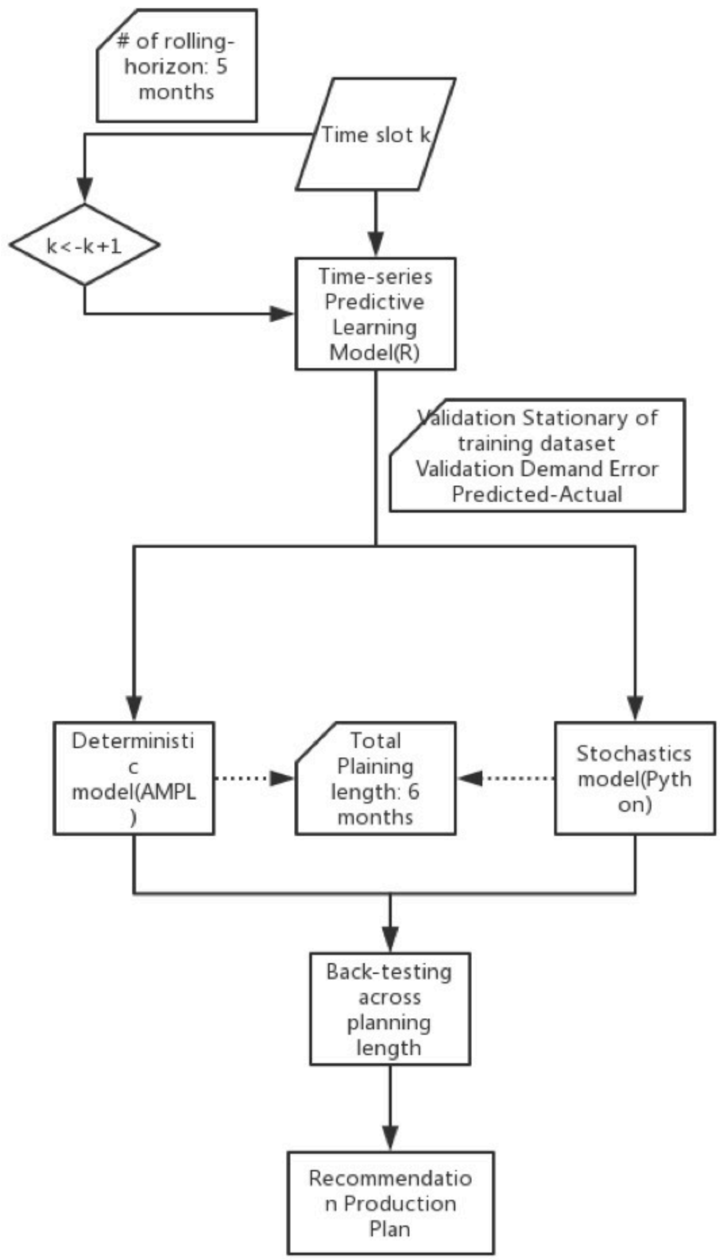
Inventory Time Series Analysis Project

Shuting Chen, Qian Wu

1. Introduction

To manage inventory of electric equipment, time series rolling horizon inventory model is used to make order plan and to optimize the objective, which is to minimize the expected cost of holding inventory plus expected cost of lost sales. The below chart shows the process of this project.

Figure 1 Workflow of project



2. Data Manipulation

2.1 Input Dataset

The original data set is ELECEQUIP, which contains information about the demand of electrical equipment over a ten-year period. The demand dataset is divided into two segments, first 5 years for training, and the rest for validation.

2.2 Data cleaning and training (R)

Use tsclean function to identify and replace outliers and missing values in time series training dataset. In order to exam the trend stationary for train data, we use kpss.test function to test All the p-value of kpss.test for each iteration are greater than 0.1, the result shows that the training datasets are stationary. So, we can use auto.arima function to fit training dataset then predict the demand, we choose time slot as 5 which gives us the predict demand for the next 5 months. And we include one more month in our train dataset each time and predict demand in the next 5 months. The times we repeat depends on how many months we want to validate. And this time, we choose 6 months to evaluate the performance of our model through validating the costs generated by inventory holding or costs of lost sale.

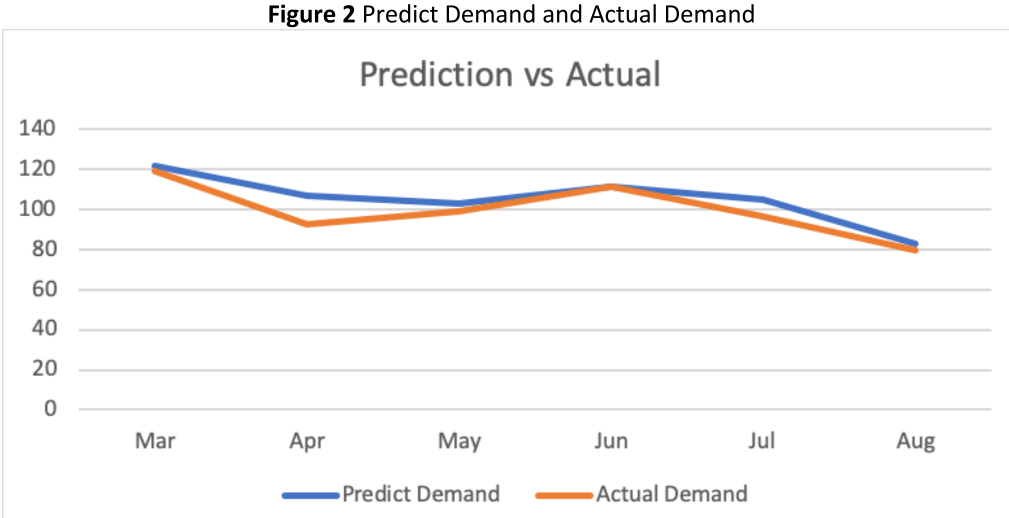
2.3 Output Data

Two output files will be generated after doing the prediction. First file is the predict demand of the next 5 months, and each time we repeat the data training process, it appends a new row which represents the predicted demand for the next 5 months from the end time of train data. Second file is the error terms, associated with the time-series model and obtained during the training phase, which is also being appended each time we repeat the data training process. The error term will be used in the stochastics model.

2.4 Output Data Validation

Before we input the predict demand data into the optimization model, we do the error term validation to check the difference of predict demand and actual demand. Comparison is shown in figure 2. Since the difference of them are small, the performance of prediction is good enough and predict demand data can be used to build optimization model.

Figure 2 Predict Demand and Actual Demand



3. Deterministic Model Creation

The objective function is to minimize the costs produced by the expense of holding extra inventory and lost sales. The constraints are about limitation on order number and relationships between start inventory and end inventory. The optimization program [1] is expressed as the following:

$$\begin{aligned} &\text{Min E } [\textstyle\sum_{t=0}^{T-1} h_t x_t + b_t z_t] \\ \text{s.t.}\quad & y_{t+1} - x_t - \Delta_t = 0 \\ & y_{t+1} \leq R_{t+1} \\ & \Delta_t \leq U_t \\ & -y_t + x_t \geq -D_t \\ & y_t + z_t \geq D_t \\ & x_t, y_t, z_t \geq 0 \end{aligned}$$

y_0 is given and represents the start inventory of first month, y_t represents the start inventory of t month, x_t represents the end inventory of t month, Δ_t represents the order plan of t+1 month, z_t represents lost sales of t month, R_{t+1} represents storage capacity of t+1 month, U_t represents delivery capacity of t month, the last three constraints can also be expressed as $x_t = Max(0,\ y_t - D_t)$, $z_t = Max(0, D_t - y_t)$, which also represents that the cost is either generated by the cost of holding inventory or lost sales.

By using the predict demand and start inventory as input of our deterministic model build in AMPL, we can calculate the actual cost of the next month. Besides, repeat this process 6 times we can get the back-testing cost of 6 months to evaluate the performance of deterministic model.

4. Stochastics Model Creation

The difference between the stochastics model and the deterministic model is that the stochastics model introduces the error terms, which is associated with the time-series model. Considering the error term, the stochastic program [1] could be expressed as the following:

$$\begin{aligned} &\text{Min E } [\textstyle\sum_{t=0}^{T-1} h_t x_t(\widetilde{w}) + b_t z_t(\widetilde{w})] \\ \text{s.t.}\ & y_{t+1}(w) - x_t(w) - \Delta_t(w) = 0 \\ & y_{t+1}(w) \leq R_{t+1}(w) \\ & \Delta_t \leq U_t \\ & -y_t(w) + x_t(w) \geq -D_t(w) \\ & y_t(w) + z_t(w) \geq D_t(w) \\ & x_t(w), y_t(w), z_t(w) \geq 0 \end{aligned}$$

5. Comparison of two model

After the deterministic and the stochastic program are built, AMPL and python are used to get final results. In this part, we compare the results gotten from two programs. We can see from the below chart that the cost of two model generated are pretty close each month, except for the June.

Figure 3 Cost generated by Deterministic model and Stochastics model

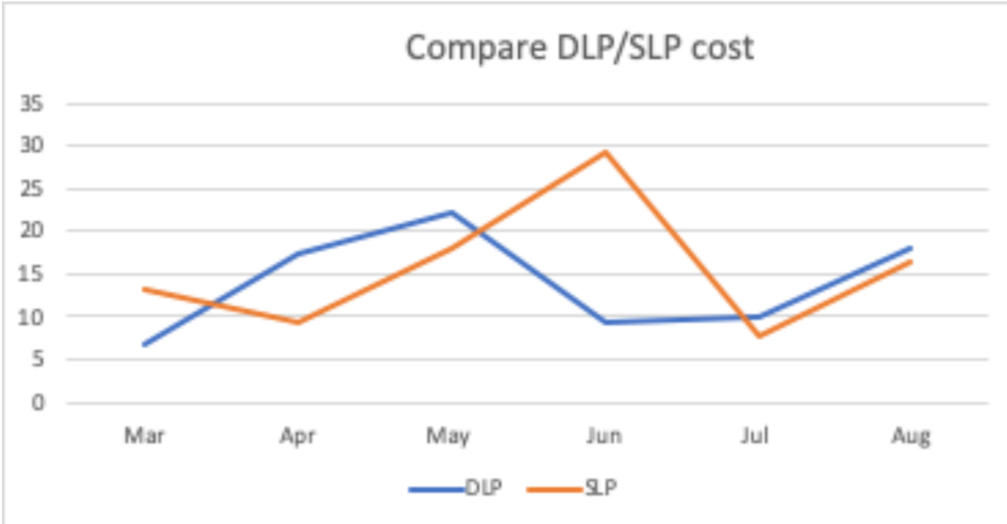


Table 1 Monthly Back-Testing Costs

	Mar	Apr	May	Jun	Jul	Aug
DLP cost	6.80	17.26	22.14	9.38	9.92	18.03
SLP cost	13.03	9.44	18.08	29.35	7.57	16.33

And we want to know it is generated by cost of holding or cost of lost sales. Then we also plot the below chart to show us the relationship between start inventory and actual demand.

Figure 4 Begin Inventory and Actual Demand of DLP

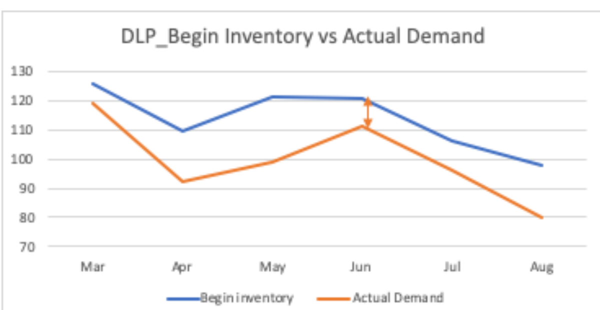
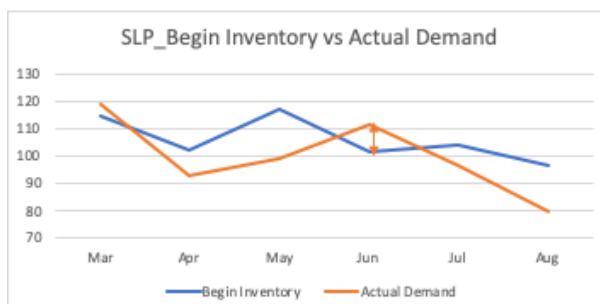


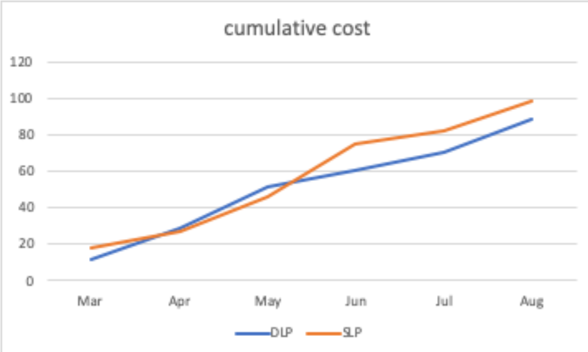
Figure 5 Begin Inventory and Actual Demand of SLP

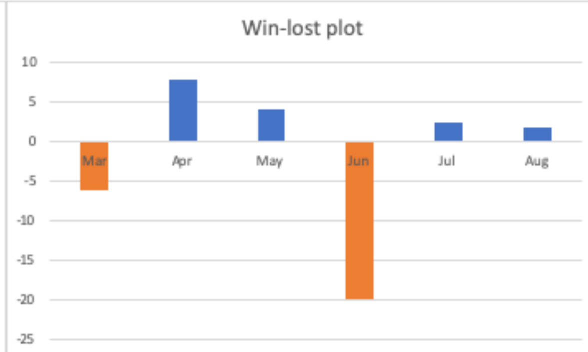


Since we know that the holding cost will be generated when the start inventory is greater than actual demand, cost of lost sales will be generated when the start inventory is smaller than actual demand. From the chart we know that except for February the cost of deterministic model generated are all the holding cost. However, for stochastics model, in February, March and June the costs are all generated by lost sales. And even the difference of start inventory and actual demand are almost same amount, we know that the coefficient of cost of lost sale per unit is three times of cost of holding, so that can explain the high cost in June generated by lost sales in stochastics model.

To further compare the two model, we also plot the cumulative cost and win lost plot.

Figure 6 Cumulative cost of DLP and SLP **Figure 7** Win-lost plot of DLP and SLP





In conclude, looking at the cumulative cost we know that the cost generated by using the stochastics model is higher than the deterministic model, but combine with win lost plot we know that only for March and June the deterministic model wins. To be specific, the deterministic model wins a lot in June. So far, by using the 6 months for validate the cost it generated, we still hard to tell which model is perform better, and it should be validated in the longer period.

Reference

[1] Jiajun Xu. "ELECEQUIP." *Record*, 2019
https://core.isrd.isi.edu/chaise/record/#1/Core:Instance/RID=W28P

Appendix

Table 2 Detail information of Deterministic model

Deterministic model	Feb	Mar	Apr	May	Jun	Jul	Aug
Begin inventory	101.38	125.86	109.72	120.89	120.52	106.05	97.75
Actual demand	103.05	119.06	92.46	98.75	111.14	96.13	79.72
End Inventory	0.00	6.80	17.26	22.14	9.38	9.92	18.03
Holding cost	0.00	6.80	17.26	22.14	9.38	9.92	18.03
Lost sales	1.67	0.00	0.00	0.00	0.00	0.00	0.00
Lost sales cost	5.01	0.00	0.00	0.00	0.00	0.00	0.00
Total cost	5.01	6.80	17.26	22.14	9.38	9.92	18.03
Order	125.86	102.92	103.63	98.38	96.67	87.83	

Table 3 Detail information of Stochastics model

Stochastics model	Feb	Mar	Apr	May	Jun	Jul	Aug
Begin inventory	101.38	114.72	101.90	116.83	101.36	103.70	96.05
Actual demand	103.05	119.06	92.46	98.75	111.14	96.13	79.72
End Inventory	0.00	0.00	9.44	18.08	0.00	7.57	16.33
Holding cost	0.00	0.00	9.44	18.08	0.00	7.57	16.33
Lost sales	1.67	4.34	0.00	0.00	9.78	0.00	0.00
Lost sales cost	5.01	13.03	0.00	0.00	29.35	0.00	0.00
Total cost	5.01	13.03	9.44	18.08	29.35	7.57	16.33
Order	114.72	101.90	107.39	83.28	103.70	88.49	