for each plant } i$$

A plant can produce only those products it is configured for

$$x_{ij} \leq r_i y_{ij} \quad \text{for each plant and product } i, j$$

Suppose that we know the demands in advance as  $D_j = d_j$  (Capital letters used for random variables, small letters for observations from associated random variables). Then we can add the demand constraint as

$$\sum_i x_{ij} \leq d_j \quad \text{for each product } j$$

However demands are not known in advance in general. Product to plant assignment decisions  $y_{ij}$  are strategical so they are made against uncertain demand. On the contrary, production levels  $x_{ij}$  are determined more frequently and we can safely assume that they are determined after observing the demand. In summary, the chronology of our decision making and demand observation is as: Decide on  $y_{ij}$ , Observe  $D_j = d_j$  and Decide on  $x_{ij}$ .

Due to the specific sequence of decisions and demand observation we adopted,  $y_{ij}$  must be independent of demand while  $x_{ij}$  can depend on the demand. To make ideas concrete, suppose that demands are likely to be either  $d_j^1$  or  $d_j^2$ . More simplistically, we may say that the observed demand depends on a coin toss after choosing  $y_{ij}$ . If head comes up, demand will be  $d_j^1$ , otherwise  $d_j^2$ . However, whether head or tail comes up,  $y_{ij}$  will be the same:  $y_{ij}$  cannot anticipate the outcome of the coin toss. This limitation is very common in decision making under uncertainty and leads to *Nonanticipatory* decision variables, e.g.  $y_{ij}$ . On the other hand,  $x_{ij}$  is chosen after the demand so it is not need to be nonanticapatory: For different demand observations, different  $x_{ij}$  can be used. For example if head comes up, we use  $x_{ij}^1$ , otherwise  $x_{ij}^2$ .

Each outcome of the demand is named as a scenario. Suppose that scenarios are indexed with  $k$  and use  $p^k$  to denote the probability of scenario  $k$ . Let us see now how the formulation changes

$$\begin{aligned} \max - \sum_{i,j} c_{ij} y_{ij} + \sum_{i,j} p^k m_j x_{ij}^k \\ \sum_j x_{ij}^k \leq r_i \quad & \text{for each plant } i \quad \text{and for each scenario } k \\ x_{ij}^k \leq r_i y_{ij} \quad & \text{for each plant and product } i, j \quad \text{and for each scenario } k \\ \sum_i x_{ij}^k \leq d_j^k \quad & \text{for each product } j \quad \text{and for each scenario } k \end{aligned}$$

By introducing probabilities  $p^k$  into the objective, objective function becomes the expected total contribution to margin minus the plant configuration cost. The meaning of constraints do not change much but they must be enforced for each scenario. The linear program given above is an example of a Stochastic Linear Program. Note that going from a deterministic setting to a stochastic one both the number of constraints and variables have increased significantly. The biggest issue in Stochastic Linear Programming is to solve large Linear Programs using decomposition techniques ([1]). The plant flexibility problem presented here is a simplification of the model given by [4].

## 5.1 Anticipatory vs nonanticipatory variable

When you are making decisions, some decisions are made before uncertainty resolution, some are made after uncertainty resolution. Nonanticipatory decisions happen before uncertainty resolution; anticipatory decisions happen after uncertainty resolution.

Now let us put these concepts to work. If demands are uncertain, they are observed when customers want products. Shipment quantities are anticipatory decision variables in a make-to-order system. In this system, shipments happen after demands. On the other hand, plant configuration happens way before

demands. Thus, configuration decisions are nonanticipatory with respect to demand. Here we have sequence of events: Configure (nonanticipatory decision) - Find out demand (uncertainty resolution) - Ship (anticipatory decision).

When Toyota decides on a recall, the recall decision is nonanticipatory with respect to the root cause of the problem. The root cause can be deformation of pedal springs as opposed to software problems or humidity build up that increases friction in the gas pedal. Once the uncertainty around the root cause is resolved, Toyota has to decide on how to fix the root cause. The decision regarding the way the root cause is fixed is anticipatory with respect to the root cause. Here the sequence of events: Recall (nonanticipatory decision) - Find root cause (uncertainty resolution) - Fix root cause (anticipatory decision).

This distinction between anticipatory/nonanticipatory decisions is very important. In management, it is natural that only a few decisions must be made simultaneously. Many decisions are staged over time. The decisions that are made earlier (before uncertainty resolution) are nonanticipatory.

*A word of caution on etymology:* If you look at the dictionary, you will see that “anticipate” comes from *anti* and *capare*. *Anti* means before and *capare* means to take in Latin. To anticipate is to take something “before” it happens. Both “predict” and “forecast” have the same concept. Predict comes from *prae* and *dicere*, so it is about saying something “before” it happens. Forecast comes from *fore* and *cast*, so it is about putting forth something “before” it happens. All these three verbs, anticipate, forecast and predict, have the same construction and they all include the word “before”. However, anticipatory variables relate to decision that happen “after” uncertainty resolution. This terminology can be confusing if you take the word “before” in anticipate too literally. Perhaps it is best to ignore the word “before” in anticipate.

## 6. Exercises

1. **Aggregated computing capacity:** The SOM currently has about 100 instructors, each of whom is given a laptop computer to use in the classroom. These laptops have course slides and associated software (access, excel, etc.) used during a lecture. A recent proposal is advocating for hosting these slides and software on a central computer. According to this proposal, the instructors will use a keyboard and a computer screen in the classroom to connect to the central computer to project their slides and to use the software. Each of the laptops currently has 1.8 GHz processing capacity and 75 GB storage capacity. Suppose that if 100 laptops are connected appropriately (with an efficient parallel processing architecture) the total processing capacity will become 180 GHz while the total storage capacity becomes 7500 GB. First note that the capacity demanded by each instructor is random. Then provide a verbal argument for why the central computer can serve the instructors as well as 100 laptops even it has less capacity than 180 GHz and 7500 GB. Use the concept that extreme values cancel each other in aggregation.
2. **Deterministic capacity expansion with infinite horizon:** Replicate Figure 3 by changing the following parameters (a and b are independent):
  - a)  $f(x) = x^{0.2}$ .
  - b)  $r=20\%$ .
 Discuss how  $x^*$  changes in each case. Intuitively explain if you expected these changes, why?
3. **Capacity expansion with finite horizon:** While drawing Figure 3, we had  $f(x) = \sqrt{x}$ ,  $r = 0.05$ , and  $\delta = 1$ . We keep the same parameters but would like to compute the discounted expansion cost only over  $T = 100$  time periods. Note that we have considered an infinite horizon in Figure 3, but now we are switching to consider a finite horizon of  $T = 100$ .
  - a) Let  $N(x)$  be the number of expansions made up to and including time  $T$ . For example, if  $x = 25$ , we would make  $N(x = 25) = 5$  expansions over the horizon. These expansions are made at times 0,

25, 50, 75 and 100. If  $x = 30$ , we make  $N(x = 30) = 4$  expansions over the horizon, at times 0, 30, 60 and 90. In general,

$$N(x) = \left\lfloor \frac{T}{x/\delta} \right\rfloor + 1,$$

where notation  $\lfloor z \rfloor$  indicates rounding down the number  $z$ . For example,  $\lfloor 2.2 \rfloor = \lfloor 2.7 \rfloor = 2$ . Explain the formula for  $N(x)$  in English. Specifically, explain why we round down  $T\delta/x$  and why we add 1.  
 b) The cost  $C^T(x)$  over the finite horizon of  $T$  becomes

$$C^T(x) = f(x) \sum_{k=0}^{N(x)-1} \exp(-rkx/\delta).$$

Use the identity

$$\sum_{k=0}^{N-1} a^k = \frac{1 - a^N}{1 - a} \quad \text{for } 0 < a < 1$$

to simplify  $C^T(x)$ , i.e., write  $C^T(x)$  without the sum  $\sum$ .

c) Draw  $C^T(x)$  found in b) to find the optimal size of expansion  $x$ .

d) Observe that optimal expansion size differs with the horizon covered by expansions. In practice, do we ever see infinite horizons in decision making? How do decision horizons for a company's stockholders and its managers differ, which is longer? What are the undesirable consequences of a manager's short-sightedness? Think of examples from daily news.

4. **Location with exchange rate uncertainty:** Consider the facility location formulation with fixed infrastructure cost  $f_i$  for plant  $i$ .

$$\text{Minimize } \sum_{i=1}^{14} f_i a_i + \sum_{i=1}^{14} \sum_{j=1}^{18} c_{ij} x_{ij}$$

s.t.

$$\sum_{i=1}^{14} x_{ij} = d_j$$

$$\sum_{j=1}^{18} x_{ij} \leq C_i a_i$$

$$x_{ij} \geq 0, a_i \in \{0, 1\}.$$

This formulation does not have the variable and parameter names that you are used to. You can guess that  $a_i$  denotes the indicator variable for opening up plant  $i$ .

a) What does  $C_i$  denote, how many plant locations and markets are there in this formulation?

b) Suppose that all of our plants and markets are in the same country except for plant 1. Since Plant 1 is in a foreign country, the production cost of producing there is random in terms of the currency of our home country. This production cost randomness makes the production plus transportation costs of the form  $c_{1j}$  random. Despite this randomness, we need to decide on plants to operate now, next we will see the exchange rates and then we will determine the production/transportation quantity  $x_{ij}$ . Is the production/transportation decision anticipatory with respect to exchange rate randomness? Is plant location decision anticipatory with respect to exchange rate randomness?

c) Modify the formulation above to incorporate two exchange rate scenarios: low cost  $c_{1j}^l$  and high cost  $c_{1j}^h$  that happen with equal probabilities. Alter anticipatory and nonanticipatory variables appropriately.

5. In many industries, such as Semiconductor and Automobile, subcontracting is becoming a popular way of capacity management. A strong motivation for subcontracting is usually the cost advantage.
  - a) Explain, how can subcontractors provide products at a lower cost than in-house manufacturing costs?
  - b) Given this cost advantage, what might be the disadvantages of subcontracting?
  - c) Comment on whether the disadvantages in (b) are more pronounced in Semiconductor or Automobile industry?
6. For a swimming suit manufacturer the demand is given by

$$D_i = \left\{ \begin{array}{ll} \mu i & \text{if } i \in \{1, 2, 3, 4, 5, 6\} \\ \mu(13 - i) & \text{if } i \in \{7, 8, 9, 10, 11, 12\} \end{array} \right\}$$

where  $i$  denotes a month, e.g.  $i = 1$  is January.

- a) Graph this demand month by month over a year. What is the average demand per month?
  - b) Suppose that the swimming suit manufacturer incurs  $m$  dollars to adjust the workforce to produce  $\mu$  more or  $\mu$  fewer swimming suits. Compute the workforce adjustment cost in terms of  $m$  to follow a chase strategy.
  - c) Suppose that customers can be given  $f/\mu$  dollars of discount per unit to pull their demand 6 months in advance. Assume that discounts are used to set the actual demand exactly equal to the average demand. For example if we give a total discount of  $5f/2$  in January we can pull  $5\mu/2$  units of demand from July. What is the total cost of workforce adjustments and discounting in terms of  $m$  and  $f$  if discounts are given in the winter months: December, January and February?
  - d) Compare your answer to b and c. For what values of  $f$  and  $m$  or a relationship between these two parameters, discounting to motivate forward buying yields lower costs?
  - e) Repeat b,c,d) when workforce adjustment has a fixed cost of  $K$ .
7. Read the Stochastic Capacity Planning section.
    - a) Suppose that the chronology of decision making and demand observations are as; Observe the demand, Decide on plant to product assignment (configurations) and production levels. Determine anticipatory and nonanticipatory variables, and provide a Linear Programming formulation.
    - b) **The value of advance information:** First let  $z^*$  be the optimal value of the formulation in the section. Second, consider the formulation in a) where there is exactly one formulation for each scenario i.e. for each demand observation. For example, if there are 10 scenarios, there will be 10 formulations, which differ from each other only by the demand data. Let  $z_k^*$  be the optimal objective value with demand scenario  $k$  and define  $z^*(advance)$  as the average (expected) value of  $z_k^*$ , i.e.  $z^*(advance) = \sum_k p^k z_k^*$ . Define the value of advance information as  $z^*(advance) - z^*$  and argue that this value is nonnegative to conclude that

$$z^*(advance) \geq z^*.$$

Hint: Start with the formulations, one for each scenario, eventually yielding the average objective value  $z^*(advance)$ , combine them into a single formulation. Note that the only difference between the *combined formulation* with objective value  $z^*(advance)$  and the *original Stochastic Linear Programming formulation* with objective value  $z^*$  is a constraint (which?). Finally, use the fact that adding a constraint to a maximization Linear Program cannot increase its optimal objective value.

8. Read the Stochastic Capacity Planning section. Consider two demand scenarios  $d_j^1$  and  $d_j^2$  given for all products  $j$  such that  $d_j^1 > d_j^2$  for all products. Suppose that we set  $D_j = d_j^1$  and  $D_j = d_j^2$ , and solve the initial LP to obtain objective values  $z_1^*$  and  $z_2^*$ . How does  $z_1^*$  compare against  $z_2^*$ , why?

9. Recall the 2 product and 3 plant example in the Stochastic Capacity Planning section. We want to measure the net benefit of reconfiguring plant 2 to produce B. Suggest a step by step framework using the Stochastic Linear Program given in the section to measure the net benefit.
10. Read the Stochastic Capacity Planning section. Suppose that there is a current plant to product assignment given by  $y_{ij}^0$ , i.e. Plant  $i$  has the tooling to produce Product  $j$ . However, we do not know if additional Plant configurations can increase the profit.
  - a) Let  $y_{ij}(y_{ij}^0)$  be the assignment after additional configurations (if any) made starting from  $y_{ij}^0$ , modify the Stochastic Linear Program to find optimal  $y_{ij}^*(y_{ij}^0)$ .
  - b) Let  $z^*(y_{ij}^0)$  be the optimal solution if the initial assignment is  $y_{ij}^0$ . Mathematically argue that  $z^*(y_{ij}^0) \geq z^*(\bar{y}_{ij}^0)$  if  $y_{ij}^0$  and  $\bar{y}_{ij}^0$  are two initial configurations such that  $y_{ij}^0 \geq \bar{y}_{ij}^0$  for all plants  $i$  and products  $j$ . Express this result in English as well.
11. Refer to the solved exercise dealing with PlaToy. Now suppose that different dolls can be produced at different factories but each doll is produced at only one factory. For example, Betsy and Vicky can be produced at Plano while Wendy is produced at Richardson. But it is not possible to produce any one of the dolls at two locations. For example, Betsy **cannot** be produced both at Plano and at Richardson. In addition, there is an additional \$10000 administrative cost incurred at a factory if any one of the new dolls are produced at the factory: If 2, 1, 0 factories are used the cost is \$20,000, \$10,000, \$0. Formulate an MILP.
12. When the number of students suddenly increase at the School of Management in the Spring term, what instruction capacity management strategy can be used? Chase strategy or level strategy, or should the instruction capacity be maintained significantly over the regular teaching demand levels? Answer this question twice, once from the perspective of the dean of the school, and once from the perspective of UT system administrators.
13. Download flexman\_question.xls, which includes the data pertaining to question 8.4 on p.236 of the textbook. Read the question statement.
  - a) The question does not mention any raw material costs for the routers or switches. What assumptions do you need to make to conclude that these costs are sunk (irrelevant): i) No backlog, ii) No subcontracting, iii) No layoffs, iv) No new hires, v) Constant raw material costs over the year.
  - b) Compute the monthly regular time and overtime capacity in the spreadsheet.
  - c) By respecting the capacities computed in b), find a production plan that minimizes the total relevant cost while ending december with exactly the same level of inventories at the beginning of January. Report the production quantities, provide a printout of your final spreadsheet with the objectives, decision variables and constraints clearly written.
14. Consider the product-to-plant assignment of major automobile manufacturers in North America in scaggregate.ppt.
  - a) Which company owns the closest car manufacturing plant to UTD and what does that plant produce?
  - b) On the US map, consider a rectangle whose corners are the states of North Dakota, Oklahoma, Arizona, Montana (NOAM rectangle). How many car manufacturing plants are there in the NOAM rectangle? (Hint: What can NOAM stand for other than the first letters of the states?) In 1-2 sentences, explain what could be the underlying reason(s) behind the automobile manufacturing activity in the NOAM rectangle?
  - c) Consider the Toyota Plants in the US from west to east: 1: Nummi, 2: Long Beach, 3: San Antonio, 4: Blue Springs, 5: Princeton, 6: LaFayette, 7: Georgetown. Also consider the products of these plants:



A: Corolla, B: Tacoma, C: Hino, D: Tundra, E: Highlander, F: Avalon, G: Camry, H: Solara, I: Sequoia, J: Sienna. List the product to plant assignment variables  $y_{i,j}$  for  $i \in \{1, 2, \dots, 7\}$  and  $j \in \{A, B, \dots, J\}$  that currently have value of 1.

d) Count the number of brands/types of cars/trucks GM is producing in the US and compare that number against 10 for Toyota. Although Toyota has larger sales, it seems to be producing fewer number of types of cars/trucks than GM does. Would you expect a decrease in the number products in GM's product portfolio in an economic downturn, explain?

## 7. Appendix: Continuous Compounding of Interest Rates

We are all familiar with the periodic compounding of the interest rates. For example the net present value of  $A$  invested today would become

$$A(1 + r)$$

in a year with an interest rate of  $r$ .

Now suppose that a year is divided into two 6 month periods each with interest rate of  $r/2$ . Then the value of  $A$  at the end of a year is

$$A(1 + r/2)^2.$$

For the purpose of generality, say that a year is divided into  $m$  periods of equal length and at the end of each period interest accrues periodically. In that case the future value of  $A$  is

$$A(1 + r/m)^m.$$

Continuous compounding is basically dividing a year into more and more periods of equal length and applying periodic compounding. As we increase the value of  $m$ , the future value of  $A$  becomes

$$\lim_{m \rightarrow \infty} A \left(1 + \frac{r}{m}\right)^m = A \left[\lim_{m \rightarrow \infty} \left(1 + \frac{r}{m}\right)^m\right] = Ae^r.$$

The last equality follows from Calculus. When  $r = 1$ , the equality is used as the definition of the number  $e$ . Another definition for  $e$  is the following series expansion

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} + \dots$$

Having established the future value, we can directly write the net present value of  $B$  to be received at the end of a year as

$$\frac{B}{e^r} = Be^{-r} = B \exp(-r).$$

If  $B$  is received  $x$  years later, the present value would be

$$\underbrace{B \exp(-r) \exp(-r) \dots \exp(-r)}_{x \text{ times}} = B \exp(-rx).$$

## 8. Appendix: MRP/OPT/JIT

### 8.1 MRP

MRP stands for material requirements planning. The name is reminiscent of the limited capabilities old MRPs which had little functionality beyond figuring out a schedule of raw material requirements. However, modern day MRP systems come along with various modules associated with many business functions from scheduling to accounting and to inventory control.

The first step of MRP is to look at the MPS and decide which, how many and when components are needed to meet MPS. This process is called MRP explosion. During the explosion MRP uses a network representation of product assembly called Bill of Materials (BOM). Suppose TexBag is company in manufacturing bags, then its BOM may look like Figure 5. According to the BOM a bag is made of two main parts: a body and a strap. A body is manufactured by sewing 12 zips on a leather body. A strap is made by attaching two hooks to the ends of a leather strap. Next to each operation, we also have its manufacturing lead time. Lead times should not be confused with actual processing times. Lead times include transportation and wait-in-queue times as well as processing times. For example, assembly of a strap to a body can be done in 5-6 minutes but the lead time can be a day mainly because of waiting. In summary, BOM is a diagram for representing parts used in assembly and manufacturing operations, and lead times of these operations.

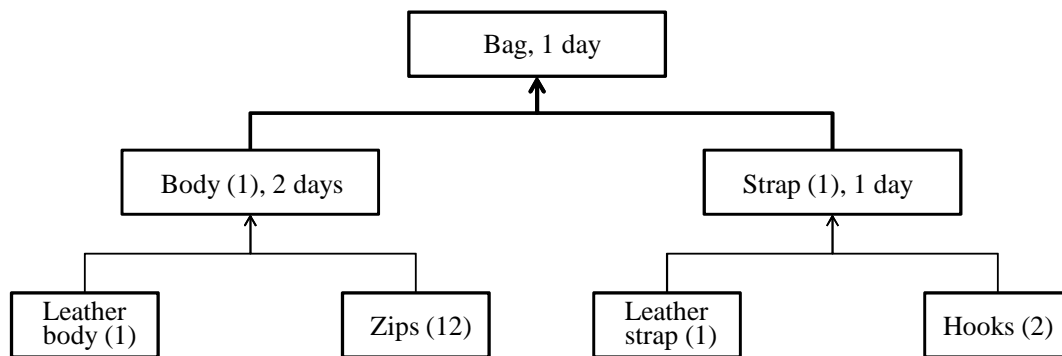


Figure 5: BOM of a bag.

Now suppose that we need to deliver 50 bags on friday morning. Then we must have 50 bodies and 50 straps ready by thursday morning, we obtain thursday morning by backing up by 1 day assembly lead time. To have 50 bodies on thursday morning, we must have 50 leather bodies and 600 zips on tuesday morning. To have 50 straps on thursday morning, we must have 50 straps and 100 hooks on wednesday morning. One can continue by backing up in time by operation lead times while computing how many parts are needed for a certain number of assemblies. This process is called MRP explosion.

Once MRP explosion reaches the raw material level (leaves of the BOM tree), a schedule of how many parts are needed and at what dates becomes available. This schedule of raw material requirements is passed to Procurement department which orders raw materials from upstream companies in the supply chain. Raw material acquisition schedule becomes an MPS for those upstream companies. MRP explosion information helps to plan operations so that correct amount and type of inputs are ready for operations and enough time is allocated to operations. For discrete product manufacturers whose BOM has multiple levels, MRP explosion must be computerized. Computerization not only expedites the process but also disciplines it and avoids manual errors. Enterprise resource planning (ERP) applies MRP logic to SCs, over various production facilities of a firm or several firms.

MRP is the first conceptual representation that realized the underlying network structure in complex production systems. It provides a systematic and coordinated way to schedule a large number of items over this network. MRP is basically a specialized database system, so all advantages of database systems apply to MRP as well. MRP streamlines information; it quickly makes correct information available. Streamlining can foster standardization and further product development. For example, BOM is relevant for product developers because it usually is a starting point for design improvements. Many design improvements

come as elimination or standardization of various parts. Suppose that TexBag is also manufacturing suitcases and uses zips in the suitcases. If the design team can standardize zips (say their length) then the same zip can be used in bags and suitcases. Such a standardization decreases the complexity of processes; one fewer product to name, buy, store, transport.

A major drawback of MRP is its failure to incorporate capacity restrictions appropriately during the planning process. MRP first makes the plan with explosion and then checks for if there is enough production capacity. If the capacity is enough, the plan is accepted. Otherwise, it has to be tweaked. MRP does not have the capability to allocate the scarce capacity to products.

Another drawback is self fulfilling prophecy of lead times in MRP. Determining assembly and production lead times is a tricky issue. MRP primarily aims to keep the production going on while efficiency is taken as a secondary objective. Thus, it puts in ample slack time into lead times even for short operations. Later on, these long lead times become standard and short operations really take longer than they should. This can be explained with a industrial psychological point of view: Those expectations that are set low actually decreases the performance.

As a summary, we conclude that MRP is not the ultimate solution for production planning. It must be supported with some decision making mechanisms. Also MRP data and performance levels should be monitored closely to avoid self fulfilling prophecies.

## 8.2 OPT

Optimized Production Technology (OPT) attempts to overcome MRP's insensitivity to capacity. OPT divides resources into two categories: bottleneck resources and nonbottleneck resources. Bottleneck resources are those resources that limit the production volume severely. For example in a serial production system, the machine with the smallest capacity is the bottleneck resource. Once bottleneck resources are identified, aggregate planning is geared entirely for the bottlenecks. In other words, it is assumed there is enough capacity at nonbottleneck resources and they are assigned infinite capacity. Naturally, this approach simplifies the planning process by focusing the attention on a few of the resources. The exact steps of how one obtains an aggregate plan with OPT is propriety information. However, it is general understanding that OPT software uses heavily tested network based heuristics for planning.

The main disadvantage of OPT is the concept of *shifting bottlenecks*. When the production volume and the mix of products are known, we can find the bottlenecks in a system. However, the aggregate planning exercise is done at least over several months and the volume or the mix may change from one week to another. Different volumes or mixes can lead to different bottlenecks. When that happens, we say that the bottleneck is shifting, i.e., it differs from one week to another. It is not clear how OPT handles this situation because it relies on a clearly identified stationary bottleneck.

OPT focuses on bottleneck machines and ignores nonbottlenecks during planning. Thus, OPT provides a plan for a production system that approximates the actual production system. In order for an OPT plan to work, it is necessary to have plenty of nonbottleneck resources. When the cost of nonbottleneck resources are not small, the OPT plan may have a high cost. This restricts the usefulness of OPT for cases when nonbottleneck resources are expensive.

Both MRP and OPT are centralized decision making tools but in SCs many parties take part. The rigid logic of MRP and OPT cannot accommodate multiple decision making parties, each serving its own objective. Even if all players agree on a single objective, centralized decision making leads to overwhelming bureaucracy: inputs and data must be transmitted up to a single decision maker and each decision must be transmitted down to interested parties. These transmissions increase the likelihood of error occurrences. Moreover, since the parties and the decision maker are substantially apart, the decision maker cannot detect errors in the data. Similarly, the parties cannot detect the flaws in the decisions. Transmissions also delay the decision making process, that is decision can be made much faster locally.

### 8.3 JIT

JIT (just-in-time) differs from MRP and OPT in its philosophy of decentralization. Decisions and improvements are made locally. For example, JIT proponents argue that workers know enough about their job to make correct decisions about it. Industrial psychologists also support JIT by saying that worker empowerment increases the worker performance. Naturally, passing decisions to workers, decreases the coordination among different groups of workers.

The coordination among different units is achieved with Kanban system. Kanban is basically a work order that generally travels upstream in the SC. It indicates what is to be done (produced) in what amount (number of units). The Kanban's are driven from MPS. The important distinction between JIT and MRP or OPT is that no production takes place at a workcenter before a Kanban card reaches that work center. Therefore, JIT works in the pull mode whereas MRP and OPT are more of push types.

Some of the JIT aspects are given below:

- Set up time/cost reduction by taking set up time out of the production cycle, modifying jigs and fixtures, etc.
- Cycle time, lot size reduction and higher customer responsiveness.
- Quality improvement. Concepts such as total quality management has come from these efforts.
- Improving supplier relations by reducing the number of suppliers and strengthening the communication and coordination with the remaining suppliers.
- Process improvement is encouraged. Especially factory workers are encouraged to report their suggestions and prize money is given to good suggestions.

JIT philosophy is quite different from production planning and inventory control in that it attempts to remove problems rather than planning around them. Quality specialists term this as a *root cause* approach. They would like to find the real problem that causes the symptoms. For example, unreliable suppliers can be a result of wrong choice of suppliers or the lack of coordination with the suppliers. In this case, JIT will attempt to increase supplier reliability whereas an inventory control based approach would study the best plan for the given levels of unreliability. Perhaps the success of JIT is due to its nature of modifying the rules of the game rather than playing with these rules.

Discussion Questions:

1. Put the following industries in an order of the highest benefactors of MRP application to the lowest: Pharmaceutical, Catering, Automobile, Semiconductor, Health. Briefly explain how you came up with the order.
2. Remember the last time you have been to a Taco Bell restaurant. Do you think Taco Bell uses a modular BOM, explain why by giving examples.
3. Suppose that we are TexBag's supplier of Leather Straps and Hooks (see Figure 5). Further suppose that the leather strap and hooks are our only products and we need 4 labor hours and 1 labor hour to manufacture 1 leather strap and 1 hook respectively.
  - a) Suppose that TexBag is our sole customer, aggregate our two products into one generic product so that we can handle labor force planning problem as a single product problem. How many labor hours does the generic product require?
  - b) Suppose that we have another customer than TexBag called ArkBag which buys only leather straps. Can we still stick with our answer to (a)? If not, do we have enough information to modify our answer to (a)? If not, what additional information is needed?

## References

- [1] J.R. Birge and F. Louveaux (1997). Introduction to Stochastic Programming. Springer-Verlag, New York.
- [2] S.E. Bodily and J.R. Freeland (1988). A simulation of techniques for forecasting shipments using firm orders-to-date. Journal of Operational Research Society Vol.39 No.9: 833-846.
- [3] V.M. Guerrero and J.A. Elizondo (1997). Forecasting a cumulative variable using its partially accumulated data. Management Science Vol43. No.6: 879-889.
- [4] W.C. Jordan and S.C. Graves (1991). Principles on the benefits of manufacturing process flexibility. Technical report GMR-7310, General Motors Research Laboratories, Warren, MI.
- [5] A.S. Manne (1961). Capacity expansion and probabilistic growth. Econometrica 29: 632-649.