

**EFFORT ESTIMATION MODEL FOR FUNCTION POINT
MEASUREMENT**

By

KOH TIENG WEI

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfilment of the Requirement for the Degree of Master of Science**

January 2007

**Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment
of the requirement for the degree of Master of Science**

EFFORT ESTIMATION MODEL FOR FUNCTION POINT MEASUREMENT

By

KOH TIENG WEI

January 2007

Chairman : Associate Professor Hj. Mohd Hasan Selamat

Faculty : Computer Science and Information Technology

Software Cost Estimation is a practical process that is applied to find out basically the effort and development time requirements for a software product, which is going to be developed. The process starts with the planning phase activities and refined throughout the development. Various cost estimation models and methods are available to be used for software development process. The COCOMO, an algorithmic model is one example. Although it is hard to predict the exact size especially in terms of lines of code (LOC) of the project at the early stage, the COCOMO model takes LOC as an input to compute the project's effort.

Nowadays, software developers recognize the importance of the realistic estimates of effort to success management of software projects and having a realistic estimates at an early stage of a project life cycle which allow project managers and development organizations to manage resource effectively. This research work has generated an algorithmic effort estimation model for function points measurement. The function-point measurement metric, invented by Allan Albrecht of IBM in the middle of the 1970s was intended to help all the practitioners measure the size of a computerized business information system. Such sizes are needed as a component of a measurement of productivity in system development and maintenance activities and as a component of estimating the effort needed for such activities.

Generally, most of the algorithmic models were generated based on historical projects, and therefore, the same methodology has been applied in this research work. The total of 2265 historical business information systems collected by

International Software Benchmarking Standard Groups (ISBSG) has been used to generate and validate the model using statistical technique, known as linear regression analysis. We proposed that a project's functionality in the terms of function points has an approximately linear relationship to the final effort computation. The investigation had carried out on between the functionality and final effort before the model was generated.

Results show that there is an approximately linear relationship between the project's functionality and final effort that is required in the development. On the other hand, the limits with 95% level of confidence have been attached together with the model to ensure that the actual effort always falls within the predicted boundary. However, we suggest that further investigation should be carried on for the technical complexity factors in function point analysis in order to increase the accuracy of the forecasted project's effort.

**Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai
memenuhi keperluan untuk ijazah Master Sains**

MODEL PENGANGGARAN USAHA BAGI PENGUKURAN MATA FUNGSI

Oleh

KOH TIENG WEI

Januari 2007

Pengerusi : Profesor Madya Hj. Mohd Hasan Selamat

Fakulti : Sains Komputer dan Teknologi Maklumat

Penganggaran kos perisian adalah proses praktikal yang telah digunakan untuk mengetahui usaha dan masa pembangunan yang diperlukan dalam pembangunan sesuatu perisian. Proses ini mula daripada aktiviti fasa perancangan dan diperincikan lagi di sepanjang masa pembangunan. Terdapat pelbagai sumber bagi model penganggaran kos and kaedah digunakan dalam proses pembangunan perisian. Contohnya, COCOMO merupakan satu model algoritmik. Walaupun agak susah untuk meramalkan saiz sebenar terutamanya dalam bentuk *lines of code* (LOC), tetapi model COCOMO ini masih menggunakan LOC ini sebagai input untuk memperolehi usaha sesuatu projek.

Kini, pembangun-pembangun perisian berpendapat bahawa penganggaran realistik adalah penting untuk mengurus sesuatu projek perisian dengan berjaya dan dapat memperolehi anggaran yang realistik pada awal pembangunan supaya membolehkan pengurus projek dan organisasi pembangunan mengurus sumber

dengan lebih cekap. Penyelidikan ini telah menghasilkan satu model penganggaran usaha algorimetik untuk pengukuran mata fungsi. Matrik mata fungsi, merupakan hasil kajian daripada Allan Albrecht dari IBM dalam pertengahan 1970 yang cuba untuk membantu dalam mengukur saiz sistem maklumat niaga. Saiz ini merupakan komponen dalam mengukur produktiviti pembangunan sistem dan aktiviti pemuliharaan, ia juga sebagai komponen dalam menganggar usaha yang diperlukan bagi aktiviti-aktiviti tersebut.

Umumnya, kebanyakan model algoritmatik dibangunkan berdasarkan projek-projek yang lepas. Oleh yang demikian, kaedah yang sama digunakan dalam kerja penyelidikan ini. Sejumlah 2265 sistem maklumat niaga bersejarah yang dikumpulkan oleh *International Software Benchmarking Standard Groups* (ISBSG) telah digunakan untuk menghasil dan mengesahkan model ini dengan menggunakan teknik statistik yang dikenali sebagai analisis regresi linear. Kami telah mencadangkan bahawa fungsi sesuatu projek dalam bentuk mata fungsi mempunyai hubungan yang hampir linear dengan pengiraan usaha akhir. Penyelidikan lanjut ke atas hubungan antara fungsian dan usaha akhir telah dijalankan.

Keputusan menunjukkan bahawa terdapatnya hubungan yang hampir linear antara fungsi sesuatu projek dengan usaha akhir yang diperlukan dalam pembangunan. Di samping itu, tahap keyakinan had-had 95% telah digunakan bersama model ini untuk memastikan usaha sebenar akan sentiasa dalam

sempadan yang dianggarkan. Walau bagaimanapun, kami mencadangkan penyelidikan lanjut perlu dijalankan ke atas factor-faktor kekompleksan teknikal di dalam analisis mata fungsi supaya dapat meningkatkan lagi ketepatan bagi usaha projek yang diramalkan.

ACKNOWLEDGEMENTS

Getting this research done did not only involve my effort, but numbers of people have helped me, either directed or indirectly. Hence, the following people deserve special mention for their passion, assistance and kind effort in making this thesis reality.

Foremost is Associate Professor Hj. Mohd Hasan Selamat, as my committee supervisors, who had been guiding me in doing and completing the research work. These guidance, advice, and support had been invaluable to this research. Besides, my sincere thanks and deepest gratitude to my co-supervisor Associate Professor Dr. Abdul Azim Abdul Ghani, who had given his time in providing his technical support and suggestions. The same goes to Associate Professor Dr. Azmi Jaafar and Professor Dr. AHM Rahmatullah Imon who had been providing me the statistical knowledge and giving me an extra confidence for my research outcome.

Also a great thanks to Faculty of Computer Science and Information Technology, the university library and University Putra Malaysia that provided the working environment for performing this work.

The same goes for a dear friend of mine known to me as my “earthly angel” – her tireless effort to unearth the philosophy about life has also opened up my mind and hence giving me a balanced set of mind.

I am grateful to a friend who wants to remain anonymous. He is a down-to-earth and humble people. His willingness to share and his advice have given me confidence in completing this thesis. Without him, this research work will not be possible.

Last but not least, I also wish to thank to my group of friends who are always there for me, be it good or bad days. Their indirect contributions go well beyond what you will see in these printed pages. They are those whom I knew during my school, university days and working places. I may not mention their names but they are always in my heart. Thank you.

I certify that an Examination Committee met on 5th January, 2007 to conduct the final examination of Koh Tieng Wei on his Master of Science thesis entitled “A Function Point-Based Effort Estimation Model Using Regression Analysis Approach” in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulation 1981. The Committee recommends that the candidate be awarded the relevant degree.

Members of the Examination Committee are as follows:

Md. Nasir Sulaiman, PhD

Associate Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Chairman)

Azmi Jaafar, PhD

Associate Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Internal Examiner)

Rusli Abdullah, PhD

Senior Lecturer

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Internal Examiner)

Harihodin Selamat, PhD

Associate Professor

Faculty of Computer Science and Information Technology

Universiti Teknologi Malaysia

(External Examiner)

HASANAH MOHD.GHAZALI, PhD

Professor / Deputy Dean

School of Graduate Studies

Universiti Putra Malaysia

Date:

This thesis submitted to Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee are as follows:

Hj. Mohd Hasan Selamat, M.Phi.

Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Abdul Azim Abdul Ghani, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

AINI IDERIS, PhD
Professor / Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 10 MAY 2007

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

KOH TIENG WEI

Date: 5th January 2007

TABLE OF CONTENTS

	Page
DEDICATION	ii
ABSTRACT	iii
ABSTRAK	v
<i>ACKNOWLEDGEMENTS</i>	<i>vii</i>
APPROVAL	ix
DECLARATION	xi
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LISTS OF ABBREVIATIONS	xiv

CHAPTER

1 INTRODUCTION	
1.1 Research Background	16
1.2 Problem Statements	18
1.3 Research Objectives	20
1.4 Research Scope ad Limitation	21
1.5 Organisation of the Thesis	21
2 LITERATURE REVIEW	
2.1 Introduction	24
2.2 Measurement Methods	25
2.3 Software Size Measures	26
2.4 Measurement Scales	30
2.5 Software Effort Estimation Models	32
2.5.1 The COCOMO Model (Constructive Cost Model)	33
2.5.2 Putnam's SLIM Model	36
2.6 Software Functionality	37
2.7 Objective and Application of Software Measurement	38
2.8 Function Point Analysis	40
2.8.1 Objective of Function Point Analysis	41
2.8.2 Function Point Components	42
2.8.3 Function Point Complexity Weights	44
2.8.4 Function Point Complexity Factors	47
2.8.5 Function Point Counting Process	48
2.8.6 Function Point Application	51
2.9 Extended Function Point Analysis Techniques	52
2.9.1 Mark II Function Point	52
2.9.2 IFPUG Version 4.1	55
2.9.3 Full Function Points	56

2.9.4	COSMIC Full Function Points	57
2.10	Limitations of Function Point (IFPUG version)	58
2.10.1	Weakness in Adjustment Factors	59
2.11	Statistical Regression Analysis	61
2.11.1	Regression Techniques	61
2.11.2	Linear Regression	63
2.11.3	Minimizing Sum-of-Squares	64
2.11.4	Slope and Intercept	65
2.11.5	R^2 , Measure of Goodness-of-Fit of Linear Regression	66
2.12	General Application of Regression Analysis	68
2.13	Application of Regression in Software Engineering Cost Estimation	68
2.14	Statistical Inadequacies in Estimating	69
2.15	Reliability of Model Inputs	71
2.16	Summary	72
3	RESEARCH METHODOLOGY	
3.1	Introduction	74
3.2	Model-Building Process	74
3.2.1	Data Collection and Preparation	76
3.2.2	Reduction of Predictor Variables	77
3.2.3	Model Refinement and Selection	78
3.2.4	Model Validation	79
3.3	Data Analysis	80
3.3.1	Mean Magnitude of Relative Error (MMRE)	80
3.3.2	Ratio of Average Error	81
3.3.3	Correlation Coefficient	81
3.3.4	Error Limits	82
3.4	Data Transformation	82
3.5	Linear Regression Analysis	84
3.6	Research Tool	85
3.7	Summary	85
4	FUNCTION POINT ESTIMATION MODEL DEVELOPMENT	
4.1	Introduction	87
4.2	Preliminary Checks on Data Quality	87
4.3	Establishing Training Sample and Testing Sample	88
4.4	Preliminary Data Analysis	89
4.4.1	Frequency Distribution	89
4.4.2	Measure of Dispersion	91
4.4.2.1	Range	92
4.4.2.2	Sample Variance	92
4.4.2.3	Sample Standard Deviation	93
4.4.3	Linear Correlation	93
4.5	Data Transformation	94

4.6	Linear Regression Analysis	96
4.7	Constructing a Confidence Interval for β_1	99
4.8	Hypothesis Testing	101
	4.8.1 One-Tailed Hypothesis Test for the Slope of the Regression Line	101
4.9	Confidence Interval for Regression	104
	4.9.1 Constructing a Prediction Interval for $Y_x = x_0$	104
4.10	Summary	106
5	RESULTS AND DISCUSSIONS	
5.1	Introduction	108
5.2	Results	108
	5.2.1 Test Sets – Sample B	109
	5.2.2 Total Datasets – Sample C	109
5.3	Results Discussion	112
5.4	Summary	113
6	CONCLUSIONS AND FUTURE WORKS	
6.1	Conclusions	114
6.2	Capabilities of the Proposed Effort Estimation Model	115
6.3	Research Contribution	116
6.4	Suggestion for Further Work	118
REFERENCES/ BIBLIOGRAPHY		120
APPENDICES		125
BIODATA OF THE AUTHOR		136

LIST OF TABLES

Table	Page
2.1 General system characteristics	29
2.2 The basic COCOMO 81 model	33
2.3 Object point productivity	35
2.4 EI (Longstreet, 2002)	45
	45
2.5 EO and EQ (Longstreet, 2002)	45
2.6 ILF's and EIF's (Longstreet, 2002)	46
2.7 Function Point Complexity Weights (Longstreet, 2002)	46
2.8 The Internal Complexity Factors (Longstreet, 2002)	48
	48
4.1 Preliminary Statistical Studies for Effort and Function Size	
	90
4.2 Transformation Ladder and Guide	95
4.3 Values of the Constants	97
4.4 Analysis of Variance	98

LIST OF FIGURES

Figure		Page
2.1	Functionality recognized in function point analysis	
		29
2.2	Slope and Intercept	65
2.3	R² Value for Different Batch of Dataset	66
2.4	R² Computation	66
3.1	Strategy for Building a Regression Model	75
3.2a	Right Skewness form transformation	
		83
3.2b	Left Skewness form transformation	
		83
4.1a	Frequency Distribution of 315 Projects' Effort	90
4.1b	Frequency Distribution of 315 Projects' Function Size	91
4.2	Correlation between Function-Size and Effort	93
4.3	Line of Best Fit for Transformed Data	96
4.4	Normal P-P Plot of Regression Standardized Residual	102
4.5	One Tailed Critical Region and Critical Value	103
5.1	Scatter Plot of Sample B Using Proposed Model	109
5.2	Error Curve of Sample B	110
5.3	Scatter Plot of Sample C Using Proposed Model	111
5.4	Error Curve of Sample C	111

LIST OF ABBREVIATIONS

COBOL	Common Business Oriented Language
COCOMO	Constructive Cost Model
COSMIC	Common Software Measurement International Consortium
DET	Data Element Types
EI	External Inputs
ELOC	Effective Lines Of Code
EQ	External Inquiry
EO	External Output
EIF	External Interface File
FFP	Full Function Point
FTR	File Type Reference
FP	Function Point
FPA	Function Point Analysis
FTE	Full Time Equivalent
GSC	General System Characteristics
ILF	Internal Logical File
IFPUG	International Function Point Groups
ISBSG	International Software Benchmarking Standards Group
ISO	International Organization for Standardization
KLOC	Thousands of Lines Of Code
LOC	Line Of Code

MIS	Management Information System
MMRE	Mean Magnitude of Relative Error
NCLOC	Non-Commented source Line Of Code
RET	Record Elements Type
R²	Correlation Coefficient
SELAM	Software Engineering Laboratory in Applied Metrics
SLIM	Software Life Cycle management
TCF	Technical Complexity Factors
TUFP	Total Unadjusted Function Point
UFP	Unadjusted Function Point
UFPC	Unadjusted Function Point Count
VAF	Value Adjustment Factor