

Utilizing Data Mining and Factor Analysis for Identifying Activity Base Costing Cost Drivers in Iranian Bank

Ali Asghar Anvary Rostamy^{1,*}, Tahereh Aliheidri Bioki², Farideh Bakhshi Takanlou³, Amin Anvary Rostamy⁴

¹Professor, Tarbiat Modares University (TMU), Chair Department of Accounting, Tehran, Iran

²PhD Student of Industrial Engineering, Yazd University, Yazd, Iran

³Accounting Student, University of Applied Sciences & Technology Tehran, Iran

⁴Department of Software Engineering, Amirkabir University of Technology, Tehran, Iran

*Corresponding Author: anvary@modares.ac.ir

Copyright © 2013 Horizon Research Publishing All rights reserved

Abstract Most of firms are interested to know about the real cost of products and services they provide. In order to measure the real cost, several methods and techniques have been provided in literature. One of the most important and prominent methods is the Activity Based Costing (ABC) method. Determination of cost drivers in ABC is regarded as a very difficult and tiresome task. Although various methods have been introduced for this purpose but each method suffers from certain disadvantages. This paper introduces the Data Mining (DM) and Factor Analysis (FA) methods to improve the effectiveness of the ABC method. The proposed combined methods let decision makers to consider several cost drivers respect to their effects on costing, simultaneously. In 2010, we have applied the proposed method in Export Development Bank of Iran: EDBI (one of the Iranian leading banks). The results confirmed the usefulness and effectiveness of the proposed methods for bank ABC system and its economic decisions.

Keywords Cost Driver, Activity Based Costing, Data Mining, Factor Analysis

1. Introduction

All the methods used for identifying, exploration and determination of the products and services costs are called costing (Babad and Balachandran, 1993). Each of the developed methods in this context meets specific part of organizational requirements in a way that choosing a product of service costing methods depends on the management information requirements and their ultimate objective. However, what's so important these days is that organizations need an approach which in addition to simplicity and inexpensiveness provides the most information regarding the processes, activities and services for managers. Therefore, the traditional costing methods seem inappropriate (Kaplan and Anderson, 2007).

Activity-Based Costing (ABC) is a method which in addition to the above mentioned benefits meets a greater range of managers' requirements. The main idea behind ABC is that a cost object (e.g. products, services) is generated by activities which require and consume resources, for instances, labor and equipment (Demeere and Stouthuysen, 2009).

According to Lievens et al.' opinion (2003), an ABC system is a two stage cost allocation process. In the first stage, organizational activities are determined and the indirect costs are allocated to activity centers. This stage of cost allocation process is carried out using resource Cost Drivers (C.Ds) which represent the activity's usage of indirect cost. In the second stage, the costs in each activity center are assigned to cost objects using activity C.Ds. Therefore, the activity cost consumption for each product or service is determined by activity C.Ds through the second stage. These steps provide a great deal of information relating to the processes and activities in addition to the cost of products and services (Schniederjans and Garvin, 1997).

Under finding of Bhimani and Pigott (1992), Cooper and Kaplan (1992), Gordon and Sivester (1999), Innes (1990), and Turney and Anderson (1989), many potential advantages of the ABC systems include: (1) helping to determine non value-added activities, (2) improving the ability of managers to make pricing, production, and investment decisions through the provision of more accurate product and process costs, and (3) provision of the required conditions in order to improve the market share and better control and monitoring over the costs. Despite of ABC considerable advantages and power in providing the required information for managers, it suffers from certain limitations that may cause significant deviances in the obtained results. One of the limitations lies in choosing the cost driver. Inaccurate selection of the cost driver may lead to confusing results. Therefore, in this paper, in addition to the introduction of the Data Mining (DM) method, application of the Factor Analysis (FA) for cost driver determination has been described

using an example.

2. Literature Review

2.1. Cost Driver Determination Methods

Cost driver is a measurable and logic variable which can be used to determine the amount of the resource consumption by each activity and the activity consumption by cost objects (Zafar, 1998). Actually, what is important in every costing method is the selection of the cost driver. Since in contrast to the traditional costing methods, multiple cost drivers are applied in two stages in ABC and this leads to more accurate and specific allocation of the costs. Thus, proper selection of the cost drivers has a significant impact on the model's reliability (Khademi and Aliheidari Bioki, 2009). Several methods for cost driver determination have been proposed. In 2001, Prof. Kocakulah at South America Indiana University made a comparison between business loan profitability of a bank using ABC and traditional costing method. It should be noted that in this paper, cost driver determination method is of a great importance which in this paper has been examined using the T-test correlation regression coefficient (Kocakulah and iekmann, 2001). In 1993, Blanchandar used the Greedy Algorithms to optimize the cost driver determination process (Blanchandar,1993). In paper of Schniederjans and Garvin notice that one year later, Dekin used the logical relationships for this purpose and in 1997, Schniederjans applied the AHP method for the optimized selection of the cost driver (Schniederjans and Garvin, 1997).

2.2. Cost Driver and Data Mining

DM is the process of deducing the unknown, exact and potentially useful information out of the data. DM is considered as a set of techniques used for information identification or decision making knowledge from the pieces of data in a manner that this information can be used in decision making, forecasting, prediction or estimation contexts (Giudici, 2003).

According to the researches, DM is related to costing methods from two perspectives. First, in which stages of costing methods can we use DM techniques? Second, how obtained results from costing method can be used in DM projects?

In 2003, Kim and Han developed an article with the title of "Application of the Genetic Algorithm and Neural Networks in Activity Based Costing". The main focus of the paper was on the optimum selection of the cost driver using the Genetic Algorithm and considering a nonlinear cost function using the Neural Networks (Kim and Han, 2003).

With respect to the second question, application of costing methods in the field of DM, a limited number of researches have been conducted that some of them propose

a costing model for DM projects. Masand, & Shapiro (1996), Domingos (1998), Boehm (2000), and Putnam (2000) tried to measure the cost of DM activity and its profitability in areas such as CRM but, none of them proposed a model for cost determination prior to the launch of the data mining project. In 2008, Menasalvas addressed this issue. In their proposed model, a stage was added to the DM process. Using this model and at the business definition stage, the total process cost is estimated and in the case of feasibility, the associated budget is allocated (Menasalvas, 2008)

Reviewing on other studies shows that there are several costing methods such as logical method, Cause-and-Effect method, Analytical Hierarchy Process (AHP), Delphi and Regression methods. In this paper, use of the DM and FA processes as a method for cost driver determination has been studied by a case.

2.3. Factor Analysis

FA is a generic name that is predicated to a set of multi-variable statistical techniques with the goal of defining the hidden structure within data. Generally, factor analysis addresses the analysis of the correlation structure among a number of variables through the definition of a set of common hidden elements called factors. Factor analysis enables the researcher first to identify the independent structure factors and then determine a justifiable limit for each variable. Thus, the factors and related descriptions for each variable are determined and then, the initial applications of the factor analysis i.e. abstraction and data volume reduction are obtained. In data abstraction, the hidden factors are revealed and then, these factors are interpreted and subsequently, data are described through fewer numbers of variables. Data volume reduction is accomplished through the score calculation for each hidden factor and replacing it with the initial variables (Thompson, 2004).

A century ago, Carlos Spearman published an article about the factor analysis in an American Psychological Magazine. Since the current mathematical factor analysis methods assume one, two or more factors with predetermined structure, until mid 90's, interpretation of the factors was conducted smoothly. The main purpose of the FA was to validate the obtained structure through the experimental solutions and quantitative analysis of the tests. The only multiple factor analysis proposed in 1935 by Tartson made it possible to derive the predetermined factor structures. However, all of the multiple FA methods such as Centroid method are complex and time consuming methods and in some cases result in different factor structures which in turn lead to a wave of despair especially in efforts made for meaningful interpretations of the factors. Thus, this facilitates the articulation of more practical and easier ideas (Lee Daniel, 2000).

Failure in factor interpretation even led to "Functionalism" in some ideas and theories such as

Tartson’s who considered the factor analysis as a scientific approach for the validation of the process nature related hypotheses (Lee Daniel, 2000).

Around 1950, reputation of the FA was indemnified by its proponents due to three wrong (erroneous) assumption devised by some authors. First it seemed that FA has been considered something more than just an ordinary statistical approach. Secondly, they considered the problems as abstract and implicit issues even if there were more appropriate experimental approaches to the problem. Ultimately, the set of variables was very large and extensive. Therefore, their efforts were focused on to extend its application far beyond its ability. Hotlin in his principal combinations showed that this deadlock enabled the calculation of the orthogonal factors unit matrix. Although this method requires a demanding mathematical calculation, but development of the computer systems may resolve this problem and soon, many researchers were attracted towards it (Harman, 1967). The emergence of this method was a renaissance for the FA method and was recognized as a structured research method in almost every science fields.

3. Materials and Methods

3.1. Data Mining for Determining Cost Driver

Nowadays, one of the major issues that businesses are facing with is the rapid data generation. Explosive growth of the stored data is the major reason that necessitates the need for novel technologies and automated tools in order to transform a great amount of data into knowledge and information (Khademi and Aliheidari Bioki, 2009). Data mining as a proper solution to this problem is an automated process to extract patterns that reflect the tacit knowledge hidden in data warehouses, large databases and other enormous data repositories (Hand et al, 2001). Different steps have been proposed by researchers and DM is considered as an entire process in all of them. This process has been summarized in six major steps: 1) Business problem recognition, 2) data understanding, 3) data pre-processing, 4) modeling, 5) model evaluation, and 6) model interpretation and expansion (Hand et al, 2001).

A research conducted in 2008 at Export Development Bank of Iran (EDBI) titled as “Design of a banking service costing structure based on a DM based ABC method”. EDBI was established on July 10th 1991, and subsequently initiated its activities in September 1992. EDBI acts as Iran’s Ex-Im bank and plays a pivotal role in providing financing facilities and banking services to Iranian exporters and the buyers of the goods and services of Iranian origin. The algorithm is illustrated in Figure 1 has been applied for cost driver determination (Khademi and Aliheidari Bioki, 2009). This algorithm has been articulated based on the six-step DM process.

In this paper, in addition to introducing the application

of DM in ABC method, it’s been tried to apply the FA method for cost driver determination.

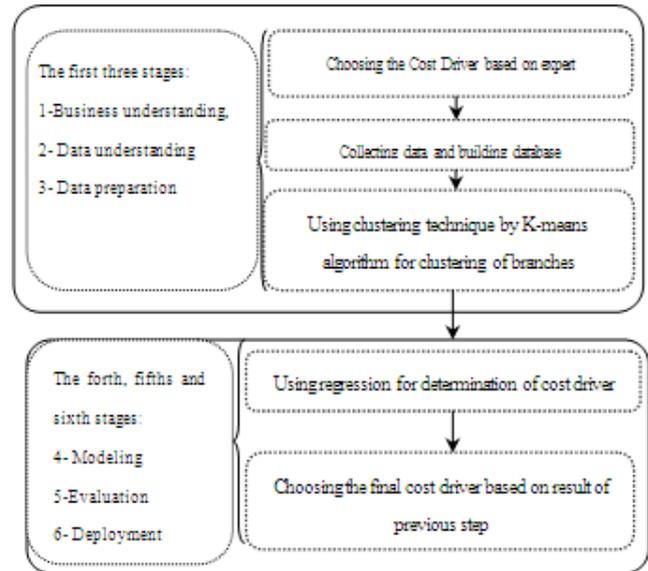


Figure 1. DM process for determining cost driver (Khademi & Aliheidari Bioki, 2009)

3.2. Factor Analysis for Determining Cost Driver

One of the developments of the FA is the factor matrix. It’s also called the factor pattern matrix. This matrix preserves the factor loadings. Factor loadings are the correlation between factors and variables which is expressed by Formula 1:

$$\frac{W_{ij}}{S_j} \sqrt{\lambda_i} \tag{1}$$

In formula 1, W_{ij} denotes the weight of j^{th} variable at i^{th} factor; S_j denotes the variation of the j^{th} variable and λ_i represents the specific value of the i^{th} factor. If the result of the formula is a high score, it means that there’s a significant correlation between the factor and variable. Loadings are used when there’s not a significant correlation between factors and variables (Harman, 1967).

The fundamental assumptions in FA are of the conceptual nature than statistical nature. From the statistical point of view, deviation from normality is allowed to a point that the observed correlations are reduced. Normality is valid through the statistical test being significant, but these tests are not practiced typically.

In addition to the statistical base for the correlation of the data matrix, the researcher must make sure that there’s a significant correlation for data matrix in order to determine the ability of the factor analysis. If through the assessment, there are no significant correlations greater than 0.3 then the FA may prove inefficient.

There are various criteria to determine the number of the extracted factors. Eigen value criterion, variance

Now, in this paper it's been tried to consider the overall effects of the initial selected cost drivers for the costing purpose in addition to the introduction of the FA method. In this section, results of the implementation of the FA method for determining the level 1 cost drivers (Hardware and software cost and cost of branches) are provided and for one of the cost groups are described in four steps.

Step 1: Selection of the primal variables

The first step in FA method is to select the primal variables (primal cost drivers). In this step, primal variables have been chosen based on expert's viewpoints. These variables are provided at Table 2 for each cost group.

Table 2. List of the primal variables selected for factor analysis

Code	Primal variables for hardware and software costs
T_r	Total of Records
T_e	Total of Entries
T_{cr}	Total of Cash Records
Code	Primal variables for cost of branches
T_r	Total of Records
T_e	Total of Entries
T_{ea}	Total Entry of Current Account
T_{en}	Total Entry of None Current Account
T_{nr}	Total of Non Cash Records
T_{cr}	Total of Cash Records

Data used for the FA are 2007 which have been retrieved from the Oracle data bases using the SQL Server software and have been analyzed using the SPSS software for each branch and monthly intervals.

Step 2: Suitability test for FA application

For the suitability of the data used by the FA, KMO and Bartlett test have been utilized. The obtained results for each testing method and each cost group are provided in Table 3.

Table 3. Test results for KMO Bartlett for each group of variables

FACTOR/VARIABLES T_r T_e T_{cr}		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.775
Bartlett's Test of Sphericity	Approx. Chi-Square	1738.655
	df	3
	Sig.	.000
FACTOR/VARIABLES T_r T_e T_{ea} T_{en} T_{nr} T_{cr}		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.872
Bartlett's Test of Sphericity	Approx. Chi-Square	6410.507
	df	15
	Sig.	.000

As shown in Table 3, the values for the KMO test are 0.775 and 0.872 respectively and since these values are greater than the proposed value (0.5), hence, use of FA as a size reduction technique is allowed. Also, result of the

Bartlett test proves a significant correlation among the variables.

Step 3: Factor derivation

In this step, the Scree Plot criterion and Eigen value greater than 1 criterion have been used for determining the number of the factor derivations. As shown in Figure 4, number of the selected factors for two factor groups is 1. However, one cannot certainly determine the number of the factor score for sure. But it's possible to use other criteria such as explained variance or Eigen value.

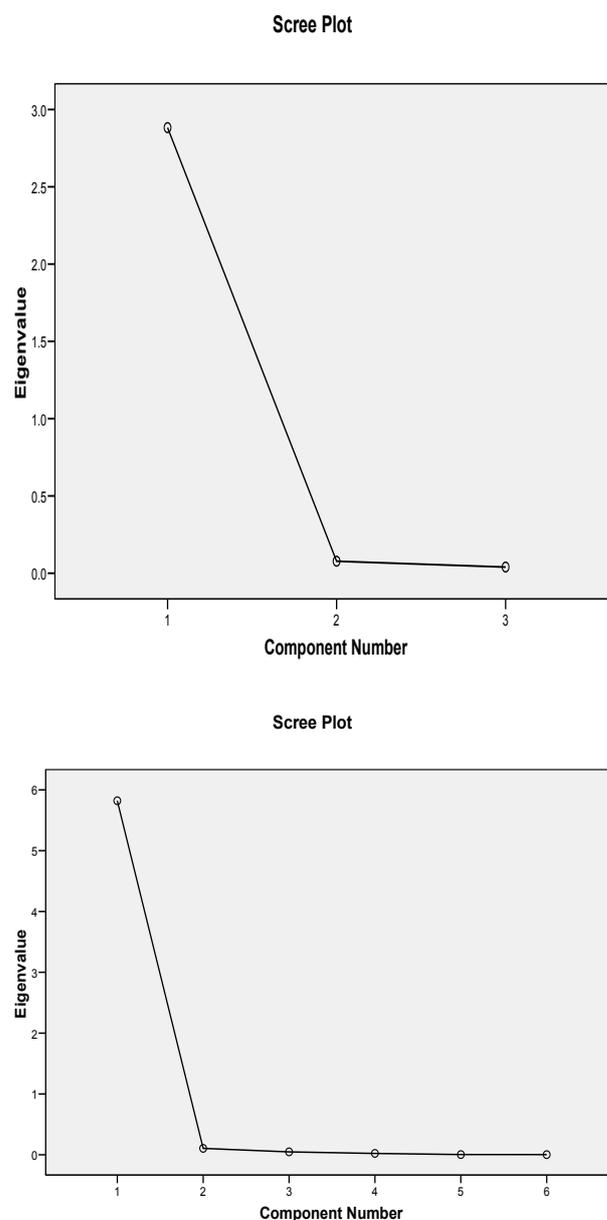


Figure 4. Scree Plot diagram for determining number of factors

Since there's need to derive one factor for each two cost groups, the other two criteria will not be discussed. It means that derivation of a factor for each group according to the descriptive variance and Eigen value in Table 4 are

definite. Table 4 shows that one factor can properly explain 97.002 and 96.08 percent of the variance for all variables.

Step 4: Application of the factor scores

When the objective of FA is to replace the set of variables with a new but limited set of variables for further analyses, then it's essential to calculate the same values or factor scores and replace the original values with these new scores. Various statistical soft ware's such as SPSS is useful for calculating and applying the factor scores in analyses. In this paper, since the objective of the factor analysis is to replace the original variables with a new set of original variables combination; factor scores will be calculated and stores in Excel software for further

calculations of the activity cost as the cost driver. Table 5 provides a part of the output and input of FA process.

According to expert opinions, this proposed method has some advantages. Because of using a lot of data which is related to cost group, this method can be better than traditional methods for selecting cost drivers. We presented the result of this study to more than 20 experts and ask them to compare reliability of the proposed model to traditional method. They confirmed that the result of proposed method outperforms the traditional method. The other important advantages of the proposed method are having effect of several cost drivers simultaneously, and its ability of having several scenarios in short time with low cost.

Table 4. Total variance descriptive

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.883	96.089	96.089	2.883	96.089	96.089
2	.078	2.585	98.674			
3	.040	1.326	100.000			
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.820	97.002	97.002	5.820	97.002	97.002
2	.105	1.747	98.749			
3	.048	.794	99.543			
4	.022	.359	99.902			
5	.003	.058	99.960			
6	.002	.040	100.000			

Table 5. Factor analysis process inputs and outputs

[FA]	T_{cr}	T_{nr}	T_{an}	T_{ea}	T_e	T_r	NUM
4.010	1,401	3,401	1,722	6,316	18,946	3,206	1
5.330	1,349	4,418	2,934	7,867	25,940	4,140	2
5.775	2,028	5,145	3,330	9,118	24,036	4,292	3
5.190	1,419	4,178	3,087	7,662	24,773	3,976	4
5.190	1,392	4,130	3,008	7,620	24,421	4,098	5
5.170	1,603	4,502	2,833	7,853	25,501	3,825	6
4.795	1,478	3,975	2,755	7,153	21,953	3,771	7
5.058	1,374	3,989	2,955	7,421	23,863	3,836	8
5.254	1,472	3,820	2,886	7,077	24,223	3,661	9
4.856	1,542	4,142	2,416	7,111	20,937	3,216	10
.
.
0.343	100	167	183	211	630	76	319
0.303	105	177	232	227	651	84	320

5. Conclusions and Final Remarks

What's so important at every costing method is to determine and select proper and effective cost drivers. Since in the ABC, in contrast to traditional costing methods, several cost drivers are applied at two steps, this results in more accurate allocation of the costs to activities. Therefore, proper selection of the cost drivers significantly affects the reliability of the model. Because cost allocation plays an important role in organizational decision making, determination of the cost drivers through the proposed method is also very important in a way that in attention to this problem may cause significant deviations in ABC results.

In this paper, two methods of cost driver determination have been introduced. First, DM process that is focused on optimum use of raw data collected from daily operations and stored in data bases of a bank and using the hidden knowledge within this data may help to improve the results of the ABC method. Additionally, in this method, one or two ultimate drivers are selected among these original cost drivers while; the FA method has been introduced to highlight the importance of effects of the total cost drivers on decision making. It means that selection of the cost driver through the proposed method enables us to consider several cost drivers with maximum effects for costing purpose. The advantage of the proposed method over traditional methods is that the effects of multiple cost drivers are considered simultaneously. According to the application scope of the multivariable techniques and on the other hand, need for improvement of ABC method and also the commitment for determining the actual cost at firms and organizations may lead to opportunities for further research and development.

REFERENCES

- [1] P. Armstrong. 2002. The costs of activity-based management, *Accounting, Organizations and Society*, Vol. 27, No. 1-2, 99-120.
- [2] Y. M. Babad, V. Balachandran. 1993. Cost driver optimization in activity-based costing. *The Accounting Review*, Vol. 68, No. 3, 563-575.
- [3] K. M. Baird, G. L. Harrison, R. C. Reeve. 2004. Adoption of activity management practices: A note on the extent of adoption and the influence of organizational and cultural factors, *Management Accounting Research*, Vol. 15, No. 4, 383-399.
- [4] A. Bhimani, D. Pigott. 1992. Implementing ABC: A case study of organizational behavioral consequences. *Management Accounting Research*, Vol. 3, No. 2, 119-130.
- [5] C. Combes, N. Meskens, C. Rivat, J. Vandamme. 2008. Using a KDD process to forecast the duration of surgery, *International Journal of Production Economics*, Vol. 112, No. 1, 279-293.
- [6] M. Comelli, P. Fèniès, N. Tchernev. 2008. A combined financial and physical flows evaluation for logistic process and tactical production planning: Application in a company supply chain, *International Journal of Production Economics*, Vol. 112, No. 1, 77-95.
- [7] R. Cooper, R. S. Kaplan, L. S. Maisel, E. Morrissey, R. M. Oehm. 1992. From ABC to ABM. *Management Accounting*, Vol. 70, No. 10, 54-57.
- [8] N. Demeere, K. Stouthuysen, F. Roodhooft. 2009. Time-driven activity-based costing in an outpatient clinic environment: Development, relevance and managerial impact. *Journal of Health Policy*, Vol. VI, Issue 1(15), 296-304..
- [9] P. Everaert, W. Brugemann, G. Sarens, S. Anderson, Y. Levant. Cost modeling in logistics using time-driven ABC experiences from a wholesaler. Ghent University. *International Journal of Physical Distribution and Logistics management*, Emerald Group Publishing Limited (2008), Vol. 38, No.3: 172 – 191.
- [10] P. Giudici. Applied data mining statistical methods for business and industry, E-book, Faculty of Economics, University of Pavia, Italy, 2003.
- [11] L. A. Gordon, K. J. Silvester. 1999. Stock market reactions to activity-based costing adoptions, *Journal of Accounting and Public Policy*, Vol. 18, 229-251.
- [12] D. Hand, H. Mannila, S. Padhraic. 2001. Principles of data mining, The MIT Press, Cambridge, August 2001.
- [13] H. Harman, H. Modern. factor analysis. 2 Editions, University of Chicago Press, 1967.
- [14] J. Innes, F. Mitchell. Activity based costing: A review with case studies. Chartered Institute of Management Accountants, London, 1990.
- [15] R. Kaplan, S. Anderson. 2007. The innovation of time-driven activity-based costing, *Journal of Cost Management*, Vol. 21, No. 2, 5-15.
- [16] H. Khademi, T. Aliheidri Bioki. The costing structure design in bank services based on ABC with data mining approach, Master Thesis: Yazd University, 2009.
- [17] A. H. Khataie, A. A. Bulgak, J. J. Segovia. 2011. Activity-based costing and management applied in a hybrid decision support system for order management, *Decision Support Systems*, Vol. 52, No. 1, 142-156.
- [18] K. Kim, I. Han. 2003. Application of a hybrid genetic algorithm and neural network approach in activity-based costing. *Expert Systems with Applications*, Vol. 24, 73-77.
- [19] M. C. Kocakulah, D. Iekmann. 2001. Implementing activity based costing to measure commercial loan profitability, *The Journal of Bank Cost & Management Accounting*, Vol. 14, No 2, 3-15.
- [20] Lee Daniel D.; Sebastian Seung H. 2000. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, Vol. 9, 515-521.
- [21] Y. Lievens, W. Van Den Bogaert, K. Kesteloot. 2003. Activity-based costing: A practical model for cost calculation in radiotherapy. *International Journal of Radiation Oncology Biology Physics*, Vol. 57, No. 2, 522-35.
- [22] E. Menasalvas, O. Marba. 2008. A cost model to estimate the effort of data mining projects (DMCoMo), *Information Systems*, Vol. 33, 133-150.
- [23] Z. Qingge. 2012. A new activity-based financial cost management method, *Physics Procedia*, Vol. 33, 1906-1912.
- [24] J. Ratnatunga M. S. C. Tse, K. R. Balachandran. 2012. Cost management in Sri Lanka: A case study on volume, activity and time as cost drivers, *The International Journal of Accounting*, Vol. 47, No. 3, 281-301.

- [25] M. Schniederjans, T. Garvin. 1997. Using the analytic hierarchy process and multi-objective programming for the selection of cost drivers in activity-based costing. *European Journal of Operational Research*, Vol. 100, 72-80.
- [26] K. Stouthuysen, M. Swiggers, A. Reheul, F. Roodhooft. 2010. Time-driven activity-based costing for a library acquisition process: A case study in a Belgian University, *Library Collections, Acquisitions, and Technical Services*, Vol. 34, No. 2-3, 83-91.
- [27] B. Thompson. 2004. *Exploratory and confirmatory factor analysis: Understanding concepts and applications*, American Psychological Association (APA).
- [28] P. Turney, B. Anderson. 1989. Accounting for continuous improvement, *Sloan Management Review*, Vol. 30, No. 2, 37-47.
- [29] U. Zafar. 1998. Activity based management for procurement: case and teaching not, *Journal of Accounting Education*, Vol. 5, 414-433.