ORIGINAL RESEARCH

A study of differences by industry using factor models influencing software development estimates

Tsuyoshi Shida*, Kazuhiko Tsuda

Graduate School of Business Sciences, University of Tsukuba, Tokyo, Japan

Received: March 17, 2018 **Accepted:** September 16, 2018 **Online Published:** September 29, 2018
DOI: 10.5430/air.v7n2p34 **URL:** https://doi.org/10.5430/air.v7n2p34

ABSTRACT

Recently, IoT and AI/machine learning have attracted attention, and software development has been a critical activity for the companies that use IT. The investment in IT has been increasing, and it varies with the industry. In addition, software development has become complex with the growing sophistication in the target applications; therefore, it is a challenging task for the software vendors to prepare an accurate estimate. Consequently, the estimates grossly deviate from the true value. In this paper, we propose a method based on the previous research that uses the factors related to productivity of software development to find factors that affect the estimation of man-hours. We analyzed the parameters among populations using two factors and simultaneous analysis of multiple populations from nine industries. We used two-factor models extracted from "the study of software estimation factors extracted using covariance structure analysis" and verified the method by applying five constraints, including factor load amount and error variance, simultaneously for the nine industries. As a result, it was possible to separate industries with large factor variance and those with small factor variance. Moreover, it was possible to separate industries with large correlation coefficient between factors and industries with small factor correlation coefficient. For industries with small variance of factors, the factors are consistent within the industry, and in industries with large correlation between factors; the relationship between the two factors is more relevant. In other words, we could find out the relationship of factors influencing software estimation for each industry type. In addition, the variance of these two factors and the correlation coefficient between the factors were grouped, and a cluster analysis was performed. It was found that there was a difference in the estimate for each group of Business-to-Business and Business-to-Customer industry groups. Based on these results, while preparing software estimates, IT vendors would capture the characteristics of the factors for each type of industry and clarify the influential factors of fluctuation by being conscious of the productivity fluctuation factors related to the two factors.

Key Words: Software estimation, Multiple group structural equation modeling, Productivity fluctuation factors, Structural equation modeling, Factor analysis, Covariance structure analysis, Cluster analysis

1. INTRODUCTION

Recently, IoT and AI/Machine learning have attracted attention, and software has ever been a critical activity for the companies that use IT. The investment in IT has steadily been increasing, and the magnitude varies with the industry. We studied the Corporate IT Trend Survey 2018[1] published by Japan Users Association of Information Systems (JUAS) with regard to the IT budget allocated by the companies listed under the first section of the Tokyo Stock Exchange. It is observed that 40.7% of the user companies in 2018 have increased their IT investment compared to 2017 setting a new record in the past decade. Particularly, in construction/civil

*Correspondence: Tsuyoshi Shida; Email: s1645006@u.tsukuba.ac.jp; Address: Graduate School of Business Sciences, University of Tsukuba, 3-29-1, Otsuka, Bunkyo-ku, Tokyo 112-0012, Japan.

engineering industry, the diffusion index (ratio of increase minus ratio of decrease) for IT budget exceeded by 9.3 points compared to 27 and 17.7 points in 2017. This investment trend is attributed to the pressing need for effective utilization of IT to balance the pr evailing shortage in human resources owing to the 2020 Tokyo Olympic. In addition, using the IT Investment Trend Survey 2017[2] published by ITR, we analyzed the predicted industry trend in the product/service fields for the Japanese companies. The survey predicts high investments for virtualization (network, server, and storage) in infrastructure/device fields, and for IoT (Internet of Thing) or M2M (Machine To Machine) in manufacturing as well as information and communication industries. IoT is a system that the information of objects can be collected via network. In addition, M2M is a communication technology where information can be mutually transferred between machines without human intervention. In other words, M2M is a system where a machine controls another machine. As for the application fields, we can expect investments in manufacturing or service industry (for BL/data analysis for public), information and communication industry (for application developments for smartphones), building and real estate industry (for business management system), and finance and insurance industry (management of enterprise contents). The enterprise applications for improved service are expected to be in B-to-B (Business To Business) or B-to-C (Business To Customer) depending upon the KGI (Key Goal Indicator) that varies with industry. Accordingly, the budget estimation for software development is a challenging task for the vendors.

It is extremely difficult for the vendors of software development to prepare an accurate estimate, especially in the initial phase. Consequently, the estimates grossly deviate from the true value[3] which is attributed to the uncertain factors that cannot be accurately accounted in terms of man-hours. The guidebook on estimation for software development[4] shows significant discrepancy between the estimated cost and the actual expenditure for a typical waterfall development project. In addition, the white paper on software development (2016-2017)[5] reports scales of measures and types of estimation for software development. It is observed that the software line of code (SLOC) is the method widely used for estimation (55%), while the Function Point (FP) has a share of 26%. The SLOC and FP methods together account for more than 80% of usage.

"Identifying factors affecting software development cost and productivity"[6] used data from 50 projects carried out at one of the largest banks in Sweden to identify the factors that have an impact on software development cost. Overall, 31 factors were studied. Correlation analysis using one-way ANOVA and bivariate regression analysis were adopted, and five factors that affect the software development cost were identified; however, in this research, the productivity fluctuation factors in one type of business are narrowed down, and they are not compared by industry type. Shida et al.[7] investigated the factors contributing to the discrepancies in estimation of software development cost. They focused their study on a few productivity-related factors that were not considered under the SLOC and FP methods, and they eventually extracted two factors: management and capability in the upstream process and the ease of development; however, this study was based on single population comprising nine industries, and it did not consider multi-populations within an industry for comparison. It is well known that the parameters of single population are not always consistent with those among multi-populations. Accordingly, as per the model using which the above two factors were extracted, we need to validate a hypothesis that the parameters of multi-populations will be different, and we can consider a method to analyze per industry. In other words, we set a null hypothesis that all parameters would be different if we respectively analyze per industry, and an alternate hypothesis that all parameters would be equal if we analyze multi-populations simultaneously. Accordingly, if we compare parameters among multi-populations, we need to validate a hypothesis that we wish to compare after communalizing the comparison frame.

Accordingly, this study statistically analyzes the parameters among multi-populations while comparing the compatibility between the model that assumes that parameters among multi-populations of nine industries are all equal and the model that assumes that those are not equal.

We will validate the difference among industries by extracting variance of factors and correlation coefficient between factors per industry based on the result. In other words, while estimating IT vendor companies, we can prepare detailed estimates by capturing the characteristics of each industry type.

2. SOFTWARE DEVELOPMENT MAN-HOURS ESTIMATION METHOD

The estimation of man-hours for software development has a long history, and many methods exist.[8–12] Among the several methods, the COCOMO II model[13, 14] is adopted for preparing software development estimates based on fluctuation factors. The following is a previous study on COCOMO II model and software estimation factors.

2.1 COCOMO II

The COCOMO II model was developed and analyzed on the achievement of a project in the latter half of 1990. It re-

sponds to the characteristics of a project including its unique features. It accumulates the achievement of a project and makes fine adjustments so as to match it. The COCOMO II model has three parts. The first one is called the early design model used at the initial stage of a project, and it is possible to estimate even when the accuracy of estimation of software man-hours is low. The second one is called the post architecture model used for the estimation when the architecture of the new system is decided based on the tendency that the estimation accuracy tends to increase as the phase of the project occurs later. These two models consider new program development, revised development of existing systems including packages, and reuse development that replaces the existing systems on another platform. The third one is the application composition model.

This model is suitable for development of screens and preparation of reports using spreadsheets (e.g. Excel), GUI, etc. It has a simple mathematical formula that applies weights to the complexity of the screens and reports to be developed and divides it by development productivity. The expressions of the early design model and post architecture model to calculate the man-hours are as per the following equations 1, 2, and 3.[12, 13]

(1) Early Design Model

$$PM = A \times Size^E \times \prod_{i=1}^{7} EM_i \quad (1)$$

$$E = B + 0.01 \times \sum_{j=1}^{5} SF_j \quad (2)$$

Note. Size: Adjustment Development Scale; PM: Person Months; A: 2.94; EM: Effort Multiplier; SF: Scale Factors; B = 0.91

(2) Post Architecture Model

$$PM = A \times Size^E \times \prod_{i=1}^{17} EM_i \quad (3)$$

PM, A, E, and EM are the same as the ones in the Early Design Model.

EM, which is also known as cost driver, is a factor that cannot be calculated from the scale of SLOC and FP. The COCOMO II model definition manual[14] defines seven factors related to man-hours: personnel capability (PERS), product reliability and complexity (RCPX), required reuse (RUSE), platform difficulty (PDIF), personnel experience (PREX), facilities (FCIL), and schedule (SCED). The users apply the grades, namely "extremely high", "very high", and "high" for these man-hour factors. From equations 1 to 3, we find that the EM is an important factor.

2.2 Extraction method of factors in software estimation by covariance structure analysis

It is possible to estimate the man-hours for software development by defining variable factors in productivity as variables in COCOMO II model; however, we cannot find a factor by variable factors in productivity of a project. Accordingly, Shida et al.[7] assumed parameters that may have an impact on software quotation under the extraction method of factors in software quotations by Structure Equation Model (SEM) based on variable factors in productivity. They revealed the relationship between factors extracted and variable factors in productivity upon the results. Figure 1 shows the results whereby the two major factors for estimation are management and capability in the upstream process (factor 1) and ease of development (factor 2). In addition, it also shows the variable factors in productivity that has an impact on these two major factors. It shows that the factor 1 has a strong relationship with two variable factors in productivity such as "experience and capability in project management" and "experience and capability in analyst management". It also shows that the factor 2 has a strong relationship with four variable factors in productivity: Functionality, compatibility of platform, clarity and stability of order requirements, and carry-over and stability of advanced model; however, the factors extracted were the factors under single population, even though the factors in development estimation were clarified based on the extraction method of factors in software estimation by Covariance Structure Analysis.[7] The comparison among multi-populations per industry has not been clarified. Accordingly, this study analyzes difference among companies by comparing under multi-populations by software factor model extracted by Covariance Structure Analysis.

3. Analysis Target Data

3.1 Experimental data

The data used in this study was collected by the Economic Research Institute of the General Foundation from 2001 to 2014 by using questionnaires. The data was collected from 344 IT vendors representing small, medium, and large enterprises. The total number of data samples was 2,008 including those with missing records. After removing the data containing the missing records by using the listwise method, 1,721 data were actually available for use in the study.

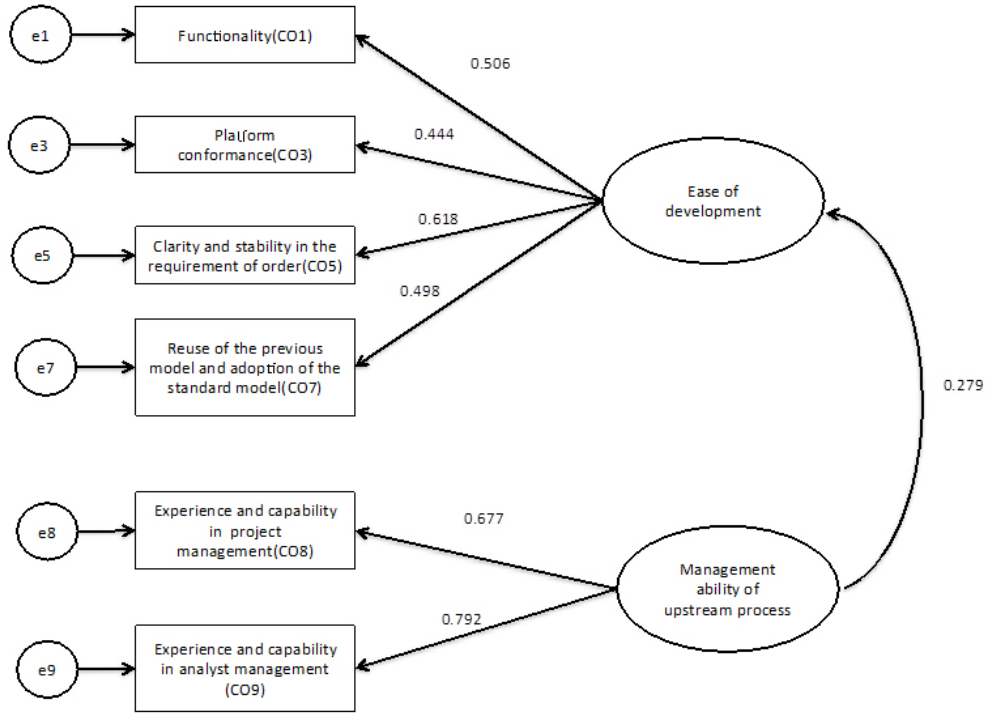

Figure 1. Covariance Structure Analysis Result

Table 1. Explanation of Productivity Fluctuation Factor[15]

Fluctuation factor	Description
Functionality (CO_1)	・Security and access control (security) ・Difficulty ・Connection with other systems ・High precision calculation
Dependability requirement (CO_2)	・System down (fault tolerance) ・Recovery time from the system down recovery) ・Failure rate (maturity)
Platform conformance (CO_3)	・Conformance of the platform (Needs, Performance etc.)
Schedule request in development (CO_4)	・Severity of development schedule constraints
Clarity and stability in the requirement of order (CO_5)	・Stability is the frequency of specification changes that occur in the factors of IT order companies in the software order period ・Clarity of order is the clarity of the order specification in ordering time
Participation frequency of user (CO_6)	・Involvement of the IT order company (user)
Reuse of the previous model and adoption of the standard model (CO_7)	・The level to which can be diverted to a similar system that is already developed in the system of the target business
Experience and capability in project management (CO_8)	・The level of experience and ability required for project management
Experience and capability in analyst management (CO_9)	・The level of experience and ability of analysts (Including business experience)
Experience and capability in system engineer and programmer (CO_{10})	・The level of experience and ability of development technology methods ・The level of programming language and tool of experience and ability ・The level of business experience and ability ・The level of experience and ability of development methodologies ・The level of platform experience and ability

3.2 Setting the productivity fluctuation factor

The man-hours for software development is most influenced by the software size and the implementation function; however, other than these, there is a factor related to the productivity of the software development influencing the man-hours.[3] The factors assumed to influence the productivity of development were set by a committee established by the "Economic Research Association" that consists of over ten members who are experts in software development such as the major IT vendors in Japan. In addition, this committee defined ten productivity fluctuation factors based on the quality characteristics of software defined in the Japanese Industrial Standards (JIS) X 0129-1 and other literatures. In addition, the scale of productivity fluctuation factor was set by the Likert method, and each factor was set in five level classifications (one to five). Table 1 shows the explanation of productivity fluctuation factors.

3.3 Industry targeted

The study covered nine industries. Table 2 shows the target industries, their sample sizes, and their relative frequencies. The manufacturing companies (410), distribution businesses (316), and financial businesses (236) topped the list in the descending order of sample size, while the other industries had smaller sample size, e.g. electricity, gas, heat supply, and water industry (96) and construction industry (49). The service industry includes restaurants, accommodation, medical care, welfare, education, and learning support business. The distribution industry includes the transportation, postal, wholesale, and retail. Others include real estate, goods rental business, academic research, and professional and technical service.

Table 2. Sample Size and Relative Frequency by Industry Type

Industry type	Sample size	Relative Frequency
Manufacturing Industry	410	24%
Distribution Industry	316	18%
Finance Industry・ Insurance Inductor	236	14%
Public Service	206	12%
Service Industry	184	11%
Information and Communication Industry	126	7%
Others	98	6%
Electricity, Gas, Heat Supply and Water industry	96	6%
Construction industry	49	3%

3.4 Extracted factor and observation variable

We construct a relationship model using two factors extracted from covariance structure analysis and the productivity variation factors which are six observation variables. From the result of the calculation of the constituent concept score from the two-factor model, the ease of development and the management ability of the upstream process are correlated. In an environment where the development is less complex, the management ability of the upstream process tends to be high; therefore, we performed multiple group structural equation modeling (Multi-group SEM) for each type of industry under these conditions, and we verified the differences between industries.

4. Evaluation Experiment

4.1 Multiple group structural equation modeling

Multi-populations indicate that there are multiple populations whose structures are compared among each other. When the same factor is assumed in multiple populations, the factors are invariant, and it is said that factor invariance holds. Multi-group SEM confirms the degree of application of the model after imposing constraints on the parameters among each group. It is an analytical method to examine whether the groups are equal or different.

In single population analysis, two factors extracted from confirmatory factor analysis (CFA) and six observation variables were extracted. Using this model structure, the Multi-group SEM was conducted in nine industries (nine populations). It basically compares the following five models (model1 to model5) that are considered to be its structures. Eventually, a model, whose conformity is within the allowable range and the constraint condition is the most severe, is adopted.[16]

4.2 Configural invariance model (model1)

Configural invariance model indicates that the factor structure is the same in multi-populations. In other words, the same observation variables are handled among the supposed populations, and the same configuration concept can be assumed.

4.3 Weak invariance model (model2)

Weak invariance model is obtained by adding equality constraints so that the factor loading is the same among the multiple populations. If the nature of the factor is appreciably different, comparing the factor averages would not be meaningful; therefore, to establish the constraint of factor loading in Multi-group SEM, the constraint must be satisfied at the minimum when factor averages are compared.

4.4 Strong invariance model (model3)

Strong invariance model is a model in Multi-group SEM with equality constraints whereby, apart from the factor loading of the weak invariance model, the intercepts also must be equal among the populations, if the difference in sample mean

of the observation variables is indicated as a factor average difference.

4.5 Strict factorial invariance model (model4)

In addition to the equality constraint in the strong invariance model, the strict factorial invariance model has an error variance of observation variables which is considered to be equal among the populations as well.

4.6 Complete invariance model (model5)

It is a model in which all parameters excluding the variance of factors are equal among populations; therefore, in this model, in addition to the strict factorial invariance model, the factor average values also impose equal constraints between populations.

In order to examine the configural invariance and measurement invariance from the above information, we used the model of Figure 1 derived from covariance structure analysis. The results are shown in Table 3. For the fitness index of the model, the following adaptability indexes of the model were used.

Table 3. Result of Multiple Group Structural Equation Modeling

Goodness-of fit	DF	AIC	BIC	Chisq diff	CFI	RMS EA
fit.configural (model1)	72	26,446	27,378		0.946	0.073
fit.loadings (model2)	104	26,444	27,201	61.621	0.924	0.072
fit.intercepts (model3)	136	26,420	27,004	40.974	0.917	0.065
fit.residuals (model4)	184	26,382	26,703	57.242	0.910	0.059
fit.means (model5)	200	26,364	26,598	14.316	0.911	0.056

Adaptability index of model:

Comparative Fit Index (CFI)

$$CFI = 1 - \frac{max(N-1)f_{ML} - df, 0)}{max(N-1)f_0 - df_0, 0)} \quad (4)$$

Root Mean Square Error of Approximation (RMSEA)

$$RSMEA = \sqrt{max(\frac{f_{ML}}{df} - \frac{1}{N-1}, 0)} \quad (5)$$

Akaike Information Criterion (AIC)

$$AIC = \chi^2 - 2df \quad (6)$$

Bayesian Information Criterion (BIC)

$$BIC = \chi^2 - \log(n)\, df \quad (7)$$

CFI and RMSEA are indicators for evaluating the fitness of the model. The closer the CFI value is to 1, the higher the fitness of the model is. The formula for the goodness of fit of CFI is shown in Equation 4. When the value is less than 0.05, RMSEA judges that the model is good for the data, and that the model is not good when the value is 0.1 or more. The formula for the goodness of fit of RMSEA is shown in Equation 5. The AIC and BIC are indicators of goodness of fit between multiple models, respectively, and their low values indicate good models.[16] The expressions of goodness of fit of AIC and BIC are shown in Equations 6 and 7. Based on the above indicators, we evaluated to determine the model that offers the best fit. From Table 3, the fit.means (model5) (CFI of 0.911 and RFMEA of 0.056) showing the lowest AIC and BIC for the most stringent constraint condition was chosen. Based on the selected model5, we created a model with four groups of "Factor loading", "intercept of the observed variables", "residual of the observed variables", and "average of the observed variables" considered together as one group. The analysis tool used the R SEM package and executed the Multi-group SEM. Table 4 shows the results.

Table 4. Factor variance and correlation coefficient

Industry Type	Factor Variance1	Factor Variance2	Factor correlation coefficient
Finance industry, insurance inductor	0.35	0.12	0.33
Information and communication industry	0.35	0.20	0.30
Public service	0.47	0.17	0.32
Service industry	0.38	0.24	0.23
Real estate industry, others	0.43	0.23	0.05
Electricity・Gas・Heat supply・water supply industry	0.44	0.15	0.16
Distribution industry	0.31	0.19	0.42
Construction industry	0.25	0.07	0.17
Manufacturing industry	0.29	0.16	0.70

Next, it was verified from three groups of "variance value of factor 1", "variance value of factor 2", and "correlation coefficient" whether it is divided into several groups. The verification method standardized the data of "variance value of factor 1", "variance value of factor 2", and "correlation coefficient", and then performed the cluster analysis, which is a method to collect similar clusters from a group in which things of different natures are mixed and to classify objects.

It is a hierarchical method of collecting clusters sequentially from the most similar individuals based on the similarity or dissimilarity (distance) between individuals.

First, the hierarchical cluster analysis method obtains the distance or similarity from the data. Next, the coffen matrix is obtained from the selected cluster analysis method. Subsequently, a dendrogram is drawn from the coffen matrix. In this study, the Ward's method was chosen as a method for distance between clusters.

In the Ward's method, grouping is performed by a combination in which the variance within the group is small and the variance among the groups is large. The distance formula between groups of the hierarchical clustering method is as follows.

$$d(C_1, C_2) = L(C_2 \cup C_2) - L(C_1) - L(C_2) \quad (8)$$

where L(C) is the sum of the squares of the distances from the center of gravity.

$$L(C) = \sum D\,(x, gc)^2 \quad (9)$$

gc is the center of gravity of C, D(x, gc) is the Euclidean distance between x and gc.

Figure 2 shows the dendrogram as a result of the cluster analysis, whereby the lower the coupling, the closer the relationship is. In the case of Figure 2, it was divided into two large clustering.

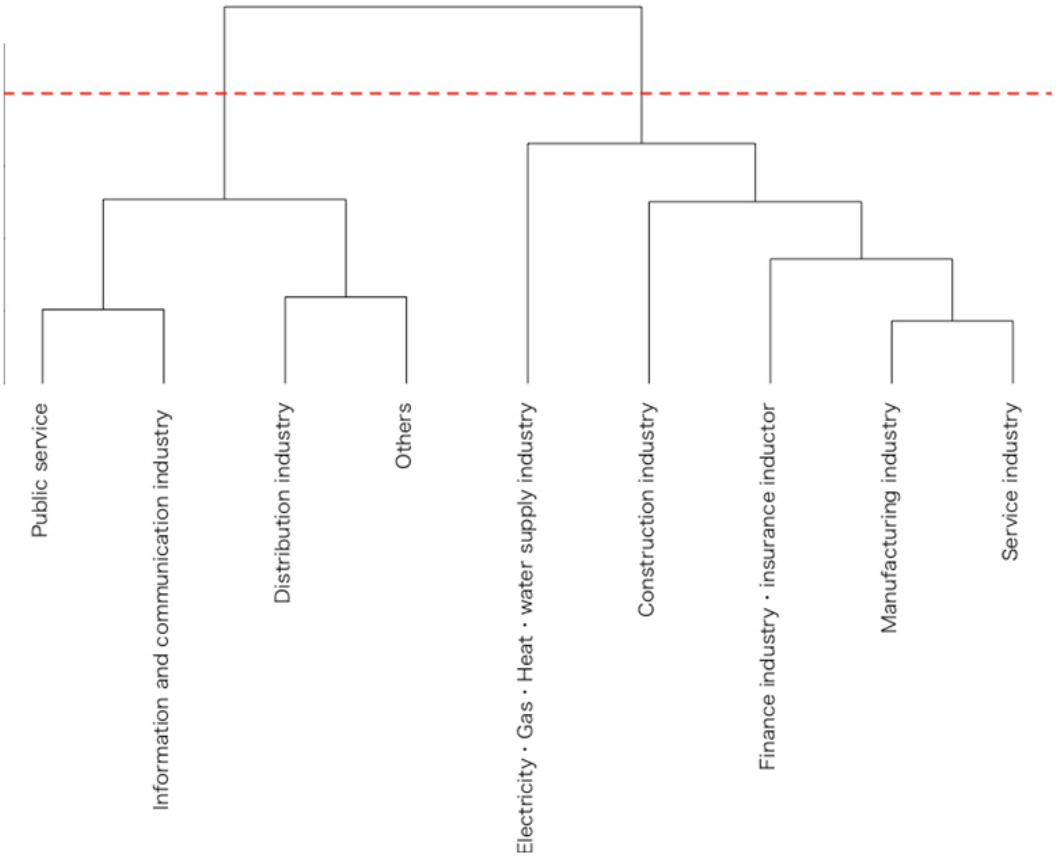

Figure 2. Industry Dendrogram

5. Discussion

Table 4 shows that the variances of factor 1 are low for the three industries: construction industry (0.25), manufacturing industry (0.29), and distribution industry (0.31), compared with the other six industries. This implies that the management and capability in the upstream process of these three industries is more consistent than the other industries. On the other hand, the variances are high for the three industries: public service (0.47), electricity gas, heat supply, and water (0.44) and real estate industry, the others (0.43). Accordingly, this implies that the management and capability in the upstream process has a height difference per project. As one of these results, it is assumed that there are variations in "project management experience and ability" and "analyst experience

and ability", which are observed variables of factor 1. In the case of projects of public institution such as "public service" and "electricity, gas, heat supply, and water", it is assumed that the development scale of software is large. As a result, it is assumed that the complexity of the system development is high.

It is found that the total variance of factor 2 is lower than the total variance of factor 1. This result indicates that there is little error in the ease of development in any industry. In other words, the estimation error in the lower process is smaller than that in the upstream process. This is attributable to a scenario whereby the invisible factor becomes clear as the development process proceeds to the lower process. Especially, in the case of construction industry (0.07) and finance and insurance industry (0.12), the variances are lower than the other industries. This implies that all the projects in these two industries have consistency. Especially, the factor 1 variance (management and capability in the upstream process) is low for construction industry thereby implying that this industry has the highest consistency in both upstream and downstream processes. In other words, this industry has a small estimation error. In addition, the finance industry is expected to have a strong relationship with "clarity and stability in the requirement of order" which is the observation variable of factor 2. As the software development conforms to law, it is conceivable that it is a stringent environment where clarity of requirements and standardization are of utmost significance.

On the other hand, the variance is high in the service industry, information and communication industry, and real estate industry, others. This implies that these industries have difference in height for ease of development between projects.

Next, the correlation coefficient of construction industry is low between factor 1 and factor 2 when we analyze the correlation coefficient per industry, that is, in case of this industry, it does not always show strong relationship with ease of development, even though the management and capability in the upstream process is high. This observation implies that separate IT vendors are in charge of the upstream and the downstream processes. On the other hand, the correlation coefficient of manufacturing industry is extremely high at 0.7 implying that ease of development would be higher for this industry in case of a strong management and capability in the upstream process, and also the ease of development may improve further by assigning an analyst or a project manager with high management and capability in the upstream process. Next, the factor 1, factor 2, and correlation coefficient between factors were grouped into three feature quantities and analyzed using cluster analysis to confirm their classification. As a result, it is observed that they are roughly divided into two clusters. It was assumed that service industry, manufacturing industry, electricity, gas, heat supply, and water industry, and construction industry were classified as Business-To-Business (B-To-B) system, whereas the information and communication industry, public service, distribution industry, and others were classified as Business-To-Customer (B-To-C) system. In the B-To-B system, the service industry was very similar to the manufacturing industry. In addition, it turned out that the cluster group, finance/insurance industry, construction industry, electricity/gas/heat supply/water supply industry were similar in order. In the B-To-C system, the distribution industry and others industry (referred to as cluster1) were similar. In the same manner, the public service and the information and communications industry (referred to as cluster2) were similar. Next, it turned out that the cluster1 groups and the cluster2 groups were similar. An enterprise system was divided into two kinds: the backbone system (B-To-B) and the web system (B-To-C) used by general consumers. In other words, it is assumed that estimation of software development cost depends on the characteristics of B-To-B system and B-To-C system. The B-To-B mainly processes the backbone systems such as office processing, accounting, and inventory management. In these systems, it is considered that factors of factor 2 such as platform conformance, clarity and stability in the requirement of order, reuse of the previous model, and adoption of the standard model are greatly influenced.

On the other hand, in the B-To-C system, the IT ordering company would provide an online system to service the general consumers. In other words, it is necessary to provide functions with high usability and services on a continuous basis. As a result, it is considered that these factors have substantial impact on estimation.

It is inferred from the above discussions that it is critical for the IT vendors (companies) to capture the characteristics of the industries while estimating the man-hours for software development. In addition, it is considered that estimations of increased accuracy are possible by being conscious of the productivity fluctuation factors related to ease of development and management capability of upstream process.

6. CONCLUSION

In this study, we used a two-factor model extracted from CFA of previous study and analyzed the differences by industry type. First, we verified the invariance of the model from the equality constraint model. For the invariance of the model, we compared the five invariant models and identified the "complete invariance model" as the most appropriate model based on the model index results. Then, from the model with

the same parameters selected, the factor loading, intercept of observation variables, residual of observation variables, and average of latent variables were grouped and the analysis was performed using Multi-group SEM. Consequently, factor 1, factor 2, and correlation coefficient between the factors were extracted. Subsequently, using the cluster analysis, differences between industries were examined, and it was possible to find the differences in software estimates from the variance and correlation coefficient between factors, for the factors classified according to the B-To-B and B-To-C system.

In this study, we could use factor models that influence software estimation for each type of industry, and we could capture the characteristics and differences for each industry from simultaneous analysis in multiple populations and cluster analysis. For future research, we propose to apply the estimation model of this research to real-time applications for validation purposes.

REFERENCES

[1] Japan Users Association of Information Systems(JUAS), Corporate IT trend survey 2018. 2018. Available from: http://www.juas.or.jp/cms/media/2018/01/it18_yosan.pdf

[2] Survey on domestic enterprise IT investment trend 2016. Available from: https://www.itr.co.jp/company/press/161019PR.html

[3] Japan Users Association of Information Systems(JUAS), Corporate IT trend survey. 2016.

[4] Information-technology Promotion Agency, Japan Software Engineering Center (IPASEC), Software Development Quotation Guidebook.

[5] Information-technology Promotion Agency,Japan (IPA), software development data hakusho (2016-2017).

[6] Tsuyoshi S. A Study of software estimation factors extracted using covariance structure analysis. KES 2017 (20th International Conference on Knowledge Based and Intelligent Information and Engineering). 2017.

[7] Lagerström R, Würtemberg LM, Holm H, et al. Identifying factors affecting software development cost and productivity. Software Quality Journal. 2011; 20(2): 395-417.

[8] Keung J, Kitchenham B, Jeffery R. Analogy-X: Providing Statistical Inference to Analogy-Based Software Cost Estimation, IEEE Trans. on Software Eng. 2008; 34(4): 471-484. https://doi.org/10.1109/TSE.2008.34

[9] Walkerden F, Jeffery R. An Empirical Study of Analogy-based Software Effort Estimation. Empirical Software Engineering. 1999; 4(4): 135–158.

[10] Ohsugi N, Tsunoda M, Monden A, et al. Effort Estimation Based on Collaborative Filtering, Proc.5th International Conference on Product Focused Software Process Improvement(Profes2004), Kyoto. 2004: 274-286.

[11] Miyazaki Y, Terakado M, Ozaki K, et al. Robust Regression for Developing Software Estimation Models. Journal of Systems and Software. 1994; 27(1): 3-16. https://doi.org/10.1016/0164-1212(94)90110-4

[12] Kurihara EI, Motoei A, Makoto N. Adjusting Estimated Software Development Effort using Cost Drivers Evaluation Data. Information Proessing Society of Japan. 2010; 2010-SE-167(2).

[13] Boehm BW, et al. Software Cost Estimation with COCOMO II, Englewood Cliffs, NJ: Prentice-Hall. 2000.

[14] COCOMO II Model Definition Manual The Center for Software, USC(2000).

[15] General Foundation Economic Research Association: Analysis of software development data repository. 2015.

[16] Hideki T. Covariance Structure Analysis - Structural Equation Modeling. 1998.