A Study on Cost Drivers Based on Principal Component Analysis

Zhan Yonghong, Xiong Yanchao
School of Management, Wuhan University of Technology, Wuhan, P.R.China, 430070
(E-mail: yh19998@sina.com, xiongyanchao3210@163.com)

Abstract The application of activity-based costing system (ABCS) is more and more prevalent, while it's somehow restrained by the difficulties and high expenditures of information collection and data processing. Thus it's really interesting to find a method to scientifically overcome such defects. Taking XAMC Company's process of activity as an example, The anthors have set up a principal component analysis (PCA) based model to optimize and simplify the cost panel data under the rule of minimizing the loss of data and resources, in which the paper has proposed a brand new perspective on selection and combination of cost drivers, namely the method of dimensionality reduction in ABC.
Key words Activity-based costing; Principal component analysis; Cost driver; Dimensionality reduction

1 Introduction

Activity-based costing (ABC) has been well applied in developed countries since 1980s, especially in manufacturing industry, because it can assign the indirect costs more accurately and rationally than traditional cost accounting methods. In ABC, all activities which consume resources are identified, and costs are accounted into corresponding activities according to their cost drivers. Actually, the basic principle of ABC is matured, and in the practice of accounting in the modern industry of western countries, the ABC method is really widely used, while the outstanding difficulty puzzling both practitioners and theorists is it's somehow costly to collect corresponding information and process the data. Thus in this paper, we attempt to construct a PCA based model to solve this problem somehow, by which the work of cost accounting is expected to be more simplified and efficient.

2 Literature Review

The essence of ABC is that activities consume resource and products consume activities. In ABC, activity pools are established and cost drivers are selected, with which cost can be calculated accurately. However, like a sword with double edges, it greatly increase the expenditures in collecting, storing and processing information (accounting data); meanwhile, errors might occur in the process of data collecting, reporting, evaluating because of the incompatibility of different accounting criterions and systems. Moreover, it's really time-consuming to conduct ABC facing a large amount of accounting data, thus it can not meet the requirement of timely management decision very well. So it's meaningful to find a way to reduce or combine some trivial data, with which the accuracy of accounting would not decrease sharply. Babad & Balachandran had empirically estimated the cost caused by the merger of two different cost drivers, through which it was found that the merger of related cost drivers was completely non-destructive [1]. Similarly, Wang also proposed that merger of relative cost drivers based on correlation analysis is really beneficial to increase the efficiency of cost accounting and can be satisfied with the accuracy of cost information [2]. Homburg established a linear programming model, in which the object function is the information cost minimization constrained by number of cost drivers and accuracy of cost calculating; However, this model can not solve the problem of cost allocation very well [3]. Schniederjans & Garvin has also set up a theoretical multi-objective programming model to select the cost drivers; anyway it's hard to be figured out in the practice of accounting [4]. Zhang employed the principle of cooperative game, argued that the cost with none or indeterminate drivers could be assigned to corresponding activities by introducing sharply value [5]. Whereas Li, et al., employed matrix theory to prove that the cost calculating errors almost did not occur when combining multi cost drivers [6]. Afterwards, Wang, et al., established a cluster analysis model to categorize cost drivers according to the strength of their correlations, based on which typical cost drivers could be determined [7].

Reviewing all above literatures, except literature [7], none of them has considered the correlation of cost drivers well, which might lead to errors occur in the process of combining cost drivers? However, we argue that PCA is available in solving the existed defects of ABC method. Our basic logic is firstly analyzing the correlation of cost drivers, and select typical ones from them, then assign them to relative activities, with which the correlated cost drivers could be combined and dimensionality of calculation could be reduced, thus the cost of implementing ABC could be decreased and it is convenient to be applied in practice.

3 PCA Model of Cost Drivers

By means of PCA, total cost drivers could be classified into several non-correlated compositive statistic indicators. In this paper, we use PCA to select typical cost drivers from cost drivers' pool, and identify the main factors through analyzing the internal structure of sample matrix. Assume a business has m kinds of products, and the original cost drivers were $D_1, D_2, \ldots\ldots, D_n$. Firstly, eliminate the correlated components, then gather the differences in the components left and classify it again to create several new critical cost driver sets. Denote the new compositive cost drivers as $d_1, d_2, \ldots\ldots, d_p (p \le n)$, and then we have:

$$\begin{cases} d_1 = a_{11} D_1 + a_{12} D_2 + \ldots\ldots, \quad + a_{1n} D_n \\ d_2 = a_{21} D_1 + a_{22} D_2 + \ldots\ldots, \quad + a_{2n} D_n \\ \qquad \ldots\ldots \\ d_p = a_{p1} D_1 + a_{p2} D_2 + \ldots\ldots, \quad + a_{pn} D_n \end{cases} \tag{1}$$

In formula (1), coefficient a_{ij} is determined according to the following principles:

(1) d_i and d_j ($i \ne j$, $i, j = 1, 2, \ldots, p$) are independent with each other; (2) d_1 stands for the maximum variance within all variances created by any linear combination of $D_1, D_2, \ldots\ldots, D_n$; d_2 stands for the maximum variance within all variances created by any linear combination of $D_1, D_2, \ldots\ldots, D_n$, but is not correlated with d_1; d_p stands for the maximum variance within all variances created by any linear combination of $D_1, D_2, \ldots\ldots, D_n$, but is not correlated with $d_1, d_2, \ldots\ldots, d_{p-1}$. $d_1, d_2, \ldots\ldots, d_p$ is named as the 1^{st}, 2^{nd}… and p^{th} principal component of the original variable $D_1, D_2, \ldots\ldots, D_n$.

Denote the sample data set as: $D_1' = (D_{1l}, D_{2l}, \ldots, D_{pl}), l = 1,2,\ldots, n$, then the sample covariance matrix can be depicted as:

$$\hat{\Sigma} = \left\{ \frac{1}{n-1} \sum_{l=1}^{n} (D_{il} - \overline{D}'_i)(D_{jl} - \overline{D}'_j) \right\}_{p \times p} \tag{2}$$

Through solving the equation $\left| \hat{\Sigma} - \lambda I \right| = 0$, we can obtain its Eigen values as: $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_p \ge 0$; Then constructing its eigenvector as :

$$u_i = (u_{1i}, u_{2i}, \ldots, u_{pi})' \tag{3}$$

Again, figuring out the rate of accumulated contribution and obtaining exact number of principal components referred to cost drivers; finally, working out the marks of k principal components of cost driver by inputting the medium value into formula (4), and then order them according to the marks.

$$D_i^* = D_i' - \overline{D'} = (D_{1i} - \overline{D_1}, D_{2i} - \overline{D_2}, \ldots, D_{pi} - \overline{D_p})' \tag{4}$$

Table 1 The Activity Pools and Cost Drivers

Serial number	Activity pool	Cost driver	Code
1	Molding	Component Types	D_1
2	Smelting & Casting	Smelted Iron Types	D_2
3	Mat.-Handling	Mat.-Handling Labor Hours	D_3
4	Punching	Punching Machine Hours	D_4
5	Modeling	Modeling Labor Hours	D_5
6	Painting	Painting Machine Hours	D_6
7	Assembling	Assembling Labor Hours	D_7
8	Setup	No. of Setup	D_8
9	Machine-Moving	Moving Distance	D_9
10	Components-Tooling	Components-Tooling Labor Hours	D_{10}
11	Order-Processing	No. of orders	D_{11}
12	Maintaining	Maintaining Labor Hours	D_{12}
13	Inspecting	Inspecting Labor Hours	D_{13}
14	Selling & Customer-Relations	No. of Customers	D_{14}

4 An Empirical Study

4.1 The data

Referring to literature [8] and taking the data of Company XAMC as a sample, we can test our model. In XAMC, there are 6 kinds of products denoted as $P_1, P_2, P_3, P_4, P_5, P_6$ respectively, and there are 14 activities involved in the activity pool, which can be denoted as: $D_1, D_2, ..., D_{14}$ respectively. The details are shown in Table 1.

From the original activity cost data of XAMC, we can obtain the amount of cost drivers consumed by each product, to construct the bill of activity (BOA) shown in Table 2.

Table 2 Bill of Activity in XAMC

Cost drivers	P_1	P_2	P_3	P_4	P_5	P_6
D_1	3500	3600	3200	1400	320	
D_2	97410	58441	15015	3519	983	0
D_3	2660	4454	660	355	663	0
D_4	307	59	372	178	200	0
D_5	2130	3242	1118	465	893	112
D_6	950	853	950	160	127	60
D_7	30620	16516	6414	220	310	220
D_8	420	210	170	140	85	59
D_9	1635	1374	408	151	255	61
D_{10}	6829	3201	1247	3760	5175	1458
D_{11}	32	22	13	10	12	21
D_{12}	2520	780	650	370	210	90
D_{13}	1517	1172	420	256	330	185
D_{14}	27	22	15	8	5	3

4.2 The analysis

By means of SPSS13.0, we can analyze the original data and obtain Table 3, Table 4 and Table 5.

Table 3 Correlation Matrix

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	D_{12}	D_{13}	D_{14}
D_1	1.00	.772	.743	.478	.807	.964	.780	.774	.806	.166	.444	.686	.781	.935
D_2	.772	1.00	.790	.252	.788	.754	.996	.950	.978	.593	.859	.933	.990	.946
D_3	.743	.790	1.00	-.093	.985	.666	.748	.608	.895	.335	.583	.529	.861	.817
D_4	.478	.252	-.093	1.00	.051	.552	.316	.489	.188	.277	-.055	.489	.205	.395
D_5	.807	.788	.985	.051	1.00	.752	.757	.631	.898	.327	.534	.550	.861	.852
D_6	.964	.754	.666	.552	.752	1.00	.782	.756	.778	.116	.494	.698	.755	.908
D_7	.780	.996	.748	.316	.757	.782	1.00	.962	.964	.577	.862	.952	.978	.948
D_8	.774	.950	.608	.489	.631	.756	.962	1.00	.885	.638	.754	.988	.911	.921
D_9	.806	.978	.895	.188	.898	.778	.964	.885	1.00	.550	.797	.849	.997	.954
D_{10}	.166	.593	.335	.277	.327	.116	.577	.638	.550	1.00	.427	.667	.594	.430
D_{11}	.444	.859	.583	-.055	.534	.494	.862	.754	.797	.427	1.00	.797	.824	.694
D_{12}	.686	.933	.529	.489	.550	.698	.952	.988	.849	.667	.797	1.00	.883	.867
D_{13}	.781	.990	.861	.205	.861	.755	.978	.911	.997	.594	.824	.883	1.00	.947
D_{14}	.935	.946	.817	.395	.852	.908	.948	.921	.954	.430	.694	.867	.947	1.00

Table 3 is a data matrix after standard processing, from which it can be found that D_1 is strongly correlated with D_2、D_3、D_5、D_6、D_7、D_8、D_9、D_{13}、D_{14}, and correlated with D_{12}. Thus it can be concluded that there are direct correlations between many variables, which shows that information overlapping exists among them. From Table 4 we can extract 3 principal components, comparing with the original 14 cost drivers, the working load of calculation is sharply decreased. Moreover, from Table 5, D_1, D_2, D_3, D_5, D_6, D_7, D_8, D_9, D_{11}, D_{12}, D_{13}, D_{14} has a high loading weight on the first principal component respectively, which demonstrates that the first principal component can nearly reflect the information of the cost drivers; D_4 has a high loading weight on the second principal component, which demonstrates that the second principal component can nearly reflect the information of this cost drivers; D_4 has a high loading weight on the second principal component, which demonstrates that the second principal component can nearly reflect the information of this cost drivers; again D_{10} has a high loading weight on the third principal component, which demonstrates that the third principal component can nearly reflect the information of this cost drivers. Summarizing the above analysis, we can conclude that the three principal components can almost reflect the information of all cost drivers; therefore it's rational to substitute the original 14 variables with the 3 ones.

Table 4 Total Variance Explained

	Initial Eigen values			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.416	74.398	74.398	10.416	74.398	74.398
2	1.514	10.817	85.215	1.514	10.817	85.215
3	1.385	9.896	95.111	1.385	9.896	95.111
4	.600	4.282	99.393			
5	.085	.607	100.000			
6	2.10E-015	1.50E-014	100.000			
7	1.22E-015	8.75E-015	100.000			
8	2.50E-016	1.79E-015	100.000			
9	1.22E-016	8.71E-016	100.000			
10	8.13E-017	5.80E-016	100.000			
11	3.55E-017	2.54E-016	100.000			
12	-7.34E-017	-5.24E-016	100.000			
13	-2.56E-016	-1.83E-015	100.000			
14	-4.69E-016	-3.35E-015	100.000			

Table 5 Component Matrix (a_{ij})

Cost Drivers	Component		
	1	2	3
D_1	.857	.130	-.482
D_2	.986	-.051	.143
D_3	.822	-.479	-.207
D_4	.339	.909	-.175
D_5	.843	-.359	-.291
D_6	.840	.204	-.480
D_7	.985	.018	.132
D_8	.944	.244	.170
D_9	.984	-.166	.025
D_{10}	.545	.169	.674
D_{11}	.781	-.240	.362
D_{12}	.910	.274	.278
D_{13}	.986	-.130	.094
D_{14}	.984	.046	-.162

4.3 The test

Through the above analysis, which merger 14 cost drivers into three compositive cost drivers, we can compare the activity costs assigned by the new 3 cost drivers with the ones assigned by the original 14 cost drivers, the corresponding results are shown in Table 6. It can be found that, after merger of cost drivers, the relative error of indirect costs assigned to products P_1, P_2, P_3, P_4, P_5 and P_6 are -0.00081, -0.04335, 0.022834, 0.033098, 0.093425, 0.052895 respectively comparing with the results obtained before the merger of cost drivers. Again, from Table 4, the three new compositive cost drivers can explain 95.11% of all original 14 cost drivers, therefore, it demonstrates that the PCA based activity-based costing method can greatly simplify the analysis, as well satisfy the requirement of accuracy.

Table 6 before and after the Mergence of the Indirect Costs of Product Distribution List Errors

Compositive Cost Drivers	P_1	P_2	P_3	P_4	P_5	P_6
d_1	0.302452	0.278161	0.307452	0.041613	0.046968	0.023355
d_2	0.27509	0.052867	0.333333	0.159498	0.179211	0
d_3	0.315136	0.147716	0.057545	0.173512	0.238809	0.067282
Indirect Costs after Merger	641931.59	411262.2	203414.5	102541.4	108521.2	40977.72
Indirect Costs before Merger	642452.06	429899.1	198873.4	99256.17	99248.87	38919.09
Absolute Error	-520.47	-18636.9	4541.14	3285.23	9272.33	2058.63
Relative Error	-0.00081	-0.04335	0.022834	0.033098	0.093425	0.052895

5 Conclusions

A PCA based method is proposed to merge correlated cost drivers into several typical principal components, with which the work load of calculating activity-based cost could be sharply reduced, while the accuracy could be completely met. Additionally, in this paper, we have also conducted an empirical study as an example to instruct practitioners how to use professional software, such as SPSS or SAS to carry on the analysis work. Generally speaking, the analytical results are quite satisfied which demonstrates the basic logic of this research is correct and convincible. By the way, our model can only work well based on a perfect ABC accounting system and abundant data; otherwise, it's hard to be implemented in practice.

References

[1] Babad, Y. M., Balachandran, B. V. Cost Driver Optimization in Activity-based Costing [J].The Accounting Review, 1993, 68(3): 563-575

[2] Wang Pingxin. On the Homogeneity of Cost Factors [J]. Quantitative and Technical Economics of China, 1999, (5):44-46 (In Chinese)

[3] Homburg, C. A Note on Optimal Cost Driver Selection in ABC [J]. Management Accounting Research, 2001, (12): 197-205

[4] Schniederjans J, Garvin T. Using the Analytic Hierarchy Process and Multi-objective Programming for the Selection of Cost Drivers in Activity-based Costing [J]. European Journal of Operational Research, 1997, 10(1):72-80

[5] Zhang Yingjian. Research on Improving Activity-Based Costing Based on Cooperative Game Theory [J]. China's Township Enterprises Accounting, 2007, 1（27）：45-46 (In Chinese)

[6] Li Buxi, Wang Pingxin, Chen Lin. Study on Cost Driver Combination Theory in Activity-based Costing [J]. Systems Engineer Theory& Practice, 2007, 27(4): 47-53 (In Chinese)

[7] Wang Fangjun, Chang Hua and Huang Kan. Using Clustering Analysis for the Selection and Combination of Cost Drivers in Activity-Based Costing [J]. Management Review, 2009, 9(21): 94-99(In Chinese)

[8] Wang, P., Jin, Q., Lin, T. How an ABC Study Helped a China Stated-owned Agricultural Machine Company to Keep Competence [J]. Cost Management, 2005, (11/12): 39-47