대한산업공학회/한국경영과학회 2002 춘계공동학술대회
한국과학기술원(KAIST) 2002 년 5월 3 일-4 일

계절유행상품 재고관리를 위한 판매예측
Sales Forecasting for Inventory Control on Seasonal fashion product

안봉근
대구광역시 남구 대명동 2139 계명대학교 경영학부

요약

계절유행상품의 수요는 연중 성수기가 길지 않고 매년 유행과 제품디자인 변화가 심한 경향이 있어 수요예측에 과거의 판매정보의 유용성이 크지 않다. 성수기 초반의 수요가 연간 수요결정에 매우 중요하며 후반부 수요가 급격히 감소하는 특성이 있다. 반면 이월상품의 잔존가치가 매우 낮지만 매출마진이 높아 수요예측의 정확도에 따라 수익률이 큰 영향을 받는다. 이러한 이유로 기존의 수요예측방법을 계절상품에 적용하기에 무리가 따르며 예측오차의 비용이 매우 커서 계절상품 관리에 이용할 수 없다. 본 연구에서 성수기를 하위기간으로 구분하여 시즌 초반부 수요발생시점을 측정하여 초반부 기간별 수요량을 구하고 이를 근거로 기간 누적수요 비율을 quantile regression에 의거 추정하여 기간별 수요량과 전체 수요량을 예측하는 방법을 제시하고 모의자료를 사용하여 이 모형의 우수성을 평가하였다.

T, time of sales is a random variable representing the time (time elapsed) when an item is sold. Note that the sales of seasonal fashion product are highly dependent on the time of the season. The sales pattern in a season can be well represented with a distribution of f(t), pdf (probability density function) of T which belongs to a family of distribution with two parameters (θ_1, θ_2) then F(t), cdf (cumulative density function) of T where

$$F(x \mid \theta_1, \theta_2) = \int_{-\infty}^{x} f(t)dt$$

We are more interested in the number of units sold rather than in the points of time which items are sold on. Note that available data are sales quantity, the numbers of units sold to date the point of time, which is censored to the right. If cdf of T is estimated with the observations available only at the early periods of season, season total sales quantity and the sales quantity of periods to come in the season can be forecasted.

Define following notations.

N : season total sales quantity of the item, which is to be forecasted.

X_k Sales quantity at period (week) k of time interval $[t_{k-1}, t_k]$

S_j : The number of unit sold to date period j or point of time t_j, where

$$S_j = \sum_{i=1}^{j} X_i$$

p_j : Sample sales proportion to the total seasonal sales to date period j where

$$p_j = S_j / N .$$

π_j : Population sales proportion to the total seasonal sales to date period j

$\bar{\pi}_j = F(t_j \mid \bar{\theta}_1, \bar{\theta}_2)$: an estimated value from the cdf with estimated parameters $(\bar{\theta}_1, \bar{\theta}_2)$

$\hat{N}_{(i)} = S_i / \bar{\pi}_i$: Forecasted season total sale made at period i

$\hat{S}_j = \hat{N}\,\bar{\pi}_j$: Forecasted value of S_j

η_p : Population quantile of order p

$\hat{\eta}_p$: sample quantile of order p

Observations on sales S_t are available for periods $j = 1, 2, \cdots, k$. We follows the definition of season by Hertz and Schaffir (1960) that the season begins at period such as $a = \min(j \mid F(t_j) \geq 0.05)$ and ends at period $b = \min(j \mid F(t_j) \geq 0.95)$ Season for an product is defined as beginning in the period in which sales for the product to date comes to 5 per cent of season total, and ending in the week when sales to date reach 95 per cent of season total sales.

1. Relation of Quantile to Distribution Parameters

The family of distribution is assumed to be given based on the historical sales data. The parameters of the distribution may change across seasons but remain fixed in a season. Two parametric distribution models are used to describe the pattern of sales behavior. The family of distributions is versatile enough to covers a wide variety of sales trend and also provide a closed form of solution to estimating parameters of cdf for given quantiles and proportions. Normal, lognormal, Weibull and extreme value distributions satisfy theses requirements. Quantile or log quantile is expressed with function of the proportion or transformed proportion with coefficients of parameters, as are shown in followings.

(1) Normal Distribution

Let η_p be a population quantile of order p and z_p be the pth percentile of standard normal variate. For a normal distribution with mean and standard deviation (μ, σ), η_p can be expressed as follows:

$$\eta_p = \mu + \sigma\ z_p. \qquad (1)$$

(2) Log normal Distribution

cdf of log normal distribution is

$$F(t) = \Phi[(\ln(t) - \mu)/\sigma] \qquad (2)$$

where $\Phi(\cdot)$ is cdf of standard normal distribution and μ and σ are the mean and standard deviation of log transformed time. Then

$$\ln(\eta_p) = \mu + \sigma\ z_p. \qquad (3)$$

(3) Weibull Distribution

cdf of Weibull distribution is

$$F(t) = 1 - \exp[-(\frac{t}{\alpha})^{\beta}] \qquad (4)$$

and the population quantile η_p is determined as

$$\ln(\eta_p) = \ln\alpha + \frac{1}{\beta}\ln(-\ln(1-p)) \qquad (5)$$

(4) Extreme Value Distribution

Unlike to Lognormal and Weibull distributions that are symmetrical or skewed to the right, extreme value distribution is skewed to the left with relatively short tail. The cdf of extreme value is

$$F(t) = 1 - \exp\{-\exp[(t-\gamma)/\delta]\} \qquad (6)$$

and the population quantile η_p is determined as

$$\eta_p = \gamma + \delta\ln(-\ln(1-p)) \qquad (7)$$

2. Estimation of season total sale

Season total sale, N is required to obtain the sample proportion $p_j = S_j/N$, which is used to estimate the parameters of distribution. Hence estimating N is very critical for accuracy of forecasting sales pattern and sales quantity for each period of the season. The season total sale is estimated from the sales quantity to date the period i and estimated population proportion $\pi_i = F(t_i \mid \bar{\theta}_1, \bar{\theta}_2)$ as

$$\hat{N}(i) = S_i/\pi_i \qquad (8)$$

Observations on sales quantity S_t are available for periods $j = 1,2,\cdots,k$. Note that $S_i = S_{i-1} + X_i$ and the most recent observation $S_k = \sum_{i=1}^{k} X_i$ is a sufficient statistic for $(X_1,\cdots,X_k)$ which includes the information on all the sales of previous periods of the season. Stability of the estimated total sale is affected by the variation of sales proportion. Note that the size of deviation of $\hat{N}_{(i)}$ is determined by relative deviation of sales proportion, which is the coefficient of variation (CV).

$$CV = \sqrt{Np(1-p)}/Np = \sqrt{(1/p-1)/N} \qquad (9)$$

(9) shows that CV of sample proportion is a decreasing function of proportion and suggests that the largest sales proportion provides better estimates of season total sale. Also Herts and Schaffis (1960) provide the empirical result of CV behavior very similar to the result described by (9).

Sample quantile of order p, $\hat{\eta}_p$ asymptotically follows a normal distribution (Rohatgi 1976) with mean $E(\hat{\eta}_p) = \eta_p$, population quantile and variance $\sigma^2(\hat{\eta}_p)$

$$\sigma^2(\hat{\eta}_p) = \frac{p(1-p)}{N f^2(\eta_p)} \tag{10}$$

Empirical investigation shows that the variance of sample quantile (10) is nonincreasing of p for the normal and lognormal distributions.

It is reasonable to use the sales quantity and estimated proportion at the most recent period, k for estimating the season total sale.

$$\hat{N} = S_k / \bar{\pi}_k \tag{11}$$

Note that $\bar{\pi}_k$ is obtained from the estimated cdf which is not from sample observations.

3. Estimating parameters of distribution

Sample proportion of sales quantity to total sales is computed as

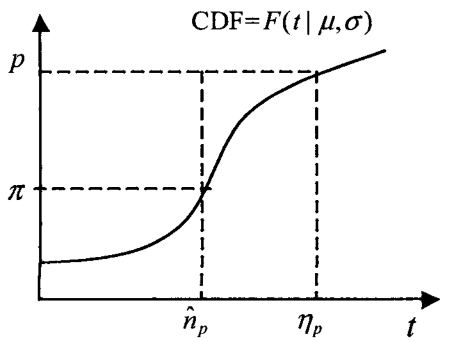
$$p = \frac{1}{N}\sum_{k=1}^{N} \mathrm{I}(t-\omega_k) = \frac{S}{N} \tag{12}$$

where ω_k is a point of time on which the kth unit is sold and $\mathrm{I}(y)=1$ for $y \geq 0$, otherwise $\mathrm{I}(y)=0$. Then S is the number of units sold to date t, and t is a sample quantile of order p, which is denoted as $\hat{\eta}_p$.

In Figure 1, the relation between quantile and proportion is given by

$$p - \pi = f(c)(\eta_p - \hat{\eta}_p) \tag{13}$$

Figure 1 Proportion and Quantile for Normal Distribution



where $f(\cdot)$ is a pdf and $(c-\hat{\eta}_p)(c-\eta_p) \prec 0$. (13) indicates that the difference between sample and population proportions can be minimized with the minimum of weighted difference between sample quantile and a population quantile. The jth sample quantile and sample proportions, $(\hat{\eta}_{p_j}, p_j)$ which are obtained from given season total sale N and observations $S_j\,(j=1,2,\cdots,k-1)$, can be transformed to $(\hat{\eta}_{p_j}, \eta_{p_j})$ as seen in Figure 1. Note that the population quantile, η_p is expressed with parameters of distribution and quantile of standardized variate or percentile as shown in (1), (3), (5) and (7). The parameters of cdf, for example, (μ, σ) of normal distribution can be estimated with a regression model based on the (k-1) observations.

$$\hat{\eta}_p = \mu + \sigma z_p + \varepsilon_p \tag{14}$$

The variance of $\hat{\eta}_p$ is dependent on p, as given (10) so that a weighted least square method is used in estimating parameters so as to minimize

$$\mathrm{SSE} = \sum w_p(\hat{\eta}_p - \mu - \sigma z_p)^2 \tag{15}$$

where $w_p = 1/N\sigma^2(\hat{\eta}_p)$, $\sigma^2(\hat{\eta}_p)$ is the variance of sample quantile of order p, given by (10). Parameters of distribution are not available in initial estimation and then the parameters estimated from unweighted regression model is used to obtain the weight w_p at initial stage.

4. Estimation Procedure

In practice, season total sale is not given and only sales information of the first part of season are available whereas the season total sale is used to obtain sample sales proportions which are needed for estimating parameters of distribution. The estimating procedure consists of following three steps: when family of distribution is given. (i) Estimating season total sale and the compute sample proportion of sales quantity (ii) Find parameters of distribution using weighted regression method (iii) Evaluate the estimation result and determine whether the procedure is reiterated.

Suppose that sales quantity observations for k periods are available.

(i) Estimation of season total sale

- Estimate the population proportion for the most recent period, week k, from the distribution with parameters estimated from previous iteration.

$$\bar{\pi}_k = F(k \mid \bar{\theta}_1, \bar{\theta}_2)$$

- Estimate season total sale N

$$\hat{N} = {S_k}/{\bar{\pi}_k}$$

Season total sale of previous season is used at initial iteration when π_k is not available.

- Compute sample proportions for period $j = 1,2,\cdots,k-1$

$$p_j = {S_j}/{\hat{N}}$$

(ii) Run weighted regression model for (k-1) observations on sample quantiles and population quantile of order p, expressed with parameters of distribution. Note that the parameters estimated at previous iteration is used to evaluate regression weight, w_p as in (15).

(iii) Evaluate the estimated parameters.

- Obtain weighted mean squared error,

$$\mathrm{MSE} = {\mathrm{SSE}}/{(k-3)} \qquad (16)$$

where SSE is defined as (13) and k is the number of observations. Note that the kth observation is used to estimate the season total sale N and simple regression model is used.

- Compute the ratio of decrease in $\mathrm{MSE}_{(i)}$ for current ith iteration.

$$\mathrm{Ratio} = (\mathrm{MSE}_{(i-1)} - \mathrm{MSE}_{(i)})/\mathrm{MSE}_{(i-1)}$$

- If the ratio is negative or very small for example, 0.005 stop the procedure. Otherwise repeat the procedures (i)-(iii)

5. Evaluation of model adequacy

1) Family of Distribution

The distribution of time is assumed to be known except the parameters. The family of distribution is chosen based on the historical data. The assumption on the family of distribution is checked on the following chi-square statistic after estimating the parameters of the distribution.

$$\chi^2 = \sum_{i=1}^{k-1} \frac{(X_i - O_i)^2}{O_i} \qquad (17)$$

where $O_i = \hat{N}(\bar{\pi}_i - \bar{\pi}_{i-1})$ is the expected sale and X_i is the sales quantity for the period i. Note that degree of freedom of the chi-square is df=k-3 since two parameters ($\bar{\theta}_1, \bar{\theta}_2$) are estimated. If P-value of the statistics is very small, other family distribution is employed for the estimation. 1% significance level is advised in that the value of chi-square statistic is approximately proportional to the season total sale, N which may be very large and tends to be very sensitive to small relative difference between O_i and X_i.

2) Residual Analysis

Residual analysis is useful for identifying unusual observations, which may cause large value of the chi-square statistic (17) and no adequate distribution family fits the data. The observation is excluded in the estimation to remove its effect on the estimation if its residual e is

$$|e| \geq t^*_{0.05,k-3}\sqrt{\mathrm{MSE}} \qquad (18)$$

where $P(t_{k-3} \geq t^*_{0.05,k-3})$ =0.05 and (k-3) is a degree of freedom of t distribution. Note that consideration on the sensitivity of chi-square test makes conservative position on evaluating the unusual observation in (18) with significance level 0.10.

3) Bimodal distribution

When a distribution of historical sales data of an item shows bi-mode, no parametric distribution fits the data. Careful investigation on the bimodal data is required to group into homogeneity observations if the sales are from the markets heterogeneous in terms of marketing channels and customers and sales time distribution are different in family of distribution or parameters. Separate estimations are made for the grouped sales data respectively and the season total sale of the item is simply sum of the respective season total sales for the grouped data.

6. Numerical Investigation and Discussion

1) Data

Data are obtained from MS EXCEL uniform random number function. 10,000 uniform random numbers are then inverted into ω, normal random times with mean 15 weeks and standard deviation 6 weeks and into lognormal random times with log mean 2.5 and log standard deviation 0.5. Sales quantity to date week j, S_j is the number of inverted random numbers ω such as $\omega \le j$ and defined as

$$S_j = \sum_{k=1}^{10000} \mathrm{I}(j - \omega_k) \qquad (19)$$

where $\mathrm{I}(y) = 1$ for $y \ge 0$, otherwise $\mathrm{I}(y) = 0$, and ω_k is the time which the k^{th} unit is sold. <Table 1> shows the cumulative sales quantity to date the week.

Numerical analysis of this part was conducted with MS EXCEL spreadsheet without any programming. Hence relevant persons in industry can easily adopt the forecasting method suggested in this paper for season products without difficulty.

2) Period of beginning season and earliest period of forecasting made

Season is assumed to start in the period in which for the item to date 5 per cent of season total, and to end in the period when sales proportion to date reach 95 per cent of season total sale. The season total sale of previous season, assuming given as 9,000 units, is used for initial estimate for current season total. Hence season appears to start at week 5 of which the sales proportion is 5.1 percent. The observations for the periods of preseason are unstable so that preseason observations are not used in estimation.

The more early the forecasting is made with reasonable level of accuracy, the more valuable to decision making on fashion products. while the more observations are, the more accurate the forecasting. Especially the most recent observation, which is used as a base for estimating season total sale, is stable enough to provide a good estimate of season total. Experience in numerical investigations shows that the forecasting made at the week with about 30 per cent sales proportion provide reasonably good estimate of season total sale for various family distributions of the sales time. In this numerical investigation, the week 11 is the earliest period at which the forecasting is made.

3) Evaluating estimation and model adequacy.

Assume that all the information available in the estimation are 7 observations (j, S_j) ($j = 5, 6, \cdots, 11$), sales quantity between week 5 and week 11 in <Table 1> among which first 6 observations are used in quantile regression model to estimate parameters of distribution. Note that the previous season total sales, 9,000 units is used at initial iteration to evaluate sample sales proportions. In case that the family of sales time distribution is not known, fit the data to four parametric distribution models: normal, lognormal, Weibull and extreme value distribution, respectively and then choose the best model with the smallest chi-square statistic.

<Table 2> shows the estimation result where the observed sales quantities from normal sales time are fitted to quantile regression models for four the parametric distributions. The estimated parameters ($\bar{\theta}_1, \bar{\theta}_2$) are presented in the table on the order of (μ, σ) for normal and lognormal distribution, (α, β) for Weibull distribution defined in (4), and (γ, δ) for extreme value distribution in (6). The season total sale is forecasted at week 11, that is, the forecasted total sale at week 11 is computed by $\hat{N}_{(11)} = s_{11}/\bar{\pi}_{11}$. In addition to the estimated parameters of distribution and chi-square statistic, estimated sales proportions are obtained from the cdf with the estimated parameters, $\bar{\pi}_j = F(j \mid \bar{\theta}_1, \bar{\theta}_2)$ and estimated sales quantity to date week j is computed by $\hat{S}_j = \hat{N}\bar{\pi}_j$ Chi-square statistics of (17) with degree of freedom (k-3)=4 and corresponding P-values are computed.

<Table 1 > Sales quantity to date the week from Normal Random Time of Sales

Week (j)	5	6	7	8	9	10	11	12	13	14	15
Sales(S_j)	459	646	909	1,211	1,552	1,983	2,544	3,109	3,723	4,351	4,958

Week (j)	16	17	18	19	20	21	22	23	24	25	26
Sales(S_j)	5,619	6,264	6,857	7,444	7,939	8,361	8,748	9,054	9,299	9,501	9,651

. <Table 2 > **Forecasted Total sale and Estimated Parameters for Normal Random Time of Sales**

Observations		Normal Dist		LogNormal Dist		Weibull Dist		Extreme Value	
Week(j)	S_j:	π_j	$\hat{S}_j$	π_j	$\hat{S}_j$	π_j	$\hat{S}_j$	π_j	$\hat{S}_j$
5	459	0.05163	477	0.00004	471	0.04929	459	0.05771	471
6	646	0.07255	670	0.00007	707	0.07448	693	0.07759	633
7	909	0.09942	918	0.00009	988	0.10504	978	0.10393	847
8	1,211	0.13292	1,227	0.00012	1,315	0.14070	1,310	0.13852	1,129
9	1,552	0.17347	1,602	0.00016	1,684	0.18102	1,685	0.18338	1,495
10	1,983	0.22114	2,042	0.00019	2,094	0.22543	2,099	0.24062	1,962
11	2,544	0.27553	2,544	0.00024	2,544	0.27327	2,544	0.31201	2,544
χ^2 (P-value)		5.104 (0.16431)		16.498 (0.00090)		17.842 (0.00047)		20.774 (0.00012)	
Est. Parameters		14.463	5.808	8.856	1.847	17.929	2.337	14.209	3.262

Chi-square statistics for lognormal, Weibull and extreme value distributions are significantly large and indicate strong evidence that these parametric distributions are not suitable for the sales time of given data. Normal distribution with the smallest chi-square value with P-value 0.1631 is accepted as a sales time distribution and the season total sale is forecasted to be 9,233 unit, which is close to 10,000 units of normal sale times simulated randomly from normal with parameters of mean 15 weeks and standard deviation 6 weeks, comparable with the estimated normal distribution parameters of mean 14.463 and standard deviation 5.808 in <Table 2>.

4) Effect of Sales Proportion for Forecasting point of Time on Forecasting

<Table 3> shows the estimated value of mean and standard deviation for normal random time and forecasted season total sales by the week, the point of time on which the forecasting is made with observations for week 5 through week j in the <table 1>. Note that S_j is the sales quantity to date week j, given in the <Table 1>The forecasted total sales at week 9 and 10 is considerably different from the true season total sales 10,000 units and these large forecasting errors are not unusual and are expected in that the corresponding sales proportion π_j are 0.1552 and 0.1882, respectively not large enough for good forecasting. The forecasting tends to be improved as the time of forecasting passes where the sales proportion increases. Our experience from the numerical investigation shows that the reliably forecasting is obtained at the week with sales proportion of 0.3. Fisher and Raman (1996) suggested that after early 20% sales proportion demands are observed, forecast accuracy can be dramatically improved compared with forecast without initial sales information. Their result, however, did not report the measure of accuracy and still appears to have considerable forecasting errors.

The forecasting at week 15 with the proportion 0.4958 is 9,978 units which is very close to the true total sales. Forecasting at the time with sales proportion about 0.5 consistently provides, in our numerical investigation, very good estimated total sales for various values of parameters regardless of family of distribution.

<Table 3> Estimation by weeks, point of forecasting time

Week of forecast	S_j	π_j	Estimated Values		
			Mean	Std Dev	Tot Sales
9	1,552	0.1552	14.208	5.628	8,751
10	1,883	0.1883	13.721	5.561	7,879
11	2,544	0.2544	14.463	5.808	9,233
12	3,109	0.3109	15.255	6.008	10,575
13	3,723	0.3273	15.261	6.003	10,541
14	4,351	0.4351	15.253	6.005	10,376
15	4,958	0.4958	15.046	5.945	9,978

References

Bardford, J. W. and P. K. Sugrue. 1990. A Bayesian Approach to the Two-Period Style-Goods Inventory Problem With Single Replenishment and Heterogeneous Poisson Demands. *Journal of Operations Research Society*. **41**, 211-218.

Fisher, M, and A. Raman. 1996. Reducing The Cost Of Demand Uncertainty Through Accurate Response To Early Sales. *Operations Research* **44**, 1, 87-99.

Hertz, D.B. and K. H. Schaffir. 1960. A Forecasting For Management of Seasonal Style-Goods Inventories. *Operations Research* **8**, 45-52.

Ravindran, A. 1972. Management of Seasonal Style-Goods Inventory. *Operations Research* **20**, 2, 265-275. New York.

Rohatgi, V. K. 1976. *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley & Sons Inc. New York