Xiamen Airline Demand Forecasting & Capacity Control

Course Name: Pricing and Revenue Management

2021.6.25

Catalogue

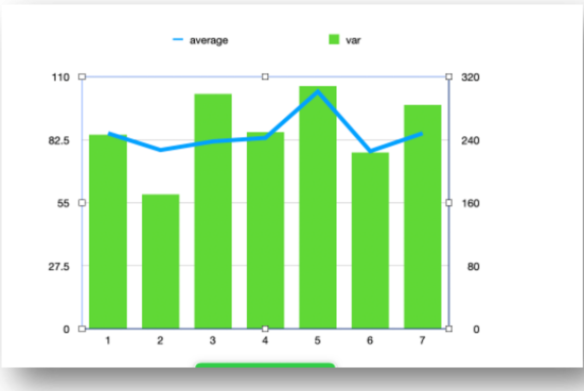
1.Descriptive Statistics

After we received the information, it has several dimensions: DOW(day of week), festival, Policy, seasonal. And our descriptive statistics will be spread from these aspects.

1.1 DOW

From DOW, we want to know whether DOW influenced the sales. And we use the mean and standard deviation to analysis it. For example, if Tuesday has a positive influence on sales. The average sales of Tuesday should be higher than other days and the standard deviation should be small. So in chart 1 we calculated the mean and standard deviation of each DOW and found that the lowest average is 77.55 of Saturday and the highest is 103 of Friday. But only average can't explain the whole situation. We decided to use $\frac{mean}{standarddeviation}$ to estimate it.

	average	var
1	85.3452380952381	246.445639701664
2	78.0348837209302	170.622298221614
3	81.8	298.139325842697
4	83.3333333333333	249.61797752809
5	103.532467532468	308.146958304853
6	77.55	223.744303797468
7	85.3452380952381	283.983250073465



But I have known recently that the KS test can test whether two examples are come from the same total distribution.

1.2 Festival

According to common sense, holidays will have a certain impact on travel. For example, for cities with developed tourism industry, when the mid-long holiday comes, more tourists will fly to the city at the beginning of the holiday, and more passengers will fly away at the end of the holiday.

However, different holidays have different effects on air ticket demand. We consider three different scenarios: first, there is a positive impact on demand, second, there is no obvious impact on demand, and third, there is a negative impact on demand, we denote them as F1, F2 and F3 respectively. There are specific dates for these three different festivals in the document. Based on these data, we analyze the impact of holidays on demand from the following dimensions.

F1	F2	F3
2019/2/7	2018/2/15	2018/2/26
2019/2/11	2018/2/16	2018/2/27
2019/2/12	2018/2/17	2018/2/28
2019/2/13	2018/4/5	2018/6/18
2019/2/14	2018/4/29	2018/9/23
2019/4/7	2018/4/30	2018/10/5
2019/5/1	2018/6/16	2018/10/6
	2018/6/17	2019/2/16
	2019/2/5	2019/2/17
	2019/2/6	2019/6/9
	2019/4/5	2019/10/5
	2019/4/6	2019/10/6
	2019/6/7	2019/10/7

Table 1-1: the festivals of two years

In order to better reflect the differences in holiday ticket sales in the sample, we calculated the average daily ticket sales for five different fares and the average ticket sales for 31 periods of 60 days prior to departure of each flight, and obtained the following table.

fareclass	30	29	28	27	26	25	24	23	22	21	20
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.03
3	0.23	0.39	0.66	0.84	1.81	2.64	4.37	4.63	5.03	5.47	5.95
4	0.07	0.13	0.28	0.37	1.29	2.06	3.69	3.91	4.23	4.63	5.03
5	0.11	0.22	0.47	0.65	1.98	3.19	5.74	6.12	6.70	7.33	8.03
fareclass	19	18	17	16	15	14	13	12	11	10	9
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.03
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
fareclass	8	7	6	5	4	3	2	1	0		sum
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		43
2	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01		14
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		15
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		11
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		3

Table 1-2: overall average sales

1.3 Ticket sales

The following table shows sales for different types of holidays and for the overall sample:

average sales	F1	F2	F3	whole
fare-class 1	88	37	72	43
fare-class 2	18	14	16	14
fare-class 3	8	13	10	15
fare-class 4	1	18	0	11
fare-class 5	0	7	0	3
sum	116	89	98	87

Table 1-3: ticket sales comparison

It is obvious that festival 1 has a optimistic effect on ticket sales, and festival 2 influence little. However, it does not make sense that sales in festival 3 (which has a pessimistic effect on ticket sales) is higher than overall average sales.

1.4 Policy limitation

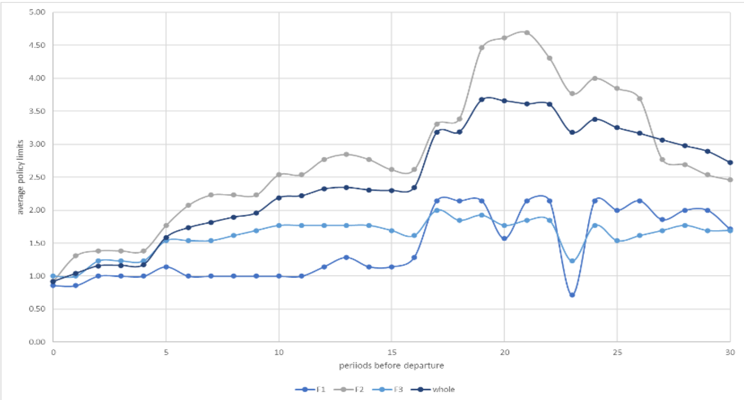


Fig 1-1: policy limitation in different type of festivals

To test whether policy restrictions affect ticket sales for these different types of holidays. We calculated the average of the policy limits for the tickets of different types of holidays in 31 periods before departure, and obtained the following chart:

It can be clearly seen that the policy restrictions of F2 in each period are close to those of the sample as a whole, while F1 and F3 tend to sell higher price tickets in terms of policy restrictions.

The policy restrictions of F2 are basically the same as those of non-holidays, so in terms of the overall sales of tickets, the sales of F2 are roughly the same as those of the sample as a whole. For F1, since festivals promote the demand for air tickets, policy restrictions tend to increase the price of air tickets, but the sales of tickets for F1 still increase significantly compared with the sample as a whole. In terms of the dimension of policy restrictions, F3, which has a negative effect on air ticket sales, does not lower the price of air tickets, but instead raises the price. However, the sales volume of air tickets is still higher than the sample as a whole.

1.5 Proportion of different tickets

According to the data, there is not much difference between F2 and the sample as a whole, while F1 and F3 are close. The majority of tickets sale of F1 and F3 both belong to fare-class 1, which accounts for more than 70% of sales, while F2 and the sample as a whole are significantly smaller on this metric.

1.6 Policy

For the policy, for every day, there will be 30 DBA and each DBA will have a limitation. And our destination is to analyze the policy XM airline use. And from the chart we can conclude that at the beginning of each day, the policy limitation will be comparatively high. And the tendency will be ascending as first. After get the highest part, the num began to descending. Why XM airline use this policy? Since at the beginning of each date. The customers who book the ticket will be little and in order to amplify the sales, XM airline will sale more cheap classes. However, as the date is approaching zero, more and more people need to book the ticket and as a result, XM airline want to increase the total sales and they limit the cheaper class.

1.7 Seasonal

(1) Daily Sales

We add up the sales of all classes to get daily sales and form a scatter chart：

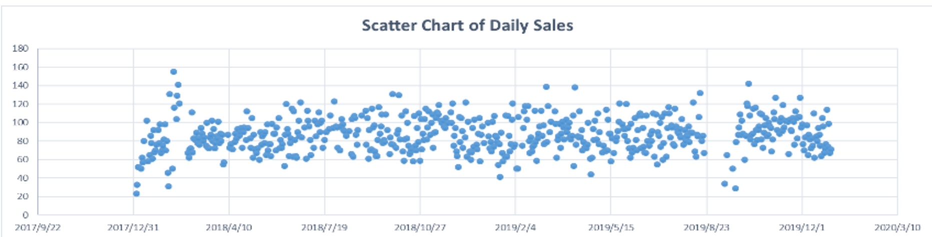


Fig 1-2: Scatter chart of daily sales

We found that the daily sales were around 60-120, the flight density was moderate, and there was a significant gap only around September 2019. Therefore, we can find that the overall demand for this flight tends to be stable, but there are still time differences. Therefore, we need to further explore the differences of each month or quarter for better flight management.

(2) Monthly Sales

By aggregating the monthly sales, we can see that there are certain trends in monthly sales in 2018 and 2019. Sales, for example, tend to peak at the end of the year and lower in August and September. However, the overall trend is not clear enough, so we further summarize the quarterly data to explore.

(3) Quarterly Sales & Yearly Sales

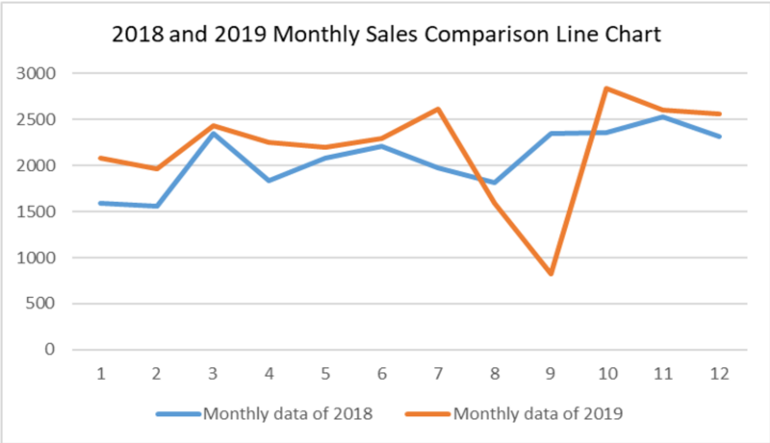


Fig 1-3: Monthly sales comparison line chart

In terms of quarterly data, sales of the airline rose steadily from the first quarter to the second quarter, followed by a slow or declining growth trend in the third quarter and a peak season in the fourth quarter. We can guess that because this flight is a business flight and the company enters the settlement period at the end of the year, the business travel increases, while the third quarter is due to the holiday, so the business travel decreases. In response, we can add more flights at the end of the year or raise prices appropriately based on supply and demand. And the off-season can lower prices to increase revenue.

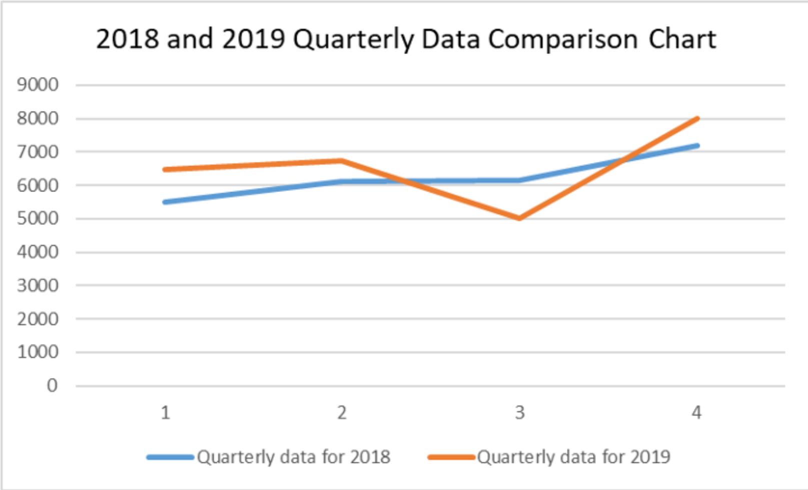


Fig 1-4: Quarterly data comparison line chart

At the same time, we can find that, except for August and September of the third quarter, the sales volume of other months in 2019 is higher than that of 2018. If we can combine the data of previous years, we can find the trend of the overall sales volume. We guess with the increase of people's economic level, the plane has become one of the choices for people to travel, xiamen to Shanghai route distance is suitable at the same time the price is suitable, so more and more become the choice between the two people, so we can pay attention to the construction of this route in the future. At the same time, since the customers are mainly business people, the construction of services is conducive to obtaining loyal customers and converting customer value.

2.Averaging Method

As we know from the data, we have the sales data of different Fareclasses of Xiamen Airlines in 2018 and 2019. We need to obtain the most accurate consumer demand based on the existing sales data. We have adopted three different ways to unconstrain sales, and finally we will compare them to find a more accurate demand. First, we chose Averaging method (AM). We will use AM to liberate the demand for existing sales data, and then forecast the demand by exponential smoothing method and moving average methods, and then introduce policies to get the sales data we predicted. Finally, we compare the predicted sales data with the actual sales data. The accuracy of AM method is reflected by the error. As for the deficiencies of this method and the improvement methods, we will mention later.

2.1Averaging Method Introduction

AM method, that is, by dividing the specific data into a small number of periods, and then using the sum of the specific data in each period to represent the value of this period, and then judging whether each stage is constrained, by converting the constraining value into (1) the observable value of the constraining value itself; (2) The mean value of the relevant unconstraining value, the larger one to realize demand liberation. Because the data we already have has been processed into 31 periods, we don't need to process it.

2.2Averaging Method Operation

(1) Defining restricted and unrestricted time points

（For example：AM_EXCEL01_fare1 缺失值及初始值 average）

As we know, the restriction on non-restriction is determined by the policy. For example, when the policy is 3, the shipping space larger than 3 is not open, that is, the Fareclss4 and 5 are restricted, and the sales amount is 0, while consumers can only buy FareClass1, 2 and 3. Therefore, the function of Excel can be expressed as IF (FareClass <=policy, original value of FareClass, 1000), that is, IF it is not limited, we assume that its sales volume is its real demand. First of all, 1000 is used instead. The 1000 substitution is used to distinguish between zero actual demand and zero restricted sales. Then we used the replacement function of Excel to change the node corresponding to 1000 into the missing value, so as not to affect the calculation of the following average value. As shown in the figure below, our table contains both missing and unrestricted values.

14	13	12	11	10	9	8	7
3	2	0					
1	3	0	3				
1	1	1	1	1	0	1	2
1	1	1	1	3	1	1	4
2	1	0	1	2	1	0	1
0	1	0	2	0	1	0	3
1	1	1	1	1	1	0	2
2	3	0	2	1	1	0	0

Table 2-1: Missing and unrestricted values

(2) Select the restricted value to replace the object

(For example：AM_EXCEL01_ fare1 AM modified)

Because we know that the average restricted should use the limited its observed value of the larger one, because we are all restricted to 0, so the average value must be greater than the observation, so we directly use limited each column of the average restricted replacement for the same column, namely the missing value.

Therefore, we first need to calculate the AVERAGE value of each column. For example, in Table Fare1, the training set is converted into missing value by limiting value and the data is obtained by using Excel AVERAGE at the end

(3) Replace restricted values

(For example：AM_EXCEL01_ fare1 AM modified:)

As shown in the table, we replaced all missing values with the unrestricted worth average values of the corresponding columns, that is, demand liberation was completed.

2.3Forecasting Demand

(1) Simple Moving Method

(For example: AM_EXCEL01_fare1 移动平均预测)

For each column of data, the moving average method is adopted to make demand prediction, in which the interval is 2, and the moving average prediction in Table Fare1 is obtained.

date	fareclass	30	简单移动	29	简单移动
2018/1/3	1	1	#N/A	0	#N/A
2018/1/4	1	0	0.5	0	0
2018/1/5	1	1	0.5	1	0.5
2018/1/8	1	1	1	1	1
2018/1/9	1	1	1	1	1
2018/1/10	1	1	1	0	0.5
2018/1/11	1	2	1.5	1	0.5
2018/1/14	1	0	1	0	0.5
2018/1/15	1	1	0.5	0	0
2018/1/17	1	0	0.5	0	0
2018/1/18	1	0	0	0	0
2018/1/19	1	2	1	1	0.5
2018/1/20	1	0	1	0	0.5

Table 2-2: Simple Moving Method

(2) Exponential Smoothing Method

(For example: AM_EXCEL01_ fare1 简单指数平滑法)

Similarly, we perform an exponential smoothing method for each column to predict the demand, where the damping coefficient is 0.2.

date	fareclass	30	简单指数平滑法	29	简单指数平滑法
2018/1/3	1	1	#N/A	0	#N/A
2018/1/4	1	0	1	0	0
2018/1/5	1	1	0.2	1	0
2018/1/8	1	1	0.84	1	0.8
2018/1/9	1	1	0.968	1	0.96
2018/1/10	1	1	0.9936	0	0.992
2018/1/11	1	2	0.99872	1	0.1984
2018/1/14	1	0	1.799744	0	0.83968
2018/1/15	1	1	0.3599488	0	0.167936
2018/1/17	1	0	0.87198976	0	0.0335872
2018/1/18	1	0	0.174397952	0	0.00671744
2018/1/19	1	2	0.03487959	1	0.001343488

Table 2-2: Exponential Smoothing Method

2.4 Policy Management

（For example：AM_EXCEL02_引入政策后的移动平均销量）

We put the resulting predictions back into the same table. As before, we use the IF function for the limit value under the policy control. When the policy is less than the corresponding FareClass, we change it to 0, so as to get the predicted sales volume.

date	fareclass	30	29	28	27	26
2018/1/3	1	0	0	0	0	0
2018/1/3	2	0	0	0	0	0
2018/1/3	3	0	0	0	0	0
2018/1/3	4	0	0	0	0	0
2018/1/3	5	0	0	0	0	0
2018/1/4	1	0.5	0	0	0	1.5
2018/1/4	2	0.5	0	0	0	1.5
2018/1/4	3	0	0	0	0	0
2018/1/4	4	0	0	0	0	0
2018/1/4	5	0	0	0	0	0
2018/1/5	1	0.5	0.5	0.5	0	2

Table 2-3: Predicted sales volume

(1) Comparison

We sum up the total sales volume of each day after the forecast, and compare it with the actual sales volume to get AM_EXCEL02_预测销量和 vs 实际销量.

date	实际销售	移动平均每一天求和	|Dt-Ft|	|Dt-Ft /Dt	（Dt-Ft）^{2}
2018/1/3	23	0	23	1	529
2018/1/4	33	29.05985866	3.940141	0.119398	15.52471
2018/1/5	52	38.30587992	13.69412	0.263348	187.5289
2018/1/8	50	61.80769422	11.80769	0.236154	139.4216
2018/1/9	62	58.17608104	3.823919	0.061676	14.62236
2018/1/10	57	58.73619731	1.736197	0.03046	3.014381
2018/1/11	80	81.70046032	1.70046	0.021256	2.891565
2018/1/14	102	88.24502715	13.75497	0.134853	189.1993
2018/1/15	58	85.05457607	27.05458	0.466458	731.9501
2018/1/17	65	58.28604936	6.713951	0.103292	45.07713
2018/1/18	86	75.70535714	10.29464	0.119705	105.9797
2018/1/19	78	68.38232179	9.617678	0.123304	92.49973
2018/1/20	61	76.16769047	15.16769	0.248651	230.0588

AM+移动平均	
MAE	13.92745
MAPE	0.165254
MSE	309.4251
RMSE	17.59048

Table 2-4: Comparison

Through the error calculation methods such as MAE, we can calculate the errors of the results obtained by two different prediction methods under AM algorithm, as shown in the figure below. Thus, it can be found that the results under AM and moving average are more likely to be close to the actual demand.

	MAE	MAPE	MSE	RMSE
AM+ Simple Moving Method	13.92745	0.165254	309.4251	17.59048
AM+ Exponential Smoothing Method	1278.37866	14.2111911	54553750.2	7386.05106

Table 2-5

(2) Optimization idea

We can divide the data into training set and test set to confirm the validity of each method more accurately. At the same time, simulation can be carried out through simulation after obtaining the predicted demand. However, due to some problems in the application of simulation technology, we did not implement the two ideas.

3. Expectation-Maximization Method

3.1 Assumption

The data are independent of each other, and the whole obeys normal distribution.

Cannot be applied to time-series data with trend or seasonality.

$\{Z_1, Z_2, Z_1, \ldots, Z_M, Z_{M+1}, \ldots, Z_{M+N}\}$ unordered observation constrained $Z_1, Z_2, Z_1, \ldots, Z_M$ else unconstrained.

3.2 Formulation of Expectation-Maximization Method

(1) Step 0 (Initialize)

$$\sigma^{(0)} = \sqrt{\frac{\sum_{i=M+1}^{M+N}(Z_i-\mu^{(0)})^2}{N}} \qquad \mu^{(0)} = \frac{\sum_{i=M+1}^{M+N} Z_i}{N}$$

(2) Step 1 (E-step)

Replace the censored data:

$$\hat{z}_i^{(k-1)} = E(X \mid X \geq X^{b_i}) \quad X \sim N(\mu^{(K-1)}, \sigma^{(K-1)})$$

(3) Step 2 (M-step)

Maximum-likelihood method to updata $censored\ Z_i$ and obtain $(\mu^{(k)}, \quad \sigma^{(k)})$

(4) Convergence test

$$\sum_{i=1}^{m} \log P(x^{(i)}; \theta^{j+1}) \geq \sum_{i=1}^{m} \log P(x^{(i)}; \theta^{j})$$

3.3 Demand Forecasting

(1) Historical Sales Data

We used the historical 590 days sales data as the original data. After data preprocessing, we used SPSS to get the unconstrained demand and the processed parameters in the end.

(2) Data Pre-processing

The first step in data processing is to categorize the raw data by FareClass to produce five sales tables.

Date	Fareclass	···
2018/1/3	1	···
2018/1/3	2	···
···	···	···
2019/12/31	4	···
2019/12/31	5	···

Date	Fareclass	···
2018/1/3	1	···
2018/1/4	1	···
2018/1/5	1	···
···	···	···
2019/12/31	1	···

Table 3-1

The second step is to compare the divided sales data with the policy data, delete the data constricted by the policy, and form the sales table with missing values.

date	fareclass	30	29	28	27	26
2018/1/3	3					
2018/1/4	3					
2018/1/5	3					
2018/1/8	3			0	1	1
2018/1/9	3			0	0	1
2018/1/10	3					
2018/1/11	3					
2018/1/14	3					
2018/1/15	3		0	0	1	1
2018/1/17	3					
2018/1/18	3					
2018/1/19	3					
2018/1/20	3					
2018/1/21	3					
2018/1/22	3	0	0	0	0	2
2018/1/23	3	0	0	1	0	1

Table 3-2

(3) Data Processing

By importing the data into SPSS, the missing value prediction method was selected as EM method, the probability distribution was set as normal distribution, and the number of iterations was 10000 times.

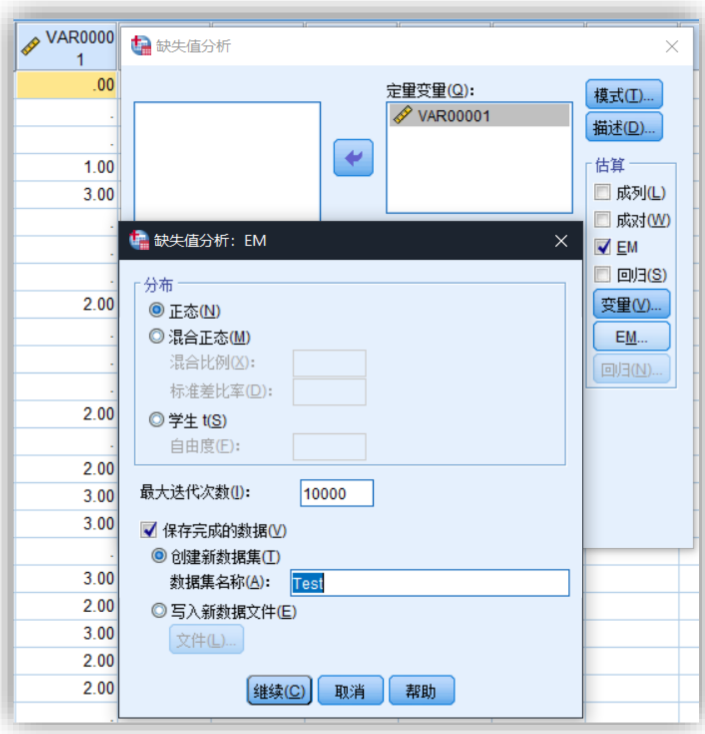


Fig 3-1: SPSS

(4) The Processed Parameters

After SPSS processing, we got the unconstrained demand of five Fareclasses. We take the data of FareClass3, DBA=25 as an example to make the prediction through SPSS.As you can see, January 4, 2018 is the missing value in the DBA=25 column. Therefore, we can get the statistical result, that is, among the 590 data, 107 data are missing values. After prediction by EM method, the mean value is 1.4783, and the standard deviation is 1.09745.

单变量统计

	个案数	平均值	标准 偏差	缺失 计数	缺失 百分比	极值数[a] 低	极值数[a] 高
VAR00001	483	1.4783	1.09745	107	18.1	0	24

a. 超出范围 (Q1 - 1.5*IQR, Q3 + 1.5*IQR) 的个案数。

Table 3-3: The statistical results

4.Double Exponential Smoothing Method(DES)

4.1 DES Introduction

DES uses two smoothing constants: one for smoothing the base component of the demand pattern and a second for smoothing the trend component. It's a good way to use it when the data shows a trend. And it's also useful in the constrained demand data. In this method, we have six parameters as below.

A_t: actual cumulative demand in period t;

F_t: the exponentially smoothed base component for period t;

T_t: the exponentially smoothed trend component for period t;

FIT_t: the forecast of cumulative demand including trend for period t;

$\boldsymbol{\alpha}$: base smoothing constant;

$\boldsymbol{\beta}$: trend smoothing constant.

DES's formulas:

$$FIT_t = F_t + T_t \qquad (1)$$

$$F_t = FIT_{t+1} + \alpha(A_{t+1} - FIT_{t+1}) \qquad (2)$$

$$T_t = T_{t+1} + \beta(F_t - FIT_{t+1}) \qquad (3)$$

$$\min_{\alpha,\beta} \sum_{T=I}^{B}(A_t - FIT_t)^2 \qquad (4)$$

$$FIT_0 = F_B + BT_B \qquad (5)$$

In the normal situation, we just calculate F_t and T_t, and then add them together which is FIT_t.

And then using FIT_t to work FIT_0 out.

4.2 DES Method in XM Airline

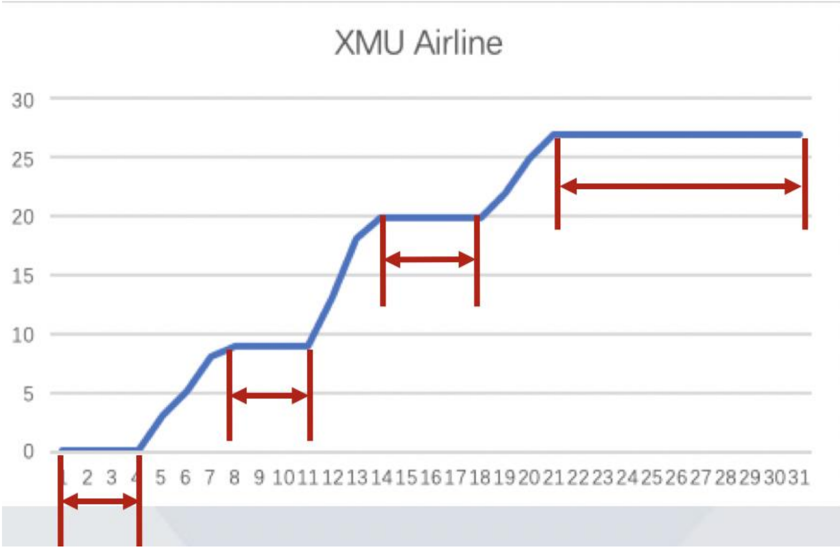


Fig 4-1: XM Airline Cumulative Booking Data

In the XM airline case, we need to forecast several time series. So that we have to change the method. Our reference using the closest historical sales data to calculate the demand. But it's useless in our case, because sometimes the historical data is too short to do a forecast.

We found a new method to solve this problem which is using cumulative historical sales data to forecast the demand.

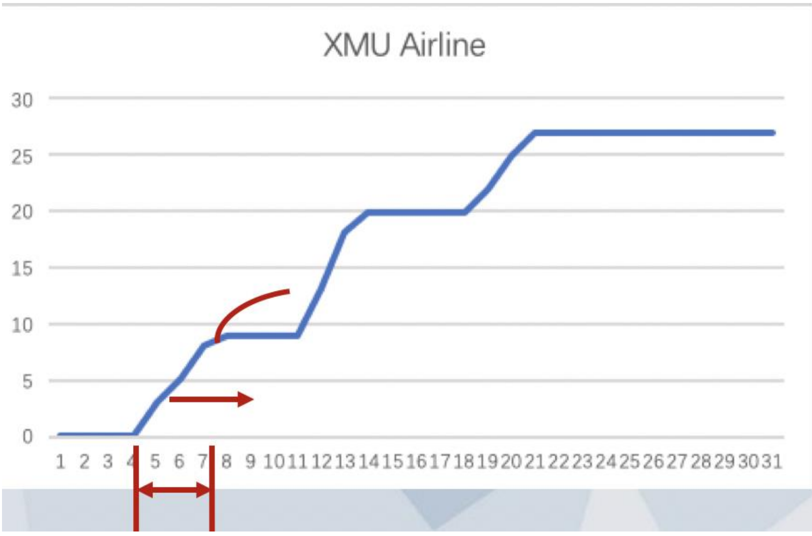


Fig 4-2: XM Airline Cumulative Booking Data

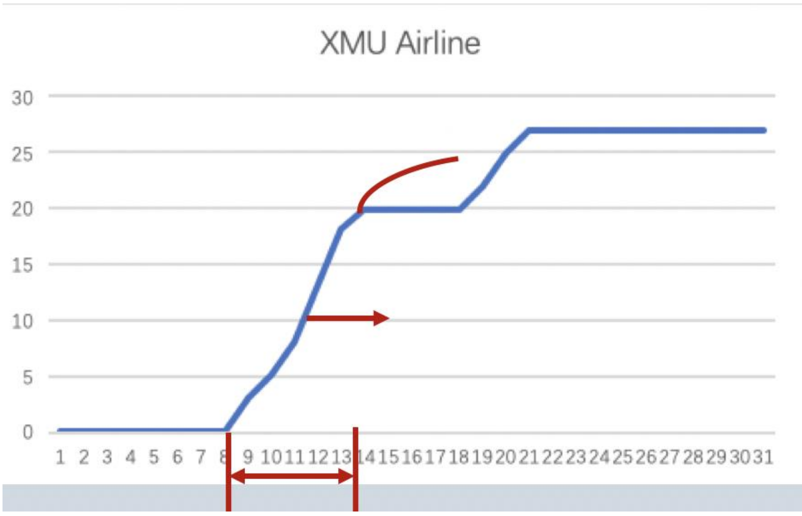


Fig 4-3: XM Airline Cumulative Booking Data

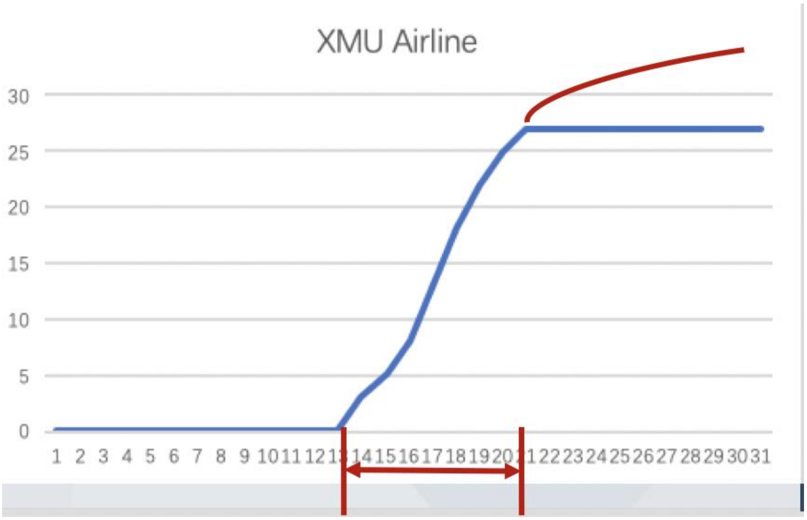


Fig 4-4: XM Airline Cumulative Booking Data

These three pictures show the calculation process. At the first time, we use historical data to get the demand. After that, we will face second constrained time period. We drop first constrained data, and using total historical sales to calculate the demand. And then, do it again and again until we finish all time period. It means that we think the demand put off if the policy limits the booking.

This new method obviously has some problems. For example, some customers will choose another airline or transport. Our forecast demand will be enlarged. But we believe this method is suitable in this case. There are two reasons, first it's better than the old method showed in the reference. Second most the demand will put off if the day is far away from departure.

4.3 Realizing DES method

First step, we transform additional data to cumulative data. Because cumulative data is more accurate in the DES method. Second, we use SPSS to help us to forecast the demand. At last we get the cumulative forecast result, and then transforming it to additional data back.

date	fareclass	30	29	28	27	26	25	24
2018/10/9	3	1	0	0	0	2	1	3
2018/10/10	3	0	0	0	0	2	4	4
2018/10/11	3	1	0	0	0	5	2	4
2018/10/12	3	0	0	2	1	2	4	5
2018/10/13	3	0	0	1	0	1	1	1
2018/10/14	3	0	0	1	0	5	0	0
2018/10/15	3	1	0	0	0	1	1	5
2018/10/16	3	0	1	0	0	1	0	3
2018/10/17	3	0	1	0	0	2	2	3
2018/10/18	3	1	0	1	0	2	3	4
2018/10/20	3	0	0	0	0	1	1	2
2018/10/21	3	0	0	0	0	4	0	0
2018/10/22	3	0	1	1	0	3	2	3

Table 4-1: Additional Data

date	fareclass	30	29	28	27	26	25	24
2018/10/9	3	1	1	1	1	3	4	7
2018/10/10	3	0	0	0	0	2	6	10
2018/10/11	3	1	1	1	1	6	8	12
2018/10/12	3	0	0	2	3	5	9	14
2018/10/13	3	0	0	1	1	2	3	4
2018/10/14	3	0	0	1	1	6	6	6
2018/10/15	3	1	1	1	1	2	3	8
2018/10/16	3	0	1	1	1	2	2	5
2018/10/17	3	0	1	1	1	3	5	8
2018/10/18	3	1	1	2	2	4	7	11
2018/10/20	3	0	0	0	0	1	2	4
2018/10/21	3	0	0	0	0	4	4	4
2018/10/22	3	0	1	2	2	5	7	10

Table 4-2: Cumulative Data

模型		11	12	13	14	15	16
@43146-模型_35	预测			19	22	25	28
@43147-模型_36	预测		18	21	24	27	30
@43195-模型_73	预测	6	7	9	10	12	13
@43268-模型_132	预测			12	14	16	17
@43272-模型_136	预测			2	3	4	4
@43274-模型_138	预测			1	2	2	2
@43312-模型_163	预测			11	14	16	19
@43313-模型_164	预测		5	7	8	10	11
@43316-模型_166	预测		5	6	8	9	10

Table 4-3: Result

5.RM simulation

5.1 Process

As we have got the real demand of each flight ticket in each period by EM, AM and DES methods, and we assume that demand in same period obey a specific normal distribution. Thus, we can calculate the mean value and variance according to the forecast demand as our estimators. Once we get the average and std, we can generate a set of random number which obeys the normal distribution with the parameters we estimate.

Then we should modify our simulated data because there are some negative numbers which are impractical. We replace minus by zero to eliminate the impact of negative numbers. Obviously, zero represent no demand as minus with no practical significance. (because the frequency of minus and their values are small, so this processing does not have a noticeable effect on the final result).

date	30	29	28	27	26
2018/1/3	0.51	-0.94	-0.13	0.10	3.04
2018/1/4	0.25	0.36	0.00	-0.54	1.48
2018/1/5	0.25	0.87	0.23	-0.24	0.22
2018/1/8	0.86	-0.23	1.15	1.73	1.88
2018/1/9	0.69	0.56	0.58	-0.15	2.81
2018/1/10	-0.26	0.25	0.62	-0.91	2.35
2018/1/11	-0.40	-0.37	0.03	0.52	3.40
2018/1/14	0.16	-0.01	0.21	-0.55	4.11
2018/1/15	0.76	0.08	0.63	0.31	2.09
2018/1/17	0.36	0.86	-0.02	0.15	3.11
2018/1/18	0.13	0.17	0.53	0.31	2.20

date	30	29	28	27	26
2018/1/3	0.51	0.00	0.00	0.10	3.04
2018/1/4	0.25	0.36	0.00	0.00	1.48
2018/1/5	0.25	0.87	0.23	0.00	0.22
2018/1/8	0.86	0.00	1.15	1.73	1.88
2018/1/9	0.69	0.56	0.58	0.00	2.81
2018/1/10	0.00	0.25	0.62	0.00	2.35
2018/1/11	0.00	0.00	0.03	0.52	3.40
2018/1/14	0.16	0.00	0.21	0.00	4.11
2018/1/15	0.76	0.08	0.63	0.31	2.09
2018/1/17	0.36	0.86	0.00	0.15	3.11
2018/1/18	0.13	0.17	0.53	0.31	2.20

Table 5-1: Modified process

To exam if the forecast of demand is valid, a useful way is to apply the same policy on the demand we forecast. To simplify such process, we make a worksheet can intuitively show whether the tickets of various fares are open on a specific day in the past, the value of 0 represent close and vice versa.

For example：forecast demand of fare-class 3 in period 30, 1/3 2018 is 0.51, and policy shows that fare-class 3 was closed at that time, so the forecast sales equal 0.

date	30	29	28	27	26
2018/1/3	0.51	0.00	0.00	0.10	3.04
2018/1/4	0.25	0.36	0.00	0.00	1.48
2018/1/5	0.25	0.87	0.23	0.00	0.22
2018/1/8	0.86	0.00	1.15	1.73	1.88
2018/1/9	0.69	0.56	0.58	0.00	2.81
2018/1/10	0.00	0.25	0.62	0.00	2.35
2018/1/11	0.00	0.00	0.03	0.52	3.40
2018/1/14	0.16	0.00	0.21	0.00	4.11
2018/1/15	0.76	0.08	0.63	0.31	2.09
2018/1/17	0.36	0.86	0.00	0.15	3.11
2018/1/18	0.13	0.17	0.53	0.31	2.20

3	30	29	28	27	26
2018/1/3	0	0	0	0	0
2018/1/4	0	0	0	0	0
2018/1/5	0	0	0	0	0
2018/1/8	0	0	1	1	1
2018/1/9	0	0	1	1	1
2018/1/10	0	0	0	0	0
2018/1/11	0	0	0	0	0
2018/1/14	0	0	0	0	0
2018/1/15	0	1	1	1	1
2018/1/17	0	0	0	0	0
2018/1/18	0	0	0	0	0

date	30	29	28	27	26
2018/1/3	0.00	0.00	0.00	0.00	0.00
2018/1/4	0.00	0.00	0.00	0.00	0.00
2018/1/5	0.00	0.00	0.00	0.00	0.00
2018/1/8	0.00	0.00	1.15	1.73	1.88
2018/1/9	0.00	0.00	0.58	0.00	2.81
2018/1/10	0.00	0.00	0.00	0.00	0.00
2018/1/11	0.00	0.00	0.00	0.00	0.00
2018/1/14	0.00	0.00	0.00	0.00	0.00
2018/1/15	0.00	0.08	0.63	0.31	2.09
2018/1/17	0.00	0.00	0.00	0.00	0.00
2018/1/18	0.00	0.00	0.00	0.00	0.00

Table 5-2: Applied policy

By this way, we can get the total demand of all fare-class flight, the figure below shows the result:

EM result (sales)

simulated sales	fare1	fare2	fare3	fare4	fare5	sum
2018/1/3	17	6	6	0	0	29
2018/1/4	13	8	0	0	0	21
2018/1/5	25	2	0	0	0	27
2018/1/8	23	7	9	0	0	40
2018/1/9	26	10	8	0	0	44
2018/1/10	29	20	9	0	0	57
2018/1/11	54	14	5	0	0	74
2018/1/14	46	9	7	0	0	62
2018/1/15	41	8	16	5	0	70
2018/1/17	51	16	8	0	0	75
2018/1/18	42	13	10	0	0	65

EM result (demand)

simulated demand	fare1	fare2	fare3	fare4	fare5
2018/1/3	59	32	39	39	30
2018/1/4	40	35	45	32	25
2018/1/5	59	31	38	37	24
2018/1/8	48	33	42	39	28
2018/1/9	41	32	32	31	28
2018/1/10	49	40	46	34	24
2018/1/11	54	29	35	42	29
2018/1/14	46	26	36	34	32
2018/1/15	41	25	41	30	25
2018/1/17	51	31	39	39	27
2018/1/18	42	30	40	36	22

Table 5-3: Result set

5.2 Result

Based on the same assumption that demand in same period obey a specific normal distribution. We get estimators by using EM methods. Now, we assess the validation of each methods by following indicators: MAE, MSE and MAPE.

And the figure below shows the result:

Error of EM method

Date	Actual sales	Forecasting sales	forecast error(et)	The absolute value of the prediction error(|et|)	The square of the prediction error(et^2)	percentage error	The absolute value of the percentage error			
2018/1/3	23	29	-5.63	5.63	31.73	-24.49	24.49		MAE	24.163504
2018/1/4	33	21	12.26	12.26	150.29	37.15	37.15		MSE	878.8754
2018/1/5	52	27	24.78	24.78	614.06	47.65	47.65		MAPE(%)	29.469982
2018/1/8	50	40	10.44	10.44	109.02	20.88	20.88			

Table 5-4: Error analysis

5.3 Improvement

As far as now, we ignore the impact of DOW, festival and meeting on actual demand. But in our descriptive part, all factors mentioned above influent demand obviously. Considering the feasibility, we added the demand impact of DOW in our model. The data unconstrained by the EM method was used as the experimental group. We divide all days in the data (590) into seven groups: Monday, Tuesday, etc. Then, calculate different periods' means and standard deviation of each group.

weekday	30	29	28	27	26	25
1	0.2409639	0.1325301	0.4819277	0.253012	1.3253012	1.2168675
2	0.2619048	0.1785714	0.2857143	0.1547619	0.8214286	0.6309524
3	0.2209302	0.1627907	0.1395349	0.127907	0.6395349	0.6162791
4	0.1777778	0.0888889	0.1555556	0.0888889	0.8777778	0.6
5	0.2	0.1111111	0.2	0.1	0.8222222	0.6555556
6	0.2727273	0.2857143	0.5454545	0.4155844	1.7402597	1.6990909
7	0.225	0.1875	0.1625	0.125	0.6	0.5

Table 5-5: Average and standard deviation

The grid highlighted by green border represents that the average of demand of fare-class 1 of period 26 on Saturday is 1.740.

In our new model, we assume that the demand of same fare-class of all Mondays, Tuesdays and etc. in a same period obey the normal distribution, and the parameters included average and standard deviation are estimators given by EM method. Based on this assumption, we use matlab to generate a series of random number which means the simulated demand.

The result is showed below:

date	simulated sales	actual sales	error	percentage	square of error			
2018/1/3	23	23	0	1.87%	0		MAE	18
2018/1/4	20	33	13	38.43%	161		MAPE	22.31%
2018/1/5	27	52	25	48.29%	631		MSE	529
2018/1/8	39	50	11	21.71%	118		RMSE	23
2018/1/9	47	62	15	24.88%	238			

Table 5-6: Result after improvement

After the improvement, the value of error is obviously reduced. As you can see, compared to the result given by EM method before perfection, MAE reduced 25%, the rest of indicators decrease as well. So, we guess that once we take factors of festival and meeting into consideration, the error will decrease more obvious. However, it is hard for us to find a feasible way to process it.

6. The Hindsight Policy-Certain Demand

6.1 Explanations & Assumptions

We used AM, EM and DES to forecast the real demand, and finally we chose the optimized EM method as the basis for our follow-up, that is, we used its results as our specific demand or the distribution of demand. In our hindsight we use the demand identified by EM as the basis for our optimal policy. We want to get a benchmark in this way that we can use as a reference for our other methods later on. At the same time, we propose the following assumptions:

(1) Capacity per flight is 163;

(2) The original ticket price was 1000 RMB. We represented a discount by the middle value of each discount, and finally got each Fareclass ticket price as follows:

Fareclass	Discount range	The midpoint of the discount	Final discount price
1	[0.9,1]	0.95	950RMB
2	[0.75,0.8]	0.775	775RMB
3	[0.6,0.75]	0.675	675RMB
4	[0.4,0.6]	0.5	500RMB
5	[0.3,0.4]	0.35	350RMB

6.2 Process

Since we have 590 days of data, Hindsight needs to optimize each day separately, which is a tedious process, so we randomly choose five days for linear programming solution. They are January 3, 2018; March 1, 2018; May 2, 2018; July 1, 2018; and September 1, 2018.(Demand Data in Hindsight Policy.xlsx)And we assume that if the fareclass is open on a certain day , we accept all of its demand.

For Hindsight solution, we use linear programming. First, the variables are described as follows:

(1) $i{:}\, the\ fareclass\ number, 1 \le int\ i \le 5;$

(2) $j{:}\, the\ day\ before\ \ arrive, 0 \le int\ j \le 30;$

(3) $x_{ij}{:}\, the\ state\ of\ fareclass\ i\ on\ j\ day\ before\ arrive. x_{ij}\ equals\ to\ 1\ or\ 0.\ (\ 1 = open\ , 0 = closed)$

(4) $x_{ij}{:}\, the\ state\ of\ fareclass\ i\ on\ j\ day\ before\ arrive. x_{ij}\ equals\ to\ 1\ or\ 0.\ (\ 1 = open\ , 0 = closed)$

(5) P_i: the ticket price of fareclass i;

(6) $Sales_i$: the total ticket sold of fareclass i;

(7) D_{ij}: the demand of fareclass i on j day before arrive;

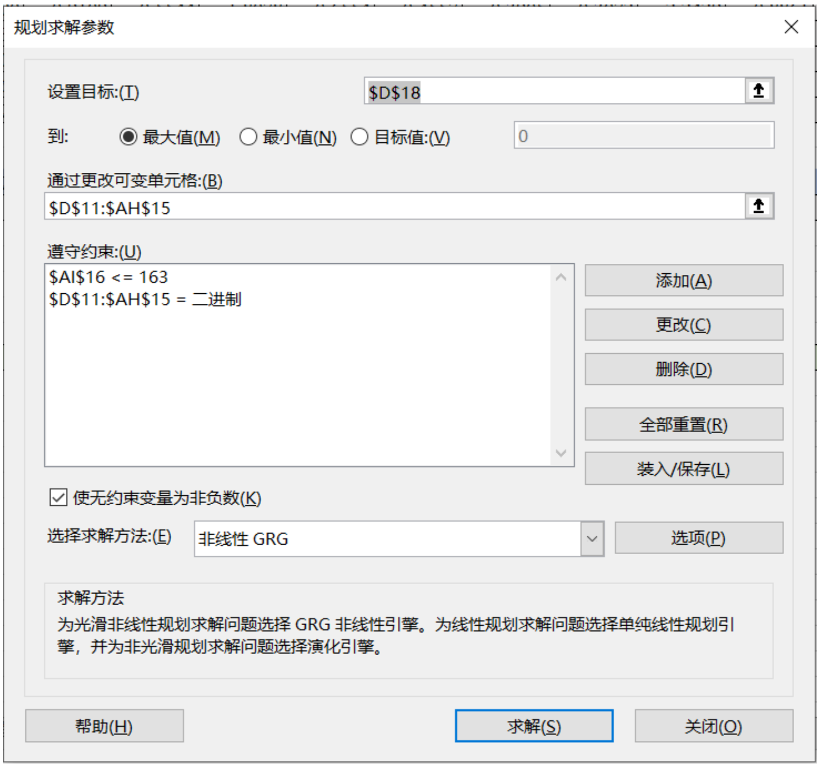
The objective function is:

$$max: \sum_{i=1}^{5} \sum_{j=0}^{30} P_i D_{ij} x_{ij}$$

s.t.

$$\begin{cases} x_{ij} = 1 \ or \ 0 \ , 1 \leq int \ i \leq 5 \ 0 \leq int \ j \leq 30 \\ 0 \leq \sum_{i=1}^{5} Sales_i \leq 163 \end{cases}$$

We put this linear programming constraint into Excel and get the optimal policy and revenue. We use the average revenue of these five days to present the benchmark of hindsight policy and it is 115224.5123 RMB.



Date	Sales	Revenue(RMB)

2018/1/3	162.8292	113859.4808
2018/3/1	162.8298	115374.4912
2018/5/2	162.0435	112470.2104
2018/7/1	162.9874	122770.7387
2018/9/1	162.7642	111647.6402

6.2 Shortcoming & Thinking

Actually, this method is only for Hindsight and can only be used as a reference, without practical significance.

What's more, since there is 0 in the certain demand provided by EM, when facing 0, the x_{ij} given by linear programming equals to 0. Considering the principle that high-priced cabins must be open when low-priced cabins are open in this case, we need to manually modify the situation that high-priced cabins are 0, so as to get the actual optimal policy.

However, you can still see some useful results with the Hindsight results. By seeking the maximum policy for the five days, we can find that our policies are mostly 5 and 4 when maximizing profits (the distribution statistics of the maximum policy for the five days are shown below), which means that Xiamen Airlines can open more low-price seats if it wants to maximize profits. We will further verify this finding in future methods. At the same time, we suspect it may have something to do with the commercial nature of the flight. Business attribute flight focuses more on customer experience, service quality and so on, rather than on the final revenue.

Date\Policy	Policy0	Policy1	Policy2	Policy3	Policy4	Policy5
2018/1/3	0	0	0	0	25	6
2018/3/1	0	1	0	7	17	6
2018/5/2	0	0	0	0	22	9
2018/7/1	1	0	1	4	25	0
2018/9/1	0	0	0	0	17	14

7.Revenue Maximization Policy

7.1 Assumptions

Implementing this strategy requires us to make some assumptions in advance:

1) The simulation results are reliable.

2) Prioritize high-priced demand.

3) The demand of each FareClass will not change as the policy changes.

7.2 Data Process

1) Obtain Demand Data

The original data for our policy control is based on the demand data of each FareClass for 590 days obtained from the last simulation.

Demand

DOW	simulated demand	fare1	fare2	fare3	fare4	fare5
4	2018/1/3	59	32	39	39	30
5	2018/1/4	40	35	45	32	25
6	2018/1/5	59	31	38	37	24
2	2018/1/8	48	33	42	39	28
3	2018/1/9	41	32	32	31	28
4	2018/1/10	49	40	46	34	24
5	2018/1/11	54	29	35	42	29
1	2018/1/14	46	26	36	34	32
2	2018/1/15	41	25	41	30	25
4	2018/1/17	51	31	39	39	27
5	2018/1/18	42	30	40	36	22
6	2018/1/19	43	36	41	35	24
7	2018/1/20	48	29	30	35	26
1	2018/1/21	42	24	33	38	26
2	2018/1/22	49	37	29	40	25
3	2018/1/23	42	34	38	42	22
4	2018/1/24	49	33	42	40	21
6	2018/1/26	55	25	39	24	25
7	2018/1/27	53	33	34	31	29
1	2018/1/28	52	33	39	40	27
2	2018/1/29	48	34	42	37	27
3	2018/1/30	49	28	39	31	26
4	2018/1/31	54	31	34	37	26
5	2018/2/1	44	35	41	25	23
6	2018/2/2	49	32	37	38	29
7	2018/2/3	46	33	41	30	27
1	2018/2/4	49	36	45	43	24
2	2018/2/5	55	34	32	40	27
3	2018/2/6	46	28	43	39	24
4	2018/2/7	50	31	42	33	29
7	2018/2/10	40	34	37	35	26
1	2018/2/11	56	33	42	33	24
2	2018/2/12	41	27	39	34	25
4	2018/2/14	54	35	47	38	21
5	2018/2/15	41	32	44	42	19
6	2018/2/16	44	33	38	39	21
7	2018/2/17	52	31	45	38	28
2	2018/2/26	40	28	41	39	25
3	2018/2/27	54	31	34	38	24
4	2018/2/28	43	32	40	37	24
5	2018/3/1	59	40	42	32	28
6	2018/3/2	52	29	35	44	27

2) Accumulated demand data

the accumulated demand data is obtained by summing-up these demand data according to FareClass1-5.It can be seen that only Fareclss4 and Fareclass5 exceed their seat capacity.That is to say, according to our previous assumption, the first three FareClass requirements can be met. Now we only need to consider selecting a certain number of passengers from FareClass 3 and FareClass 4 to reach the capacity upper limit.

Cumulative Demand

fare1	fare2	fare3	fare4	fare5
59	91	129	168	199
40	74	119	151	175
59	90	128	165	189
48	81	123	163	190
41	73	105	137	165
49	89	135	169	193
54	83	118	160	189
46	72	108	142	173
41	66	107	137	163
51	82	121	161	187
42	71	112	148	171
43	79	120	156	179
48	77	107	141	167
42	67	99	138	164
49	87	115	155	181
42	76	114	156	177
49	82	123	163	184
55	80	119	143	168
53	86	120	151	180
52	85	124	164	192
48	82	124	161	188
49	78	116	148	173
54	86	120	156	182
44	79	120	145	168
49	80	117	156	185
46	79	120	150	178
49	84	129	172	196
55	89	121	161	188
46	74	117	156	180
50	81	123	156	185
40	75	112	147	173
56	89	131	164	188
41	67	106	140	165
54	88	136	174	195
41	73	118	159	178
44	76	114	153	175
52	82	127	165	192
40	68	109	148	173
54	84	119	157	180
43	75	115	152	176
59	98	141	173	202
52	81	116	160	187

3) Mark the accumulated demands

The cumulative requirements are marked so that we can filter out the number of FareClass that need to be met but cannot be met in full.

Tag of Cdemand>163				
fare1	fare2	fare3	fare4	fare5
0	0	0	1	0
0	0	0	0	1
0	0	0	1	0
0	0	0	0	1
0	0	0	0	1
0	0	0	1	0
0	0	0	0	1
0	0	0	0	1
0	0	0	0	0
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	1	0
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	1	0
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	1	0
0	0	0	0	1
0	0	0	1	0
0	0	0	0	1
0	0	0	0	1
0	0	0	1	0
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	1	0
0	0	0	0	1

4) Calculate number of seats to satisfied

we subtracted 163 seat capacity from the demand, so as to get how many seats each Fareclass still has to satisfy the seat capacity.

Number of additional demand				
fare1	fare2	fare3	fare4	fare5
0.0000	0.0000	0.0000	33.8564	0.0000
0.0000	0.0000	0.0000	0.0000	12.1633
0.0000	0.0000	0.0000	34.8982	0.0000
0.0000	0.0000	0.0000	0.0000	0.4656
0.0000	0.0000	0.0000	0.0000	26.3823
0.0000	0.0000	0.0000	28.2128	0.0000
0.0000	0.0000	0.0000	0.0000	3.1680
0.0000	0.0000	0.0000	0.0000	21.4635
0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	2.4875
0.0000	0.0000	0.0000	0.0000	14.7518
0.0000	0.0000	0.0000	0.0000	7.4615
0.0000	0.0000	0.0000	0.0000	21.6250
0.0000	0.0000	0.0000	0.0000	25.1978
0.0000	0.0000	0.0000	0.0000	7.6051
0.0000	0.0000	0.0000	0.0000	7.3298
0.0000	0.0000	0.0000	0.0000	0.0548
0.0000	0.0000	0.0000	0.0000	19.9339
0.0000	0.0000	0.0000	0.0000	11.8091
0.0000	0.0000	0.0000	39.0920	0.0000
0.0000	0.0000	0.0000	0.0000	2.4669
0.0000	0.0000	0.0000	0.0000	15.4339
0.0000	0.0000	0.0000	0.0000	6.7353
0.0000	0.0000	0.0000	0.0000	18.1302
0.0000	0.0000	0.0000	0.0000	7.4220
0.0000	0.0000	0.0000	0.0000	12.5174
0.0000	0.0000	0.0000	33.5472	0.0000
0.0000	0.0000	0.0000	0.0000	2.0029
0.0000	0.0000	0.0000	0.0000	6.8502
0.0000	0.0000	0.0000	0.0000	6.9642
0.0000	0.0000	0.0000	0.0000	16.4307
0.0000	0.0000	0.0000	31.7421	0.0000
0.0000	0.0000	0.0000	0.0000	23.3501
0.0000	0.0000	0.0000	27.3284	0.0000
0.0000	0.0000	0.0000	0.0000	3.5924
0.0000	0.0000	0.0000	0.0000	9.8162
0.0000	0.0000	0.0000	36.2384	0.0000
0.0000	0.0000	0.0000	0.0000	14.7462
0.0000	0.0000	0.0000	0.0000	6.3644
0.0000	0.0000	0.0000	0.0000	10.9179
0.0000	0.0000	0.0000	22.0191	0.0000
0.0000	0.0000	0.0000	0.0000	2.9712

5) Meet the seat capacity

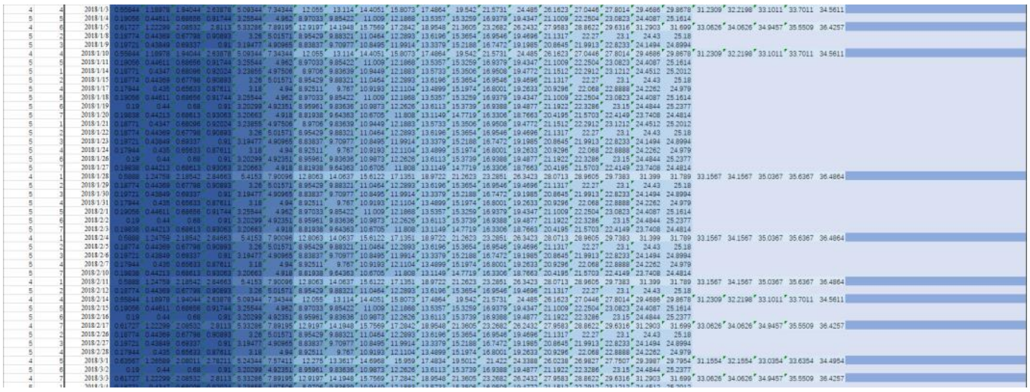
According to the data classified by day of week obtained by our simulation, we made it meet the seat capacity of each day through accumulation

DOW	1	2	3	4	5	6	7	8	9	10	11	12
1	0.588795	0.658795	0.937831	0.661205	2.568675	2.485663	4.905301	1.25747	1.548434	1.522892	1.837108	2.29012
2	0.695	0.64119	0.769524	0.609048	2.638095	2.509524	4.870476	1.243571	1.409524	1.47381	1.689524	2.205238
3	0.581279	0.62186	0.749535	0.645349	2.383721	2.30814	4.445814	1.108023	1.267674	1.297674	1.549767	1.965116
4	0.558444	0.631333	0.750667	0.698333	2.454667	2.25	4.711556	1.059	1.291111	1.402222	1.679111	2.055556
5	0.635667	0.630222	0.814222	0.702	2.461333	2.330667	4.700889	1.086667	1.335111	1.262222	1.524444	2.017778
6	0.617273	0.605714	0.862338	0.725974	2.521558	2.559091	5.027792	1.275065	1.562078	1.527273	1.670649	2.405714
7	0.604125	0.5375	0.86125	0.65175	2.328	2.1665	4.5525	1.20125	1.25275	1.335	1.528	1.9365

DOW	13	14	15	16	17	18	19	20	21	22	23	24
1	2.022771	3.057229	1.729036	0.889157	0.777831	1.660723	0.39	1.367711	1	0.88	0.6	0.849639
2	1.949167	2.9175	1.680952	0.892857	0.741905	1.624762	0.380714	1.363333	1	0.881429	0.6	0.86
3	1.8	2.717791	1.665116	0.89186	0.759535	1.626047	0.383488	1.35907	1	0.871163	0.586047	0.86
4	2.031111	2.911889	1.677333	0.882222	0.756889	1.667111	0.399222	1.363111	0.988889	0.881333	0.6	0.86
5	1.920778	2.916778	1.685	0.958889	0.768	1.648	0.396778	1.36	1	0.88	0.6	0.86
6	1.907662	2.975065	1.715065	0.903896	0.769351	1.658701	0.408701	1.363636	1	0.883117	0.605195	0.874805
7	1.867	2.739625	1.6805	0.87875	0.7565	1.6125	0.375375	1.343	1.0125	0.883	0.605	0.86175

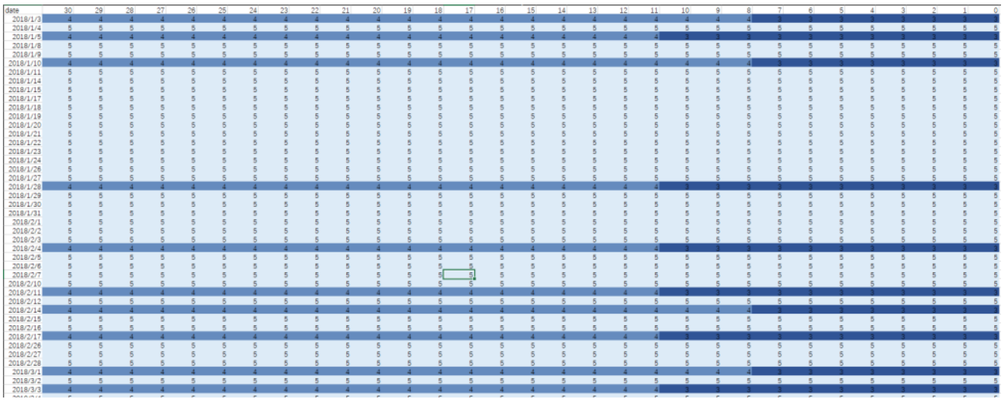
	1	2	3	4	5	6	7	8	9	10
1	0.187711	0.246988	0.246265	0.239277	2.318313	1.736506	3.995542	0.865783	1.108554	1.243373
2	0.187738	0.255952	0.234286	0.230952	2.351071	1.755714	3.938571	0.928929	1.163214	1.242857
3	0.197209	0.241279	0.254884	0.216628	2.284767	1.714884	3.928721	0.871395	1.139767	1.14186
4	0.179444	0.255556	0.221333	0.219778	2.303889	1.76	3.985111	0.841889	1.152333	1.191111
5	0.190556	0.255556	0.240444	0.230889	2.338	1.706556	4.008333	0.883889	1.154778	1.177778
6	0.19	0.25	0.24	0.23	2.292987	1.720519	4.036104	0.876753	1.150909	1.275325
7	0.198375	0.24375	0.244	0.2445	2.276	1.711375	3.901375	0.82425	1.026875	1.1375

	11	12	13	14	15	16	17	18	19
1	1.38494	1.777349	1.600241	2.526386	1.673976	1.14	0.83	1.33	0.75
2	1.330357	1.745714	1.589286	2.515	1.662024	1.138333	0.83	1.33	0.75
3	1.346512	1.88093	1.528372	2.451279	1.666047	1.126744	0.831977	1.326163	0.75
4	1.379444	1.707556	1.602667	2.463222	1.666222	1.138444	0.820778	1.337444	0.752778
5	1.348889	1.790222	1.612	2.496778	1.666222	1.149556	0.831889	1.326333	0.752778
6	1.348701	1.762597	1.564935	2.548831	1.704545	1.136364	0.821429	1.334416	0.753247
7	1.306875	1.657	1.55875	2.435625	1.65325	1.15075	0.844625	1.325875	0.740625

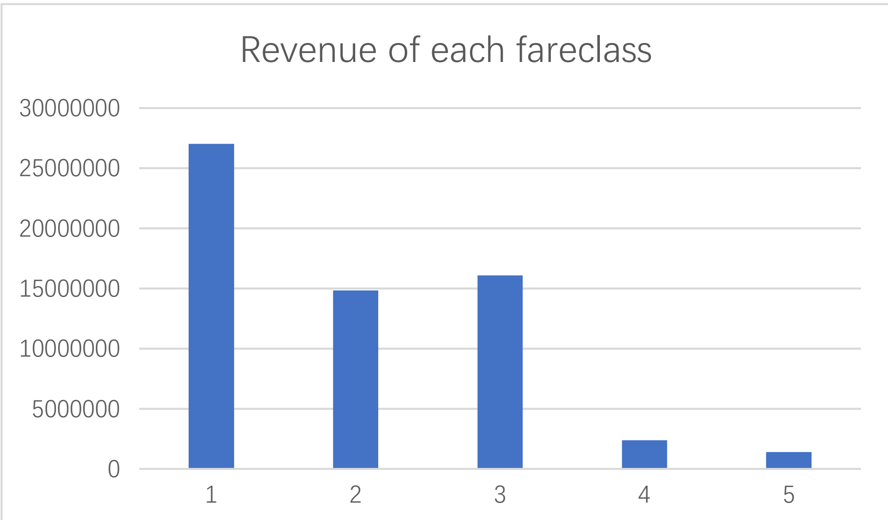


7.3 Policy

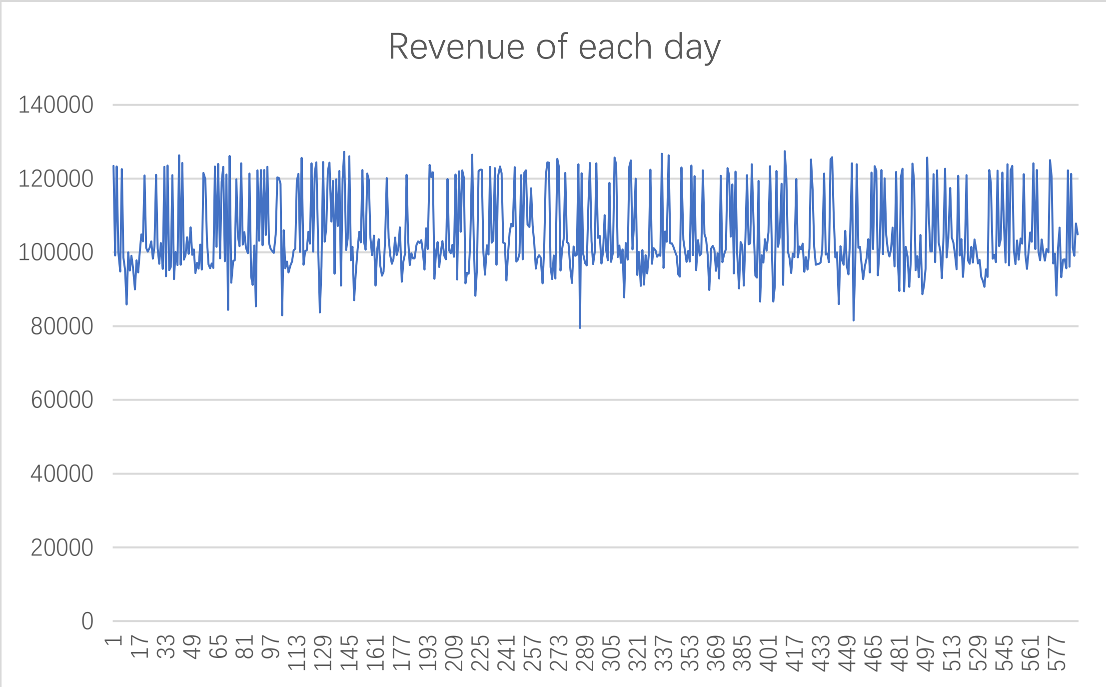
After a series of data processing, we get the policy under specific conditions. The characteristics of this policy are as follows: First, it can fully meet the demand of high price tickets. The second is a strategy that maximizes a company's revenue to a certain extent. The calculation of the revenue is described in the remaining sections.



7.4 Revenue



When we calculate revenue according to the policy, we can see that the revenue of each fareclass is gradually decreasing. This is because we first meet the demand of high ticket price.



Revenue of each day is within the range of 80,000-120,000, and the average revenue of each flight is 104651.32.

8. Expected Marginal Seat Revenue (EMSR)

8.1 Protection level

Accept a request if and only if its revenue exceeds the expected marginal seat revenue (EMSR), or bid price, displacement cost, opportunity cost (the expected loss in future revenue from using the capacity now rather than reserving it for future use). EMSR is suitable for two-class model, but XM airline has five classes. So that we use EMSR-b to calculate protection level in XM airline case. Actually, EMSR and EMSR-b are similar.

Step 1: EMSR-b lets the lowest price class as a part, and the aggregates the others as a part. After that we change multiple-class model into two-class model. We will use EMSR's formulas to work protection level out.

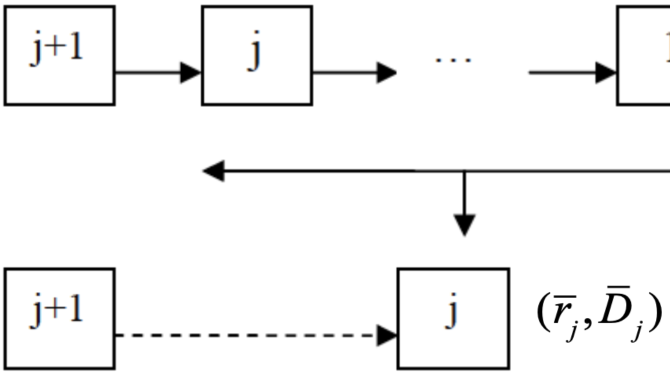


Figure 8-1 EMSR-b Introduction

$$\bar{D}_j = \sum_{k=1}^{j} D_k, \quad \bar{r}_j = \frac{\sum_{k=1}^{j} r_k \mathrm{E}[D_k]}{\sum_{k=1}^{j} \mathrm{E}[D_k]}$$

$$\mathrm{P}(\bar{D}_j > y_j) = \frac{r_{j+1}}{\bar{r}_i}$$

$$\mu = \sum_{k=1}^{j} \mu_j, \quad \sigma^2 = \sum_{k=1}^{j} \sigma_k^2.$$

Figure 8-2 EMSR-b formula

Step 2: And then we will repeat the Step 1 until we get n-1 protection level. The XM airline result as follow:

Step 1

Protection Level	fare1&2&3&4
rj	741
r(j+1)/rj	0.47
Dj	156.72
Standard Deviatio	9.74
y(1&2&3&4)	157

Step 2

Protection Level	fare1&2&3
rj	811
r(j+1)/rj	0.62
Dj	121.08
Standard Deviatio	8.46
y(1&2&3)	119

Step 3

Protection Level	fare1&2
rj	880
r(j+1)/rj	0.77
Dj	80.66
Standard Deviatio	7.27
y(1&2)	75

Step 4

Protection Level	fare1
rj	950
r(j+1)/rj	0.82
Dj	48.22
Standard Deviatio	6.10
y(1)	43

Figure 8-3 Protection level calcution process

Protection Level	fare1	fare2	fare3	fare4	fare5
fare1&2&3&4	157				
fare1&2&3	119				
fare1&2	75				
fare1	43				

Figure 8-4 Protection level result

In order to compare EMSR-b policy and old policy, we calculate each fare class protection level. And then comparing with simulated sales. Simulated sales mean that we impose old policy into forecasting demand. And then calculating their revenue.

	fare1	fare2	fare3	fare4	fare5
Protection Level	43	33	43	39	6
Simulated sales	45	15	16	11	3

Figure 8-5 Different policy's sales

	Revenue	
Simulated Demand	125104	Up Bound
Protection Level	115284	New Policy
Simulated Sales	71440	Reality

Figure 8-6 different policy's revenue

From these tables, we can easily find that new policy performs much better than old policy. And the reason is that new policy gets more revenue from fare class 2&3&4&5.

I think XM airline's salesmen know there are some demand in lower price class. But they reject this part of demand. I think the first reason is that reality is complex. Our new policy ignores buy up, buy down and customer choice and so on. Second reason is If XM airline often sells the ticket at a low price, it is hard to sell ticket at a high price in the future. So that our new policy may be a good suggestion for them, but not a obviously good method in the reality.

8.2 EMSR-b under DOW

In the previous part, we got same protection level for every day. In this part, we want to make an improvement, so we recalculate protection level under day of week.

fare1&2&3&4	Monday	Tuesday	Wednesda	Thurday	Friday	Saturday	Sunday
rj	752	735	730	733	733	757	729
r(j+1)/rj	0.47	0.48	0.48	0.48	0.48	0.46	0.48
Dj	174.32	147.77	135.35	142.90	142.50	181.83	133.40
Standard Deviation	8.55	8.21	7.91	8.10	8.16	8.75	8.10
y(1&2&3&4)	175	148	136	143	143	183	134
fare1&2&3	Monday	Tuesday	Wednesda	Thurday	Friday	Saturday	Sunday
rj	818	810	806	807	808	821	807
r(j+1)/rj	0.61	0.62	0.62	0.62	0.62	0.61	0.62
Dj	137.84	112.12	101.90	108.34	108.00	145.40	99.82
Standard Deviation	8.02	7.53	7.26	7.43	7.46	8.26	7.36
y(1&2&3)	136	110	100	106	106	143	98
fare1&2	Monday	Tuesday	Wednesda	Thurday	Friday	Saturday	Sunday
rj	884	879	876	877	877	885	877
r(j+1)/rj	0.76	0.77	0.77	0.77	0.77	0.76	0.77
Dj	94.60	74.04	66.19	70.90	71.07	100.85	65.06
Standard Deviation	7.00	6.49	6.23	6.37	6.37	7.25	6.33
y(1&2)	90	69	62	66	66	96	60
fare1	Monday	Tuesday	Wednesda	Thurday	Friday	Saturday	Sunday
rj	950	950	950	950	950	950	950
r(j+1)/rj	0.82	0.82	0.82	0.82	0.82	0.82	0.82
Dj	58.78	43.99	38.35	41.27	41.33	63.56	37.88
Standard Deviation	5.75	5.22	4.95	5.04	5.01	5.95	5.07
y(1)	54	39	34	37	37	58	33

Figure 8-7 Protection level under DOW

	fare1	fare2	fare3	fare4	fare5
Monday	54	36	46	27	0
Tuesday	39	30	41	38	15
Wednesda	34	28	38	36	27
Thurday	37	29	40	37	20
Friday	37	30	39	37	20
Saturday	58	37	47	20	0
Sunday	33	27	37	36	29

Figure 8-8 EMSR-b under DOW's result

From the sales data, we found week is important factor in XM airline demand forecasting. So that same protection level for each day is unreasonable. We believe protection level under DOW will performs much better than same protection level.

8.3 EMSR-b's shortcoming

EMSR-b's assumption is that low-fare demand arrives before high-fare demand. But in XM airline case, low-fare demand and high-fare demand arrived at the same time. They arrived at the begin of booking period. It means EMSR-b will ignores some low-fare demand when we use it in the XM airline case.

9.EMSR_b With Dynamic

9.1 EMSR-b with dynamic

After we have used the EMSR-b to solve the allocation. However, we find that if we can do every EMSR-b after we confirmed one class allocation. We can get the optimal allocation for every class.

The formula can be rewritten for every class:

$$fromclass_1 toclass_n$$

$$foreveryclass_i fromi = ntoj$$

We first combine them as the first class, and from j-1 to 1 as the class two.

After that, we can use the two class module to solve the problem.

Have assumed that every class is normally distributed and independent, we can use the price as the weight to calculate the $\mu and \sigma$:

$$\mu_1 = \frac{\sum_{i=n}^{i=j} \mu_i w_i}{\sum_{i=n}^{i=j} w_i} \quad \mu_2 = \frac{\sum_{i=j-1}^{i=1} \mu_i w_i}{\sum_{i=j-1}^{i=1} w_i} \quad \sigma_k^2 = \sum_{i=n}^{i=j} w_i^2 \sigma_i^2 (k = 1,2)$$

As a result, at the beginning, from $class_i fromi = 1ton$, we can calculate the protection level.

One of the assumption that we have assumed is that we from low class to high class. After we received the actual sales. We can use the capacity minus the sales and minus the class in order to use the method again.

Strength：

Improvement compared with the former method is that we use the information we have got to optimize the revenue. Since the target is simple: $\pi(B_i)$. We can use the $latter\pi(B_i) - former\pi(B_i)$ to evaluate the value of the information.

10.Revenue Comparison

10.1 Process

This part includes four steps, says, first step is to take the estimates of the demand obtained in the previous phase as the real demand distribution, above up, we chose the estimates given by EM method which considered the effect of DOW. Second step is to generate the random variates which obey the normal distribution we assumed before. Third step required us to modify the raw data given by the previous step so that it can be interpreted in reality (as there are some negative numbers and almost all numbers are decimal). The last step is to apply the sales policy to get the revenue of every flight.

Now, we will detail each step in the following section.

10.2 Generate random variates

As we had unconstrained the data and got the real demand by several methods. In the newest method, we use EM algorithm and take the factor of DOW into account, then, we get the mean and standard deviation of demand of each day in week.

weekday	30	29	28	27	26	25	24	23	22	21	20
1	0.2409639	0.1325301	0.4819277	0.253012	1.3253012	1.2168675	2.4698795	0.5273494	0.6626506	0.8313253	1.0361446
2	0.2619048	0.1785714	0.2857143	0.1547619	0.8214286	0.6309524	1.5238095	0.2130952	0.297619	0.4047619	0.3390476
3	0.2209302	0.1627907	0.1395349	0.127907	0.6395349	0.6162791	1.1744186	0.1766279	0.2209302	0.2662791	0.255814
4	0.1777778	0.0888889	0.1555556	0.0888889	0.8777778	0.6	1.3333333	0.1988889	0.3266667	0.3444444	0.3662222
5	0.2	0.1111111	0.2	0.1	0.8222222	0.6555556	1.4	0.21	0.2666667	0.3705556	0.3328889
6	0.2727273	0.2857143	0.5454545	0.4155844	1.7402597	1.6990909	3.2433766	0.6053247	0.9064935	0.7779221	0.9350649
7	0.225	0.1875	0.1625	0.125	0.6	0.5	1.225	0.13625	0.1975	0.1875	0.1625
1	0.5082337	0.3751775	0.6121025	0.4644266	1.1487898	1.0482063	2.0442504	0.6985282	0.8156564	0.9476451	1.1092272
2	0.4423118	0.4434453	0.4802466	0.3638498	0.7630111	0.7569718	1.5482302	0.4110902	0.5967737	0.8231448	0.6817264
3	0.4446059	0.4017744	0.3807976	0.3359451	0.7180495	0.7695469	1.1292044	0.4632636	0.4446059	0.4887825	0.464911
4	0.4126586	0.2861776	0.3644639	0.2861776	0.9219883	0.8585328	1.08099	0.3714667	0.5564333	0.5641163	0.6736628
5	0.4292664	0.3160303	0.4787624	0.3016807	0.8814923	0.6730963	1.2702296	0.40878	0.5363212	0.5829937	0.6490048
6	0.5533716	0.4547163	0.6795016	0.6145313	1.3018468	1.479673	2.6532669	0.7655641	0.9462304	0.7678359	1.0174486
7	0.4766709	0.423779	0.3712364	0.3688847	0.9084693	0.7291403	1.1904153	0.3441349	0.5056579	0.423779	0.4038972

Figure 10-1: The mean and standarad deviation of demand of each day in weel.

Since we had assumed that the demand in each DBA of each day in week obey the normal distribution, thus, we generate a set of demand observations to represent the actual demand of each fare class (we use the own function of Matlab to do this). The result is as follows:

date	30	29	28	27	26	25	24	23	22	21	20
2018/1/3	0.0645	0.4694	-0.0528	0.3427	0.0986	-0.7270	0.1600	-0.0743	0.6209	-0.0927	0.1562
2018/1/4	-0.1197	-0.0036	0.8323	0.0801	1.1553	0.4045	0.9628	0.0729	-0.7774	1.1490	1.2555
2018/1/5	0.3337	0.1080	-0.1926	0.3106	2.2702	0.0355	1.2184	-0.9027	1.4019	-0.2345	2.8323
2018/1/8	0.1159	0.4594	0.5665	-0.2160	0.5277	1.5865	1.7176	0.9927	-0.5080	0.0920	0.4265
2018/1/9	-0.1207	0.6972	-0.0829	-0.0924	0.3283	2.4196	0.6540	0.7529	0.0489	0.4427	-0.0022
2018/1/10	-0.7392	0.1195	0.1896	-0.3501	2.9200	0.7782	1.1823	0.8149	-0.6183	1.1147	0.4049
2018/1/11	-0.0335	0.2520	1.2342	0.1396	0.2905	0.8115	2.2912	-0.1268	0.1094	0.2396	0.5179
2018/1/14	-0.2311	-0.0416	-0.3808	0.4793	2.7340	0.5480	4.1270	0.1355	-0.4096	1.3218	0.0541
2018/1/15	0.1301	0.1004	-0.0916	0.9467	1.5100	0.4439	2.9674	-0.0673	-0.5452	1.9137	0.2788
2018/1/17	0.4872	-0.3982	0.4403	0.0635	1.0737	1.2072	3.4012	-0.0576	-0.1527	0.7962	-0.0671
2018/1/18	0.6897	0.2612	1.0119	0.2392	0.2248	-0.0874	2.9440	1.1659	0.8672	-0.2696	0.6339
2018/1/19	-0.3927	0.4910	0.6668	0.6646	1.3181	-0.9879	1.0291	0.6363	2.4548	0.7521	1.6950
2018/1/20	-0.9899	0.1126	-0.0850	-0.0687	-0.5261	0.4431	0.8533	0.4625	-0.2440	-0.0587	-0.8245
2018/1/21	1.3706	0.3325	0.8952	0.0540	1.9954	-0.7514	2.1449	-0.2950	-0.8109	-0.0039	1.2130
2018/1/22	0.2581	-0.7915	0.7945	-0.3015	0.0085	0.6235	3.0365	0.1778	0.0482	-0.9306	0.1209
2018/1/23	1.5337	0.2181	0.4312	0.1214	0.0590	0.2371	2.8646	1.2053	-0.1637	0.7163	0.5719
2018/1/24	-0.6795	0.2920	-0.7496	-0.4061	1.0812	-0.1058	2.5463	0.2902	-0.5505	1.4341	0.1171
2018/1/26	-0.2693	0.2055	-0.2952	-0.6653	0.7559	3.6889	2.8175	0.4998	0.9923	0.2538	1.2524
2018/1/27	-0.2975	0.3733	-0.0867	0.0360	0.5555	-0.2762	2.9502	-0.1993	-0.8805	0.5349	0.3062
2018/1/28	-0.1370	0.0091	1.1939	0.2684	3.1948	1.2481	2.0327	1.0240	-0.0045	-0.6019	3.0022
2018/1/29	0.3101	-0.2009	0.3563	0.1390	0.3347	0.8571	4.7456	-0.0201	-0.2108	-0.0639	0.0829

Figure 10-2: The simulated demand of each fare class in each day.

10.3 Modify the data

As the above, a negative demand or the decimal form of the demand for air tickets is not reasonable in reality, so there should be some modification for raw data to process the following operations.

1. Eliminate the negative

A negative value is reasonable in all normal distribution, however, in a normal distribution with a nonnegative mean, the probability to incur a negative value is small, thus, we roughly eliminate the effect of negative "demand" by replace negative random variates with 0. Considered that this process would influence the results within a little scale.

2. Eliminate the decimal

Almost all random variates are decimals, it should be cautious to handle on them. To minimize the effect, we take the following measures to eliminate decimals.

- Perform Gaussian rounding of the demands of each DBA.
- Add up the fractional parts of demand of each DBA chronologically, once the value of sum exceeds 1, add 1 to the demand of current DBA.
- After the above modifications, we have following results:

date	30	29	28	27	26	25	24	23	22	21	20
2018/1/3	0	1	1	1	2	1	1	1	3	2	1
2018/1/4	0	0	0	0	1	1	1	1	1	1	1
2018/1/5	0	0	1	1	1	4	4	1	2	2	2
2018/1/8	0	1	1	1	1	1	2	1	1	1	1
2018/1/9	0	1	1	1	1	1	1	1	1	1	1
2018/1/10	0	0	0	0	0	0	1	1	1	1	1
2018/1/11	0	0	0	0	0	1	2	1	1	1	1
2018/1/14	0	0	0	2	1	4	6	1	1	1	2
2018/1/15	0	0	0	1	1	1	3	1	2	1	1
2018/1/17	0	0	1	1	1	1	1	1	1	1	1
2018/1/18	0	0	0	0	2	1	2	1	1	1	1
2018/1/19	1	0	1	2	1	3	5	1	2	1	2
2018/1/20	0	0	0	0	1	0	2	1	1	1	1
2018/1/21	0	1	1	1	2	2	5	1	2	2	2
2018/1/22	0	0	0	0	1	2	3	1	1	1	1
2018/1/23	0	1	1	1	2	1	2	1	1	1	2
2018/1/24	0	0	0	0	1	1	3	1	2	1	1
2018/1/26	0	0	1	2	2	2	3	1	2	2	1
2018/1/27	1	0	0	0	0	0	1	1	1	1	1
2018/1/28	0	0	0	0	2	1	3	1	1	2	1

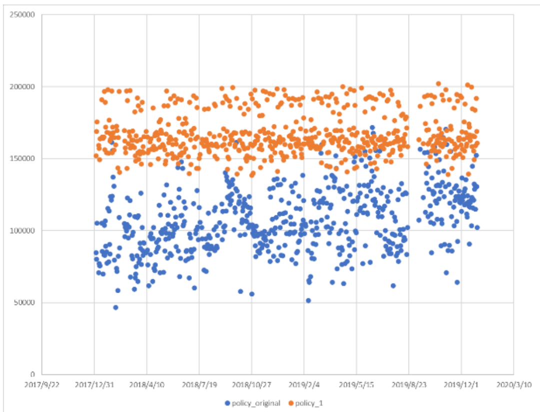
Figure 10-3: Demand after modification

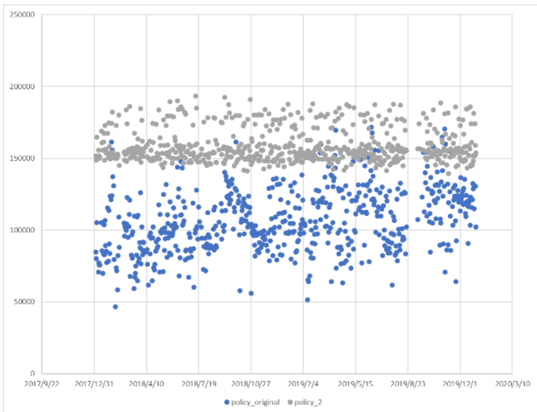
10.4 Revenue comparison

We get the policy from hindsight and EMSR_b respectively, in the following, we denote policy given by hindsight and EMSR_b as *policy_1* and *policy_2*, and original policy is denoted as *policy_original*.

Three sets of demand are generated in the previous step. Under the same level of demand, different sales policies are applied, and we calculate the average total revenue under each policy.

Figure 4: Revenue comparison of each policy.





Both *policy_1* and *policy_2* have led to huge increases in revenue. However, it is tough to have such progress in reality, we attributed the huge promotion in revenue to the ignorance of customers' behaviors of buy-down and no-show. Both will lead to a higher revenue in results. Thus, there are still room for improvement in the model to make it more realistic.