

**VARIATION IN PROJECT
PARAMETERS AS A MEASURE OF
IMPROVEMENT IN SOFTWARE
PROCESS CONTROL**

This thesis is presented for the degree of

Doctor of Philosophy

**in the
School of Electrical, Electronic & Computer Engineering,
The University of Western Australia**

By

Terence Leslie Woodings BSc, DipComp, FQSA, FACS

September 2005, revised June 2006

Declaration

I declare that this thesis is my own account of my research and contains as its main content, work which has not previously been submitted for a degree at any tertiary education institution.

Terence L Woodings

Perth, Western Australia. 30 September 2005

Some of the material in this thesis has been previously published in six papers presented at conferences. All had T L Woodings as the sole author. They were:

- "Meta-metrics for the Accuracy of Software Project Estimation", *Proceedings WA Workshop on Information Systems Research (WAWISR)*, Murdoch University, Perth, November 1999.
- "Exploring the Aetiology of Software Project Development Diagnostics", *Proceedings 12th ESCOM Conference on Software Control and Metrics*, London, April 2001.
- "Upon the Variance of Estimation", *Proceedings 11th University of WA Computer Science Research Conference*, Mandurah, Western Australia, September 2001.
- "Limiting Factors of Estimation Accuracy and Project Risk", *Papers from COCOMO-17 International Forum*, University of Southern California, Los Angeles, October 2002.
- "The Mechanical Improvement of Estimates during Projects", *Proceedings of the Australian Conference on Software Metrics (ACOSM)*, Melbourne, November 2002.
- "Software Project Estimation - How much Improvement is Possible?", *Proceedings 14th Annual UK Software Metrics Conference*, Wolverhampton, September 2003.

© Copyright, Terence Woodings, 2005.

Abstract

The primary tool for software process control is the project plan, with divergence from the schedule usually being the first indication that there are difficulties. Thus the estimation of the schedule, particularly the effort parameter, is a central element of software engineering management. Regrettably, estimation methods are poorly used within the software industry and accuracy is lacking when compared with other engineering disciplines. There are many reasons for this. However, the need to predict project effort remains, particularly in situations of tendering for contracts.

The broad objective of this research is the improvement of project control by means of better estimation. It focuses on the development of a practical approach by which software engineers may systematically improve their estimation processes in order to obtain a more effective framework for management. To improve project planning, managers need to have an understanding of the theory and practical techniques of estimation together with an appreciation of the sources of error and bias. The error in the prediction of a project parameter is investigated as the result of the variation in two distinct (estimation and actual development) processes. Improvement depends upon the understanding, control and then reduction of that variation.

A strategy for the systematic identification of the sources of greatest variation is developed - so that it may be reduced by appropriate software engineering practices. The key to the success of the approach is the statistical partitioning of the Mean Square Error (of the estimate) in order to identify the weakest area of project control. The concept is proven with a set of student projects, where the estimation error is significantly reduced. The conditions for its transfer to industry are discussed and a systematic reduction in error is demonstrated on five sets of commercial project data.

The thesis concludes with a discussion of the linking of the approach to current estimation methods. It should also have implications for the statistical process control of other projects involving small sample sizes and multiple correlated parameters.

Contents

Declaration	(ii)
Abstract	(iii)
Table of Contents	(iv)
List of Figures	(viii)
List of Tables	(ix)
List of Key Assertions and Concepts	(x)
Acknowledgments	(xi)
Chapter 1: An Introduction to the Problem and Structure of the Thesis	1
1.1 The Current Need for Improvement in Project Control	1
1.2 The Nature of Software Project Estimation	3
1.3 Obstacles to Effective Project Control	7
1.4 The Hypotheses	17
1.5 Contributions of this Thesis to Current Knowledge	18
1.6 The Structure of the Thesis	20
Chapter 2: A Review of Software Estimation Improvement	23
2.1 Metrics for the Development of Software Systems	24
2.2 Complexity, Requirements and Assumptions of Project Estimation	30
2.3 A Taxonomy of Estimation Methods	37
2.4 Comparative Studies of Software Estimation	48
2.5 Industrial Practices in Software Estimation	49
2.6 Major themes for Improving System Estimation	53
2.6.1 The Use of Project Databases for evaluation and calibration	53
2.6.2 Improving measures of System Size	54
2.6.3 Improvement by means of new Estimation Models	60
2.6.4 Improvement by means of new Calibration Methods	64
2.6.5 Tailoring of Estimation Methods	66
2.7 Statistical Process Control and the Theory of Constraints	67
2.8 Conclusions on Estimation Improvement Approaches	68

Chapter 3: Some Concerns regarding Variation in Project Planning	71
3.1 Human Factors in Estimation and Control	73
3.1.1 Human Variation in the Development Process	74
3.1.2 Human Variation in the Estimation Process	75
3.2 Error and Bias in the Estimation of Projects	76
3.2.1 The Natural Optimism of Estimators	76
3.2.2 Concern as to how the Estimates will be used	80
3.2.3 Self-fulfilling Prophecies and Past-Project Statistics	82
3.3 The Multi-dimensional Aspects of Project Success	87
3.4 Difficulties in partitioning the estimation error at a milestone	90
3.5 The Adequacy of Proxies for Measuring System Size	93
3.6 The Issue of Requirements Creep during a Project	97
3.7 The use of Negotiation to Decide Project End-points	99
3.8 Models of the Estimation Activity	100
3.9 Conclusions on Controlling Project Parameters	105
 Chapter 4: The Theory of Measuring Process Variation and Improvement	 107
4.1 Assumptions and Constraints	108
4.1.1 General Assumptions for Commercial Projects	109
4.1.2 Special Assumptions for student Projects	111
4.1.3 The True Size of a Software System	112
4.2 Meta-metrics for Estimation Accuracy	116
4.3 Defining a Strategy for Improving Software Process Control	123
4.4 Partitioning the Variation in Estimation Error	127
4.4.1 A First Step towards Partitioning the MSE	127
4.4.2 A General Theory Incorporating the Size Effect	129
4.4.3 The Case when Estimation is a Mechanical Function of Size	134
4.4.4 The Case for Projects of Practically Constant Size	136
4.4.5 Interpreting the Analyses	138
4.5 Calibration of Estimation Models	142
4.6 GEM – A General Estimation Method and Research Framework	145
4.7 Partitioning the Variation in the Development Process	152
4.8 Conclusions on Theories of Partitioning & Reducing Error	153

Chapter 5: An Analysis of the Improvement in Estimate Variance	155
5.1 The Source of the Data	155
5.2 Constraints and Assumptions in the Use of the Student Data	157
5.3 A Linear Predictor of Student Effort	158
5.4 Results of Partitioning the Variation – A Pilot Study	163
5.5 Partitioning the Variation – 2001 to 2004	164
5.6 Calibrating the Estimates	168
5.7 Conclusions on the Value of the Analysis of Estimation Variance	173
Chapter 6: The Application of the Techniques to Industry	175
6.1 The Sources of the Data	175
6.2 The Effect of Size Variation	177
6.3 A Comparison of Student Projects with Industrial Data	178
6.4 Showing the Industrial Applicability of the Approach	185
6.5 Process Improvement by Combining Estimation Methods	189
6.6 Process Improvement by Revision of the Estimation Model	191
6.7 Process Improvement by Means of Training	194
6.8 Conclusions on Commercial Estimation Process Improvement	196
Chapter 7: Conclusions and Opportunities for Further Research	199
7.1 Findings within a Current Industrial Context	199
7.2 Avenues for Further Research	205
7.2.1 Improved Metrics for Software System Size	205
7.2.2 The Need for a Theoretical Distribution of Estimation Error	205
7.2.3 Alternative Meta-metrics of Estimation Accuracy	206
7.2.4 An Experimental Intervention in Project Control Methods	207
7.2.5 Links to Current Estimation Tools	207
7.2.6 Application to other Engineering Disciplines	207
7.3 Conclusion	208

References	211
Appendix A: Material used in the Student Trials	231
Details of Project Requirements	232
Form used for Data Collection	261
Appendix B: Data Listings and Sources	263

List of Figures

1.1	Hoops in the sky – adjusting course for a successful conclusion	9
1.2	The reduction in uncertainty during a project	11
1.3	The project S curve and corrections to the plan	12
1.4	Variation in effort for a change over a project’s duration	15
2.1	The inter-relationships of various subjects of measurement	27
3.1	The influence of risks on project planning	72
3.2	Self-sustaining bias in repositories of project statistics	85
3.3	A project’s four dimensions as a hydraulic system	89
3.4	Orthogonal partitioning of performance at a milestone	91
3.5	Changes in the measurement of system size	96
3.6	A model indicating the trade-off of estimation factors	102
3.7	The revised iterative model of estimation practices	103
4.1	Sources of error emanating from the functional size of projects	113
4.2	$Pred(q)$ as a proportion of projects with error less than q	117
4.3	EQF as the ratio of error area to (<i>actual x time</i>) for three intervals	121
4.4	Improvement in archery as a two-stage process	123
4.5	Variation versus bias as contributors to error	125
4.6	Sources of variation in software projects	125
4.7	Estimation as a two-stage process	134
4.8	GEM and project estimation as a cascade of proxies	146
5.1	Distribution of estimated and actual effort over the duration of a project	159
5.2	<i>Effort</i> as a linear function of <i>Estimate/Time-available</i>	161
5.3	<i>Effort</i> as a function of <i>Estimate/Time-available</i> over five years of data	162
6.1	<i>Effort</i> and <i>size</i> for a subset of the Boehm 1981 data	179
6.2	Components of variation for estimation in the two Prechelt groups	196

List of Tables

2.1	Two examples of the classification of estimation methods	47
4.1	Application of accuracy metric <i>Pred</i> to a simulated project distribution	118
4.2	PAMESE – equations for partitioning the <i>MSE</i> of projects of varying size	133
4.3	The set of equations for partitioning the <i>MSE</i> of projects of constant size	137
4.4	Simulations of the partitioning of variation	139
4.5	Simulations of the use of the predictor set	141
5.1	Summary of statistics from the 2000 course data	164
5.2	Partitioning the error of initial estimates of effort (2001 to 2004)	165
5.3	Partitioning the error of estimates of effort after three weeks (2001-2004)	167
5.4	Partitioning the error of 2002 initial estimates of effort after calibration	168
5.5	Partitioning the error of 2002 3-week estimates of effort after calibration	169
5.6	Partitioning the error of 2003 initial estimates of effort after calibration	170
5.7	Partitioning the error of 2003 3-week estimates of effort after calibration	171
5.8	Reduction in estimation error after three weeks and calibration	172
6.1	Comparison of student projects with commercial data	181
6.2	Partitioning of variation for student and commercial projects	182
6.3	Basic statistics of the Kemerer analysis	186
6.4	Partitioning of variation for the four methods in the Kemerer study	187
6.5	Extension of Kemerer analysis with improvement using calibration	188
6.6	Partitioning of variation for 3 estimation models using the Maxwell data	192
6.7	Susceptibility of RRMSE to the effects of atypical projects	193
6.8	Basic statistics on the estimations for the two Prechelt groups	195
6.9	Partitioning of variation for the Prechelt data-set	195

List of Key Assertions and Concepts

1.3	Successful projects result from understanding allowable variation	10
1.3	Variation is the result of project control plus environmental change	10
1.4	The aim of the research – estimation and project control	17
1.4	The three Hypotheses for this thesis	18
2.3	Estimation methods explained by taxonomy with six dimensions	47
2.3	Research on transfer of knowledge from past to future projects	47
3.0	Meta-risks of project planning as distinct from project development	71
3.2	Project data repositories as a part of a self-perpetuating bias	84
3.3	An estimate must cover all project success factors	89
3.5	Currently no definition (and no metric) for the true size of a system	96
4.2	Meta-metrics should reward frequent and early revision of estimates	122
4.3	An ability to partition error as a pre-requisite to effective project control	126
4.3	Process improvement as a systematic search for areas of least certainty	127
4.4	PAMESE equations partition variation into estimation vs development	132
4.6	Expectation at each milestone: estimate increases, uncertainty decreases	147
4.6	GEM as a self-improving estimation tool with no inherent variation	149
4.6	The total project variation may be partitioned to determine weak phases	151
5.3	Effective effort as a function of estimates and time to deadline	160
5.5	Variation in the software process is a lower bound on estimation error	166
6.3	Estimation improvement as a process of accommodating extra factors	184
6.3	Relevant factors and covariance of estimation & development processes	185
6.8	Partitioning variation in project repositories for key factors	197
7.1	The three Hypotheses for assessment	200

Acknowledgments

Numerous friends and colleagues have helped and influenced me in this research. In particular, I would like to express my special appreciation to:

Associate Professor **Gary Bundell** – my principal supervisor, School of Electrical, Electronic & Computer Engineering – who, in the most patient manner possible, helped me to discover that I did not know half as much about the subject as I thought I did. Dr **Rachel Cardell-Oliver** – my co-supervisor, School of Computer Science & Software Engineering – who established a high standard and provided many valuable suggestions on the thesis in both form and content. Professor **Robyn Owens**, Dr **Nick Spadaccini**, Dr **Lyndon While** and Mr **Peter Jones** – who set challenging expectations and gave support at critical moments. Associate Professor **Richard Thomas** – who urged me, persistently and correctly, to dig deeper. Mr **Kevin Vinsen** and the staff of **ADI** – who freely provided details on the latest practices. Dr **Robin Milne** – who commented on some of the early statistics. Associate Professor **Jim Everett** – who provided excellent advice when I started the research. Mr **Colin Smith**, my first software manager – who introduced me to the capabilities of the I.T. industry. Dr **David Goodall**, CSIRO – who exposed me to the addictive thrill of research. Mr **Norman Stenhouse** – who showed me that the statistics of people are always more interesting than those of products. Mr **Alex Reid** – who taught me computer science and good management by always encouraging his staff to experiment with new methods. Professor **Jeff Rohl** – who encouraged me to enquire into software production as a science rather than an occupation. The professional team of programmers of the **WARCC Applications Group** – who completed an amazing range of projects and provided much feedback on my management style. Mr **Roger Dickinson**, Dr **Michael Wheatley** and the staff of **DUIT** – who daily grapple with these software development problems in inspiring and innovative ways. The students of **RDCS 423** at UWA – who, in tackling the lift controller project, provided the data and the insights on software teams in action. **Mum, Dad, Jean, Barbara** and **Ross** – who never questioned the obsession but simply gave unqualified support. And **Simon** – whose own research inspired me to try likewise.