

Is This Cost Estimate Reliable? – The Relationship between Homogeneity of Analogues and Estimation Reliability

Naoki Ohsugi¹, Akito Monden¹, Nahomi Kikuchi², Michael D. Barker¹
Masateru Tsunoda¹, Takeshi Kakimoto¹ and Ken-ichi Matsumoto¹

¹Graduate School of Information Science, Nara Institute of Science and Technology

²Software Engineering Center, Information-technology Promotion Agency, Japan

{naoki-o, akito-m, mbarker, masate-t, takesi-k, matumoto} @is.naist.jp

n-kiku@ipa.go.jp

Abstract

Analogy-based cost estimation provides a useful and intuitive means to support decision making in software project management. It derives a cost estimate required for completing a project from information about similar past projects, namely the analogues. While on average this method provides a relatively accurate cost estimate there remains a possibility of large estimation errors. In this paper, we empirically tested the hypothesis that “using more homogeneous analogues produces a more reliable cost estimate” using a software engineering data repository established by the Software Engineering Center (SEC), Information-technology Promotion Agency, Japan. This testing showed that low and high homogeneity projects had a large variation in estimation reliability. For instance, the difference was 22.9% ($p = 0.021$) in terms of percentage to get accurate estimates (better than Median of Magnitude of Relative Error).

1. Introduction

Formal cost and effort estimation methods have been proposed both as the basis for project scheduling and the baseline for bidding on external contracts. Formal estimation methods derive a cost estimate from a project’s observable characteristics or *project features* [3]. Typical examples include function points, work period and engineers’ skill levels. These project features are measured together with the actual cost required to complete a project; then, this information is stored in a project data repository for use in deriving new cost estimates.

On average, these methods provide an accurate cost estimate; however the accuracy varies enormously

between individual projects. For example, Toda et al. [23] estimated the cost of testing for 68 projects in the ISBSG data repository [13]. In their case, a stepwise regression model accurately estimated 30 projects with 30% or less relative errors, while estimates for 13 other projects had more than 100% errors. This suggests project managers must face difficulty in trusting the individual estimate even though the average estimation error is low. If they knew the *estimation reliability* or uncertainty for an individual project, it would accelerate the diffusion of formal estimation methods.

In this paper, we experimentally test a hypothesis about the reliability of *analogy-based cost estimation* “*more homogenous analogues for a project produce a more reliable cost estimate.*” This hypothesis is based on the fact that analogy-based methods typically use an average (or a weighted-sum) of the analogues’ costs to derive the target’s cost estimate. We hypothesize that the estimated cost is unreliable when the analogues’ costs vary widely. For instance, consider the following two cases: the analogues’ costs in person-hours are {2900, 3000, 3100}; or they are {500, 2000, 6500}. Although in both cases the estimated cost is 3000 (by averaging analogues’ costs), the latter seems to be less reliable. In this paper, the homogeneity of analogues is evaluated with a statistical measure, *Coefficient of Variation* of the analogues’ costs. Then, the relationship between these measures and the estimation reliability is investigated experimentally.

To test the hypothesis, we implemented an analogy-based cost estimation method by combining advanced features of the typical analogy-based tools *ESTOR* [15], *ANGEL* [18] and *ACE* [26]. At first, the implemented method searches a project data repository for analogues by the same manner of ANGEL; i.e. Euclidean distance between completed projects and the target is calculated. Once the analogues are determined,

a linear size extrapolation is performed by the same manner of ACE; i.e. individual analogues' costs are adjusted along a project size feature such as function points. Then, the cost estimate is calculated by averaging analogues' adjusted costs. The Euclidean distance offers the benefits of clarity and consistency over human judgement alone. The linear size adjustment attempts to account for the influence on cost of the difference in size between the target and the completed projects [26]. The utilized method combines these advantages to make accurate estimations.

We applied this method to a project repository called the *SEC data repository* [19] established by the Software Engineering Center (SEC), Information-technology Promotion Agency, Japan. As of March 2005, the SEC data repository consists of 1009 projects from 15 Japanese software companies. We selected 14 project features available at the time of estimation (at the end of architecture design phase) and that contained less missing values than the others. Then, we extracted for estimation 104 projects having all of the selected features and the completion cost (effort).

The remainder of this paper is structured as follows. Section 2 introduces the background of the study, especially about estimation reliability and analogy-based cost estimation. Section 3 explains the hypothesis to be analysed more precisely. Section 4 explains an analysis for the hypothesis testing, especially regarding the used data and the experimental procedure. Then, Section 5 presents the results; and Section 6 discusses the implications given by the results. Section 7 concludes the paper and provides directions for further research.

2. Background

2.1. Estimation Reliability

In this paper, we define estimation reliability as the probability that the estimation error becomes smaller than a threshold value. The threshold can be given by some standard evaluation criteria such as *Mean Magnitude of Relative Error (MMRE)* or *Median of Magnitude of Relative Error (MdMRE)*. As the basis of these criteria, the *Magnitude of Relative Error (MRE)* is defined as the following equation (1) indicating the degree of estimation error against the actual cost:

$$MRE = \frac{|c_a - e_a|}{c_a} \quad (1)$$

where c_a is the cost actually spent to complete a project p_a , and e_a is its cost estimate. MRE is usually calculated for many projects in a data repository under a standard evaluation process such as cross-validation [5]. Then, its mean and median values are used as evaluation criteria; *Mean MRE (MMRE)* and *Median of MRE (MdMRE)*.

Estimations reliability is usually assessed by exploring reasons or risk factors for estimation error, and then checking whether the target project has the potential risks. Risk factors have been investigated by questionnaire-based approaches [17], [25], [20] and by statistical analysis [11], [7], [10]. For example, Lederer and Prasad [12] reported that one of the most important risks is having “overlooked tasks” in measurement. In this approach, if the possibility of having the risk is high, the estimation reliability is considered low. One of the difficulties with this method is that the assessment of some kinds of risks is not easy in principle. For example, the possibility of having “overlooked tasks” in past measurement is often not clear at the time of making an estimate. In comparison with these previous studies, our study focuses on an easy-to-assess aspect of a project, the homogeneity of analogues, to avoid the difficulty of assessment.

There also have been several attempts to build formal methods to assess the estimation reliability systematically [2], [9]. These studies proposed methods to demonstrate the estimation reliability with a confidence interval of the cost estimation. A confidence interval consists of a maximum and a minimum bound of cost estimate and its confidence level. Jørgensen and Sjøberg [9] proposed a method to identify the confidence interval using the distribution of previous estimation accuracy. Angelis and Stamelos [2] developed a tool identifying the confidence interval by repeating accuracy evaluation with bootstrap sampling. Our study shares their goal to provide a systematic and objective way to assess the estimation reliability for an individual project, but the approach is different. Our contribution is to provide practitioners with clear and intelligible evidence of estimation reliability so that they become more confident in trusting/distrusting the derived cost estimate. We believe it is useful to develop several different types of reliability assessment methods to promote the use of formal estimation methods in the field.

2.2. Analogy-based Cost Estimation

We focus on *analogy-based cost estimation*, which is an application of *Case-Based Reasoning (CBR)* in software engineering [26]. CBR emerged as a major

research area within artificial intelligence [6], and was first introduced to software development cost estimation by Mukhopadhyay et al [15]. The process of analogy-based cost estimation is as follows. First, as inputs to estimation, observable project features of a target (ongoing) project are selected. Then, one or more analogues having similar features to the target project are found in a project data repository. Finally, a cost estimate is derived using the known costs of the analogues. Our expectation is that the analogues contain valuable information about the reliability of estimation for the target project.

We have developed a prototype tool *Trinity* to analyze the nature of analogy-based estimation. Trinity makes estimates based on techniques used in three typical tools *ESTOR* [15], *ANGEL* [18] and *ACE* [26] as follows. At first, Trinity calculates the *Euclidean distance* between the target project and the completed projects, and chooses the closest k -projects as the analogues [15][18]. Here, the number k is selected by the same method as used in ANGEL, an exploratory search to minimize MMRE for a particular project repository. Once the analogues are determined, Trinity then adjusts each analogue's cost according to function points. Finally, Trinity derives the cost estimate by averaging the adjusted costs of the analogues.

More precisely, if each project p_i is described by a set of values $\{c_i, v_{i,1}, v_{i,2}, \dots, v_{i,m}\}$, Trinity calculates the Euclidean distance $dist(p_a, p_i)$ between the target project p_a and each completed project p_i , as:

$$dist(p_a, p_i) = \sqrt{\sum_{j \in F} (v_{a,j} - v_{i,j})^2} \quad (2)$$

where F denotes a set of project features observed at the time of estimation, and $v_{i,j}$ denotes the project p_i 's j -th feature value. Then, Trinity derives the target project p_a 's cost estimate e_a , as:

$$e_a = \frac{1}{k} \sum_{i \in A} \left(c_i \frac{fp_a}{fp_i} \right) \quad (3)$$

where A denotes a set of the analogues, k denotes the number of the analogues; c_i and fp_i denote each similar project p_i 's cost and its function points. Before the calculation, all categorical features are transformed as dummy variables. In addition, all ranges of values are normalized to interval $[0, 1]$ to equalize the influence of each project feature for similarity computation [3].

Note that this paper does not propose any analogy-based estimation method. Trinity is true to its name, consisting of the great contributions of three representative studies in the area.

3. Hypothesis

The hypothesis to be tested is “*the higher the homogeneity of the analogues, the more reliable a project's estimate is.*” In this paper, we attempt to measure the homogeneity by a basic statistical metric, *Coefficient of Variation (CV)* of analogues' costs, i.e., the homogeneity is calculated as:

$$CV = \frac{\delta_A}{\mu_A} = \frac{1}{\mu_A} \sqrt{\frac{1}{k-1} \sum_{i \in A} (c_i - \mu_A)^2} \quad (4)$$

where A denotes a set of analogues, δ_A and μ_A denote the standard deviation and the arithmetic mean of the analogues' costs, k denotes the number of the analogues, c_i denotes each analogue's cost. CV is a measure of dispersion of a probability distribution, defined as the ratio of the standard deviation to the mean. It expresses how spread out the values are [16]. In addition, it allows comparison of the variation of populations that have significantly different mean values because it is a dimensionless number. Note that CV has converse relation to homogeneity. In other words, a smaller CV indicates higher homogeneity.

We consider there are two typical cases where analogues' costs become heterogeneous. The first case is that the analogues have some measured project feature causing cost uncertainty; and the second case is that the analogues have potential uncertainty in project features that are unavailable at the time of estimation. We explain these cases with three toy examples (Figure 1 to 3). Each figure assumed that the completion cost and three project features (FP, FP Counting Method and Architecture Type) were available. Also, each figure assumed that three projects were found as analogues having feature values that are very similar to each target project. Although the estimated costs were 3000 in all cases, only in the Figure 1, CV of the analogues' costs was small 0.03 (i.e. homogeneous).

Figure 1 shows a case where homogeneous analogues were found. These contribute to make high reliability estimation if our hypothesis is true.

Figure 2 shows a case where project features cause cost uncertainty. In this example, we assume the FP counting method “NESMA” is less reliable than “IFPUG” since NESMA is usually used in a more upstream development phase. And thus the accuracy of FP is low in this case. It can often happen that the measured FP differs from actual FP [8]. In this case, the project feature “NESMA” caused cost uncertainty, resulting in heterogeneity of analogues' costs. These analogues are considered to cause a low reliable estimation.

CV = 0.03

	Cost	FP	FP Counting Method	Architecture Type
Project 1	2900	198	IFPUG	Stand-alone
Project 2	3000	200	IFPUG	Stand-alone
Project 3	3100	202	IFPUG	Stand-alone

Figure 1. The analogues are homogeneous

CV = 0.88

	Cost	FP	FP Counting Method	Architecture Type
Project 4	1000	198	NESMA	Stand-alone
Project 5	2000	200	NESMA	Stand-alone
Project 6	6000	202	NESMA	Stand-alone

Figure 2. The analogues have project features causing cost uncertainty

CV = 0.88

	Cost	FP	FP Counting Method	Architecture Type	Overlooked Tasks
Project 7	1000	198	IFPUG	Stand-alone	0
Project 8	2000	200	IFPUG	Stand-alone	500
Project 9	6000	202	IFPUG	Stand-alone	4000

Figure 3. The analogues have potential uncertainty in unavailable project features

Figure 3 shows another case where heterogeneous analogues were found, but in this case the cause of heterogeneity was in the unavailable project features. In this example, both project 8 and 9 spent cost for some special purpose, namely *Overlooked Tasks*, which was not measured. This resulted in the heterogeneity of the analogues' costs. These analogues are also considered to cause a low reliable estimation.

4. Analysis for Testing Hypothesis

4.1. Overview

We experimentally tested the hypothesis described in Section 3. In the analysis, Trinity was used to estimate the total cost of each project in the SEC data repository. The accuracy and the reliability of each estimated cost were evaluated by leave-one-out cross validation [1]. Afterwards, each estimated cost was classified into three levels (*Low / Medium / High*) according to the homogeneity of analogues used for the estimation. Then, the differences of the accuracy and the reliability between the levels were tested with statistical test methods such as *Welch's t-test*, *Mann-Whitney U test* and *Fisher's exact test* [14]. At last, the dispersion of the estimation accuracy was visually shown with the *Outlier box plot* [21] to confirm that the results of the statistical test meet the analyzers' subjective evaluation.

4.2. Analyzed Data

The source of data used in the analysis was the 2005 version of the SEC data repository containing 1009 projects from 15 Japanese software companies. Projects completion dates ranged from 1996 to 2005. All of the projects are custom enterprise software developments. Each project is characterized by about 400 features, although many include a large number of missing values. The business domains of these projects include manufacturing, communications, wholesale and retail sales, finance and insurance, etc. The variety of projects in the data covers a large part of the Japanese enterprise software industry. We believe that the SEC data repository is a good test bed since it is maintained by measurement specialists who keep on improving the data quality based on follow-up surveys and interviews with companies who collected data. This data has been used previously in some studies such as a performance comparison of estimation methods [22] and the productivity analysis [24].

We selected 14 project features available at the time of estimation and that contained less missing values than the others. Then, we extracted for estimation 104 projects having all of the selected features and the completion cost (effort). We assumed that the estimation is done at the end of the architecture design phase, which is the next phase of requirement analysis. So, first, the unavailable project features at the time of

Table 1. Project Features in the Analyzed Data

Name of Feature	Type	Unit / Range / Values
Cost	Numerical	In Person Hours
FP	Numerical	Function Points
FP Counting Method	Categorical	IFPUG Unadjusted, NESMA Estimated
Objective	Categorical	New Development, Maintenance
Product Type	Categorical	Package Software, Company Specific System
Customer's Business	Categorical	Manufacturing, Communications, Wholesale and Retail, Finance and Insurance, etc.
Application Domain	Categorical	Accounting, Marketing, Customer Management, Analysis, etc.
Number of Users	Categorical	Limited, Unlimited
Commercial Package Use	Categorical	Yes, No
Main Architecture Type	Categorical	Stan-alone, Mainframe, 2Tier Client-Server, Internet, etc.
Main Platform	Categorical	Windows, Unix, Other
# of Platforms	Numerical	[1, 3]
Main Development Language	Categorical	COBOL, C, Visual C++, C#, Visual Basic, Java, etc.
# of Languages	Numerical	[1, 3]
Lifecycle Model	Categorical	Waterfall, Iterative, Others

estimation were removed from the data repository. Another point of this extraction is to avoid influences of the missing values on the analysis results.

Table 1 shows the selected project features. The column labelled *Type* contains the scale level of each feature as either “Numerical” or “Categorical”. The column labelled *Unit / Range / Values* contains each feature’s concrete values eligible for disclosure to the public. The *Cost* is the total cost required for completing a project, and is the estimation target in this analysis. The *FP* is the function point count expressing the size of the main product developed in the project. The *Customer’s Business* is the main business of the customer in each project. The *Application Domain* is the situation or the purpose of the use of the product. In the *Number of Users*, the value of “Unlimited” means the product is for general public while “Limited” means some specific customers. The *# of Platforms* and *# of Languages* become two or more if the product was developed with a combination of some platforms and languages.

4.3. Analysis Procedure

The hypothesis was tested by the systematic procedure explained in this section. There was less possibility that the analysis results were affected by the analyzers’ decisions, because most steps in the procedure were automatically executed with Trinity. The analysis consisted of nine steps as described below:

- Step 1.* The cost of each project in the dataset was estimated with the Trinity (leave-one-out cross validation [1]).
- Step 2.* The accuracy of each estimate was evaluated as MRE via the equation (1).
- Step 3.* MMRE and MdmRE of all projects are used as the threshold values, which separate the reliable estimate from unreliable estimate.
- Step 4.* The homogeneity of analogues for each cost estimate was evaluated in terms of CV via the equation (4). Instead of each analogue’s raw cost in the equation (4), the adjusted cost obtained from the equation (3) was used.
- Step 5.* The cost estimates were categorized into three levels (Low / Medium / High) according to the calculated CV of the analogues’ costs. Note that larger CV level means lower homogeneity of analogues.
- Step 6.* For each level, each *Level’s Mean MRE (LMMRE)* and each *Level’s Median of MRE (LMdMRE)* were calculated as the average and the median of MREs of the projects included in each level. For instance, LMMRE was calculated as:

$$LMMRE = \frac{\sum_{i \in P} MRE_i}{n} \quad (5)$$

where P denotes a set of projects included in a given level, n denotes the number of these projects, and MRE_i denote the MRE of the cost estimate for a project p_i .

Step 7. For each level, *Reliability of MMRE (RMMRE)* and *Reliability of MdMRE (RMdMRE)* were calculated as:

$$RMMRE = \frac{\sum_{i \in P} r_i}{n} \times 100(\%) \quad \begin{cases} r_i = 1, MRE_i < MMRE \\ r_i = 0, MRE_i \geq MMRE \end{cases} \quad (6)$$

$$RMdMRE = \frac{\sum_{i \in P} r_i}{n} \times 100(\%) \quad \begin{cases} r_i = 1, MRE_i < MdMRE \\ r_i = 0, MRE_i \geq MdMRE \end{cases} \quad (7)$$

where the definitions of P , n and MRE_i are the same as Step 6.

Step 8. The differences of LMMRE, RMMRE, LMdMRE and RMdMRE between each two levels were statistically tested. The statistical test methods were appropriately chosen according to the type of the tested values. LMMRE was tested with Welch's t-test, LMdMRE was tested with Mann-Whitney U test, and RMMRE and RMdMRE were tested with Fisher's exact test.

Step 9. The dispersion of MRE in each level was graphed with a outlier box plot to confirm the results of statistical tests.

5. Results

Table 2 shows the relationship between estimation reliability and homogeneity. Each column denotes each degree of homogeneity (i.e. CV level), "Range of CV" row contains the minimum and the maximum value of CV in each level, "# of Projects" row contains the number of projects in each level. The other rows contain the values of LMMRE, RMMRE, LMdMRE and RMdMRE.

Table 3 shows the difference of estimation reliability between degrees of homogeneity. Each column denotes the difference between each two degrees of homogeneity. Each cell contains the difference of LMMRE, RMMRE, LMdMRE and RMdMRE between corresponding column's degrees. Each cell also contains a p -value in parentheses for the difference. For example, the rightmost bottom cell contains difference of RMdMRE between Low and High homogeneity, and p -value of the difference. If the difference was statistically significant at the 10% level ($p < 0.1$), the p -value is underlined. In addition, if a value is quoted in the body text, it is emphasized with bold font.

Figure 4 shows MRE dispersion with respect to homogeneity of analogues as an outlier box plot diagram. The horizontal axis denotes the degree of homogeneity (CV level). The vertical axis denotes

MRE, so that lower points indicate lower estimation errors, i.e. higher accuracy. Each box stretches from the 25th percentile to the 75th percentile of MRE, therefore it contains the middle half in the distribution of each level. The LMMRE is shown as a plus (+) mark on the vertical line. The LMdMRE is shown as a line across the box. A vertical line is drawn in the center of the box from the smallest MRE to the largest MRE excluding the outliers. The outliers having 1.5 times the difference between the 25th and the 75th percentile are drawn as small circles. In this case, low homogeneity (high CV level) has three extremely large outliers (5.0 or more MREs) outside of Figure 4. The numbers at the right side of the box indicate the value of the representative points including the LMMRE, the LMdMRE, the 25th and the 75th percentile, the top and the bottom of the vertical lines.

6. Discussion

For both LMMRE and LMdMRE, when the analogues' homogeneity was low, the estimation reliability was relatively lower than the others; and there were statistically significant differences ($p < 0.1$). For instance, in terms of RMMRE, the difference between medium and low homogeneity was 19.4% ($p = 0.032$); and the difference between low and high was 17.1% ($p = 0.051$). This tendency was consistent with RMdMRE, the difference between medium and low homogeneity was 15.8% ($p = 0.052$); and the difference between low and high was 22.9% ($p = 0.021$). In addition, Figure 4 visually shows clear differences of estimation accuracy between homogeneity low and the other levels. We believe these results strongly support our hypothesis.

On the other hand, statistical significance was not observed in the difference between high and medium homogeneity. In terms of MdMRE, the estimation reliability was higher when the homogeneity was high; however the relationship was converse in MMRE. Also the observed p -values show there was no difference in them. These results suggest the estimation reliability became extremely lower when the homogeneity was lower than the certain degree. So, there would be an "optimized" threshold for CV to distinguish the unreliable estimates from reliable ones. Actually, if a threshold was put between the low and the medium homogeneity, i.e., between CV level of high and medium, large differences can be observed in the estimation reliability. For instance, the difference of RMMRE is 18.25% (nearly equal to the average of 19.4% and 17.1%); and the difference of RMdMRE is 19.35% (nearly equal to the average of 15.8% and 22.9%).

Table 2. Relationship between Estimation Reliability and Homogeneity of Analogues

Homogeneity	High	Medium	Low
CV Level	Low	Medium	High
Range of CV	[0.000, 0.705)	[0.705, 0.944)	[0.944, 3.225]
# of Projects	35	34	35
LMMRE	0.747	0.607	1.515
RMMRE	77.1%	79.4%	60.0%
LMdMRE	0.364	0.499	0.770
RMdMRE	60.0%	52.9%	37.1%

Table 3. Difference of Estimation Reliability between Degrees of Homogeneity

Homogeneity	High vs Medium	Medium vs Low	Low vs High
CV Level	Low vs Medium	Medium vs High	High vs Low
LMMRE	0.140	0.908	0.768
(<i>p</i> -value)	(0.418)	(<u>0.041</u>)	(<u>0.092</u>)
RMMRE	2.3%	19.4%	17.1%
(<i>p</i> -value)	(0.294)	(<u>0.032</u>)	(<u>0.051</u>)
LMdMRE	0.135	0.270	0.406
(<i>p</i> -value)	(0.933)	(<u>0.023</u>)	(<u>0.034</u>)
RMdMRE	7.1%	15.8%	22.9%
(<i>p</i> -value)	(0.137)	(<u>0.052</u>)	(<u>0.021</u>)

** Each () contain *p*-value. The underlined *p*-values indicate the difference was statistically significant at the 10% level ($p < 0.1$). The values emphasized with bold font are quoted in the body text.

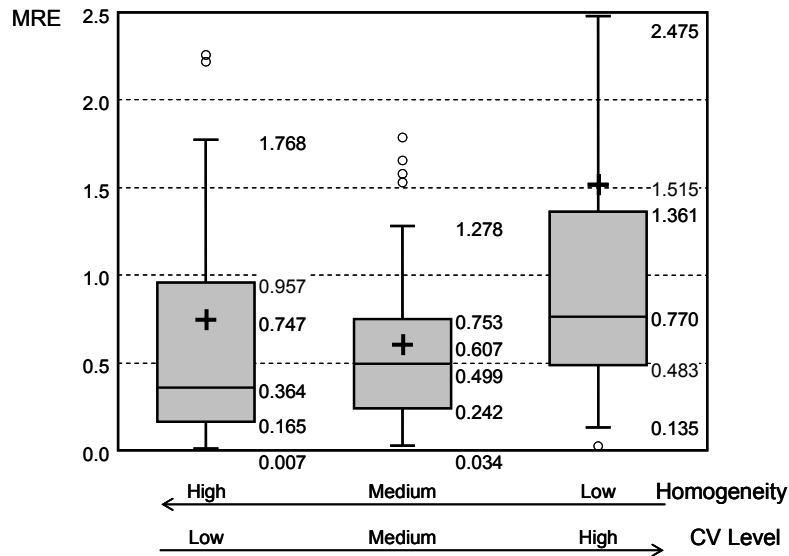


Figure 4. Outlier Box Plot of MRE Dispersion with Respect to Homogeneity of Analogues

7. Conclusion

In this paper we have experimentally tested a hypothesis “using more homogeneous analogues for a project to be estimated produces a more reliable cost estimate.” The homogeneity is evaluated with *Coefficient of Variation (CV)* of the analogues’ costs. The hypothesis was statistically tested based on cost estimates obtained by applying the analogy-based approach to the dataset containing 104 projects each having 14 project features.

As a result, a large variation in reliability was observed between high and low homogeneity level projects. For instance, in terms of *Reliability of Mean Magnitude of Relative Error (RMMRE)*, the difference between medium and low homogeneity was 19.4% ($p = 0.032$); and the difference between low and high was 17.1% ($p = 0.051$). This tendency was consistent with *Reliability of Median of Magnitude of Relative Error (RMdMRE)*, the difference between medium and low homogeneity was 15.8% ($p = 0.052$); and the difference between low and high was 22.9% ($p = 0.021$).

The results also suggested that there would be an “optimized” threshold for CV to distinguish the unreliable estimates from reliable ones. Actually, if a threshold was put between the low and the medium homogeneity, large differences can be observed in the estimation reliability. For instance, the difference of RMMRE is 18.25% (nearly equal to the average of 19.4% and 17.1%); and the difference of RMdMRE is 19.35% (nearly equal to the average of 15.8% and 22.9%).

A limitation of this study is that the hypothesis was tested based on just one dataset using one estimation algorithm. To more fully verify the results, we need to use different datasets and algorithms.

8. Acknowledgements

This work has been done under the collaboration between Nara Institute of Science and Technology and Software Engineering Center, Information-technology Promotion Agency, Japan. The collaboration is supported by the Comprehensive Development of e-Society Foundation Software program of the Ministry of Education, Culture, Sports, Science and Technology.

9. References

- [1] D.M. Allen, “The Relationship between Variable Selection and Data Augmentation and a Method for Prediction”, *Technometrics*, vol.16, no.1, pp.125-127, 1974.
- [2] L. Angelis, and I. Stamelos, “A Simulation Tool for Efficient Analogy Based Cost Estimation”, *Empirical Software Eng.*, vol.5, no.1, pp.35-68, 2000.
- [3] M. Auer, A. Trendowicz, B. Graser, E. Haunschmid, and S. Biffl, “Optimal Project Feature Weights in Analogy-Based Cost Estimation: Improvement and Limitations”, *IEEE Trans. on Software Eng.*, vol.32, no.2, pp.83-92, 2006.
- [4] B.W. Boehm, “Software Engineering Economics”, *IEEE Trans. on Software Eng.*, vol.10, no.1, pp.4-21, 1984.
- [5] L.C. Briand, V.R. Basili, and W.M. Thomas, “A Pattern Recognition Approach for Software Engineering Data Analysis”, *IEEE Trans. on Software Eng.*, vol.18, no.11, pp.931-942, 1992.
- [6] S. Dutta, B. Wierenga, and A. Dalebout, “Case-Based Reasoning Systems: From Automation to Decision-Aiding and Stimulation”, *IEEE Trans. on Knowledge and Data Eng.*, vol.9, no.6, pp.911-922, 1997.
- [7] A. Gray, S. MacDonnell, and M. Shepperd, “Factors Systematically Associated with Errors in Subjective Estimates of Software Development Effort: The Stability of Expert Judgment”, In *Proc. of the 6th Int’l Software Metrics Symposium*, pp.216-227, 1999.
- [8] ISO/IEC 24570:2005: Software Engineering - NESMA Functional Size Measurement Method v.2.2 - Definitions and Counting Guidelines for the Application of Function Point Analysis, 2005.
- [9] M. Jørgensen, and D.I.K. Sjøberg, “An Effort Prediction Interval Approach Based on the Empirical Distribution of Previous Estimation Accuracy”, *Information and Software Technology*, vol.45, no.3, pp.123-136, 2003.
- [10] M. Jørgensen, and K. Moløkken-Østfold, “Reasons for Software Effort Estimation Error: Impact of Respondent Role, Information Collection Approach, and Data Analysis Method”, *IEEE Trans. on Software Eng.*, vol.30, no.12, pp.993-1007, 2004.
- [11] A.L. Lederer, and J. Prasad, “Information Systems Software Cost Estimating: A Current Assessment”, *Journal of Information Technology*, vol.8, no.1, pp.22-33, 1993.
- [12] A.L. Lederer, and J. Prasad, “Causes of Inaccurate Software Development Cost Estimates”, *Journal of Systems and Software*, vol.31, no.2, pp.125-134, 1995.
- [13] C. Lokan, T. Wright, P.R. Hill, and M. Stringer, “Organizational Benchmarking Using the ISBSG Data Repository”, *IEEE Software*, vol. 18, no. 5, pp. 26-32, 2001.

- [14] P.S. Myles, T. Gim, “Statistical Methods for Anaesthesia and Intensive Care”, Butterworth-Heinemann, 2000.
- [15] T. Mukhopadhyay, S.S. Vicinanza, M.J. Prietula, “Examining the Feasibility of a Case-Based Reasoning Model for Software Effort Estimation”, *The Management Information Systems Quarterly*, vol.16, no. 2, pp.155-171, 1992.
- [16] K. Pearson, “On the Dissection of Asymmetrical Frequency Curves”, *Philosophical Transaction of Royal Society*, vol.A185, pp.71-110, 1894.
- [17] D. Phan, D. Vogel, and J. Nunamaker, “The Search for Perfect Project Management”, *Computerworld*, pp.95-100, 1988.
- [18] M. Shepperd, and C. Schofield, “Estimating Software Project Effort Using Analogies”, *IEEE Trans. on Software Eng.*, vol.23, no.12, pp.76-743, 1997.
- [19] Software Engineering Center, Information-Technology Promotion Agency, Japan, “The 2005 White Paper on Software Development Projects”, Nikkei Business Publications, 2005 (in Japanese).
- [20] G.H. Subramanian, and S. Breslawski, “An Empirical Analysis of Software Effort Estimate Alterations”, *Journal of Systems and Software*, vol.31, no.2, pp.135-141, 1995.
- [21] J.W. Tukey, “Exploratory Data Analysis”, Addison-Wesley, 1977.
- [22] N. Ohsugi, M. Tsunoda, A. Monden, T. Matsumura, K. Matsumoto, and N. Kikuchi, “Using Cross-company Data to Estimate the Effort of Software Development Projects”, *SEC journal*, no.5, pp.16-25, 2006 (in Japanese).
- [23] K. Toda, M. Tsunoda, A. Monden, and K. Matsumoto, “Characterizing Software Projects Unpredictable by Effort Estimation Models”, *Technical Report of IEICE*, No.SS2005-74, pp.67-72, 2005 (in Japanese).
- [24] M. Tsunoda, A. Monden, H. Yadohisa, N. Kikuchi, K. Matsumoto, “Productivity Analysis of Japanese Enterprise Software Development Projects”, In *Proc. of 2006 Int'l Workshop on Mining Software Repositories (MSR2006)*, pp.14-17, 2006.
- [25] M. Van Genuchten, “Why is Software Late? An Empirical Study of Reasons for Delay in Software Development”, *IEEE Trans. on Software Eng.*, vol.17, no.6, pp.582-590, 1991.
- [26] F. Walkerden and D.R. Jeffery, “An Empirical Study of Analogy-Based Software Effort Estimation”, *Empirical Software Engineering*, vol.4, no.2, pp.135-158, 1999.