Software Productivity Prediction by Using Structural Equation Model

Niroot Por-ngam[1], Somchai Prakancharoen[2], and Thanarak Limjitsomboon[3]
[1]Faculty of Technical Education,
Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand
[2]Faculty of Science,
Rajabhat Chandrakasem University, Bangkok, Thailand
[3]TRUE Cooperation, Bangkok, Thailand
[1]niroot@rmutt.ac.th
[2]somchai.p@chandra.ac.th
[3]thanarak_l@hotmail.com

Received: 6/8/2019
Accepted: 23/8/2019

Abstract **- Software development projects anxiety are always coverage about development time, quality and limited resource. The objective of this research is to find out the factors that have direct effect with software development productivity. Two hundred sample data were collected from fifteen small size software development firms. This project coverage on finished small size software development project which could sign off the project in about six months. Factor analysis was used to explore factors from observation variables. Structural equation model was used to confirm purposed causal model of hypotheses. Prediction equation was cross validation to measure accuracy of software productivity prediction equation. Thirty observations were used in cross validation testing. MMRE of testing was about 34.53 percent.**

Keywords **- Software Productivity; Prediction; Structural Equation Model**

I. INTRODUCTION

Software productivity is important concerns of many software development firms. Productivity meaning in this research is covered about software size, software bugs, task complete and hour of work of developed software. Good software product should have characteristics that represent prefer qualities. The software size should have a small size as there could be easily installed in computer system with less resource consumption. Software should free of bugs as much as possible. Processing time should not take a long time. Delivered software must complete all request function and non functional of customer requirements. Unfortunately, many software development projects have difference conditions and environments.

Therefore, this research tries to limit distinct of sample data by to prescribe the firm size and size of software project. Sample data were collected from fifteen small size software development firms in Thailand. The software projects were considered only less than six month development software project. The output of this research was a software productivity prediction equation. Software house should be used it to manage, on some significant attributes, oncoming software development project in order to increase their productivity.

II. RELATED THEORY AND RESEARCH

A. Theory

1) ***Factor Analysis [1]***

Factor analysis is a method that used to form latent variable which should represent as a component of related observed variables. Some attributes may not be composed in formed latent variable cause of low correlation with other attributes. Therefore, Factor analysis is a method to explore if group of observation variables are able to be composed as a new latent variable, call factor. Kaiser-Meyer-Olkin (KMO) test is a technique to test if sampling adequacy of data is suited to Factor analysis. KMO value should be, at least, greater than 0.6. Cronbach's Alpha test is used to detect how well of grouping observed variable together under either correlation or covariance consideration. Cronbach's Alpha value above 0.7 is an acceptable condition.

2) ***Structural Equation Model [2]***

Structural equation model, SEM, is a method that is used to detect a user's propose causal model if it is a suitable model. Causal model presents factors relationship to other factors. Each factor composed of its members, related observed variables. Model fitting, or suitable, is measured by many statistical test. Chi square test is an important one that "p" value should be significant less than 0.05. Root mean square error of approximation (RMSEA) value should be less than 0.05. Goodness of fit index (GFI) value should be greater than 0.9.

3) ***Cross Validation [3]***

Prediction model is created from training data. This model is either well use in other data observation prediction or not. This model should be tested with independent data observations in order to detect how accuracy of prediction. The result of testing should be conducted on difference testing finished observation data so that the average accuracy of prediction is summarized. There are many accuracy measurement techniques such as mean magnitude of relative error (MMRE).

$$MMRE = \frac{1.}{n} \frac{\sum_{i=1}^{n} |\, actual - estimated \,|}{actual} \quad (1)$$

While n is number of cross validation observations, testing, actual is a real value of class or dependent variable and estimated is an estimate value of the same testing observation dependent variable. MMRE should be multiplied by 100 to gain a percentage error of prediction.

B. Related Research

Jones [4] suggests that there are six factors that related to productivity of software development.

Factor 1: Financial (FIC) constrain which composes of software budget (Budt), programmer salaries (SALS), availability of computer resources (ACRE) and availability of computer programming tools (APTR).

Factor 2: Time constraint (TICT) which composes of development time (TTCN), amount of fine if project delay (CODY), programmer available (PGAE).

Factor 3: Software requirement - specification (SWSN) which composes of flexibility (FLXY), efficiency (EFFY) and portability (PORY).

Factor 4: Programming methodology (PGMY) which composes of use of programming generator (PGPR), prototyping-tool (PTTL), defect detecting (DFDE), documentation (COMG), programming paradigm (PGGM) (structure or object oriented) and design on maintainability / upgradability / portability (DMUP).

Factor 5: Software development environment (SWDE) which composes of organization structure, organization politic and geographic firm location.

Factor 6: Uncontrollable environment (UCET) which compose of organization size, newest of required programming language, programmer team moral.

Tyler [5] suggest that software productivity (PDTY) could measure by many metrics such as source line of code (LOCE), amount hour of software development using (HOWK), software error – defect correction (BUFD) and amount of complete requirements (TACD).

III. RESEARCH METHODOLOGY

A. Limitation

In order to prevent variety of software development firm under distinct environment and constraints then this research try to avoid diversity of software project development by setting some limitations.

First, software project development observations were chosen from only small size software development project segment so that software project development could be finished on not over six months. This condition guarantees that each project should have a similarity in financial, time constraint, software specification and program methodology.

Second, software development project were a small size of programming that could be written by only one or small team programmer.

B. Sample

Sample data of finished software development project were collected from fifteen Thailand software development firm, software house. Data were collected during May 2018 - May 2019. Respondent of questionnaire were software project manager of those finished software project sample. Two hundred observations were reserved to be a training data set. Thirty observations were used to test of model cross validation.

C. Hypothesis

1) ***Exploratory for Factor***

Factor 1: Financial (FIC) was composed of observe variables software budget (BUDT), programmer salaries (SALS), Availability of Computer Resources (ACRE) and availability of computer programming tools (APTR).

Factor 2: Time constraint (TICT) was composed of development time (TTCN), amount of fine, to pay, if project delay (CODY), programmer available time.

Factor 3: Software requirement-specification (SWSN) was composed of flexibility (FLXY), efficiency (EFFY) and portability (PORY).

Factor 4: Programming methodology (PGMY) was composed of use of programming generator (PGPR), prototyping-tool (PTTL), defect detecting (DFDE), documentation (COMG), programming paradigm (PGGM) (structure or object oriented) and design on maintainability / upgradability / portability.

2) ***Research Hypothesis-Confirmation***

H1: Factor-Financial (FIC) has direct effect to factor software productivity (PDTY).

H2: Factor-Time constraint (TICT) has direct effect to factor software productivity (PDTY).

H3: Factor-Software requirement-specification (SWSN) has direct effect to factor software productivity (PDTY).

H4: Factor-Programming methodology (PGMY) has direct effect to factor software productivity (PDTY).

D. Questionnaire

Questionnaire was designed as shown in table I. It was composed of two parts.

Part 1: request respondent to inform information about project development time, project cost, amount of firm's programmer.

Part 2: request respondent to give his opinion about important of each attribute that related to software productivity.

Level of important is Likert's five scale, Scale "1" represent less important while "5" is most important to the topic.

TABLE I
PART 2: QUESTIONNAIRE DETAIL

	Level of important				
Question	1(less)	2	3	4	5(most)
Firm constrain					
1.Software budget control (BUDT)					
2.Programmer salaries (SALS)					
3.Availability of computer resources (ACRE)					
4.Availability of computer programming tools (APTR)					
Time to complete					
5.Development time constrain (TTCN)					
6.Amount of fine if project delay (CODY)					
Software specification					
7.Software flexibility requirement (FLXY)					
8.Software efficiency requirement (EFFY)					
9.Software portability requirement (PORY)					
Program methodology					
10.Programming generator-CASE support (PGLE)					
11.Prototyping-tool using (PTTL)					
12.Software documentation management (COMG)					
13.Using of programming paradigm (PGGM)					
14.Defect detecting-software tesing (DFDE)					
Metrics used to measure productivity					
15.Source line of code (LOCE)					
16.Amount hours use in software development (HOWK)					
17.Software error –defect correction-amount error (BUFD)					
18.Amount of complete requirements of software (TACD)					

E. Descriptive Statistics

Two hundred and thirty finished software development projects were collected from fifteen software houses. All projects were detected whether they were followed research limitation or not before include them into dataset.

TABLE II
PERCENTAGE OF PROGRAMMER USING IN EACH SOFTWARE PROJECT

Amount of Programmer	Percentage
1-2	23
3-4	52
5-6	8
Outsource	17

1) ***Factor Analysis***

Purposed attributes were found out their relation by factor analysis method. Principle components analysis technique was used to extract a factor from group of attributes. Rotational method was "Variamax". KMO and Bartlett test of sphericity was chosen in order to detect adequacy of sampling of data. Result of factor analysis was presented in table III. The attribute "COMG", under factor "PGMY", has a factor loading value 0.26, less than 0.3. Hence this attribute was eliminated from factor "PGMY" member.

TABLE III
FACTOR AND ITS RELATED ATTRIBUTES

Factor / member	KMO	Cronbach	Loading	Remark
Firm constrain(FIC)	0.65	0.72		
BUDT			0.98	
SALS			0.85	
ACRE			0.78	
APTR			0.62	
Time to complete(TICT)	0.71	0.74		
TTCN			0.83	
CODY			0.72	
Software specification(SWSN)	0.63	0.81		
FLXY			0.69	
EFFY			0.61	
PORY			0.55	
Program methodology(PGMY)	0.61	0.71		
PGLE			0.94	
PTTL			0.91	
DFDE			0.77	
PGGM			-0.53	
COMG			0.26	Eliminated
Metrics-productivity(PDTY)	0.68	0.75		
LOCE			0.81	
HOWK			0.77	
BUFD			0.71	
TACD			0.66	

2) ***Structural Equation Model***

- **Proposed Structural Equation Model**

Causal model of this research was illustrated in figure 1. The model was composed of derived factors and their member. Relation between all factors, based on four research hypothesis. Since some attribute, COMG, was eliminated therefore it was disappeared from proposed structural equation model.

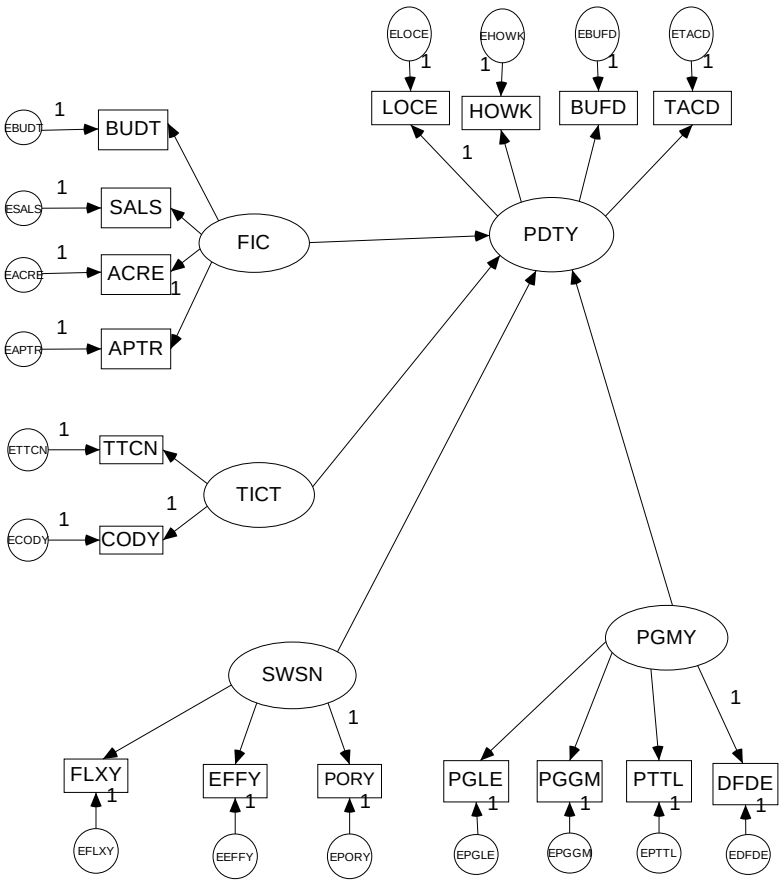

Figure 1. Proposed Structural Equation Model

- **Saturated Structural Equation Model**

Two hundred finished software projects were calculated with structural equation model method (SEM). Estimation technique is maximum likelihood. Calculation results were both presented in standardized and un standardized form. Important statistical model fit measurements were all passed statistical condition. Chi square test was an important one. The p value was significant value since its value is 0.04, should less than 0.05. Root mean square error of approximation (RMSEA) value is 0.037, should less than 0.05. Goodness of Fit Index (GFI) value is 0.94, should be greater than 0.90. Saturated structural equation model was illustrated in figure 2.

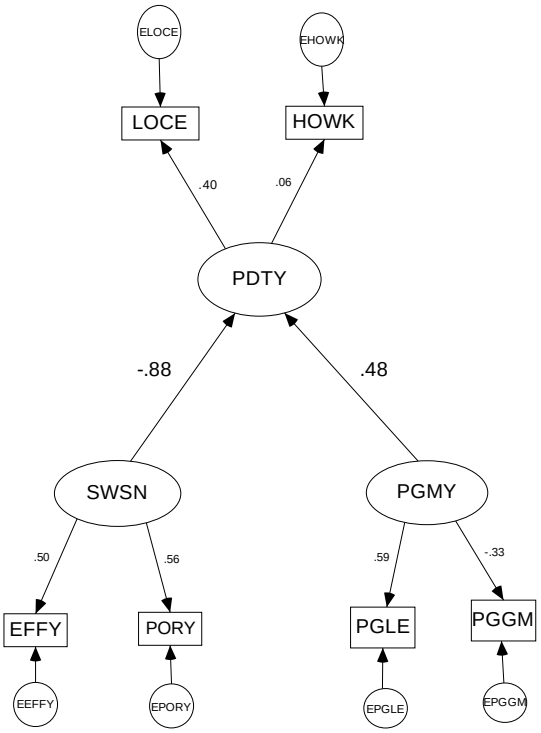

Figure 2. Saturated Structural Equation Model (Standardized Form)

Hypothesis 1, 2 were not significant direct effect to factor software project development productivity. There were only hypothesis #3 and #4 which pass model fitting significant test.

Hypothesis #3, software specification factor has a direct effect to software productivity in negative form (-0.88). Hence, high level of efficiency and/or portability of developing software should reduce productivity of software development, Therefore software project developer should took attention on these two non functional requirements.

Hypothesis #4, factor PGMY has a positive direct effect to software productivity (+0.48). Programming generator-CASE support and using of programming paradigm were two significant attributes that were composed to be factor PGMY. Attribute PGGM has a negative coefficient value (-0.33) but there will be reversed to positive element since its factor loading was negative value (-0.53), as shown in table III.

$$PDTY = 0.48PGMY - 0.88SWSN \quad (2)$$

$$\begin{aligned} PGMY &= 0.56PGLE - 0.33PGGM \\ &\quad + EPGLE + EPGGM \\ &= 0.56PGLE - 0.33PGGM + 0.473 + 0.475 \end{aligned} \quad (3)$$

$$\begin{aligned} SWSN &= 0.5EFFY + 0.56PORY \\ &\quad + EEFFY + EPORY \\ &= 0.5EFFY + 0.56PORY + 0.221 + 0.135 \end{aligned} \quad (4)$$

(Standard error of EPGLE = 0.473, EPGGM = 0.475, EEFFY = 0.221 and EPORY = 0.135).

3) ***Cross Validation***

PDTY factor was set to be as dependent variable while SWSN and PGMY were as independent variables. Test observations, thirty finished software projects, were calculated to find out their (actual) factor value. After that, prediction of PDTY was calculated from equation (2) then further compile for MRE, equation (1), with PDTY (actual).

Average of "Magnitude relative error: MRE", MMRE, were calculated said as "absolute amount of error". MMRE of this test was 0.3453 hence percentage of estimation error was 34.53 percent. Therefore, the accuracy of prediction was about 65.47 percent, 100-34.53.

IV. CONCLUSIONS

There were only four attributes or two factors that could direct effect on software productivity. Software specification factor was most important since it was a negative large coefficient. If software project factor (SWSN) is increase one unit then PDTY is treated to reduce its value about 0.88. One unit increasing of factor PGMY should positive effect about 0.48 to PDTY value increasing. Software developer should seriously aware of SWSN, important level increasing, while negotiation with customer about two non functional requirement (EEFFY and PORY) in term of reference, TOR.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Sampson Olkin, A.R. (2001). Multivariate Analysis: Overview. International Encyclopedia of the Social & Behavioral Sciences.

[2] Arbuckle, J.L. (1996). Easy Structural Equation Modeling and Confirmatory Factor Analysis (1st Ed.). Psychology Press.

[3] Devijver, P.A. & Kittler, J. (1982). Pattern Recognition: A Statistical Approach. London, GB: Prentice-Hall. ISBN: 0-13-654236-0.

[4] Jones, C. (1986). Programming Productivity. New York: McGraw-Hill Book Company.

[5] Tyler, H. (2019). 7 paces. Germany.