The Price Forecast Model of China-made Large Aircraft Based on Partial Least Squares Regression

Pei Fang[1,a],Lu Wang[2,b] and Shuhua Mao[3,c]

[1] Wuhan University of Technology, College of Computer Science & Technology, Peace Avenue 1040, Wuhan, Hubei Province, China

[2] Wuhan University of Technology, School of Economics, Luoshi Road 122, Xiongchu Avenue, Wuhan, Hubei Province, China

[3] Wuhan University of Technology, School of Science, Peace Avenue 1040, Wuhan, Hubei Province, China

[a]470810840@qq.com, [b]295276730@qq.com, [c]651332723@qq.com

Keywords: domestic large aircraft; price forecast; partial least squares regression; discount factor; mathematical model

Abstract. This article adopts partial least squares regression method to forecast the price of domestic (China-made) large aircraft C919. With analysis of the pricing background and the relationship between performance index and price of foreign aircraft, we build preliminary model. And comparing with other modeling methods by calculating average relative error, we find the preliminary model fits better. Considering the difference in environment between domestic and foreign market, we propose the concept of discount factor. And the factor is equal to 0.6529 by analyzing domestic and foreign small aircrafts of the same type. Finally through modified model we get the forecast price, namely 48.95 million dollars per aircraft. The forecasting model and discount factor have some reference value to the price forecast of domestic large aircraft.

Introduction

C919 is Chinese self-developed large aircraft. The project of it started in November, 2008. Its first flight is expected in 2014. And it's planned to obtain airworthiness certificate and put into service in 2016. When developing a new aircraft, we should consider lots of economic problems besides technique details. Among them, the questions whether and how long the aircraft manufacturer can recover the original investment outstand. To answer these questions, price forecast of the new aircraft is necessary.

Partial least squares regression (hereinafter referred to as PLSR) can screen independent variables briefly. Its rich aided analysis technique can achieve screening while modeling and overcome bad influence caused by variables' multiple correlation [1].

In literature [2,3] by Li Shouan etc(2005,2006), the comparison among PLSR, regression without screening variables and stepwise multiple regression shows that PLSR is more accurate in price forecast of military aircraft than the other two models. Compared with price forecast relied on manufacturing cost and period expenses, the similar aircraft price forecast model based on analysis of military aircraft in literature [4] by Li Shouan etc (2007) is better. The price forecast background of the domestic large aircraft C919 is very similar to military aircraft. They both have a small number of sample data and a large number of price driving factors, as well as having few types. Besides, neither of their data collection is sound and their cost of development and manufacturing isn't available to the public.

In literature [5] by LI Zhanke etc (2011), PLSR is adopted to forecast the price of C919, but the value of equalization-engineering-value-ratio is much too subjective and brief. In this article, we also apply PLSR to forecast the price of C919. And combining with China's actual situation, we deal with domestic small aircraft with PLSR, calculate discount factor, get modified model and obtain reasonable forecast price in ensured market competitiveness. Through the above work, we expect to provide some guidance for the price forecast of domestic large aircraft.

Introduction of Regression Model

General linear relationship and log-linear relationship are selected to analyze here. To choose better model, we compare the fitting precision of the two methods.

General linear regression form is as follows

$$y=\alpha_0+\alpha_1 x_1+\alpha_2 x_2+......+\alpha_n x_n \quad (1)$$

Log-linear regression form is as follows

$$\ln y=\beta_0+\beta_1 \ln x_1+\beta_2 x_2+......+\beta_n x_n \quad (2)$$

Considering modeling problem with single dependent variable y and p independent variables $x_1, x_2, ..., x_p$, the basic practice of PLSR is shown below. Primarily, extract the first component t_1 (t_1 is the linear combination of $x_1, x_2, ..., x_p$, and contains as much original information as possible. It's required that the relevance between t_1 and y reaches the maximum. Then establish the regression equation of dependent variables y and t_1. If its precision is satisfying, then the practice ends. Otherwise the extraction of the second component is needed, and it continues until reaching the required precision. Given that m components $t_1, t_2, ..., t_m$ are extracted ultimately, PLSR will establish the regression equation of y and $t_1, t_2, ..., t_m$, and then transform it into the regression equation of y and original independent variables, namely PLSR equation.

The Preliminary Model of PLSR

C919 is large aircraft, so regardless of performance or pricing, it's more reasonable to refer to other large commercial aircrafts' data than small aircrafts' and helicopters. A variety of factors considered, we choose the performance index and prices of 26 commercial aircrafts from the Boeing Company and Airbus as the basic data of PLSR modeling. The aircrafts' types and specific data are shown in the following table 1[6].

Table 1 26 Large Commercial Aircrafts' Types and Specific Data

Aircraft Type	Capacity [Number of people]	Span [m]	length [m]	Height [m]	Maximum fuel load [kg]	Maximum takeoff weight [kg]	Maximum range [km]	Maximum engine thrust [kN]	Average price [million]
737-600	110	34.3	31.2	12.6	22000	65090	6000	91.6	55
737-700	126	34.31	33.6	12.55	22000	60330	6230	116	64
737-800	162	34.31	39.5	12.55	22000	70553	5665	124	76.75
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
350-900	300	61.1	58.8	17.2	97370	265000	13900	774	240.6
380	674	79.8	73	24.1	217000	560000	15100	1208	327.4
C919	162	36.1	38.6	12.6	16800	72500	5555	124	?

In table I we omit the data of following 21 aircrafts: 737-900ER, 747-400/-400ER, 747-8, 767-200ER, 767-300ER, 767-400ER, 777-200ER, 777- 200LR, 777-300ER, 787-3, 787-8, 787-9, 318, 319, 320, 321, 330-200, 330-300, 340-300, 340- 500, 340-600. More information can be referred to the website we give in [6]. The correlation coefficient matrix is given below:

$$\begin{pmatrix}
1 & 0.9012 & 0.8948 & 0.8955 & 0.9116 & 0.9244 & 0.7230 & 0.6599 & 0.9224 \\
0.9012 & 1 & 0.9448 & 0.9420 & 0.9492 & 0.9567 & 0.8693 & 0.7494 & 0.9799 \\
0.8948 & 0.9448 & 1 & 0.8652 & 0.9045 & 0.9194 & 0.8427 & 0.6696 & 0.9589 \\
0.8955 & 0.9420 & 0.8652 & 1 & 0.9220 & 0.9141 & 0.7771 & 0.6330 & 0.9134 \\
0.9116 & 0.9492 & 0.9045 & 0.9220 & 1 & 0.9811 & 0.8447 & 0.7253 & 0.9486 \\
0.9244 & 0.9567 & 0.9194 & 0.9141 & 0.9811 & 1 & 0.8702 & 0.7193 & 0.9711 \\
0.7230 & 0.8693 & 0.8427 & 0.7771 & 0.8447 & 0.8702 & 1 & 0.6088 & 0.8954 \\
0.6599 & 0.7494 & 0.6696 & 0.6330 & 0.7253 & 0.7193 & 0.6088 & 1 & 0.7150 \\
0.9224 & 0.9799 & 0.9589 & 0.9134 & 0.9486 & 0.9711 & 0.8954 & 0.7150 & 1
\end{pmatrix}$$

From the matrix above, we can see that the correla tion is large. In this situation, y denotes the average price of aircraft, $x_1 \sim x_8$ respectively denote capacity, span, length, height, maximum fuel load, maximum takeoff weight, maximum range, maximum engine thrust and average price.

Under the control of cro ss validation, we extract th e principal components of all independent variables to do regression analysis. Then we get the following preliminary general linear PLSR model of forecast price of C919,

$$\begin{aligned} y = & -77.1711 + 0.0972x_1 + 1.2517x_2 + 1.4658x_3 + 0.1885x_4 + 0.1786\times10^{-4}x_5 + 0.8580\times10^{-4}x_6 \\ & + 0.4471\times10^{-2}x_7 + 0.7181\times10^{-2}x_8 \end{aligned} \tag{3}$$

The effect of every independent variable x_j on dependent variable y can be measured by variable importance of projection index VIP. VIP is defined [7] as follows,

$$VIP = \sqrt{\frac{p}{Rd(y;t_1,t_1,\cdots t_m)}\sum_{h=1}^{m} Rd(y;t_h)\omega_{hj}^2} \tag{4}$$

in which,

$$\begin{cases} Rd(y;t_1,t_1,\cdots t_m) = \sum_{h=1}^{m} Rd(y;t_h) \\ Rd(y;t_h) = r^2(y,t_h) \end{cases} \tag{5}$$

In the formula above, ω_{hj} is the first $j\,th$ component of ω_j. $r(y,t_h)$ is the correlation coefficient of y and t_h.

Figure 1 below is the histogram of VIP. From the figure we can see that the VIP values of all independent variables are in the nearby of value 1. This indicates that the explanatory ability of all independent variables on dependent variable is equivalent.

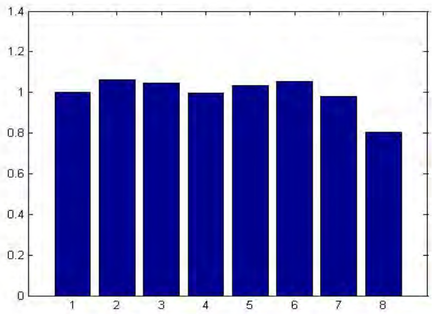

Fig. 1 Histogram of VIP

Then we calculate precision and fi nd that the average relative e rror of this regression m odel is 3.84%, which indicates that the model fits well and is feasible. However, the average relative error of log-linear regression model is 4.47%, thus showing that general linear regression is better. Figure 2 below are fitting precision maps of all samples.

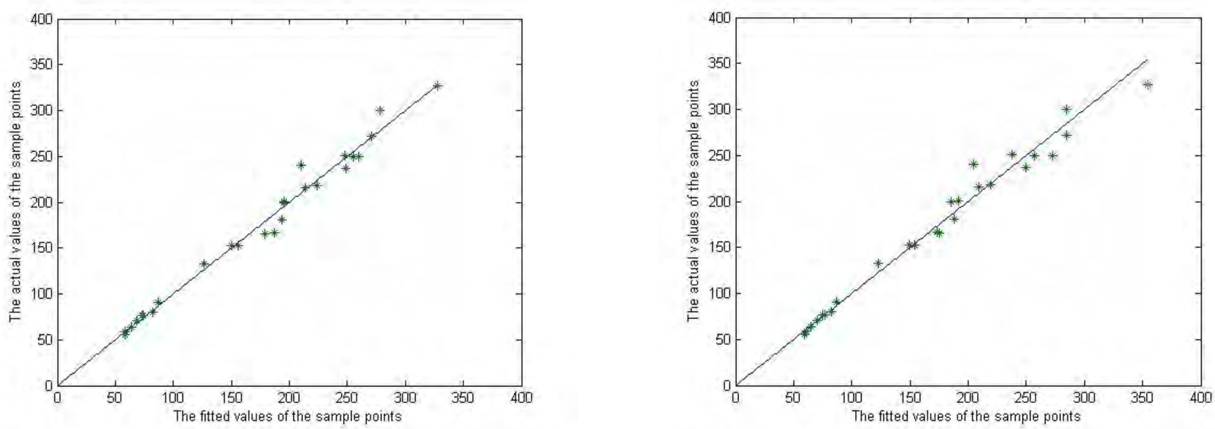

Fig. 2 Fitting precision map based on general linear regression & Fitting precision map based on log-linear regression.

Moreover, by calculation we know that the averagerelative error of least squares regression model is 3.96%, which is less precise than that of PLSR. So this article chooses the relatively most precise PLSR as the preliminary model. Feed the performance data of C919 (see Table 1) into linear PLSR equation (3) and we get preliminary value of forecast price, namely 74.97 million dollars per aircraft.

PLSR Model Modified by Discount Factor

The price of aircraft is influenced by lots of external factors, including inflation rate and the competition among aircraft suppliers. And the case exists that suppliers offer discounts to airlines while raising catalog price. The increase rate of price is not fixed and there is much uncertainty. For these reasons, the effect of time is not considered in this article.

In this article the original data to forecast the price of C919 are adopt from foreign large aircrafts' performance data and prices accordingly. During the process of forecasting, we neglect the difference between domestic and foreign market. The distinction in labor cost, manufacturing efficiency, policy and laws, marketing methods and many other factors have non-negligible impact on the pricing of aircraft. So modification of the preliminary model is needed.

As we know, China's self-developed aircraft ARJ-21 has put into production and service. ARJ-21 is a kind of small commercial aircraft used for extension operation. Once we figure out the ratio of the price of domestic aircraft to the price of foreign aircraft of the same type, then we can modify the model and obtain a more reasonable one to forecast the price of domestic aircraft.

The performance data and prices of foreign aircrafts are available. We can use these data and the original model in the previous part to forecast the price of ARJ-21 P. And the real price of ARJ-21 P' is known, then the ratio we need is $\frac{P'}{P}$. We use C to denote $\frac{P'}{P}$, and give it the meaning of discount factor. With the price got from original model multiplied by C, we gain relatively precise forecasting price.

When using PLSR to forecast the price of ARJ-21, we choose the same standard as the original model so as to get a more reasonable discount factor. Through searching, we find the data of three foreign small aircraft shown in the table 2 below [6].

Table 2 The Performance Data and Prices of Samll Aircrafts

Aircraft Type	Capacity [Number of people]	Span [m]	length [m]	Height [m]	Maximum fuel load [kg]	Maximum takeoff weight [kg]	Maximum range [km]	Maximum engine thrust [kN]	Average price [million]
CRJ-700	70	23.2	32.51	7.57	9306.91	33000	3124	61.3	36
CRJ-900	90	23.2	36.19	7.57	9306.91	36500	2778	64.5	41.6
CRJ-200	50	21.21	26.77	6.22	4653.5	21500	1825	40.98	22
ARJ-21	90	27.288	33.464	8.442	10386	40500	2225	68.14	30

Then we get the following price forecast model of small aircraft

$$\begin{aligned} y' = & -29.5357 + 0.1331x_1 + 0.7322x_2 + 0.4791x_3 + 1.0793x_4 + 0.0313\times10^{-2}x_5 + 0.0214\times10^{-2}x_6 \\ & -0.0027\times10^{-1}x_7 + 0.1035x_8 \end{aligned} \tag{6}$$

Feed the data of ARJ-21 (see Table 2) into equation (6), and get the forecasting price P, namely 45.95 million dollars per aircraft. While the real price of ARJ-21 is P', namely 30 million dollars per aircraft. Then the discount factor is $\frac{P'}{P}$, namely 0.6529. The final price forecast model is

$$M = C \times y \tag{7}$$

In which M is the final forecast price of C919, y is the preliminary forecast price, C is discount factor. Then the final forecast price is

$$M = 0.6529 \times 74.97 = 48.95 \tag{8}$$

Namely 48.95 million dollars per aircraft.

Conclusion

This article aims to forecast the price of domestic large commercial aircraft. To reach the goal, we find the relationship between performance index and price of foreign aircraft by regression methods. Though comparison we figure out that under the for ecast background, general linear PLSR is more precise than log-linear regression and least squares regression.

Combining with the real pricing situation of domestic aircraft, we modify the original model through discount factor which considers actual effects. Compared with literature [5], this article provides a more reasonable method for the quantification of equalization-engineering-value-ratio.

The final forecast price of C919 gained in this article is 48.95 million dollars per aircraft. And China's Ministry of Commerce once disclosed to the Xinhua News Agency that the market price of C919 would be less than 50 million dollars per aircraft [8]. This also proves that our result is reasonable and feasible.

Although this model barely takes the influence of time into consideration, it offers some help to the price forecast and pricing of domestic large aircraft.

References

[1] Feng Zunde, Lu Xiushan, Shi Yufeng et al. Application of partial least squares regression in data analysis of mining subsidence[J]. Transactions of Nonferrous Metals Society of China, 2005, 15(S1): 148-150.

[2] Li Shouan, Zhang Hengxi, Li Dongxia et al. Battleplan acquisition cost forecast based on partial least2squares regression[J]. Journal of Naval University of Engineering, 2005, 17(4): 64-68.

[3] Li Shouan, Zhang Hengxi, Tong Zhongxiang et al. Application of Partial Least2squares Regression to the Acquisition Cost Estimating of Military Aircrafts[J]. Acta Aeronoutica et Astronautica Sinica, 2006, 27(3): 600-604.

[4] Li Shouan, Zhang Hengxi, Song Bifeng et al. Research into Warplane Procurement Price Prediction[J]. Fire Control and Command Control, 2007, 32(1): 81-83.

[5] LI Zhanke, FU Jinfeng. Research on the cost forecast of civil transports based on partial least squares regression[J].Flight Dynamics,2011,29(6):81-83.

[6] China Holiday. 377 types of aircraft models [EB/OL]. http://www.china-holiday.com/air/flight/AircraftType/aircraftmodel.htm, 2012.

[7] Wang Huiwen, Wu Zaibing, Meng Jie. Partial Least-Squares Regerssion Linear and NonLinear Methods[M]. Beijing: National Defence Industry Press,2006:119-120.

[8] Xinhua. China-made large aircraft C919 challenge Boeing and Airbus with the price advantage[EB/OL]. http://news.xinhuanet.com/fortune/2010-11/17/c_12782905.htm, 2012.

Published by Atlantis Press, Paris, France.