

Forecasting for Supply Chain and Portfolio Management

Katta G. Murty

Department of Industrial and Operations Engineering

University of Michigan

Ann Arbor, MI 48109-2117, USA

Phone: 734-763-3513, fax: 734-764-3451

e-mail: murty@umich.edu

Webpage: <http://www-personal.engin.umich.edu/~murty/>

November 2006

Abstract

Material imbalances at some companies have been traced to the procedures they use for forecasting demand based on the usual normality assumption. In this paper we discuss a simple and easy to implement nonparametric technique to forecast the demand distribution based on statistical learning, and ordering policies based on it, that are giving satisfactory results at these companies. We also discuss an application of this nonparametric forecasting method to portfolio management.

Key words: Forecasting demand, updating demand distribution, nonparametric method, overage and underage costs, order quantity determination, news-vendor approach; returns from investment, risk, portfolio management and optimization, statistical learning.

1 Introduction

In production planning projects at computer companies (Dell, Sun), filter making companies (Pall), automobile component suppliers (Borg Warner, Federal Mogul), and others, we found that high inventories for some items, and expedited shipments to cover shortages for some others, are common occurrences at some of them. Examination of the materials requirement planning (MRP) systems used for making production and order quantity decisions at these companies has shown that a common cause for these occurrences are the procedures they use for forecasting demand based on the usual normality assumption. This paper discusses the features of a new, simpler nonparametric forecasting method based on statistical learning, and ordering and lot sizing policies based on it, implemented and working satisfactorily at these companies.

We also discuss an application of this nonparametric forecasting method to portfolio management. We then develop a model based on the principles of statistical learning to determine an optimum portfolio WRT (with respect to) a measure of risk that is closer to the common investors perception of risk.

2 Costs of High Inventories and Shortages

Models for controlling and replenishing of inventories have the aim of determining order quantities to minimize the sum of total overage costs (Costs of excess inventory remaining at the end of the planning period), and underage costs (shortage costs, or costs of having less than the desired amount of stock at the end of the planning period).

In inventory control literature, the total overage (underage) cost is usually assumed to be proportional to the overage (shortage) amount or quantity, to make the analysis easier. But in some of the companies we were told that a piecewise linear (PL) function provides a much closer representation of the true overage and underage costs. In these companies there is a buffer with limited space in which excess inventory at the end of the planning period can be stored and retrieved later at a low cost (i.e., with minimum requirements of manhours needed) per unit. Once this buffer is filled up, any remaining excess quantity has to be held at a location farther away that requires greater number of manhours for storing or retrieval/unit. Similar situation exists for underage cost as a function of the shortage amount. This clearly implies that the overage and underage costs are PL functions of the excess, shortage quantities. Determining optimum order quantities to minimize such unusual overage, underage cost functions is much harder with existing inventory control models using forecasting techniques in current literature.

An advantage of the new forecasting system discussed in this paper is that such unusual overage, underage cost functions can easily be accommodated under it.

3 Commonly Used Techniques for Forecasting Demand

In almost all inventory management problems in practice the demand during a future planning period is a random variable with an unknown probability distribution, and the models for these problems have the objective of minimizing the sum of expected overage and underage costs. Successful inventory management systems depend heavily on good demand forecasts to provide data for inventory replenishment decisions.

The output of forecasting is usually presented in the literature as the **forecasted demand quantity**, in reality it is an estimate of the expected demand during the planning period. Because of this, the purpose of forecasting is often

misunderstood to be that of generating this single number, even though sometimes the standard deviation of demand is also estimated. All commonly used forecasting methods are parametric methods, they usually assume that demand is normally distributed, and update its distribution by updating the parameters of the distribution, the mean μ , and the standard deviation σ . The most commonly used methods for updating the values of the parameters are the method of moving averages, and the exponential smoothing method.

The method of moving averages uses the average of n most recent observations on demand as the forecast for the expected demand for the next period. n is a parameter known as the *order of the moving average method* being used, typically it is between 3 to 6 or larger.

The exponential smoothing method introduced and popularized by Brown [2], is perhaps the most popular method in practice. It takes \hat{D}_{t+1} , the forecast of expected demand during next period $t + 1$, to be $\alpha x_t + (1 - \alpha)\hat{D}_t$, where x_t is the observed demand during current period t , \hat{D}_t is the forecasted expected demand for current period t , and $0 < \alpha \leq 1$ is a *smoothing constant* which is the relative weight placed on the current observed demand. Typically values of α between 0.1 and 0.4 are used, and the value of α is increased whenever the absolute value of the deviation between the forecast and observed demand exceeds a tolerance times the standard deviation. Smaller values of α (like 0.1) yield predicted values of expected demand that have a relatively smooth pattern, whereas higher values of α (like 0.4) lead to predicted values exhibiting significantly greater variation, but doing a better job of tracking the demand series. Thus using larger α makes forecasts more responsive to changes in the demand process, but will result in forecast errors with higher variance.

One disadvantage of both the method of moving averages and the exponential smoothing method is that when there is a definite trend in the demand process (either growing, or falling), the forecasts obtained by them lag behind the trend. Variations of the exponential smoothing method to track trend linear in time in the demand process have been proposed (see Holt [4]), but these have not proved very popular.

There are many more sophisticated methods for forecasting the expected values of random variables, for example the Box-Jenkins ARIMA models [1, 6], but these methods are not popular for production applications, in which forecasts for many items are required.

4 Parametric Methods for Forecasting Demand Distribution

4.1 Using Normal Distribution With Updating of Expected Value and standard Deviation in Each Period

As discussed in the previous section, all forecasting methods in the literature only provide an estimate of the expected demand during the planning period. The optimum order quantity to be computed depends of course on the entire

probability distribution of demand, not just its expected value. So, almost everyone assumes that the distribution of demand is the normal distribution because of its convenience. One of the advantages that the normality assumption confers is that the distribution is fully characterized by only two parameters, the mean and the standard deviation, both of whom can be very conveniently updated by the exponential smoothing or the moving average methods.

Let t be the current period; x_r the observed demand, and $\hat{D}_r, \hat{\sigma}_r$ the forecasts (i.e., estimates by whichever method is being used for forecasting) of the expected demand, standard deviation of demand respectively in period r ; for $r \leq t$. Then these forecasts for the next period $t + 1$ are:

$$\begin{aligned} \hat{D}_{t+1} \text{ (by method of moving averages of order } n) &= \frac{1}{n} \sum_{r=t-n+1}^t x_r \\ \hat{D}_{t+1} \text{ (by exponential smoothing method with smoothing constant } \alpha) &= \alpha x_t + (1 - \alpha) \hat{D}_t \\ \hat{\sigma}_{t+1} \text{ (by method of moving averages of order } n) &= + \sqrt{(\sum_{r=t-n+1}^t (x_r - \hat{D}_{t+1})^2) / n} \end{aligned}$$

To get $\hat{\sigma}_{t+1}$ by the exponential smoothing method, it is convenient to use the mean absolute deviation (MAD), and use the formula: standard deviation $\sigma \approx (1.25)\text{MAD}$ when the distribution is the normal distribution. Let MAD_t denote the estimate of MAD for current period t . Then the forecasts obtained by the exponential smoothing method with smoothing parameter α for the next period $t + 1$ are:

$$\begin{aligned} \text{MAD}_{t+1} &= \alpha |x_t - \hat{D}_t| + (1 - \alpha) \text{MAD}_t \\ \hat{\sigma}_{t+1} &= (1.25) \text{MAD}_{t+1} \end{aligned}$$

Usually $\alpha = 0.1$ is used to ensure stability of the estimates. And the normal distribution with mean \hat{D}_{t+1} , and standard deviation $\hat{\sigma}_{t+1}$ is taken as the forecast for the distribution of demand during the next period $t + 1$ for making any planning decisions under this procedure.

4.2 Using Normal Distribution With Updating of Expected Value and Standard Deviation Only When There Is Evidence Of Change

In some applications, the distribution of demand is assumed to be the normal distribution, but estimates of its expected value and standard deviation are left unchanged until there is evidence that their values have changed. Foote [3] discusses several statistical control tests on demand data being generated over time to decide when to reestimate these parameters. Under this scheme, the method of moving averages is commonly used to estimate the expected value

and the standard deviation from recent data whenever the control tests indicate that a change may have occurred.

4.3 Using Distributions Other Than Normal

In a few special applications in which the expected demand is low (i.e., the item is a slow-moving item) other distributions like the poisson distribution are sometimes used, but by far the most popular distribution for making inventory management decisions is the normal distribution because of its convenience, and because using it has become a common practice historically.

For the normal distribution the mean is the mode (i.e., the value associated with the highest probability), and the distribution is symmetric around this value. If histograms of observed demand data of an item do not share these properties, it may indicate that the normal distribution is a poor approximation for the actual distribution of demand, in this case order quantities determined using the normality assumption may be far from being optimal.

These days industrial environment is very competitive with new products replacing the old periodically due to rapid advancements in technology. In this dynamic environment, the life cycles of components and end products are becoming shorter. Beginning with the introduction of the product, its life cycle starts with a *growth period* due to gradual market penetration of the product. This is followed by a *stable period* of steady demand. It is then followed by a final *decline period* of steadily declining demand, at the end of which the item disappears from the market. Also, the middle stable period seems to be getting shorter for many major components. Because of this constant rapid change, it is necessary to periodically update demand distributions based on recent data.

The distributions of demand for some components are far from being symmetric around the mean, and the skewness and shapes of their distributions also seem to be changing over time. Using a probability distribution like the normal defined by a mathematical formula involving only a few parameters, it is not possible to capture changes taking place in the shapes of distributions of demand for such components. This is the disadvantage of existing forecasting methods based on an assumed probability distribution. Our conclusions can be erroneous if the true probability distribution of demand is very different from the assumed distribution.

Nonparametric methods use statistical learning, and base their conclusions on knowledge derived directly from data without any unwarranted assumptions. In the next section we discuss a nonparametric method for forecasting the entire demand distribution that uses the classical empirical probability distribution derived from the relative frequency histogram of time series data on demand. It has the advantage of being capable of updating all changes occurring in the probability distribution of demand, including those in the shape of this distribution.

Then in the following section we illustrate how optimal order quantities that optimize piecewise linear and other unusual cost functions discussed in Section 2 can be easily computed using these empirical distribution.

5 A Nonparametric Method for Updating and Forecasting the Entire Demand Distribution

In supply chain management, the important random variables are daily or weekly (or whatever planning period is being used) demands of various items (raw materials, components, sub-assemblies, finished goods, spare parts, etc.) that companies either buy from suppliers, or sell to their customers. Observed values of these random variables in each period are generated automatically as a time series in the production process, and are usually available in the production data bases of companies. In this section we discuss a simple nonparametric method for updating changes in the probability distributions of these random variables using this data directly.

Empirical Distributions and Probability Density Functions

The concept of the probability distribution of a random variable evolved from the ancient practice of drawing histograms for the observed values of the random variable. The observed range of variation of the random variable is usually divided into a convenient number of value intervals (in practice about 10 to 25) of equal length, and the relative frequency of each interval is defined to be the proportion of observed values of the random variable that lie in that interval. The chart obtained by marking the value intervals on the horizontal axis, and erecting a rectangle on each interval with its height along the vertical axis equal to the relative frequency is known as the relative frequency histogram of the random variable, or its discretized probability distribution. The relative frequency in each value interval I_i is the estimate of the probability p_i that the random variable lies in that interval, see Figure 1 for an example.

Let I_1, \dots, I_n be the value intervals with u_1, \dots, u_n as their midpoints, and $p = (p_1, \dots, p_n)$, the probability vector in the discretized probability distribution of the random variable. Let $\bar{\mu} = \sum_{i=1}^n u_i p_i$, $\bar{\sigma} = \sqrt{\sum_{i=1}^n p_i (u_i - \bar{\mu})^2}$.

Then $\bar{\mu}, \bar{\sigma}$ are estimates of the expected value μ , standard deviation σ of the random variable respectively.

We will use the phrase **empirical distribution** to denote such a discretized probability distribution of a random variable, obtained either through drawing the histogram, or by updating a previously known discretized probability distribution based on recent data.

When mathematicians began studying random variables from the 16th century onwards, they found it convenient to represent the probability distribution of the random variable by the **probability density function** which is the mathematical formula for the curve defined by the upper boundary of the relative frequency histogram in the limit as the value interval length is made to approach 0, and the number of observed values of the random variable goes to infinity. So the probability density function provides a mathematical formula for the height along the vertical axis of this curve as a function of the variable rep-

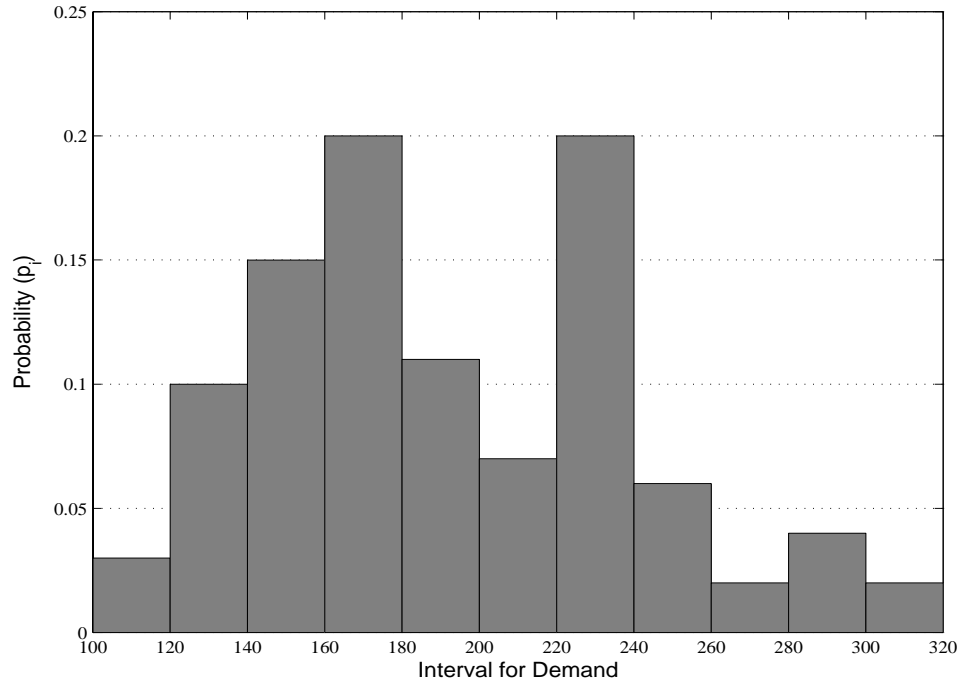


Figure 1: Relative frequency histogram for daily demand for a major component at a PC assembling plant in California.

resented on the horizontal axis. Because it is a mathematically stated function, the probability density function lends itself much more nicely into mathematical derivations than the somewhat crude relative frequency histogram.

It is rare to see empirical distributions used in decision making models these days. Almost everyone uses mathematically defined density functions characterized by a small number of parameters (typically two or less) to represent probability distributions. In these decision making models, the only freedom we have in incorporating changes is to change the values of those parameters. This may be inadequate to capture all dynamic changes occurring in the shapes of probability distributions from time to time.

Extending the Exponential Smoothing Method to Update the Empirical Probability Distribution of a Random Variable

We will now see that representing the probability distributions of random variables by their empirical distributions gives us unlimited freedom in capturing

any type of change including changes in shape, Murty [7].

Let I_1, \dots, I_n be the value intervals, and p_1, \dots, p_n the probabilities associated with them in the present empirical distribution of a random variable. In updating this distribution, we have the freedom to change the values of all the p_i , this makes it possible to capture any change in the shape of the distribution.

Changes, if any, will reflect in recent observations on the random variable. Following table gives the present empirical distribution, histogram based on most recent observations on the random variable (for example most recent k observations where k could be about 30), and x_i to denote the probabilities in the updated empirical distribution to be determined.

Value interval	Probability vector in the		
	Present empirical distribution	Recent histogram	Updated empirical distribution (to be estimated)
I_1	p_1	f_1	x_1
\vdots	\vdots	\vdots	\vdots
I_n	p_n	f_n	x_n

$f = (f_1, \dots, f_n)$ represents the estimate of the probability vector in the recent histogram, but it is based on too few observations. $p = (p_1, \dots, p_n)$ is the probability vector in the empirical distribution at the previous updating. $x = (x_1, \dots, x_n)$, the updated probability vector, should be obtained by incorporating the changing trend reflected in f into p . In the theory of statistics the most commonly used method for this incorporation is the weighted least squares method ([7]), which provides the following model (1) to compute x from p and f . In it, β is a weight between 0 and 1, similar to the smoothing constant α in the exponential smoothing method for updating the expected value (like that α there, here β is the relative weight placed on the probability vector from the histogram composed from recent observations).

$$\begin{aligned}
 & \text{minimize} (1 - \beta) \sum_{i=1}^n (p_i - x_i)^2 && + \beta \sum_{i=1}^n (f_i - x_i)^2 \\
 & \text{subject to} && \sum_{i=1}^n x_i = 1 \\
 & && x_i \geq 0, \quad i = 1, \dots, n
 \end{aligned} \tag{1}$$

x is taken as the optimum solution of this convex quadratic program. $\beta = 0.1$ to 0.4 works well, the reason for choosing this weight for the second term in the objective function to be small is because the vector f is based on only a small number of observations. Since the quadratic model minimizes the weighted sum of squared forecast errors over all value intervals, when used periodically, it has the effect of tracking gradual changes in the probability distribution of the

random variable.

The above quadratic program has a unique optimum solution given by the following explicit formula.

$$x = (1 - \beta)p + \beta f \tag{2}$$

So we take the updated empirical distribution to be the one with the probability vector given by (2).

The formula (2) for updating the probability vector in the above formula is exactly analogous to the formula for forecasting the expected value of a random variable using the latest observation in exponential smoothing. Hence the above formula can be thought of as the extension of the exponential smoothing method to update the probability vector in the empirical distribution of the random variable .

When there is a significant increase or decrease in the mean value of the random variable, new value intervals may have to be opened up at the left or right end. In this case the probabilities associated with value intervals at the other end may become very close to 0, and these intervals may have to be dropped from further consideration at that time.

This procedure can be used to update the discretized demand distribution either at every ordering point, or periodically at every r th ordering point for some convenient r , using the most recent observations on demand.

6 An Application of the Forecasting Method of Section 5 for Computing Optimal Order Quantities

Given the empirical distribution of demand for the next period, the well known newsvendor model [5, 8, 9] can be used to determine the optimal order quantity

$I_i =$ interval for demand	Probability p_i	$u_i =$ mid-point of interval
100 – 120	0.03	110
120 – 140	0.10	130
140 – 160	0.15	150
160 – 180	0.20	170
180 – 200	0.11	190
200 – 220	0.07	210
220 – 240	0.20	230
240 – 260	0.06	250
260 – 280	0.02	270
280 – 300	0.04	290
300 – 320	0.02	310

for that period that minimizes the sum of expected overage and underage costs very efficiently, numerically. We will illustrate with a numerical example. Let the empirical distribution of demand (in units) for next period be the one given above. The expected value of this distribution $\bar{\mu} = \sum_i u_i p_i = 192.6$ units, and its standard deviation $\bar{\sigma} = \sqrt{\sum_i (u_i - \bar{\mu})^2 p_i} = 47.4$ units.

Let us denote the ordering quantity for that period, to be determined, by Q , and let d denote the random variable that is the demand during that period. Then

y = overage quantity in this period = amount remaining after the demand is completely fulfilled = $(Q - d)^+ = \text{maximum}\{0, Q - d\}$

z = underage quantity during this period = unfulfilled demand during this period = $(Q - d)^- = \text{maximum}\{0, d - Q\}$.

Suppose the overage cost $f(y)$, is the following piecewise linear function of y

Overage amount = y	Overage cost $f(y)$ in \$	Slope
$0 \leq y \leq 30$	$3y$	3
$30 \leq y$	$90 + 10(y - 30)$	10.

Suppose the underage cost $g(z)$ in \$, is the fixed cost depending on the amount given below

Underage amount = z	Underage cost $g(z)$ in \$
$0 \leq z \leq 10$	50
$10 < z$	150.

To compute $E(Q)$ = the expected sum of overage and underage costs when the order quantity is Q , we assume that the demand value d is equally likely to be anywhere in the interval I_i with probability p_i . This implies for example that the probability that the demand d is in the interval 120 – 125 is = (probability that d lies in the interval 120 – 140)/4 = $(0.10)/4 = 0.025$.

Let $Q = 185$. When the demand d lies in the interval 120 – 140, the overage amount varies from 65 to 45 and the overage cost varies from \$440 to 240 linearly. So the contribution to the expected overage cost from this interval is $0.10(440 + 240)/2$.

Demand lies in the interval 140 – 160 with probability 0.15. In this interval the overage cost is not linear, but it can be partitioned into two intervals 140 – 155 (with probability 0.1125), and 155 – 160 (with probability 0.0375) in each of which the overage cost is linear. In the interval $140 \leq d \leq 155$ the overage cost varies linearly from \$240 to \$90; and in $155 \leq d \leq 160$ the overage cost varies linearly from \$90 to \$75. So, the contribution to the expected overage cost from this interval is $\$(0.115 (240 + 90)/2) + (0.0375(90 + 75)/2)$.

Proceeding this way we see that $E(Q)$ for $Q = 185$ is: $\$(0.03 (640+440)/2) + (0.10(440 + 240)/2) + [(0.115 (240 + 90)/2) + (0.0375(90 + 75)/2)] + (0.20(75$

$+ 15)/2) + [0.0275(15 + 0)/2) + 0.055(50) + 0.0275 (150)] + (0.07 + 0.20 + 0.06 + 0.02 + 0.04 + 0.02)150 = \$ 140.87$.

In the same way we computed the values of $E(Q)$ for different values of Q spaced 5 units apart, given below.

Q	$E(Q)$
195	178.00
190	162.27
185	143.82
180	139.15
175	130.11
170	124.20
165	120.40
160	121.95
155	122.60
150	124.40
145	139.70

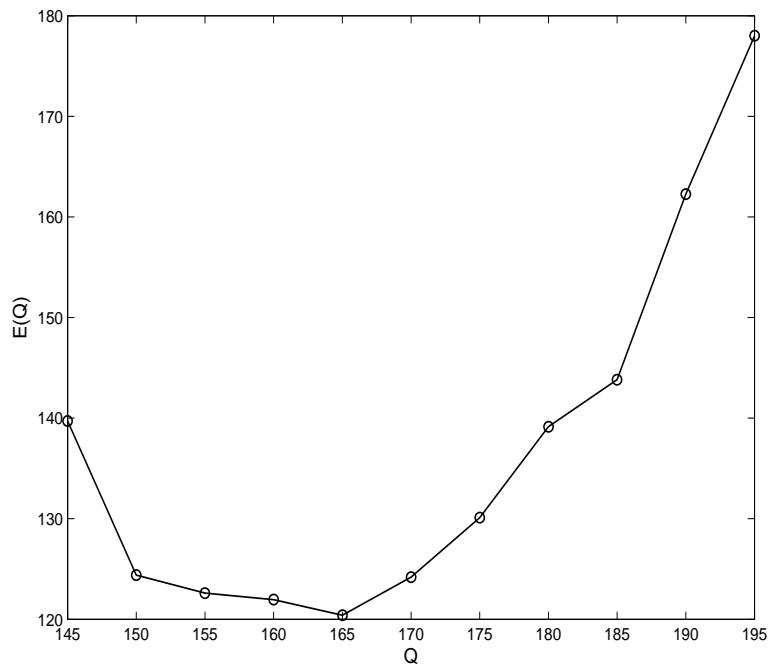


Figure 2: Plot of $E(Q)$ for various values of Q .

Figure 2 is a plot of these values of $E(Q)$. Here we computed $E(Q)$ at values of Q which are multiples of 5 units, and it can be seen that $Q = 165$ is the

optimum order quantity correct to the nearest multiple of 5. If the optimum is required to greater precision, the above calculation can be carried out for values of Q at integer (or closer) values between 150 to 170 and the best value of Q there chosen as the optimum order quantity.

The optimum value of Q can then be translated into the actual order quantity for the next period by subtracting the expected on-hand inventory at the end of the present period from it.

For each i assuming that demand d is equally likely to be anywhere in the interval I_i with probability p_i , makes the value of $E(Q)$ computed accurate for each Q . However, in many applications people make the simpler assumption that p_i is the probability of demand being equal to u_i , the midpoint of the interval I_i . The values of $E(Q)$ obtained with this assumption will be approximate, particularly when the overage and underage costs are not linear (i.e., when they are piecewise linear etc.); but this assumption makes the computation of $E(Q)$ much simpler, that's why people use this simpler assumption.

7 How to Incorporate Seasonality in Demand Into the Model

The discussion so far dealt with the case when the values of demand in the various periods form a stationary time series. In some applications this series may be seasonal, i.e., it has a pattern that repeats every N periods for some known value of N . The number of periods N , before the pattern begins to repeat is known as the length of the season. In order to use seasonal models, the length of the season must be known.

For example, in the computer industry majority of sales are arranged by sales agents who operate on quarterly sales goals. That's why demand for components in the computer industry, and demand for their own products tends to be seasonal with the quarter of the year as the season. The sales agents usually work much harder in the last month of the quarter to meet their quarterly goals, so demand for products in the computer industry tends to be higher in the third month of each quarter than in the beginning two months. As most of the companies are building to order now-a-days, weekly production levels and demands for components inherit the same kind of seasonality.

At one company in this industry each quarter is divided into three homogeneous intervals. Weeks 1 to 4 of the quarter are slack periods, each of these weeks accounts a fraction of about 0.045 of the total demand in the quarter. Weeks 5 to 8 are medium periods, each of these weeks accounts for a fraction of about 0.074 of the total demand in the quarter. Weeks 9 to 13 are peak periods, each of these weeks accounts for a fraction of about 0.105 of the total demand in the quarter. This fraction of demand in each week of the season is called the *seasonal factor* of that week.

In the same way in the paper industry demand for products exhibits seasonality with each month of the year as the season. Demand for their products

in the 2nd fortnight in each month tends to be much higher than in the 1st fortnight.

There are several ways of handling seasonality. One way is for each $i = 1$ to N (= length of the season), consider demand data for the i th period in each season as a time series by itself, and make the decisions for this period in each season using this series based on methods discussed in earlier sections.

Another method that is more popular is based on the assumption that there exists a set of indices c_i , $i = 1$ to N called *seasonal factors* or *seasonal indices* (see [8, 9]), where c_i represents the demand in the i th period of the season as a fraction of the demand during the whole season (as an example see the seasonal factors given for the computer company described above). Once these seasonal factors are estimated, we divide each observation of demand in the original demand time series by the appropriate seasonal factor to obtain the deseasonalized demand series. The time series of deseasonalized demand amounts still contains all components of information of the original series except for seasonality. Forecasting is carried out using the methods discussed in the earlier sections, on the deseasonalized demand series. Then estimates of the expected demand, standard deviation, and the optimal order quantities obtained for each period must be reseasonalized by multiplying by the appropriate seasonal factor before being used.

8 Portfolio Management

One of the most important and very widely studied problems in finance is that of optimizing the return from investments. Everyone in this world from little individual investors to Presidents and CEOs of very large corporations with annual incomes ranging to hundreds of millions of dollars; all the banks, mutual funds, and other financial institutions have great interest in this problem.

There are many different investment opportunities, but the return (also called yield, which may be positive, 0, or negative) from each varies from period to period as a random variable. The area is “data rich” in the sense that the return per unit investment in each investment opportunity in each period in the past is freely available as a time series and can be accessed by anyone.

One important problem in the area is: given a budget B (amount of money available to invest) and a list of investment opportunities $1, \dots, n$ to invest in; how to optimally divide the budget among the various investment opportunities. Once invested, that investment may be kept for several periods, and the returns from it keep accumulating period to period as long as the investment is kept. So, the important feature is that the return from investment is not in a single installment, but paid out in each period over the life of the investment. In applications, n , the number of investment opportunities under consideration, tends to be large.

Denoting by the decision variable x_i the amount of the budget allocated to investment opportunity i , for $i = 1$, to n ; the vector $x = (x_1, \dots, x_n)^T$, called a **portfolio**, is a solution for the problem. The goal of portfolio optimization, is

to characterize and find an optimum portfolio.

Let the random variable $P(x)$ denote the total return from portfolio x in a period. For each period in the past we can compute $P(x)$ using x and the available data on returns from individual investment opportunities. So, for any x , $P(x)$ can be generated as a time series. Let $\mu(x)$ = the expected value of the return $P(x)$ in a period from portfolio x .

$P(x)$ varies randomly, this variation is perceived as the **risk** (or **volatility of returns**) associated with portfolio x . Everyone agrees with treating variation in returns from period to period as a risk, but there is not a universal agreement on how to measure this risk. We will use the symbol $r(x)$ to denote this risk as a function of x . The two most important parameters for characterizing an optimum portfolio are:

$\mu(x) = E(P(x))$, the expected value of the return $P(x)$ in a period, it is a measure of the long term average return per period from portfolio x ,

$r(x)$ = a measure of risk associated with portfolio x , a suitable measure is to be selected.

There is universal agreement that an optimum portfolio should maximize $\mu(x)$. In fact some investors select a portfolio x and keep it for a long time. For such investors the period to period variation in the return $P(x)$ may not be that critical, they mainly want to see $\mu(x)$ maximized.

But the majority of investors (particularly large investors like banks, mutual funds etc.) change their portfolio periodically by selling some investments made in earlier periods at current prices, or by investing additional amounts. For these investors the period to period variation in the return is also an important factor to take into consideration. These investors not only want to maximize the long term average return, but would also like to keep the return in every period as high as possible. So, from their perspective, an optimum portfolio should maximize the expected return $\mu(x)$, and minimize the risk $r(x)$; i.e., it should achieve both these objectives simultaneously. So finding an optimum portfolio here is a **multi-objective optimization problem**.

But in multi-objective optimization, there is no concept of “optimality” that has universal acceptance. Also, the two objectives typically conflict with each other; i.e., portfolios that maximize expected return $\mu(x)$ are usually associated with high values for what ever measure $r(x)$ is chosen to represent risk.

Usually the various investment opportunities are partitioned into various sectors by their type (for example utility opportunities, banking opportunities, etc.). Then the decision makers usually impose lower and upper bounds on the amount of the budget that can be invested among investment opportunities in each sector, and may be some other linear constraints also. Suppose the system of all these constraints including the budget constraint is (here e is the column vector of all 1s in R^n)

$$\begin{aligned}
Ax &\leq b \\
e^T x &\leq B \\
x &\geq 0
\end{aligned}
\tag{3}$$

A portfolio x is said to be a *feasible portfolio* if it satisfies all the constraints in (3). Once a measure $r(x)$ for risk is selected, if x, \bar{x} are two feasible portfolios satisfying: either $\mu(x) > \mu(\bar{x})$ and $r(x) \leq r(\bar{x})$, or $\mu(x) \geq \mu(\bar{x})$ and $r(x) < r(\bar{x})$; then x is said to *dominate* \bar{x} , because it is better or the same as \bar{x} WRT (with respect to) both the objective functions in the problem, and strictly better on at least one of the two objectives.

A feasible portfolio \bar{x} is said to be a *nondominated portfolio* or *efficient portfolio* or *pareto optimum portfolio* if there is no other feasible portfolio that dominates it. In multi-objective optimization problems like this one, there is no concept of “optimality” that has universal acceptance, but clearly no investor would like a portfolio that is dominated by another one. So, we should look among efficient Portfolios for a solution to the problem. But usually there are many efficient portfolios, the set of all of them is called the *efficient frontier*.

Mathematicians would consider a multi-objective problem well solved if an algorithm is developed to enumerate the efficient frontier in a computationally efficient way. Here I can mention the entertaining Hollywood movie *A Beautiful Mind* based on the life of John Nash who received the Nobel Prize in economics in 1994 for proving that a certain type of two-objective optimization problems always have at least one efficient solution.

In a pair of efficient portfolios, if the 1st is better than the 2nd WRT the average return $\mu(x)$, then the 2nd will be better than the 1st WRT the risk function $r(x)$, so the best portfolio among these two is not defined. Given a feasible portfolio that is not efficient, an efficient portfolio better than it can be found; but there is no universally acceptable criterion for selecting the best among efficient portfolios. The challenge in portfolio optimization is to select a good measure for “risk”, and obtain a good portfolio that has satisfactory values for both the objective functions.

Besides portfolio optimization, portfolio management deals with the issues of determining how long an optimum portfolio determined should be kept, and the appropriate tools for tracking its performance while it is kept. Changing the current portfolio and adopting a new one in every period is a very labor-intensive and expensive process, that’s why once an optimum portfolio is determined in some period, most investors do not like to change it as long as it is performing upto expectations.

9 Variance as a Measure of Risk, And the Markovitz Model for Portfolio Optimization

Most of the work in finance is based on the assumption that the yields in a period from unit investments in the various investment opportunities follow a multivariate normal distribution. Let the vector of expected values in this distribution be $\mu = (\mu_1, \dots, \mu_n)^T$, and let the variance-covariance matrix in it be the symmetric positive definite matrix $\Sigma = (\sigma_{ij})$ of order n .

Then $\mu(x) =$ the expected return from portfolio x in a period is $\mu^T x$, and the variance of this return is $x^T \Sigma x$.

In statistical theory, the variance of a random variable is a well accepted measure of variation of this random variable. Since the “risk” of a portfolio stems from the variation in the returns from it from period to period, Harry Markovitz proposed in 1952 using the variance $x^T \Sigma x$ of returns from the portfolio x as the measure $r(x)$ of risk associated with it under the normality assumption. He suggested the approach of minimizing this risk function subject to the constraint that the expected fractional return per period must be \geq some specified lower bound δ , to define an “optimum portfolio”. This leads to the following classical *Markovitz portfolio model* for which he won the 1989 Von Neumann theory prize for contributions to OR of INFORMS, and the 1990 Nobel Prize in economics.

$$\begin{aligned} & \text{Minimize } x^T \Sigma x \\ & \text{subject to the feasibility conditions in (3), and } \mu^T x \geq \delta B \end{aligned} \quad (4)$$

where e is a column vector of all 1’s in R^n , and e^T is its transpose. This is a quadratic programming problem, its optimum solution is known as a *minimum variance portfolio*

The minimum variance portfolio does not have universal acceptance as the best portfolio to adopt, since it may not have good practical features. For example, suppose $\delta = 0.07$, and the minimum variance portfolio is a portfolio G with expected return fraction per period of 0.07 and variance of 0.0016. There may be another feasible portfolio H with expected return fraction per period of 0.25 and variance of 0.0018. Portfolio H which is not optimum for this model, yields a higher return than portfolio G with very high probability in every period and is definitely more desirable. Also, as pointed out in Papahristodoulou and Dotzauer [19], many investors and traders as well question whether the variance of the return $x^T \Sigma x$ is an appropriate measure of risk; and many researchers question whether the assumption that the returns from individual investment opportunities follow a multivariate normal distribution is reasonable.

Another problem with this model deals with the computational difficulties in solving it. The distribution of returns may be changing with time, and updating the distribution requires re-computation of the variance-covariance matrix using recent data at frequent intervals, an expensive operation when n is large. Also, the variance-covariance matrix will be fully dense, this makes the model (4) computationally difficult to handle if n is large.

10 Other Measures of Risk for Portfolio Optimization

While everyone perceives variation in the returns as an element of risk, no one complains if the variation is taking the returns higher; but they will definitely complain when they begin to decrease. That is, investor's reaction to the two types of variation are highly asymmetric. For this reason the variance of the return used as the measure of risk in the classical Markovitz model is not a fully appropriate measure of risk.

Several other measures of risk of a portfolio have been proposed in the literature [10 to 16, 19, 21 to 25]. We will use as a measure of risk of a portfolio x , the

probability $d(x, \delta)$ that the return $P(x)$ from it in a period is $\leq \delta e^T x$, where δ is a minimum return per unit investment per period demanded by the investor.

A good portfolio should either have as one of its objectives minimizing this risk measure $d(x, \delta)$, or keeping it \leq some specified upper limit γ for it. This measure is closely related to the Value-at-Risk (VaR) measure, and other downside risk measures and the safety-first conditions studied in [10, 12, 13, 15, 21, 22].

The expected return per period $\mu(x)$ of a portfolio x is a measure of the long term benefit of adopting it; because it measures the average return per period one can expect to get from it if the present distribution of returns continues unchanged. One model that we will consider later in Section 12 for defining an optimum portfolio is to maximize $\mu(x)$ subject to the constraint that $d(x, \delta) \leq \gamma$.

11 Portfolio Management: Tracking the Distribution of Return From a Portfolio That is Kept for a Long Time

The literature in finance has many research publications dealing with models for portfolio optimization, and we will discuss one such model based on statistical learning in the next section. But very few research publications deal with portfolio management, which also deals with tracking the performance of the optimum portfolio determined to check whether it is performing to expectation, and deciding when to change the portfolio. In this section we discuss an application of the simple forecasting method discussed in Section 5 to track the performance of the portfolio in current use.

For this, the most important random variables are the per unit return in a period from various investment opportunities. The distributions of these random variables may be changing over time, and unless these changes are incorporated into the decision making process, the selected portfolio may not be a satisfactory one for the next or for any future period.

The distribution of return from a single investment opportunity can be estimated from past data by its discretized probability distribution discussed in Section 5. This discretized probability distribution can be updated over time based on recent data by the technique discussed in Section 5, we will use the phrase “empirical distribution” to denote the updated distribution.

The updating technique discussed in Section 5 is quite convenient for updating the empirical distribution of return from a single investment opportunity, because in this case the probabilities associated with a small number of intervals need to be updated at each updating. But for studying the returns from two or more investment opportunities (2 or more random variables in a dependence relationship), its direct extension becomes unwieldy due to the curse of dimensionality. In the multivariate context, the discretized distribution breaks up the space of the vector of variables into a number of rectangles each with its associated probability. Even when the number of variables is 2, the number of these rectangles is too large, and updating their probabilities becomes impractical.

However we will see that this one-variable technique is itself a useful tool in keeping track of portfolios that are kept for long periods of time.

Suppose an investor likes to keep her/his portfolio \bar{x} unchanged as long as it is performing to his/her expectations. The value of $P(\bar{x})$ in each period can be computed directly from \bar{x} and the available data on the returns from the various investment opportunities, and generated as a time series. Using it, the distribution of $P(\bar{x})$ can be updated over time as explained in Section 5. If the distribution of $P(\bar{x})$ is estimated and maintained in the form of an empirical distribution, the expected value of return from the current empirical distribution, is an estimate of the current expected return from portfolio \bar{x} . Also, since the empirical distribution is a discretized distribution, an estimate of the risk measure $d(\bar{x}, \delta) = \text{probability that the return is } \leq \delta e^T \bar{x}$ in it can be computed very easily. From estimates of expected return, and $d(\bar{x}, \delta)$, the two measures for evaluating a portfolio, the investor can judge whether to continue to keep the portfolio \bar{x} , or look for a better portfolio to change to.

12 A Model Based on Statistical Learning to Find An Optimum Portfolio

Let \bar{x} denote the current portfolio in use.

Under the assumption that the returns from various investment opportunities follow a multivariate normal distribution, the measure of risk $d(x, \delta)$ for any portfolio x is a nonlinear function, and the problem of maximizing the expected return μx subject to the constraint that $d(x, \delta) \leq \gamma$ is a complex problem. Even if the optimum solution of this problem can be determined, since the actual distribution of the returns vector is unknown, it is not clear how good the performance of the resulting portfolio derived from the normality assumption will be in reality.

In statistical learning, instead of making assumptions about the distribution

of the returns vector, we base our decisions on knowledge derived from actual data. We now develop an MIP (mixed integer programming) model for finding an optimum portfolio based on estimates of relevant quantities obtained from actual data over the most recent m periods, for some selected m . The first model ignores the transaction costs of moving from the current portfolio \bar{x} to the optimum portfolio. Let

e_{ij} = actual return from unit investment in the i -th period from the j -th investment opportunity, $i = 1$ to m , $j = 1$ to n

$E = (e_{ij})$, an $m \times n$ matrix of data on actual returns

$E_i = (e_{i1}, \dots, e_{in})$, the i -th row vector of E .

The risk condition $d(x, \delta) \leq \gamma$ translates to the requirement that the constraint $E_i x \geq \delta e^T x$ must hold for at least t periods i , where $t = \text{ceiling of } ((1 - \gamma)m)$, the smallest integer $\geq ((1 - \gamma)m)$. Define the binary variables z_1, \dots, z_m , where $z_i = 0$ if $E_i x \geq \delta e^T x$, 1 otherwise. In terms of these binary variables, the model for finding an optimum portfolio is (5) to (10), here $L > 0$ is a positive number such that $-L$ is a lower bound for each $E_i x - \delta e^T x$.

$$\text{Maximize } \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n e_{ij} x_j \quad (5)$$

$$\text{subject to } \sum_{j=1}^n x_j \leq B \quad (6)$$

$$Ax \leq b \quad (7)$$

$$E_i x - \delta e^T x + L z_i \geq 0, \quad i = 1, \dots, m \quad (8)$$

$$\sum_{i=1}^m z_i \leq m - t \quad (9)$$

$$x_j \geq 0, z_i \in \{0, 1\} \quad \text{for all } i \quad (10)$$

Transaction costs for selling existing investments, or acquiring additional investments can also be taken into account in the model. Assuming that the transaction costs are linear, suppose c_j^+, c_j^- are the costs of acquiring additional unit investment, selling unit investment respectively in investment opportunity j , for $j = 1$ to n . Then the transaction cost for moving from current portfolio \bar{x} to portfolio x is $\sum_{j=1}^n [c_j^+(x_j - \bar{x}_j)^+ + c_j^-(x_j - \bar{x}_j)^-]$, where $(x_j - \bar{x}_j)^+ = \text{Maximum}\{x_j - \bar{x}_j, 0\}$ = additional investment in opportunity j acquired, and $(x_j - \bar{x}_j)^- = \text{Maximum}\{\bar{x}_j - x_j, 0\}$ = investment in opportunity j sold.

Assuming that the transaction cost coefficients c_j^+, c_j^- are all positive, the model for maximizing average return per period – transaction costs is to: Maximize $\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n e_{ij} x_j - \sum_{j=1}^n (u_j^+ c_j^+ + u_j^- c_j^-)$ subject to constraints (6) to (10) and $x_j - \bar{x}_j = u_j^+ - u_j^-$ and $u_j^+, u_j^- \geq 0$ for $j = 1$ to n .

The number of binary variables in either model is m , the number of recent periods considered in the model. Since the distribution of returns may be

changing over time, we will not make m too large anyway; so these models can be solved within reasonable time with existing software systems for MIP. For example, if the period is a week, and weekly return data over the most recent 6-month (26 week) period is used to construct the model, it will have only 26 binary variables, and so is quite easy to solve with software tools available in the market.

Solving the same model with different values of δ, γ generates different portfolios which can be compared with each other and the best among them selected for implementation.

The matrix E of returns is expected to be fully dense. So, when n , the number of investment opportunities considered, is large, the LP relaxations of these models will be dense and may turn out to be hard to solve with existing methods based on matrix inversion operations. New descent methods for LP not based on matrix inversion operations discussed in [18] have the potential to solve such models efficiently when n is large.

13 References

References on Forecasting and Supply Chain Issues

- [1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco, 1970.
- [2] R. G. Brown, *Statistical Forecasting for Inventory Control*, McGraw-Hill, NY, 1959.
- [3] B. L. Foote, "On the Implementation of a Control-Based Forecasting System for Aircraft Spare Parts Procurement", *IIE Transactions* 27(1995)210-216.
- [4] C. C. Holt, "Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages", ONR Memo no. 52, 1957.
- [5] S. Karlin, "Optimal Inventory Policy for the Arrow-Harris-Marschak Dynamic Model", Ch. 9 in *Studies in the Mathematical Theory of Inventory and Production*, K. J. Arrow, S. Karlin, and H. Scarf (eds.), Stanford University Press, 1958.
- [6] D. D. Montgomery and L. A. Johnson, *Forecasting and Time Series Analysis*, McGraw-Hill, St. Louis, 1976.
- [7] K. G. Murty, "Histogram, an Ancient Tool and the Art of Forecasting", Technical report, IOE Dept., University of Michigan, Ann Arbor, 2002.
- [8] S. Nahmias, *Production and Operations Analysis*, 2nd ed., Irwin, Boston, 1993.
- [9] E. A. Silver and R. Peterson, *Decision Systems for Inventory Management and Production Planning*, 2nd. ed., Wiley, NY, 1979.

References on Portfolio Optimization

- [10] F. Andersson, H. Mausser, D. Rosen, and S. Uryasev, "Credit Risk Op-

- timization with Conditional Value-at-Risk Criterion”, *Mathematical Programming B* 89(2001)273-292.
- [11] P. Artzner, F. Delban, J. -M. Eber, and D. Heath, “Coherent Measures of Risk”, *Math. Finance*, 9(1999)203-227.
- [12] S. Benati and R. Rizzi, “A Mixed Integer Linear Programming Formulation of the Optimal Mean/Value-at-Risk Portfolio Problem”, *European Journal of Operational Research*, 176/1(January 2007)423-434.
- [13] A. Cheklov, S. Uryasev, and M. Zabarankin, “Drawdown Measure in Portfolio Optimization”, *International Journal of Theoretical and Applied Finance*, 26/7(2002)1443-1471.
- [14] H. Folmer and A. Schied, *Stochastic Finance*, 2004.
- [15] D. W. Jansen, K. G. Koedijk, and C. G. de Vries, “Portfolio Selection with Limited Downside Risk”, *Journal of Empirical Finance*, 7(2000)247-269.
- [16] H. Konno and H. Yamazaki, “Mean-Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market”, *Management Science*, 37(1991)519-531.
- [17] H. Markovitz, “Portfolio Selection”, *Journal of Finance*, 7(1952)77-91.
- [18] Katta G. Murty, “A New Practically Efficient Interior Point Method for LP”, *Algorithmic Operations Research*, 1(2006)3-19; paper can be seen at the website: <http://journals.hil.unb.ca/index.php/AOR/index>.
- [19] C. Papahristodoulou and E. Dotzauer, “Optimal Portfolios Using Linear Programming Models”, *Journal of the Operational Research Society*, 55(2004)1169-1177.
- [20] A. Perold, “Large-Scale Portfolio Optimization”, *Management Science*, 30(1984)1143-1160.
- [21] R. T. Rockafellar, and S. P. Uryasev, “Optimization of Conditional Value-at-Risk”, *J. of Risk*, 2(2000)21-42.
- [22] A. D. Roy, “Safety-First and the Holding of Assets”, *Econometrica*, 20(1952)431-449.
- [23] M. G. Speranza, “Linear Programming Models for Portfolio Optimization”, *Finance*, 14(1993)107-123.
- [24] B. K. Stone, “A Linear Programming Formulation of the General Portfolio Selection Problem”, *J. Financ. Quant. Anal.*, (1973 September)621-636.
- [25] M. R. Young, “A minimax-Portfolio Selection Rule with Linear Programming Solution”, *Management Science*, 44(1998)673-683.