

The Estimating Method of Military Aircraft R&D Cost Based on Hierarchical Partial Least Squares Regression

Shuai Shao^{1,a}, Jilian Guo^{1,b}, Anwei Shen^{1,c}, Zhuojian Wang^{1,d} and Wentong Jia^{1,e}

¹Department of Aircraft and Engineering, Air Force Engineering University, Xi'an, 710038, China

^ashaoshuai7219@126.com, ^bguojilian@aliyun.com, ^czjsaw@foxmail.com, ^dmailzhuojianw1974@sina.com, ^ewen_t@163.com

Keywords: Aircraft; Hierarchical Partial Least Squares; R&D Cost.

Abstract. The requirements of using modern aircraft have been increased constantly, which results in different influencing factors of aircraft R&D and procurement costs. And the original model in the predication for R&D costs appears accretion of estimation especially for the extensive using of advanced avionics and composite materials, stealth materials. In this paper, we take the R&D costs of aircraft predict as an example and apply the method of hierarchical partial Least Squares regression for modeling and verification. Case analysis indicates that, the method proposed by this paper is feasible and the error is tiny.

Introduction

With the development of science and technology, a large number of new technologies, new materials, and new manufacturing processes are used in military aircraft design and production. Accompanied with the improvement of the aircraft's combat effectiveness, the aircraft also become more sophisticated, which leads to a constant increase in the costs of development and production. How to improve the combat effectiveness of the aircraft and keep down its costs effectively has become a common issue for all the countries in the world^[1].

In the early period of developing weapon and equipment, the method of Parametric Cost Estimating has been emphasized and a number of fruitful researches have been achieved. This method has been successfully applied in the practice by many famous companies such as Development and Procurement Costs of Aircraft^[2] (DAPCA) of Rand Company, MLCCM of Northrop Grumman Company, FR-103-USN and PRC .etc.^[3] In the model of DAPCA, the general expression of cost estimating can be shown as $y = a_0 x_1^{a_1} x_2^{a_2}$, y refers to costs, x_1 and x_2 mean weight and speed respectively, a_0, a_1 , and a_2 are correlation coefficients. The expression of parametric estimating can be obtained from the regression analysis of the same type of aircraft sample data.^[4]

In the cost prediction models, we often select several relatively important parameters as the cost-driven factors, such as weight, maximum flight speed, the maiden flight time and reliability. However, the requirements of using modern aircraft have been increased constantly, which results in different influencing factors of aircraft development and procurement costs, especially the extensive using of advanced avionics and composite materials, stealth materials. Thus weight and other factors cannot fully reflect the costs, and the original model in the predication for R&D costs appears accretion of estimation. In addition, the limited number of optional cost-driven factors is not conducive to the implementation of cost sensitivity analysis. Therefore, we must adopt more cost-driven factors to establish a new cost prediction model. For the prediction of aircraft R&D costs, the variables in reflecting the quality characteristics of the airplane are numerous, while the number of samples is significantly reduced. How to take advantage of the information implied in variables has become a problem. This issue has been further studied and discussed by Li^[4] and Zhang^[5] who applied the partial least squares method to a typical small sample multivariate data problems and achieved good results. However, the selection of variables was still relatively limited. How to solve this problem better? The Hierarchical Partial Least Squares Regression has gained

unique advantages in dealing with multivariate data, so the Hierarchical Partial least squares (Hi-PLS) regression can be used to predict the cost of aircraft R&D. In this paper, we take the R&D costs of aircraft predict as an example and apply the method of hierarchical partial Least Squares regression for modeling and verification.

Application of Hi-PLS to Aircraft Cost Estimation

Construction of Independent Variable Sub-block. In 1996, Wold, a French mathematician, put forward the Hierarchical Partial Least Squares (Hi-PLS) method based on the thoughts of multivariate hierarchical modeling.[6][7][8] The detail introduction of this method can be found in ref [6]. Currently, it is a trend to use the integrated parameters to construct the costs estimating model. True Planning can be considered as the latest software for estimating equipment costs. Despite it needs entering very complicated data, according to the most fundamental core model, the main explanatory variable used in this model is the integrated parameters for describing the advance of equipment. When estimating the costs of aircraft, in order to obtain the integrated parameters [9] of component, hundreds of input parameters for each component are needed.

This article acts the main aircraft capacity which reflects the effectiveness of aircraft as independent variable and introduces it into model, studies the description method of each kind of aircraft combat effectiveness and constructs the independent variable sub-block of costs estimating on the basis of the modeling that was put forward by Zhang^[5]. In this modeling, the effectiveness of aircraft is considered as the integrated data for estimating the costs of aircraft. The variables which reflect the aircraft R&D costs are divided into different groups according to the following standards: mobility, firepower, situation awareness and rapid targeting ability, viability and electronic warfare capability, etc. As for other parameters which have more influence on the aircraft R&D costs, for instance, empty body weight, combat radius, reliability index, ~~they~~ are divided into a group. Separate variables are shown in table 1.

Table 1 Separate variables

Team	Variables
Mobility(ξ_{mane})	Specific Excess Power $SEP(x_{\text{mane},1})$; Weight Ratio($x_{\text{mane},2}$); Maximum Instantaneous Turning Angular Velocity($x_{\text{mane},3}$)
Firepower(ξ_{fire})	The Number of Missile Pylon($x_{\text{fire},1}$); the Maximum Carrying Ability($x_{\text{fire},2}$); Air Combat Weapon Effectiveness($x_{\text{fire},3}$)
Situation Awareness and Rapid Targeting Ability(ξ_{dete})	Radar Detection Capability Coefficient($x_{\text{dete},1}$); Cockpit FOV ($x_{\text{dete},2}$); Information Coefficient ($x_{\text{dete},3}$); Sight Coefficient($x_{\text{dete},4}$)
Viability and Electronic Warfare Capability(ξ_{sur})	Wing Area($x_{\text{sur},1}$); Radar Cross Section($x_{\text{sur},2}$) Electronic Warfare Capacity($x_{\text{sur},3}$);
Others(ξ_{else})	Combat Radius($x_{\text{else},1}$), reliability index; Empty Body Weight($x_{\text{else},2}$); cycle of the first restoration($x_{\text{else},3}$); MFHBF($x_{\text{else},4}$)

Table 2 Sample Data

Type	Mobility			Firepower			Situation Awareness Ability				Viability			Others				R&D Costs(billion dollars)
	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_1	x_2	x_3	x_4	
1	170	0.928	8	7.6	1750	0.3	0.25	124	0.51	0.5	28	3.8	1.2	659	5992	900	2.1	0.032
2	225	0.9178	8	12	1750	0.51	0.41	124	0.61	0.5	25.6	9	1.2	865.5	4156	900	2	0.033
3	225	0.9178	8	6.5	1750	0.61	0.73	124	0.61	0.5	25.6	10.1	1.2	1008.3	4848	1120	2.3	0.094
4	121.6	0.7342	8	7.6	2330	0.66	0.45	124	0.61	0.5	48.6	3.5	1.2	1308	7418	1120	2.5	0.120
5	335	1.081	10	7.6	2910	0.72	0.87	124	0.61	0.5	48.6	12	1.2	1367	8044	1340	2	0.193
6	225	0.9178	10	8.7	5810	0.77	0.98	187	0.72	1	65.2	10.1	1.3	2163.5	13153	1670	2.5	0.439
7	153.5	0.7107	10	6.5	8130	0.82	0.89	187	0.93	1.6	65.2	3.8	1.3	1957	13153	2220	3.1	1.138
8	263.5	1.1728	15	8.7	6970	0.93	0.95	282	0.93	1.6	46	8.4	1.4	1780	13822	2220	3.4	3.035
9	203	0.826	15	13.1	6970	0.93	0.97	282	0.93	1.6	53.8	10	1.4	2370	14276	2220	3.4	3.668
10	360	1.3	25	14	9290	0.95	0.99	376	0.98	2.1	93.5	0.2	1.5	2370	15000	2770	5.3	24.750
11	380	1.3	25	14	7550	0.95	0.99	376	0.98	2.1	90.1	0.2	1.6	2075	15000	2770	5.3	23.000

Variable description: among these variables, some belong to basic physical performance parameters, some comprehensive evaluation parameters, some expert assessment parameters. The calculation and assessment methods are explained in detail in air-to-air combat effectiveness model. As for estimating of other types of military aircraft, some different performance parameters and grouping schemes can be used according to the design features and characteristics of the aircraft missions and tasks. The aircraft performance data can be seen in Table 2, in which the front nine parameters are regression samples and the latter two are used for models.

Building of Model. (1) Building of Base Model. As the relation between aircraft R&D costs and the explanation of variables belongs to a superposition of exponential relation, the method of logarithmic transformation should be used and therefore the linearized relational expression can be obtained. And then the method of hierarchical partial Least Squares regression for modeling should be applied. The original data need to adopt the method of logarithmic transformation correspondingly.

After adopting the method of logarithmic transformation, considering the R&D costs as the dependent variables, establish five base hierarchical partial Least Squares regression modeling according to mobility, firepower, situation awareness ability, viability and other characteristics respectively. And on the principle of cross validity, extract one component t_{mane} in the partial least squares regression modeling in which the mobility of aircraft is the independent variable. The regression relation between t_{mane} and the original variable can be showed as follows:

$$t_{\text{mane}} = 0.1426 \cdot x_{\text{mane},1} + 0.0287 \cdot x_{\text{mane},2} + 0.9894 \cdot x_{\text{mane},3} \quad (1)$$

The explanation degree of t_{mane} for costs (dependent variables) is 77.6%.

In the PLS regression model which the firepower of aircraft is the independent variable, extract one component t_{fire} . The regression relation between t_{fire} and the original variable can be showed as follows:

$$t_{\text{fire}} = 0.1613 \cdot x_{\text{fire},1} + 0.7270 \cdot x_{\text{fire},2} + 0.6674 \cdot x_{\text{fire},3} \quad (2)$$

The explanation degree of t_{fire} for costs is 90.58%.

In the PLS regression modeling which the situation awareness ability of aircraft is the independent variable, extract one component t_{dete} . The regression relation between t_{dete} and the original variable can be showed as follows:

$$t_{\text{dete}} = -0.4447 \cdot x_{\text{dete},1} - 0.5155 \cdot x_{\text{dete},2} - 0.5227 \cdot x_{\text{dete},3} - 0.5132 \cdot x_{\text{dete},4} \quad (3)$$

The explanation degree of t_{dete} for costs is 95.19%.

In the PLS regression modeling which the viability of aircraft is the independent variable, extract two components $t_{\text{sur},1}$. The regression relation between $t_{\text{sur},1}$ and the original variable can be showed as follows:

$$\begin{cases} t_{\text{sur},1} = 0.6049 \cdot x_{\text{sur},1} + 0.1654 \cdot x_{\text{sur},2} + 0.7789 \cdot x_{\text{sur},3} \\ t_{\text{sur},2} = -0.6864 \cdot x_{\text{sur},1} - 0.2612 \cdot x_{\text{sur},2} - 0.6827 \cdot x_{\text{sur},3} \end{cases} \quad (4)$$

The explanation degree of $t_{\text{sur},1}$ and $t_{\text{sur},2}$ for costs is 93.53%.

In the PLS regression modeling which other characteristics of aircraft are the independent variables, extract one component t_{else} . The regression relation between t_{else} and the original variables can be showed as follows:

$$t_{\text{else}} = 0.4835 \cdot x_{\text{else},1} + 0.4922 \cdot x_{\text{else},2} + 0.5273 \cdot x_{\text{else},3} + 0.4959 \cdot x_{\text{else},4} \quad (5)$$

The explanation degree of t_{else} for costs is 95.75%.

(2) Building of Top Model

Based on the Base Model, the Top Model of $(t_{\text{mane}}, t_{\text{fire}}, t_{\text{dete}}, t_{\text{sur}}, t_{\text{else}})$ for aircraft R&D costs can be built and two components of partial Least Squares can be obtained:

$$\begin{cases} t_{TOP,1} = -0.4031 \cdot t_{mane} - 0.4405 \cdot t_{fire} + 0.4476 \cdot t_{dete} - \\ \quad 0.3880 \cdot t_{sur,1} - 0.3062 \cdot t_{sur,2} - 0.4458 \cdot t_{else} \\ t_{TOP,2} = -0.3650 \cdot t_{mane} - 0.01050 \cdot t_{fire} + 0.3503 \cdot t_{dete} - \\ \quad 0.3737 \cdot t_{sur,1} + 0.7405 \cdot t_{sur,2} - 0.2264 \cdot t_{else} \end{cases} \quad (6)$$

The explanation degree of costs is 92.31%, while in the Base Model, the explanation degree of partial Least Squares model for the costs is 77.60%、90.58%、95.19%、93.53%、95.75% respectively. Thus we can see that, the costs can be explained by extracted component from the Top Model based on the Base model, and the explanatory ability of extracted component PLS is improved. The VIP of each variable (t_{mane} , t_{fire} , t_{dete} , t_{sur} , t_{else}) is 1.0921, 1.0668, 1.0701, 1.0710, 1.0656, 1.0654 respectively. When the VIP of mobility, firepower and situation awareness ability exceed 1, we can say that they have greater impact on the aircraft costs and making them as costs driving factors is meaningful.

The obtained top partial Least Squares regression model can be showed as follows:

$$\hat{y} = -18.4947 + 1.8807 \times t_{mane} + 0.5687 \times t_{fire} - 0.7184 \times t_{dete} + 0.2983 \times t_{sur,1} - 0.7318 \times t_{sur,2} + 0.6464 \times t_{else} \quad (7)$$

Furthermore, putting the (1)-(5) types of Base Model into the (6) type of Top Model, we can obtain the linear regression model of cost for independent variables:

$$\begin{aligned} \hat{y} = & -18.4947 + 0.2682 \cdot x_{mane,1} + 0.0541 \cdot x_{mane,2} + 1.8607 \cdot x_{mane,3} + \\ & 0.0917 \cdot x_{fire,1} + 0.4143 \cdot x_{fire,2} + 0.3796 \cdot x_{fire,3} + \\ & 0.3195 \cdot x_{dete,1} + 0.3704 \cdot x_{dete,2} + 0.3755 \cdot x_{dete,3} + 3687 \cdot x_{dete,4} \\ & -0.3219 \cdot x_{sur,1} - 0.1418 \cdot x_{sur,2} + 0.7319 \cdot x_{sur,3} + \\ & 0.3126 \cdot x_{else,1} + 0.3181 \cdot x_{else,2} + 0.3409 \cdot x_{else,3} + 0.3206 \cdot x_{else,4} \end{aligned} \quad (8)$$

Then logarithmic inverse transformation of model should be carried out:

$$\begin{aligned} \hat{y} = & (9.2869 \times 10^{-9}) \cdot x_{mane,1}^{0.2682} \cdot x_{mane,2}^{0.0541} \cdot x_{mane,3}^{1.8607} \cdot x_{fire,1}^{0.0917} \cdot \\ & x_{fire,2}^{0.4143} \cdot x_{fire,3}^{0.3796} \cdot x_{dete,1}^{0.3195} \cdot x_{dete,2}^{0.3704} \cdot \\ & x_{dete,3}^{0.3755} \cdot x_{dete,4}^{0.3687} \cdot x_{sur,1}^{-0.3219} \cdot x_{sur,2}^{-0.1418} \cdot x_{sur,3}^{0.7319} \cdot \\ & x_{else,1}^{0.3126} \cdot x_{else,2}^{0.3181} \cdot x_{else,3}^{0.3409} \cdot x_{else,4}^{0.3206} \end{aligned} \quad (9)$$

Verification of Model. In order to investigate the prediction accuracy of the regression equation, the original data is substituted into the fitting parametric model, the prediction values of the original observed values are obtained in Table 3.

Table 3 the observed values and the prediction values (billion US dollars, 2015)

Type	The observed values	The prediction values
1	0.032	0.034
2	0.033	0.051
3	0.094	0.076
4	0.120	0.077
5	0.193	0.190
6	0.439	0.599
7	1.138	1.020
8	3.035	3.319
9	3.668	3.253

Seen from Table 3, the data points of the predicted value and the observed values are substantially similar, which indicates that the effect of the prediction model is ideal. To further examine the results of the regression equation, this article use the parameters of a new developed amphitypy aircraft to verify the model calculation. The actual investment volume of Airframe is 24.75 (billion US dollars) and 23.00 (billion US dollars), the cost in predicting the results by using the model is 28.078 (billion US dollars) and 26.612 (billion US dollars). The predicted residual rates are: 15.07% and 13.45%. Generally, the estimation for R&D costs in a large system, the error is effective within 30%. The error of these two authentication types are within 20%, so the established model is feasible. At the same time, seen from the models of R&D cost, radar reflection

area index is negative, which means the radar reflection area smaller, the cost higher. The result was matched with the fact relatively.

Summary

The article takes an example of aircraft, the carrying on modeling studies of aircraft costs by using Hierarchical Partial Least Squares Regression Algorithm. During the process of instance modeling, the Hierarchical Partial least squares regression method has higher capability on information integration and filtering. In the underlying model, PLS of each independent variable sub-blocks and dependent variable can effectively extract the most useful information from these variables, and removed part of noise from the original data; on the basis of information filtering, these more effective integrated variables will be used for creating top-level model, and the results can have a better explanation.

Acknowledgements

The authors express their gratitude to the editor and the referee for their detailed comments and suggestions, which is of great use in improving the manuscript. The work is sponsored by the National Natural Science Foundation of China (71501185).

References

- [1] Tan. X., Xu. Y., Early. J., Wang. J., Curran. R. and Raghunathan. S., "A Framework for Systematically Estimating Life Cycle Cost for an Integrated Wing," AIAA Paper 2007-7809, Sep 2007
- [2] Mark V., Obaid Y and Lionel A., Fox B., John C., Jerry M., Felicia W, Carolyn W, "Impossible Certainty-Cost Risk Analysis for Air Force Systems," The Rand Corporation, MG-415, 2005.
- [3] Jia. L, Guo. J and Shen. A., "The Effects of Advanced Materials on Aircraft Production Cost," Advanced Materials Research, Vol. 710, 2013, pp. 798-803.
doi: 10.4028/www.scientific.net/AMR.710.798
- [4] Li. S, Song. B and Zhang. H. "Method for Multivariate Analysis with Small Sample in Aircraft Cost Estimation," Journal of Aircraft, Vol. 44, No. 3, 2007, pp. 1042-1045
doi: 10.2514/1.28174
- [5] Zhang, H. X., Modern Aircraft Efficiency and Cost Analysis, Aviation Industry Press, Beijing, 2001, pp. 115–121.
- [6] Wold. S, Kettaneh. N and Tjessem. K, "Hierarchical Multiblock PLS and PC Models for Easier Model Interpretation and as an Alternative to Variable Selection," Journal of Chemometrics, Vol. 10, No. 5, 1996, pp. 463~482.
- [7] Hoskuldsson. A, "Path regression models and process control optimization," Journal of Chemometrics, Vol. 28, No. 4, 2014, pp. 235~248.
- [8] Agusdinate. D and Delaurentis. A, "Adressing equity issue in multi-actor policymaking via a system of systems approach: Aviation emissions reduction case study," Journal of Systems Science and Systems Engineering, Vol. 20, No. 1, 2011, pp. 1-24.
- [9] Choi. J. W., Kelly. D., Raju. J., and C. Reidsema, "Knowledge Based Engineering System to Estimate Manufacturing Cost for Composite Structures," Journal of Aircraft, Vol. 42, No. 6, 2005, pp. 1396-1402
- [10] Malone. P., and Smoker R., "Increasing Development Cost Estimate Accuracy of Complex Systems using Advanced Tools" AIAA Paper 2012-5136, Sep 2012