



Theorizing Outliers – Explaining Variation in IT Project Performance

SUBMITTED FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY IN MANAGEMENT STUDIES

ALEXANDER BUDZIER, GREEN-TEMPLETON COLLEGE

Hillary Term 2014
Saïd Business School, University of Oxford

Abstract

IT projects are temporary organizations of strategic importance. Companies invest large amount of money, time, and resources into business-embedded IT projects in order to change and gain a competitive advantage. Extreme cases of failures were previously only analyzed as case studies, e.g., Denver Airport, London Stock Exchange Taurus, London Ambulance Service.

The research poses an important question: What is the risk of these outliers, that is markedly deviant observations of IT project performance? What causes outliers in IT project performance?

Only very few studies problematized the frequency of outliers directly.

Reported numbers range from 33% to as low as 0.2%. The variation has been explained through biases in planning processes of organizations and as artefact of data collection. An alternative explanation is that the true nature of IT projects contains more variation than commonly assumed.

A rich body of organizational, project management, and IT project management literature offers antecedents of outliers. The extant literature falls broadly into three schools of thought: (1) system-centric, (2) event-centric, and (3) process-centric theories of why outliers occurred.

System-centric explanations focus on the question of system design, based on theories of normal accidents and high reliability organizations. *Event-centric explanations* focus on how organizations respond to rare events that impact the organization, based on theories of crisis management, management of

organizational turbulence, and strategic surprises. *Process-centric explanations* focus on the role of managing uncertainty and risk over time, based on theories of man-made disasters, escalation of commitment to a failing course of action, and the normalization of deviance.

The study is based on the archival research of 4,307 IT projects from 190 organizations.

The findings show that the tail of the cost, schedule, and effort performance distributions is best fitted by a power law, with overwhelming goodness of fit. Moreover, the findings show that system-centric explanations and process-centric theories offer explanations for the thickness of the tail and the odds of an outlier occurring. In particular five variables were associated with outliers: estimated cost and duration, perceived uniqueness of the project, the qualification and motivation of the project team, and the effectiveness of monitoring and controlling.

The results show that outliers are not chance events; they follow patterns that are describable. The study showed how design factors, that are often conceptualized as system complexities, and execution factors, that are often conceptualized as the effectiveness of project processes, explain project outliers. Lastly, the thesis draws implications for research and practice.

Table of Contents

1	Introduction	11
1.1	Setting the stage	12
1.2	The Three Schools of Thoughts	15
1.3	The Thesis' Key Findings.....	17
1.4	Research Aim and Contribution to Knowledge.....	19
1.5	Structure and Summary of the Thesis	21
2	Literature Review	25
2.1	Definition of IT Projects	26
2.2	Definition of Outlier	28
2.3	How Frequent are Outliers in IT Project Performance?	30
2.3.1	Surveys of Performance of IT Projects.....	30
2.3.2	Explaining the Variation in the Data	37
2.4	Why Do Outliers Occur in IT Project Performance?.....	42
2.4.1	System-Centric School of Thought.....	45
2.4.2	Event-Centric View	62
2.4.3	Process-Centric School of Thought	67
2.4.4	Summary of the Literature Review	90
3	Methodology	97
3.1	The Survey Instrument	99
3.2	Data Collection & Preparation.....	100
3.3	Statistical methods.....	110
3.3.1	Rare Events Multilevel Logistic Regression.....	111
3.3.2	Fitting and Testing the Thickness of the Tails	114
3.4	The Sample	123
3.4.1	Sample Size	123
3.4.2	Frequency of Project and System Types.....	123
3.4.3	Size and Duration of IT Projects in the Sample	124
3.4.4	Heterogeneity in the Sample.....	126
3.4.5	Summary of the Sample Statistics	133
3.5	Summary of the Methods Chapter.....	134
4	Findings: How Frequent are Outliers in IT Project Performance?	135
4.1	Outline of the Chapter	135

4.2	Hypothesis 1: Under the Assumption of Thin Tails, Markedly Deviant Observations of IT Project Performance did not Occur	137
4.2.1	Cost Performance	137
4.2.2	Schedule Performance	139
4.2.3	Effort Performance	141
4.2.4	Benefits Performance	142
4.2.5	Summary of Findings of Hypothesis 1.....	143
4.3	Hypothesis 2: Performance of IT projects is not Normally Distributed	146
4.4	Hypothesis 3: IT Project Performance Data Fit Power Law Tails, an Extreme Value Distribution.....	147
4.4.1	Visual Fit to Common Extreme Value Distributions	147
4.4.2	Power Law Fit of the Tails	152
4.4.3	Goodness of Fit Test.....	162
4.4.4	Higher Order Moments.....	165
4.4.5	Alternative Explanations of the Power Law Tails.....	168
4.5	Summary of Findings	176
5	Findings: Why did Outliers Occur?	179
5.1	Determinants of Cost Outliers in IT Projects.....	179
5.1.1	Determinants of Common Cost Overruns in IT Projects	180
5.1.2	Determinants of the Thickness of the Upper Power Law Tail of the IT Project Cost Performance Distribution.....	187
5.1.3	Explanatory Multilevel Regression Models	195
5.2	Determinants of Schedule Performance in IT Projects	209
5.2.1	Determinants of Common Schedule Overruns in IT Projects.....	209
5.2.2	Determinants of the Thickness of the Upper Power Law Tail of the IT Project Schedule Performance	215
5.2.3	Exploratory Multilevel Regression Models.....	218
5.3	Summary of Findings Why Outliers Occurred.....	226
6	Discussion and Implications of Findings	234
6.1	How Frequent are Outliers in the Performance of IT Projects?. 234	
6.1.1	Implications of Power Law Tails for Researching Project Performance	234

6.1.2	Implications of Power Law Tails for IT Project Forecasting and Risk Modelling Practice	239
6.2	Theoretical Implications: Why did Outliers Occur in IT Project Performance?	252
6.2.1	Implications for System-Centric Theories.....	255
6.2.2	Implications for Process-Centric and Event-Centric Theories	259
6.3	Practical Implications: Principles of Project Design and Project Execution.....	264
6.4	Summary	267
7	Summary	269
7.1	How Frequent are Outliers?	269
7.2	Why do outliers occur?	275
7.3	Ideas for future research	285
8	Appendices	288
8.1	Event-Centric Explanation of Why Outliers Occur	289
8.1.1	Cognitive Perspective	289
8.1.2	Sense Making Perspective	291
8.1.3	Framing Perspective.....	293
8.1.4	Organizational Learning Perspective	299
8.2	Project Success/Failure versus Project Performance.....	304
8.3	First Draft of Variables to Include in Study	310
8.3.1	Technical data (for project as built).....	310
8.3.2	Schedule	312
8.3.3	Budget	313
8.3.4	Causes of inaccurate cost estimates	315
8.3.5	Relative project size	315
8.3.6	Estimated project size.....	315
8.3.7	Users	316
8.3.8	Actual Size Break-Down in Persondays.....	316
8.3.9	Value and benefits estimates.....	317
8.3.10	Complexity	317
8.3.11	Critical Success Factors for Senior Mgmt	322
8.3.12	Uncertainty/Dynamics of the project.....	324
8.3.13	Qualitative characteristics	324

8.4	Technical definitions	325
8.4.1	Project Types	325
8.4.2	Planning constraints	326
8.4.3	Contract Types.....	326
8.4.4	Project Lifecycle	326
8.4.5	Project Performance	327
8.4.6	Actual and Estimated Cost, Effort, Schedule, and Benefits .	327
8.5	Anchoring of Explanatory Variables and Anchoring of Variables	329
8.6	Maximum Likelihood Estimation of a Power Law	333
8.7	Testing Power Laws in Disaggregating Observations.....	343
8.8	Correlation Table of Explanatory Variables	345
8.9	Principal Component Analysis of the Explanatory Variables ...	346
8.10	Determinants of Effort Outliers IT Projects	350
8.10.1	Determinants of Effort Overruns in IT Projects	350
8.10.2	Determinants of the Thickness of the Upper Power Law Tail.....	352
8.10.3	Exploratory Multilevel Regression Models.....	354
8.11	Participant Information	357
8.12	Implications of Power Law Tails for (IT Project) Management Theory.....	361
8.12.1	System-centric explanations	361
8.12.2	Event-centric explanations.....	362
8.12.3	Process-centric explanations.....	364
9	Bibliography.....	366

List of Figures

Figure 1 Boxplot Project Performance (actual/estimate, LN Scale).....	145
Figure 2 Boxplot Project Performance (Linear Scale, truncated at $0 \leq$ actual/estimate ≤ 5).....	145
Figure 3 QQ-plots Project Cost Performance	149
Figure 4 QQ-plots Project Schedule Performance	150
Figure 5 QQ-plots Project Effort Performance	151
Figure 6 Binned LN-LN Histogram Project Cost Performance	153
Figure 7 Binned LN-LN Histogram Schedule Performance	154
Figure 8 Binned LN-LN Histograms Project Effort Performance	155
Figure 9 Power Law Fit Cost Performance (Log-Log Plot of Inverse Cumulative Distribution Function (CDF) for cost overruns $x \geq x_0 = 1.087$, $\alpha = 2.5$).....	157
Figure 10 Power Law Fit Schedule Performance (Log-Log Plot of Inverse Cumulative Distribution Function (CDF) for schedule overruns $x \geq x_0 = 1.091$, α $= 3.1$)	158
Figure 11 Power Law Fit Effort Performance (Log-Log Plot of Inverse Cumulative Distribution Function (CDF) for effort overruns $x \geq x_0 = 1.562$, $\alpha=2.7$).....	158
Figure 12 Uncertainty of the Fit of the Cost Performance Distribution	160
Figure 13 Uncertainty of α and x_0 of Cost Performance Distribution	160
Figure 14 Uncertainty of the Fit of the Schedule Performance Distribution	161
Figure 15 Uncertainty of α and x_0 of Schedule Performance Distribution	161
Figure 16 Uncertainty of the Fit of the Effort Performance Distribution	162
Figure 17 Uncertainty of α and x_0 of Effort Performance Distribution.....	162
Figure 18 Goodness of Fit Test Project Cost Performance.....	163
Figure 19 Goodness of Fit Test Project Schedule Performance.....	164
Figure 20 Goodness of Fit Test Project Effort Performance	164
Figure 21 The sample mean and standard deviation for a heavy-tailed power law distribution with $\alpha=2.5$ for different sample sizes, for each sample size variance and mean are the estimates from the same synthetic data.....	167
Figure 22 The sample mean and standard deviation for a fat-tailed power law distribution with $\alpha=3.1$ for different sample sizes, for each sample size variance and mean are the estimates from the same synthetic data.....	167
Figure 23 Power Law QQ-plots for Different Project Types.....	172
Figure 24 QQ-plots Before and After Organizational Change	174
Figure 25 Correlation between estimated project size and cost overruns (blue line indicates the median at which the two groups are split in the boxplots)...	185
Figure 26 Correlation between actual project size and cost overruns (blue line indicates the median at which the two groups are split in the boxplots)	185
Figure 27 Cost Overruns in Relation to 'If the True Cost of the Project were Known...'	185
Figure 28 Prediction Cost Performance Outliers	203
Figure 29 Predicted Average Probability of a Project Being a Cost Outlier	205
Figure 30 Predicted Probability of the Occurrence of Outliers for Given Project Cost and Schedule Estimates	206

Figure 31 Rare Events Bias in Logistic Regression Explaining Cost Outliers ...	207
Figure 32 Correlation between Actual Duration and Schedule Overruns	211
Figure 33 Schedule Overruns and Realistic and Reliable Schedules.....	212
Figure 34 Schedule Overruns and Qualification and Motivation of the Team.	213
Figure 35 Schedule Overrun and Projects that Measured Benefits	214
Figure 36 Predictions of Average Probability of Schedule Outlier.....	224
Figure 37 Rare Events Bias in Logistic Regression Explaining Cost Outliers ...	225
Figure 38 Comparison of Thickness of Power Law Tails for 5 Selected Variables.....	231
Figure 39 Main Effects of Logistic Regression to Explain Occurrence of Outliers in IT Projects	232
Figure 40 Rare Events Bias in Multilevel Logistic Regression	233
Figure 41 Convergence of the Mean in Cost Overrun Samples	237
Figure 42 Observed Average Maximum Cost Overrun ($x_{\max}$) Depending on the Sample Size (n)	238
Figure 43 Inverse Cumulative Distribution of Cost Overruns $\Pr(X > x)$ on Linear-Linear and Log-Log Plot	240
Figure 44 Comparison of the Tails of Commonly Used Distributions in Risk Modelling.....	242
Figure 45 Comparison of Thickness of Power Law Tails for 5 Selected Variables.....	280
Figure 46 Main Effects of Logistic Regression to Explain Average Probability of Outliers in IT Projects	281
Figure 47 Restated Framework of Explanations of Outliers in IT Project Performance	284
Figure 48 Screeplot of the Principal Component Analysis	348
Figure 49 Factor Loadings for the First Five Dimensions Extracted by the Principal Component Analysis	349
Figure 50 Prediction of Average Probability of Effort Outliers Occurring	356

List of Tables

Table 1 Overview of Literature of Why Outliers Occur in IT Project Performance	92
Table 2 Final Selection of Variables to Explain IT Project Outliers	96
Table 3 Selected survey items and anchoring (full table in Appendix 8.5).....	99
Table 4 Overview of the Two Analyses Used to Identify Why Outliers Occurred in IT Project Performance	111
Table 5 Project Sizes in Previous Studies	125
Table 6 Intraclass Correlation Coefficients of Potential Hierarchical Dependencies in IT Project Cost Performance	128
Table 7 Summary of Project Performance Distribution	144
Table 8 Anderson-Darling Test for Normality	146
Table 9 Maximum Likelihood Power Law Fit.....	157
Table 10 Higher Order Moments of the Cost, Schedule, and Effort Performance Tails.....	165
Table 11 Goodness of Fit Test Power Law Tails in the Organizational Change Example	175
Table 12 Non-Parametric Wilcoxon Test for Location Shift in Cost Overrun Distribution	181
Table 13 Test for Thickness of the Upper Tail of the Cost Overrun Distribution	189
Table 14 Correlations and Sample Sizes of Fixed Effects in Model to Explain Common Cost Overruns	196
Table 15 Fixed Effects Model to Explain Common Cost Overruns (LN transformed).....	196
Table 16 Fixed Effects Multilevel Logistic Regression for Cost Outliers	199
Table 17 Correlation between Fixed Effects	200
Table 18 Logistic Regression for Cost Outliers	201
Table 19 Multilevel Logistic Regression for Cost Outliers	203
Table 20 Non-Parametric Wilcoxon Test for Location Shift in Schedule Overrun Distribution	209
Table 21 Test for Thickness of the Upper Tail of the Schedule Performance Distribution	215
Table 22 Fixed Effects of the Multilevel Model to Explain Common Schedule Overruns (LN Transformed)	220
Table 23 Correlation between Fixed Effects	220
Table 24 Parsimonious Explanatory Fixed Effects Model	221
Table 25 Multilevel Fixed Effects of Model for Schedule Outliers.....	222
Table 26 Summary of Findings.....	228
Table 27 Power Law Exponents and Higher Order Moments of the Key Performance Dimensions of IT Projects	236
Table 28 Power Law Tails of Key IT Project Risks.....	243
Table 29 Power Law Tails of Estimated and Actual Cost of IT Projects	245

Table 30 Summary of Results of Multilevel Logistic Regression Models: Why do Outliers Occur?.....	252
Table 31 Project and system types (Stahlknecht and Hasenkamp, 1997)	325
Table 32 Simplified Project Phases	326
Table 33 Anchoring of Variables in Survey Instrument.....	329
Table 34 Goodness of Fit Test for Different Subsamples of Cost Performance	343
Table 35 Goodness of Fit Test for Different Subsamples of Schedule Performance	344
Table 36 Determining Dimensions for Imputation of Missing Values	346
Table 37 Eigenvalues and Percentage of Variance Extracted by Principal Component Analysis	346
Table 38 Non-Parametric Wilcoxon Test for Location Shift in Effort Overrun Distribution.....	351
Table 39 Test for Thickness of the Upper Tail of the Effort Performance Distribution.....	353
Table 40 Explanatory Fixed Effects Model for LN(Effort Overruns)	355
Table 41 Parsimonious Explanatory Fixed Effects Model for the Odds of Effort Outliers Occurring	355

1 Introduction

Kmart was the retailer in the United States. It opened its first discount store in 1962, only months ahead of Wal-Mart. In 1994, nearly at its centennial, Kmart struggled. The stores were perceived as being dated and the company announced the closure of 110 out of its 2,800 stores. Moreover, Kmart had failed to invest in its IT systems. Kmart lacked state-of-the-art supply chain planning and execution systems, which led to low availability in stores, which in turn eroded sales. It was slowly being edged out by competitors with superior IT, such as Target and Wal-Mart.

To turn around, Kmart experimented with novel store concepts and re-branding, it also invested heavily in IT. In 2000 the company launched a USD 1.4bn IT modernization effort, despite posting a loss of USD 244m. Kmart, which still ran mainframes, decided to build a monolithic solution by heavily customizing TradeMatrix, a commercial off-the-shelf retail management solution. However the heavy customization made maintenance and operations of the system prohibitively expensive. Mid-2001, when the project finished, Kmart was forced to write-off USD 310m for two distribution centers and parts of its new IT system. In 2001, Kmart posted USD 1.3bn loss on USD 6.7bn of sales.

To rid itself of the high operations cost Kmart started a USD 600m project to implement Manhattan Associates, a standard supply chain management

system. By October 2001, the second project had run into difficulties and Kmart wrote off an additional USD 130m on its IT projects.

In early 2002 an accounting scandal engulfed Kmart and its IT projects continued to struggle. Kmart filed for chapter 11 on January 22, 2002. The bankruptcy judge not only put the struggling project on but also barred Kmart from making any new software acquisitions. In 2004, Kmart merged with Sears and immediately switched to Sear's IT systems. Kmart finally emerged from bankruptcy in May 2006 and by 2007 it had closed nearly 900 stores, laid off more than 67,000 employees, and sold its foreign subsidiaries (Isidore 2004, ICMR 2003, Charette 2005, Konicki 2002).

1.1 Setting the stage

The Kmart case illustrates the impact of failing IT projects¹. Cases of markedly deviant performance, like Kmart's, attract scholarly and public attention. Many more have been documented, e.g., the London Ambulance System (Dalcher 1999), the London Stock Exchange's Taurus System (Drummond 1997), the Denver Airport Baggage Handling System (Montealegre and Keil 1998).

However, two questions arise: How frequent are these cases of failures really? And what is causing them?

¹ In the context of this thesis IT projects are understood as projects, i.e., temporarily purposeful organization set up to carry out a task, in this context investing into information and communication technology. Often, IT projects deliver business change through changes to the organization's information technology. IT projects in this study are conceptualized from the viewpoint of the project sponsor. IT project performance is conceptualized through cost overruns, schedule delays, and benefits shortfalls.

The first question is highly contested. In 1994, the Standish Group, a consultancy, published a performance survey of IT projects. The survey reported that only 16% of IT projects succeeded. Project had an average cost overrun of +189% and an average delay of +222%. The report portrayed an industry in which project failure is the norm. However, subsequent academic studies failed to replicate the findings. The studies raised strong suspicions not only about the results itself but questioned whether it is possible to benchmark IT projects across organizations at all (Glass 1997, Sauer et al. 2007, Eveleens and Verhoef 2010, Jørgensen et al. 2012).

The second question, what is causing IT project failure has been studied through case studies and surveys. The thesis will show in its literature review that theorizing has taken place within three schools of thought: system-centric, event-centric, and process-centric explanations. Each school of thought has rich theorizing within its theoretical base but rarely have explanations spanned multiple schools of thought. This thesis is a first attempt at theorizing across schools of thought.

The goal of this study is not provide a full explanation of the subject but to improve existing research and methods in several ways.

First, using documentary evidence instead of self-reported surveys will provide a more unbiased view of the performance of IT projects and the occurrence of markedly deviant observations, in other words the risks of implementing IT projects.

Secondly, research face the trade-off between the number of observations and the richness of variables, which is even more aggravated when studying rare events (King and Zeng 2001). One research strategy is to sample on the dependent variable (for example in Keil 1995). However, to ensure that the results are representative, the thesis will not sample on the dependent variable. Moreover, the study design will supplement the large-scale study of IT project performance with a more in-depth survey of fewer IT projects, for which more explanatory variables are collected, that test antecedents of performance suggested by management theories.

Thirdly, most analysis strategies first exclude outliers from samples to then run statistical analyses to explain the average performance (e.g., Sauer et al. 2007; Pinto and Slevin 1987). McKelvey and Andriani (2005) best expressed the issue with this strategy:

“...most of the cases and stories are about extreme events – successes or failures – but seldom about averages. No wonder there is a disjunction [between academia and practice]: managers live in the world of extremes; researchers using statistics report findings about averages. Talk about irrelevant research! [...]

The first thing we scholars have to do is get over the idea that studying averages is good science, is the only thing relevant to organizations and offers something useful to managers. Sometimes ‘Yes’, but we think ‘Mostly No’.” (McKelvey and Andriani 2005, pp. 224-225)

To answer the McKelvey and Andriani’s critique the thesis uses a novel analysis, which compares the thickness of the tails of the project performance distribution, thus simultaneously measuring the frequency and impact of

overruns and, more importantly, avoiding the exclusion of outliers. In other words, the thesis will treat large overruns not as unusual cases but as part of the data.

Fourthly, the comparison of the thickness of the tails of the project performance distribution will be compared to a conventional rare events multilevel logistic regression. The thesis improves on previous studies that used regression analyses (e.g., Keil 1995, Kulk 2009) by accounting for organizational differences through multilevel models (Gelman and Hill 2007). Multilevel models have not been used by studies in this field.

While other possible explanations can be articulated and many will and have provided partial explanation for the repetitive deviation of outcomes of IT projects from initial plans, the goal of this research is to explore not only the risk that comes from IT projects running of course but also explore the reasons why this happens. The goal is to study the extent to which different theoretical schools of thought can contribute to the understanding of markedly deviant observations of IT project performance, in other words why outliers occur.

1.2 The Three Schools of Thoughts

The *system-centric school of thought*'s key theoretical explanations of outliers are Normal Accident Theory (Perrow 1981, 1984) and the theory of high reliability organizations (Roberts 1989, Weick and Sutcliffe 2007, Weick 2006). In short, both theories have the premise that highly complex organizations, i.e., those characterized by high interactive complexity and tight coupling, have a higher

propensity to fail. Failure can be prevented through technical designs, influencing organizational and human relationships, and attention to weak signals. In project management theories these ideas have been translated into the concepts of social and technical complexity (McFarlan 1981, Zmud 1980, Whitty and Maylor 2009, Bruijn and Leijten 2008). In the IT project management literature these concepts have been rooted in forecasting theory. The notion of complexity has been operationalized through the proxies of size and duration (e.g. Jones 2000, Sauer et al. 2007, McFarlan 1981, Zmud 1980, Jenkins et al. 1984, Wallace et al. 2004a, Brooks 1975). Other studies have operationalized complexity through assessments of difficulty or manageability of social factors (e.g., Addison and Vallabh 2002, Glass 1997, Yourdon 1996, Boehm 1991) and technical factors of IT projects (e.g. Phalp and Cox 2000, Heltewig 2008, Collopy 2007, Bergeron and St. Arnaud 1992, Heemstra 1992).

The *event-centric school of thought*'s key theoretical explanations of outliers are the notions of rare events and catastrophe and crisis management. The notion of rare events has been linked to interruptions of organizational routines (Christianson et al. 2008), strategic surprises (Ansoff 1975) and crises (Roux-Douffort 2007). In sum, these theories explain outliers as the effect of an insufficient organizational response to a rare event outside the control of the organization. Those insufficient organizational responses are the result of failure to make sense of a rare event, framing a rare event in a way that inhibits an adequate response, and failure to learn from and through rare events (Baumard and Starbuck 2005, Weick 2004, Rerup 2009, Christianson et al. 2008).

In the project management literature, the concept of rare events is translated into the notion of turbulences, which is the core of project risk management (Miller and Lessard 2000, 2001, 2008, Pich et al. 2002). This line of inquiry has failed to attract much research in project and IT project management fields because it is conceptualized to be the very core of project management.

The *process-centric school of thought*'s key theoretical explanations of why outliers occur are the theories of man-made disasters (Turner and Pidgeon 1997), normalization of deviance (Vaughan 1996), and the escalation of commitment (Ross and Staw 1986, Sleesman et al. 2012). In project management literature, the escalation of commitment is the dominant theoretical lens, which has been particularly studied as project escalation (Keil et al. 2000). In IT projects the process-centric school of thought has often been atheoretical and focused on identifying success factors of project execution and delivery (Pinto and Slevin 1987, Cooke-Davies 2002, Fortune and White 2002, Bleassi and Tukel 1996). Other theoretical explanations of why IT projects have become outliers has been the study of project abandonment which is rooted in the notion of support, i.e., projects were abandoned because they lost economic (Oz and Sosik 2000), political (Sauer 1993), or technological support (Ewusi-Mensah and Przasnyski 1991).

1.3 The Thesis' Key Findings

The framework of explanations from the extant management literature will be demonstrated by its application to the context of business embedded IT

projects. The premises of system-centric, event-centric, and process-centric explanations will be tested on a sample of 4,307 projects from 190 organizations. The data came from multiple sources. Project data were directly collected on 1,365 projects from 34 participating organizations. A desk study collected data on 289 projects from published academic studies and auditor reports. Finally data on 2,653 IT projects from the federal US government were collected. 1,029 through FOIA data for budget requests from each of the 134 agencies and 1,624 from submissions to the Office of Management and Budget.

Crucially, the findings will demonstrate that outliers happen frequently and regularly in IT project performance. 14% of the observations were an outlier of cost performance, 15% a schedule performance outlier, 10% an effort performance outlier, and 6% a benefits performance outlier. Moreover, the analysis of the tails of the cost, schedule, and effort performance distribution will show that the tails can be fitted to power laws of the form $p(x|\alpha) = \frac{\alpha-1}{x_0} \left(\frac{x}{x_0}\right)^{-\alpha}$ above a cut-off value $x > x_0$. The power law tail has the exponent $\alpha = 2.5$ for cost performance, 3.1 for schedule performance, and 2.7 for effort performance. This key finding indicates that very large overruns are infrequent but not impossible. IT project performance is, in Mandelbrot's (1997) terms fat tailed in some dimensions even heavy tailed. Outliers are so frequent that the noise drowns out the signal in the data, which is evidenced by the finding that project performance follows distributions with slowly converging means and variance, if the higher order moments converge at all.

When explaining the tails the rare events logistic regression and the comparison of the thickness of the tail will consistently show that factors from the system-centric school of thought (in particular estimated cost, estimated duration, and perceived uniqueness) and factors from the process-centric school of thought (in particular effectiveness of monitoring and controlling the project, and qualification and motivation of the project team) significantly explain the occurrence of outliers.

1.4 Research Aim and Contribution to Knowledge

The aim of the thesis is two-fold. First, to advance the understanding of the challenge of outliers, i.e., which effect the markedly deviant observations have on overall performance of IT projects. Second, to advance the understanding of the causes of outliers that draws on different organizational, project management, and IT project management theories comparing system-centric, event-centric, and process-centric explanations.

The purpose of the thesis to bridge the gap between case studies of outliers in IT project performance, which offers rich theoretical explanations, and quantitative surveys of IT project performance, which offers a view on the average risk of IT projects but have excluded outliers before the analysis.

The thesis will contribute to knowledge in three ways: (1) theoretically, (2) empirically, and (3) methodologically.

First, the thesis provides a first summary of the extant literature across the three schools of thought. The thesis compares explanations from different schools of thought head-to-head. The results will strongly show the need for future research that engages in more theorizing across schools of thought.

Second, the thesis provides a large n sample of the performance of IT projects based on documentary evidence. The power law distribution implies that higher order moments converge slowly, if at all, and that the largest observed overrun in a sample depends on the sample size. In other words, small samples underrepresent the true risk of IT projects. Moreover, the thesis does not rely on self-reported project performance but follows the practice of studies of software forecasting in Computer Science (e.g., Molloken-Ostvold et al. 2004), something that has not been done in management research.

Third, the thesis proposes a novel method of comparing the thickness of the tails of the project performance distribution to study the frequency and size of large overruns. The analysis will also problematize the false discovery rate of conducting simultaneous independent statistical tests, which again has not been done before in this field (e.g., Pinto and Slevin 1987, Cooke-Davies 2002). The thesis also shows that multilevel rare events regression models are suited to compare project performance between different organizations. Aggregating project performance from multiple organizations has been questioned as a suitable research strategy (Verhoef 2004) because of organizational biases, learning, and spill-over effects.

1.5 Structure and Summary of the Thesis

The remainder of the thesis is structured into 6 chapters that (1) present the extant literature, (2) describe the research methodology, data collection, and analyses, (3) describe the findings of how frequently outliers occurred in IT project performance, (4) describe the findings why they occurred, (5) discuss the implications of the challenge and the causes of outliers in IT project performance, and (6) provide a summary.

In chapter 2 the extant literature is reviewed. First, the chapter reviews prior quantitative surveys of IT project performance. The survey shows that studies reported large variations between and within studies. To the extent that it has been questioned whether the studies can serve as a reliable benchmark at all. The literature review then groups the explanations of why overruns and outliers happen in three schools of thought: system-centric, event-centric, and process centric theories.

With the findings from the literature as a starting point chapter 3 develops the methodology. It describes the survey instrument, the data points of the archival research, and the items and anchoring of the items used for the post-mortem analysis. The chapter also includes a documentation of the steps taken to collect the data and prepare the data for analysis. The chapter then derives the statistical formulae and provides their proofs. Lastly, the chapter describes the sample of the 4,307 IT projects in the analysis. It shows that among those projects software implementation projects were most common (90%), followed

by IT infrastructure (5%), communication technology (3%), and IT architecture (1%) projects. The most frequent types of software implemented were supply chain management software (27%), ERP software (20%), and HR management systems (10%). The projects in the sample are on average USD 24.6m (in 2010 terms), median 1.2m. The average duration is 280 days, median 182 days. The sample also is tested for heterogeneity between observations that could potentially hinder aggregation. The test analyzed program structure, geography, organization, project type, and data sources. The results strongly suggest that the organization is structuring the data and that multilevel models are needed for analysis.

Chapter 4 presents the findings from analyzing the frequency of outliers in IT project performance. To do so chapter 4 tests and rejects the hypothesis that outliers did not occur more frequently than under the assumption of thin tails. It then tests and rejects the hypothesis that the data are normally distributed. Finally, it fails to reject the alternative hypothesis that power laws can be fit to the tails of the distribution. The analysis also fails to explain the tails as a spurious result of mixing lognormal distributions, mixing distributions with same means and different variances, and aggregating over non-stationary parameters.

Chapter 5 presents the findings from analyzing the suggested causes of the outliers in IT project performance. The findings show that several factors from the system-centric school of thought are associated with the occurrence of

outliers. Those factors are estimated project size (cost and duration), social complexity (clearness of goals, organizational resistance, political intent, realistic schedules, perceived uniqueness of the project). Factors from the process-centric school of thought are linked to outliers, in particular monitor and control, benefits measurement, up-to-date plans, competency of project managers, qualification and motivation of the team, suitability of staffing, and the number of years of work experience of the most senior project manager.

Combining the results from the different analyses and accounting for the false discovery rate of simultaneously testing independent hypotheses five factors emerge that explain the occurrence of outliers: estimated cost, estimated duration, perceived uniqueness of the project, effectiveness of monitoring and controlling, and motivation and qualification of the project team.

Chapter 6, discusses the implications of the two previous chapters. First, the implications from the finding that the performance of IT projects regularly and frequently include outliers. It is concluded that outliers are not unique, idiosyncratic events. Large overruns occur infrequently but are not impossible. Observations of markedly deviant performance ought to be expected. The findings have implications for research methods, namely that outliers are not from a different universe of randomness than smaller overruns. Thus, they ought not to be excluded from analyses. The findings also have implication for the understanding, forecasting, and modeling of the risk of IT projects, namely that the largest observation strongly influences the center of the distribution. In

short, it concludes that in most state-of-the-art models risk is severely underestimated.

The chapter then discusses the implications from the explanations of why outliers occurred. The findings imply that the system-centric and the process-centric school of thoughts have merits. The first one provides key design variables (estimated cost, estimated duration, and perceived uniqueness of the project) to explain why outlier happened. The second school of thought provides key management variables (effectiveness of monitoring and controlling, motivation and qualification of the project team) to explain why outlier occurred. The findings strongly suggest that both design and execution activities contribute to the explanation of markedly deviant project performance. Finally, the chapter discusses the practical implications of the findings as well.

In chapter 7, the final conclusions of the thesis are summarized, beginning with a review of the key propositions, findings, and the evidence presented in the thesis. To conclude, the chapter provides a practitioner oriented framework of assessing risk and suggests avenues for further research.

2 Literature Review

Chapter 1 problematized the performance of IT projects. The chapter illustrated the potential impact of an out-of-control IT project with the case of Kmart, where the combination of two challenged IT projects contributed to, if not caused, the bankruptcy of a major US company with USD 6.3 billion in assets. Chapter 1 then questioned whether these extreme performances – termed outliers – are atypical for IT projects.

The literature review will document the prior academic literature for the two research questions: (1) How frequently are outliers, that is markedly deviating observations, in IT project performance? (2) Why do outliers occur in IT project performance? However, before reviewing the extant literature, it is necessary to define the two concepts of the IT project and the outlier.

2.1 Definition of IT Projects

In the extant literature, the concept *IT project*² has been used to describe different types of observations. The thesis draws on two traditions (1) software engineering and (2) information management. Both traditions used different conceptualizations, linked to different research objectives.

In the *software engineering* tradition studies mostly focused on improving the accuracy of software effort and time estimates (e.g., Jones 2000, Jones 2007, Jones 2008, Moløkken-Østvold et. al. 2004, Bergeron and St. Arnaud 1992).

These studies take a narrow definition of what constitutes an IT project. For

² Projects offer a mid-range theoretical space between the study of tasks as intentions and whole organizations as intentions (Packendorff 1995). While tasks, when defined unambiguously, predominantly pose the question of efficiency and effectiveness of carrying out the tasks; projects added layers of leadership, control, structure, and goals in terms of time, cost, and scope/quality (Packendorff 1995). A mid-level theory portrayed projects as temporary organisations. Projects, like organizations, allocated resources to unique, novel, and transient undertakings, which managed inherent uncertainty and integration to deliver a beneficial objective (Lundin and Soderholm 1995). The temporary organisation has four characteristics. It (1) is an organized course of action, (2) has a predetermined point in time where it is expected to cease to exist, (3) has some kind of performance objective, and (4) is complex enough to require organizing (Packendorff 1995, p.327). However, projects also demonstrated differences from permanent organisations. The most important difference was time. The very first study of projects as a management problem points out that: "Another unique aspect of the project manager's job is that his task is finite in duration" (Gaddis, 1959:90). Discussions of permanent organisations did not problematize time. In projects time created a sense of urgency and scarcity to motivate swift action but time also created stress and dysfunctional temporal behaviours, e.g. group think or goal completion effects (Packendorff 1995, p.327). A second important difference between temporary and permanent organizations was goal specificity (Packendorff 1995). Projects were described as an organisational unit in pursuit of a singular goal (e.g. the development of a new product) on-time, on-budget, and on-scope (Gaddis 1959). This triple constraint (Lech 2013) has been later termed the iron triangle (Atkinson et al. 1999). However, Shenhar et al. (2001) found that the abstraction of temporary organizations to singular financial goals neither captures the full complexity of organisational reality nor reflects long-term success. Additional goals of temporary organizations have been documented, e.g., perceived value of the project for the organisation and client satisfaction (Pinto and Mantel 1990, p.270, also Shenhar et al. 2001, p.705). However, Lundin et al. (1995, p.440) argued that projects differed from permanent organisations because they have a single goal whereas organisations have multiple goals. It has also been suggested that temporary organisations were setup to buffer uncertainty from the permanent organisation (Kreiner 2013). In sum, a strong argument exist that projects can be analysed as temporary organizations through organizational theories.

example in Jones (2000) the activities that constitute an IT project are specification, design, development, documentation, testing, and project management.

The *information management* tradition focused mostly on improving the delivery of IT projects to businesses (e.g., Keil et al. 2009, Keil et al. 2010, Sauer and Cuthbertson. 2003, Kulk et al. 2009, Eveleens and Verhoef 2009). These studies observed IT projects embedded in a business context and not only in a software development context. Embedding projects in a business context widened the view of what constitutes an IT project. For example, most studies of business-embedded IT projects measured the total effort of projects, which included efforts to change business operations, to provide post-implementation support, and most importantly the front-end process of project planning.

The thesis uses the second definition. IT projects are business embedded projects, in the sense of purposeful, temporary organizations where the solution to achieve the purpose requires information and communication technology. As a result the thesis also demonstrates the novel challenged faced by software engineering research to extend the scope of forecasting beyond “coding and bug fixing” as urged by Hassan et al. (2013).

2.2 Definition of Outlier

The use of the notion *outlier* is somewhat muddled in the literature. Historically it has been used to describe extreme observations of a phenomenon that deviate markedly from other observations in the sample (Grubbs 1969).

In other words, an outlier is an observation which falls beyond the expected range of potential observations. The term *outlier* is also used to describe the most common cause of markedly deviant observations – that of an erroneous observations, for instance as a result of measurement error, change in experimental conditions, transcription and translation error. In research practice the first understanding of the notion outlier, that of a markedly deviant observation, is often used to imply the second understanding, that of an erroneous observation.

Mixing both notions dates back to the very first usage by Bessel and Baeyer (1838) and Legendre (1805). Hence exclusion is the de-facto standard of dealing with outliers, often without rigorous causal analysis whether an outlier is actually the result of an error or not. To underline the difference in both meanings Tukey (1977) designated the first category *potential outliers* because exploratory statistics alone cannot establish whether an observation is or is not the result of error. Tukey (1977) also established the commonly used definition that outliers are more than 1.5 interquartile ranges away from the ends of the box in a boxplot. The interquartile range is the length of the box in a boxplot, i.e., distance between the bottom and the top quartile (Tukey 1977, p. 44). This

approximation is still commonly used in outlier detection (cf. the reviews in Beckman and Cook 1984, Hubert and van der Veeken 2008, Hodge and Austin 2004).

The thesis follows the terminology established by Anscombe (1960), Grubbs (1969), and Barnett (1978). Here, the term *outlier* denotes an observation of a project that markedly deviates from expected performance.

Tukey (1977, p. 43-44) defines a “convenient rule of thumb that picks out certain values as ‘outside’ or ‘far out’.” Tukey’s rule of thumb states that outliers are 1.5 interquartile ranges³ away from the top quartile. Similar approaches identify outliers by different distance measures, e.g., distance from the median (Anscombe 1960), standard deviation (Dixon 1950), extreme studentized deviate (Grubbs 1969). This thesis follows Tukey’s original definition that outliers are more than 1.5 interquartile ranges away from the box in a boxplot.

³ The interquartile range is the difference between the 25th and the 75th percentile, i.e., the length of the box in a boxplot or in Tukey’s terms the ‘H-spread’

2.3 How Frequent are Outliers in IT Project Performance?

To date, no surveys exist that specifically quantified the occurrence of outliers in IT project performance⁴, even though outliers were present in the data as evidenced by boxplots and mentioned in description of how data were cleansed before analysis. This thesis' focus lies on the key performance⁵ dimensions of IT projects: cost, schedule, and benefits performance. Thus the thesis reviewed existing quantitative surveys of these dimensions. Because the issue of outliers in project performance is not well covered by prior work, the thesis also reviewed studies of indicators closely related to project performance, namely inputs and throughputs of the software development process. In these studies the variables of interest were budget estimates, effort and size estimates, developer productivity, and defects in the development process. These findings will allow to make conjectures about the frequency of outliers in these related dimensions.

2.3.1 *Surveys of Performance of IT Projects*

Brooks (1975) first problematized the issue of cost overruns and schedule delays in the seminal the 'Mythical Man-Month'. Shortly thereafter, quantitative

⁴ And in the performance of projects in general.

⁵ Before later defining project performance in detail: Project performance is typically quantified in terms of three dimensions of (1) cost, (2) schedule, and (3) benefits or quality or functionality delivered. This thesis focusses on cost-benefit-analyses. Thus, performance is expressed as actual/estimated cost, schedule, and benefits either as percentage figure (actual/planned - 1) or as ratio (actual/planned). In the traditional project management literature the dimension of benefits is often linked to the differentiation between projects (which perform against quality or functionality goals) and programs (which have benefit goals). However, this thesis' data collection and prior academic work showed that NPV and payback calculations are the dominant decision rationale if project outcomes are considered at all. Thus, while benefits are a theoretically valid way to differentiate projects from programs, this thesis was not able to observe this distinction in practice.

studies, beginning with Augustine (1982), tried to critically assess the issue. More research was published as a response to the controversial Chaos Report (Standish 1994 and onwards). The literature reviewed identified 31 quantitative studies that reported project performance, more research exist that quantified the performance on the task level (e.g., Genuchten 1991, Moløkken-Østfold et al. 2004). Table 1 lists the studies and the reported sample sizes in number of projects, mean project size, success rate, mean (median) cost overruns, mean (median) schedule delays, and mean (median) benefit shortfalls.

Table 1 Overview of prior quantitative studies of IT project performance, mean (median) performance

Author	Year	Sample size	Project size	Success rate	Cost overrun	Schedule slippage	Benefits shortfall ⁶
Augustine	1982	100				+33%	
Bergeron and St. Amand	1992	89	1251 PD ⁷		+33%		
Emam and Koru	2008	288		84% (2005) 88% (2007)			
Eveleens and Verhoef	2009						
• Org X		867			+286% (-38%)		
• Org Y		168			+16% (±0%)		+10% (±0%)
• Org Z		307			±0%		
• Little	2006	121			-15%	+105% (+79%)	
Ewusi-Mensah and Przasnyski	1994	49		28%			
Heemstra	1989	2659		30%			
Jenkins et al.	1984	72	USD 103k	61%	+67% (+34%)	+22%	
Jones	2000	9155		69%			
Jones	2007			62%			
Jones	2008	8290		71%		-8%	
Keil et al.	1994 ⁸	579		62%	+18% ⁹	+22% ¹⁰	
Kulk et al.	2009	165	USD 2.2m		-4%		

⁶ A positive benefits shortfall is a shortfall, a negative benefits shortfall delivered more benefits than planned

⁷ PD = Person Days

⁸ This dataset has been reported also in Keil and Mann (1995, 1997) and Keil et al. (2000)

⁹ For non-escalated projects, in escalated projects +119%

¹⁰ For non-escalated projects, in escalated projects +103%

					(-10%)		
Lederer and Prasad	1991		Total 202 PD	40%			
Little	2006	106	329 days		+80% (+100%)		
McKeen	1983	32	USD 35k		+28% (+22%)	+18% (+22%)	
Mitchell and Zmud	1999	42	USD 3.8m			+29%	
Moløkken-Østvold et al.	2004	36	400 PD	11%	+41% (+21%)	+25% (+9%)	
Oz and Sosnik	2000	27		2.66 abandoned projects in 5 years			
Patanakul	2014	16	USD 594m		+75% (+67%)	+118% (+60%)	
Phan	1990	827			+33%		
Sauer and Cuthbertson.	2003	421		16%	+18%	+23%	+7%
Sauer et al.	2007	412			+13%	+20%	-7%
Standish	1994	365 org		16%	+189%	+222%	
Standish	1996			27%	+142%	+131%	
Standish	1998			26%	+69%	+79%	
Standish	2000			28%	+45%	+63%	
Standish	2002	80000	USD 3-5m	34%	+43%	+82%	+67%
Standish	2004	80000	USD 3-5m	29%	+56%	+84%	+64%
Standish	2006	80000	USD 3-5m	35%	+47%	+72%	+68%
Standish	2008	80000	USD 3-5m	32%	+54%	+79%	+67%
Standish	2010	80000	USD 3-5m	37%	+46%	+71%	+74%
Topping	1985	22			+40% (+26%)		

Note: Patanakul (2014) figures are calculated from data given in study, they were not reported directly

2.3.1.1 Cost Performance of IT Projects

Studies documented a wide range of project cost performance. Findings ranged from cost underruns of -4% and -15% (Kulk et al. 2009¹¹, Little 2006), to projects being on average on target (Eveleens and Verhoef 2009), to moderate cost overruns between +13% (Sauer et al. 2007) and +40% (Topping 1985, Moløkken-Østvold et al. 2004). Findings also include observations of extreme performance with average cost overruns of +286% (Eveleens and Verhoef 2009). In sum,

¹¹ The study comprises three organizations, the Little (2006) data is part of other studies by the same authors (Eveleens et al. 2012) but not this one.

these studies report large variation between the average cost overruns, with little consensus between studies.

In addition to the large amount of variation, the studies cited in table 1 also showed a large amount of variation within the samples observed by each study. The spread between median and mean project cost performance gives an indication of the skew within samples, that might be the effect of outliers, which influence the mean but not the median. Only nine studies reported a median as well as a mean. The mean fell above the median cost overrun in eight studies. The average spread was 50 percentage points with a maximum spread between mean and median of 324 percentage points (Eveleens and Verhoef 2009).

Drawing the review back to the research question, the frequency of outliers in cost performance data of IT projects, that is the frequency of markedly deviant observations, has only been reported in Kulk (2009)¹². Kulk (2009) reported that 33% of the observations of cost performance were outliers and needed to be excluded to fit a regression model to the data. The number of outliers is staggering given that outliers should not occur in more than 0.7% of the observations under the assumption of thin-tailed distributions¹³. The data show

¹² None of the studies cited in table 1 explicitly discussed the issue of outliers, except Yoon et al. (2007), who discussed total duration and total effort of projects. Outliers were only reported in boxplots (e.g., Mitchell and Zmud 1999) or mentioned in the description of how data were cleansed for statistical analyses (e.g., Kulk 2009).

¹³ While frequently questioned the assumption is commonly made. Software engineering often refers to the cone of uncertainty (McConnell 1996) as an idealised version of how estimates increase their accuracy over time (e.g., in Zappata and Chaudron 2012; Moløkken-Østfold et al. 2004; Eveleens and Verhoef 2009). An implicit implication of the cone of uncertainty is that the

an over incidence of outliers of 4,614%. In addition, two studies have looked at other indicators that are closely related to project cost performance.

Verhoef (2005) reported the observation of 8% outliers in actual and estimated cost, and 5% outliers in cost per function point. Banker and Kauffman (1991) measured software development productivity, where 5% of observations were classified as outliers (Banker and Kauffman 1991). These data, which are closely related to IT project cost performance, again show an over incidence of outliers. Here outliers are 1042% and 614% more common than given by the expectation that a maximum of 0.7% of the observations fall beyond the whiskers of the boxplot.

In sum, only one single study has reported the frequency of outliers, that is markedly deviant observations, of cost performance data of IT projects. In this

law of large numbers holds true, i.e., that the true mean and higher order moments can be estimated without statistical bias with a large enough sample, moreover that the error of estimates is symmetric and finite. This assumption is wide spread, e.g., can be found in research on financial analysis on the project portfolio (McFarlan 1981), real options evaluation of projects (for example Tiwana et al. 2007, Taudes et al. 2000, Huchzermeier and Loch 2001), and recently Bayesian updating of cost and schedule forecasts in earned value management systems (Caron et al. 2012).

An exemplary argument is provided by Zappata and Chaudron (2012). The study found with overwhelming statistical significance that the sample distributions are not normal, i.e., thin-tailed. Yet Zappata and Chaudron argued that the cone of uncertainty predicts a very wide range of potential estimates in early project stages. If one assumes a range of three standard deviations around the mean, nearly all projects in their sample fell within that range. They conclude that this made the standard normal an acceptable approximation. Kitchenham (1998) has long questioned this approach pointing out that statistical analyses based on normal distributions are inadequate in IT research. Kitchenham (1998) analysed software development productivity and Verhoef (2005) the distribution of project budgets in project portfolios. Kitchenham (1998) and Verhoef (2005) made the conjecture that most IT performance indicators are skewed and fat-tailed.

MacDonnell and Shepperd (2003) compares different estimation techniques on a skewed data set and found that least squares regression gave the most accurate results but the analysis did not include robust methods. Raffo and Kellner (2000) used simulation studies to predict defects, effort, and duration of a software development task. Despite using 'real-world data' results of the simulation were normally distributed.

study and other studies of closely related indicators outliers occurred much more frequent than the assumption of thin tails would suggest. Moreover, the review showed that the reported average cost overruns varied widely between and within the 31 studies, which measured the IT project cost performance.

2.3.1.2 *Schedule Performance in IT Projects*

Only 11 studies reported schedule performance of IT projects. The average schedule performance ranged from -8% underrun (Jones 2008) to moderate schedule delays between +18% (McKeen 1983) and +33% (Augustine 1979). In comparison, the schedule performance showed less variation between studies than cost performance.

In terms of the variation within studies, only three studies reported mean and median schedule overruns. In two the mean fell above the median (Little 2002 and Moløkken-Østfold et al. 2004). The average spread between mean and median was 26 percentage points, with a maximum spread of 58 percentage points (Patanakul 2014).

Again, only one single study reported the frequency of outliers in the schedule performance data. The distributions in Mitchell and Zmud (1999) depicted that 2% of the observations of schedule performance were outliers. The frequency of outliers suggests a 186% over incidence of outliers compared to the assumption of thin tails. Other studies have reported on outliers in dimensions closely related to project schedule performance. Grant et al. (2006) reported on the monetary value of schedule variation, which showed 0.2% outliers. Two studies

reported the frequency of outliers in actual project duration (Verhoef 2005 and Yoon et al 2007). Yoon et al. (2007) reported that 11.5% of the observations were outliers, whereas 1.5% of the observations in Verhoef (2005) were outliers. With the exception of Grant et al. (2006) where the sample showed fewer outliers than expected, these dimensions again demonstrate an over incidence of outliers of 1,543% and 114%, respectively.

In sum, the data of schedule performance show fewer outliers than the data on cost performance. Moreover the reported averages and medians indicate less variation between and within studies than the cost performance data.

2.3.1.3 Benefits Performance in IT Projects

Only three studies, apart from the Standish reports, have quantified benefits shortfalls. The reported average benefits performance ranged from an underestimation of benefits, i.e., over-delivery, by -7% (Sauer et al. 2007) to a shortfall between 7% (Sauer and Cuthbertson. 2003) and 10% (Eveleens and Verhoef 2009). No study reported the frequency of outliers in the data. Only Eveleens and Verhoef (2009) reported a median for one of the organizations observed in the sample (Org Y), where the median was 0% and the average 10% shortfall. Again, the findings point towards variation in the data, which in comparison is not as large as the variation in the cost performance data.

2.3.2 *Explaining the Variation in the Data*

The wide range of variation between studies has alerted scholars to two potential biases in the results: (1) sampling bias and (2) deliberate bias in forecasts.

First, Moløkken-Østvold et al. (2004) hypothesized that the reported averages widely varied due to differences in data collection. Studies of project performance have used two principal sources of data: the review of documentary evidence and surveys. Moløkken-Østvold et al. argued, in a review of a subset of the studies listed in table 1, that the survey method used by Sauer and Cuthbertson (2003) might have led to lower average cost overruns than studies that were based on documentary evidence. They suspected that survey respondents were biased in which projects they chose to report and thus introduced bias, which explained the smaller overruns.

Second, Eveleens and Verhoef (2009) unpacked the variation found between organizations in their sample, where the data came from documentary evidence. They found that one organization showed a tendency to deliberately underestimate, because project forecasts were deliberately made to identify the fastest possible time to market and the minimum amount of resources needed for the project. In the light of this description of estimation practice, the authors

concluded: "[prior survey] figures should not be used for benchmarking, since the biases in the data sets are unknown" (Eveleens and Verhoef 2009:983)¹⁴.

However, neither of the two explanations can fully account for the variation documented across the three dimensions of IT project performance. First, if observations are randomly sampled survey methodologies should yield representative and externally valid results. Second, deliberate organizational biases of the forecasts cannot explain the large amount of variation within organizations.

Outliers, i.e., markedly deviant observations, are an alternative explanation of the variation within and between studies. Frequently occurring outliers suggest the existence of fat tails (Mandelbrot 1960, 1961, 1963ab, Fama 1965). While rarely reported, the limited empirical evidence suggests that outliers frequently occur in all dimensions of IT project performance and in closely related variables.

Phenomena with fat tails show large variance and extreme skewness; but the most important feature is the length of the tail (Mandelbrot 1997 chapter E3). In

¹⁴ This thesis followed the same precautions as Eveleens and Verhoef (2009) to address the potential concern of political and deliberate biases in the forecasts. Whenever a credible suspicion arose that the data points were deliberately biased the data points were excluded from the analysis. In total nearly 2 projects for every project in the sample were excluded. Common deliberate misrepresentations were outright lying, measuring against latest baselines while omitting how performance was calculated, performance measurement against annualized budgets (in one organisation even quarterly project budgets)

other words, outliers, that is markedly deviant observations, occur so frequently that the noise in the data obscures, even drowns out, the signal.

Moreover, if variance is infinite the strong law of large numbers¹⁵ is violated but the weak law¹⁶ still holds true (Mandelbrot 1997). But under the condition of fat tails the weak law of large numbers requires large sample sizes for convergence because statistical variance might be infinite. In other words, sample moments (mean, variance etc.) are erratic and extremely sample dependent (Mandelbrot 1997 chapter E3). The combination of different sample sizes, which ranged from small $n < 30$ to very large samples $n > 1,000$, would result in the lack of consensus among research findings that is shown in table 1. The mean (and the variance) used to measure the performance of IT projects would only stabilize with infinite samples. That is to say that “even the largest available sample is too small” (Mandelbrot 1997, chapter E3, p. 13).

The most common strategy to deal with unruly data sets is to exclude outliers (Barnett 1978). In the extant literature on IT projects few documented whether the analyses excluded outliers, among those who mentioned the rejection of outliers are Ahmed et al. (2013), Kulk (2009), Trendowicz et al. (2006), Yoon et al. (2007), Grant et al. (2006), Barros et al. (2004), Ruhe et al. (2003), Mitchell and Zmud (1999), Banker and Kauffman (1991), Andersen (1990).

¹⁵ $\Pr\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1$

¹⁶ $\bar{X}_n \rightarrow \mu$ for $n \rightarrow \infty$

The only example in the project management literature, where fat tails were explicitly incorporated in the analysis, is found in scheduling studies, especially studies advancing the PERT method¹⁷. The studies focused on understanding the true variation of project input factors – in particular effort and time needed to complete the tasks that comprise projects. Malcom et al. (1959) developed PERT and replaced the normal distribution assumption and with a beta distribution assumption (which is the standard in PERT studies e.g., Abdelkader 2004, Hahn 2008; Keefer and Bodily 1983, Perry and Greig 1975, MacCrimmon and Ryavec 1964, Kotiah and Wallace 1973, Golenko-Ginzburg 1988, Herrerias-Velasco et al. 2011). In the original PERT paper (Malcolm et al. 1959) the beta distribution is chosen so that the standard normal is truncated to ± 2.96 , which in fact has shorter tails than the normal distribution. The assumption of thicker tails than the normal distribution were rare, e.g., uniform distribution used by Chapman and Ward (2003) or lognormal distribution used by Abdelkader (2004) and Perry and Greig (1975). Other models assumed tails whose thickness was thinner than the lognormal distribution but still thicker than the normal distribution, for example: Gamma distribution (Perry and Greig 1975), exponential distribution (Kulkarni and Adlakha 1986), Erlang¹⁸

¹⁷ PERT stands for Project Evaluation and Review Technique, a project estimation technique first developed to measure and control the development of the US Navy's Polaris Fleet Ballistic Missile Program. The method made a three point estimate for each task duration (optimistic, likely, pessimistic), it was assumed that task duration follows a beta distribution so that the weighted average of expected task duration = $4/6 * \text{likely estimate} + 1/6 * \text{optimistic estimate} + 1/6 * \text{pessimistic estimate}$ (Malcom et al. 1959).

¹⁸ A special case of the Gamma distribution.

distribution (Bendell et al. 1995; Azaron et al. 2006), and Weibull distribution (Abdelkader 2004).

In sum, the studies documented a wide range of how IT projects perform, in terms of cost, schedule, and benefits performance. The studies also showed considerable skew in the results – in short IT projects showed both variation between studies and variation within studies.

The thesis argued that outliers should be considered a symptom of fat tails, one neglected explanation of the large variation in the extant literature. The review found that most studies excluded outliers before analysis, few reported their frequency. In sum, the question how frequent outliers occur in IT project performance has not yet been successfully answered.

2.4 Why Do Outliers Occur in IT Project Performance?

The previous section addressed the question: How frequent are outliers, in the sense of markedly deviating observations, in IT project performance? This section now focusses on the causes: Why do outliers occur in IT project performance?

The literature review shows that the extant literature can be grouped in three schools of thought: (1) system-centric, (2) event-centric, and (3) process-centric explanations of why outliers occurred.

First, *system-centric explanations* problematized the question of organizational and technical design of systems. As the literature review will show system-centric theories argued that organizational complexity¹⁹ led to catastrophes and disasters, that is outliers in the sense of markedly deviant organizational

¹⁹ Complexity is a multi-faceted concept. The organizational and project management theories reviewed here understand complexity as (1) interactive complexity between the parts, including human actors, that make up a system and (2) the coupling between those parts (e.g., Perrow 1981, Weick and Sutcliffe 2007, Baccarini 1996). This idea originates from Simon's Architecture of Complexity: "Roughly, by a complex system I mean one made up of a large number of parts that interact in a nonsimple way" (Simon 1962, p. 468).

A second perspective on complexity is the notion of emergence (e.g. McKelvey 2004). In this vein, project management scholars have defined complexity theory as: "as the study of how order, structure, pattern, and novelty arise from extremely complicated, apparently chaotic systems and conversely, how complex behavior and structure emerges from simple underlying rules" (Cooke-Davies and Cicmil 2008, p. 52).

Other approaches in organizational theories view complexity through a solvability lens, often to contrast between complex and complicated system. Complexity here refers to the unsolvability of models, i.e., the inability to optimize the system (e.g. Shoemaker 1990) or the nonlinearity of systems (Anderson et al. 1999). Unsolvability and nonlinearity again have been linked the number of actors and interdependency, where complexity is the "inability to evaluate the effects of actions because too many variables interact" (Pich et al. 2002, p. 1009). Similarly, complexity has been operationalized through "variety as a proxy for complexity" (Boisot and McKelvey 2012, p. 421).

For now, the thesis uses as working definition the first concept of complexity, where complexity is defined as a system characteristic that makes the system non-simple.

performance. The research question formulated by the first school of thought is: What social and technical design factors make an organization complex and thus fragile, i.e., more likely to turn into an outlier?

Second, *event-centric explanations* emphasized the organizational response to rare events that impact an organization. The notion of an event implies that an organization cannot control the event itself but organizations control how they respond to events. Crises, disruption, and turbulences – in other words outliers – are then the result of an inadequate organizational response. The research question posed by the second school of thought is: How do organizations effectively respond to rare events to prevent becoming outliers?

Third, *process-centric explanations* emphasize how organizations operate and behave over time. Instances of markedly deviating performance are explained as the result of escalation of commitment to a failing course of action, normalization of deviance, or the accumulation of miscommunication and misunderstood errors over time. The research question posed by of the third school of thought is: How effective was the management of the IT project, in order for the project to not become an outlier?

The literature review will discuss each school of thought in turn. Each school of thought is present in organizational, project management, and IT project management literatures. This thesis builds on a substantial body of prior work, each of the schools of thought has been thoroughly discussed and reviewed.

However, little theorizing has taken place across the three different schools of thought. This is a first attempt.

2.4.1 System-Centric School of Thought

The system-centric school of thought is grounded in two organizational theory perspectives: (1) the Theory of Normal Accidents, which has a technological focus; (2) Theory of High Reliability Organizations, which has a social focus.

Building on these concepts project management theory further problematizes both perspectives through studies of technical and social project complexity.

The concepts of project complexity are further refined and operationalized in IT project management theories, which focus on complexity as a scaling factor to estimate effort, schedule, and cost of IT projects.

2.4.1.1 System-Centric, Organizational Perspectives of Why Outliers Occur

The system-centric school of thought has its historic origin in the study of technological disasters, which can be seen as an instance of markedly different performance outcomes. Normal Accident Theory first explained technological disasters as a consequence of failure of a technological system. Normal Accident Theory showed that systems with high interactive complexity, i.e., a large number of non-linear interactions, and tightly coupled systems were bound to eventually fail (Perrow 1981). Normal Accident Theory viewed systems through their structure, e.g., the structure of teams, technology, and cognition (Perrow 2004) and, based on the system structure, predict system behavior and outcomes.

Normal Accident Theory originated from the case study of the Three Mile Island nuclear accident. On March 28, 1979, a partial meltdown occurred of one of the two nuclear power reactors at the Three Mile Island facility in

Pennsylvania. It had been the worst nuclear power plant accident in the US. During the accident a stuck-open valve released large amounts of radioactive coolant into the environment. Perrow's case analysis showed that the system complexity obfuscated what was happening during the accident. The accident remained mysterious and incomprehensible to the experts on the ground while it unfolded. Three Mile Island demonstrated how system complexity prevented operator intervention that could have stopped the accident by correctly identifying that the coolant valve was stuck open and coolant was being lost. Moreover, post-accident investigation showed that the actual accident was triggered by an insignificant event, i.e., the low-probability equipment failure in the pilot-operated relief valve, which were amplified into a full system failure through the interactive complexity of the system: in this case, the interaction with auxiliary cooling pumps (Perrow 1981).

Abstracting from the Three Miles Island case, Normal Accident Theory proposed that complex technical systems create noise. Strong signals are weakened and are therefore not picked up by operators. Muffled signals fail to trigger the intervention that could have prevented the accident. Strong signals were weakened by a multitude of simultaneous warnings, e.g., in the form of audio-visual feedback from the operator interface and from warnings embedded in handbooks and standard operating procedures. False alarms were a second aspect of how signals, that were designed to send strong warnings, are weakened. Complex interactions in the system were a main source of false warnings (Perrow 1981). Later, the concept of organizational mindfulness

(Weick 2006, Weick and Sutcliffe 2006, 2007, and Vogus et al. 2014) further expanded on the notion of muffled warnings.

Additionally, Normal Accident Theory found, tight coupling between (sub)-systems quickly spread (false) alarms and warnings, which generated to noise at higher system levels, e.g., in the energy grid that the Three Mile Island facility was connected to (Perrow 1984). Tight coupling is defined as the reciprocal interdependence across different units and levels (Perrow 1984).

While Perrow's original analysis was mostly technical in nature, subsequent research has shown how organizations used buffering strategies to overcome barriers to decoupling, when external pressures from outside the organization's environment pressed for tight coupling (Sauder and Espeland 2009).

Conversely, loose coupling has been defined as the simultaneous presence of distinctiveness and responsiveness, i.e., "elements are responsive but retain evidence of separateness and identity" (Orton and Weick 1990, p. 203).

Further studies have elaborated how coupling is a mechanism to balance different approaches of managing uncertainty (Grote et al. 2008). The more technical recommendations made by Normal Accident Theory were to decouple and decentralize safety critical activities and substances, create redundancy and to reduce concentrations of organizational power that ignore regulation (Perrow 1981; Perrow 2004). Organizations minimize uncertainty through tighter coupling and cope with uncertainties through looser coupling. Organizations tighten or loose coupling through three activities: (1)

standardization vs. flexibility in routines and in the use of routines, e.g., centralized vs. localized planning, rule setting, and definition of procedures; (2) reduction vs. extension of operational autonomy of actors; and (3) avoiding vs. exploiting disturbances that change the system (Grote et al. 2008).

The limitation of Normal Accident Theory is its focus. Not all disasters are normal accidents, i.e., inevitable technological system failures. Perrow (2004) clarified the focus of Normal Accident Theory. He argued, for example, that the Challenger disaster was a management failure not a technical failure and therefore can't be explained with Normal Accident Theory. While the management system has been conceived as a subcomponent of the space shuttle system, Perrow (2004) argued that it was not a normal accident because the management system was neither tightly coupled nor had non-linear interactions with the technological system. In similar vein, Perrow argued that Bhopal and Chernobyl, two frequently discussed accidents, were caused by gross malfeasance, which means they were not normal accidents (Perrow 2004). This clarification highlighted the limitation of Normal Accident Theory: The theory's most important contribution was that it highlighted that failure is inevitable in interactively complex or tightly coupled systems. Moreover, it showed that decoupling and creating redundancies helps. It also demonstrated that high-reliability organizations do not rely on the rationale of economies of scale, to achieve safety. However, the theory's largest gap were non-technical explanations. Thus if an accident happens only an after-the-fact, causal analysis is able to establish whether or not it was a normal accident.

Even though Normal Accident Theory incorporated social aspects in explaining failures of technology systems, for instance, the theory predicted how accidents were amplified by the management system, the theory itself had little impact on organizational research (Perrow 2004). The theory gained more traction in the engineering studies, e.g., in discussions of the future of nuclear power (Ramana 2009), and safety studies, e.g., systems-theoretic accident model and processes (Leveson 2004; Leveson et al. 2009), or cognitive systems engineering (Hollnagel and Woods 1999). However, even in engineering studies the theory is not uncontested. Critics argued that the key recommendations derived from Normal Accident Theory, creation of redundancy and protection systems, were among the least cost effective solution to prevent accidents. Other solutions were preferable, e.g. replacing hazardous materials with safe ones, decoupling, and design for controllability, and reducing operator errors (Marais et al. 2004).

According to Weick (2004), Normal Accident Theory had its biggest value in understanding and unpacking the genesis of accidents at the macro-level of systems. Subsequently, the study of high reliability organizations developed a micro-level analysis and social perspective on Perrow's Normal Accident Theory (Weick 2004; Barton and Sutcliffe 2009; Weick and Sutcliffe 2007). This string of work added organizational analysis to the technological systems analysis (Sagan 2004; Leveson et al. 2009). High Reliability Organizations are defined as potentially hazardous organizations that have a track record of preventing failure and accidents (Roberts 1990). Hazardous but highly reliable organizations are characterized by the simultaneous occurrence of multiple

critical outcomes, hyper complexity, tight coupling, extreme hierarchical differentiation, large numbers of decision-makers in complex communication networks, on the one hand, and, on the other hand, by high degree of accountability, high frequency of immediate feedback about decisions, and compressed time (Roberts and Rousseau 1989). In sum, studies of High Reliability Organizations argued that safety is achieved through the design of organizational and human relationships not through the design of technical systems (Roberts 1989; Roberts and Rousseau 1989; Roberts 1990; Weick and Sutcliffe 2007).

Case studies of High Reliability Organizations acknowledged human limitations in complex socio-technological systems that worked at high levels of reliability and safety, e.g., in air traffic control, energy grids, and nuclear submarines (Turner 2005). The study of high reliability organizations has shown that successful prevention of disasters is linked to redirecting action, i.e., stopping a dysfunctional course of action that might be harmful to the organization. Picking up on weak signals that foretell a disaster or signify a dysfunctional course of action was, however, not sufficient for members of an organization to redirect their action. Redirection of action actually happened when members of an organization voiced concerns and actively sought alternative and disparate perspectives. These activities were moderated by self-interest and organizational pressures (Barton and Sutcliffe 2009). Other micro-level activities to establish High Reliability Organizations were managerial control and managing trade-offs between mutually incompatible demands of

the system, in particular, balancing the limits of human performance and retaining administrative flexibility (Turner and Pidgeon 1997). In sum, the studies of High Reliability Organizations demonstrated that markedly deviating performance outcomes were avoided because an inherently hazardous organization was able to intervene and correct a dysfunctional event unfolding.

A second line of inquiry further developed the Normal Accident Theory's finding that complex organizations muffled important warnings. The concept of organizational mindfulness explained how organizations were able to become high reliability organizations (Weick 2006, Weick and Sutcliffe 2006, 2007, and Vogus et al. 2014). Mindfulness was defined as an active state of information processing (Levinthal and Rerup 2006). To prevent disasters attention was directed towards small errors that build up over time, e.g., dysfunctional social processes, normalization of deviance (Vaughan 1996), and framing processes that erroneously categorize critical events as non-events (Weick 1988). Core enablers of mindfulness were located within the organizational culture and the leadership of the organization (Weick and Sutcliffe 2007).

The precursor to mindfulness studies was the case study of flight deck communications, which unpacked the interplay between mindfulness, then called heedfulness, and tight-loose coupling (Roberts 1994, Weick and Roberts 1993). The tighter the coupling of the social organization the more mindful the actors constructed and enacted their actions, understanding, and individual

relationship to the system. Drawing on the notion of undeveloped and developed groups Weick and Roberts argued that high reliability organizations are safe because of, and not in spite of, tight coupling (Weick and Roberts 1993, p. 377).

Nevertheless, Weick (2004) argued that high reliability organizations still produce normal accidents. If circumstances led to tighter coupling as in the Iraqi friendly fire incident (Snook 2002). The case showed that activities of high reliability organizations (e.g., specialization, training, culture, and creation of redundancies) loosened the coupling of systems. The account also showed that high reliability organizations were local adaptations, which if circumstances changed, i.e. the system components were re-joined more tightly, were prone to normal accidents (Weick 2004).

Critics have argued that the definition of high reliability organizations is tautological: an organization needs to be near error free to be qualified as a high reliability organization (Marais et al. 2004; Hopkins 2007). Thus studies of high reliability organizations that aimed to reduce errors suffered from survivor biases. Not only would regulators have shut down unsafe hazardous organizations but the organizations would also be excluded per definition from studies of High Reliability Organizations (Hopkins 2007; Marais et al. 2004). In other words, most studies of High Reliability Organizations are case studies where the researcher sampled on the dependent variable (Sagan 2004). Thus causal claims about accidents and safety might not hold true because samples

missed organizations that have followed high-reliability practices but were struck by disaster.

Furthermore, Marais et al. (2004) found fault with the common characterization of high reliability organizations as high complexity. For instance, Weick & Roberts (1993) described that operators had nearly full technical knowledge of the system and that aircraft carrier technology is stable because dominant design configurations hardly change. Both aspects are at odds with what Perrow (1981) describes as technically complex. Weick in turn (Weick 2009, pp. 20-21) has defended the characterization. He argued that high-reliability organizations have successfully reduced high complexity, which resulted in profound simplicity that should not be taken for the absence of complexity (Weick 2009).

Moreover, high reliability organizations have been characterized as carrying safety as their highest organizational goal (LaPorte and Consolini 1991; Weick and Sutcliffe 2007). But a common characteristic of complex organizations are their incongruent and conflicting goals. A case study of an aircraft carrier showed that the system put safety first in peacetime operations but once mission goals in a fighting role superseded the safety goal and accidents started to happen. The case also questioned the claim of long-term high reliability through designing a better management system (Marais et al. 2004).

A final critique of High Reliability Organizations came from the organizational learning perspective. High Reliability Organizations have been described as

using sophisticated learning processes, e.g., debriefs and lessons learned (Weick and Sutcliffe 2006). However, such trial-and-error learning, is suboptimal because learning requires costly accidents or near-misses to occur (Marais et al. 2004; Sagan 2004).

Open questions why outliers occur remain. Normal Accident Theory and the theory of High Reliability Organizations both focused the attention of researchers and managers on information difficulties, e.g., noise, weak signals. Neither Normal Accident Theory nor the study of high reliability organizations incorporated organizational politics, such as blame shifting after disasters (Pidgeon and O'Leary 2000). In a similar vein, Perrow (2004) noted that aspects of organizational power were absent from the analysis of high reliability organizations. Weick (2004) countered the argument by pointing towards the discussion of culture, which "shapes the way for power, defines power, is shaped by power, masks power, embodies power." (2004, p. 31). Overall, the study of high reliability organizations has added a social perspective to the technological perspective offered by Normal Accident Theory.

In sum, the account of normal accidents and systemic analyses of high-reliability organizations have contributed to the understanding of how designs of organizations and technologies contributed to accidents turning into disasters, which are markedly deviating outcomes of performance. Despite methodological issues the studies demonstrated that central to the social aspect of the design problem were mindful practices, which reduced the complexity of

organizational life, and design principles that tighten or loosen the coupling of organizations. The next sections will show how project management theory further expanded on the concept of complexity.

2.4.1.2 Project Management Theory: Project Complexity

The concept of project complexity resembles Perrow's (1981) system characteristics of interactive complexity and tight coupling. Stemming from the comparison of complicated and complex projects, complexity has been operationalized as a function of the number the actors in the system and the interdependencies between those actors (e.g., Whitty and Maylor 2009). It is thus similar to Simons's architecture of complexity (Simon 1976). Similar analyses of project complexity are found in the reviews by Baccarini (1996, 2004). Cicmil & Marshall (2005) argued that complexity is not only simply driven by interdependencies between actors but by the power relations that characterize those interdependencies.

A different perspective conceptualized project complexity as structural uncertainty (Williams 1999). In addition to the number of system elements, multiplicity, uncertainty, and ambiguity of goals characterized project complexity (Williams 1999). Similarly, Maylor et al. (2008) argued that project complexity was driven by structural and dynamic qualities of a project – particularly the qualities of mission, organization, delivery, stakeholder, and team.

In the same vein, complexity is often broken down into the two dimensions of technical and social complexity, e.g., in McFarlan (1981), Zmud (1980), Sauer et al. (2007), Jenkins et al. (1984), Wallace et al. (2004ab), Brooks (1975), Jiang and Klein (2000), Xia and Lee (2004). Bruijn and Leijten (2008) built upon this conceptualization of project complexity. Through a review of case studies they argued that complexity needs to be reframed. It is not complexity itself but how well complexity is managed which explains cost overruns in projects. The case studies unpacked the social and technical complexity dimensions. Bruijn and Leijten (2008) demonstrated through their case studies that *technical complexity* proved to be manageable when (a) the design was robust, (b) the technology was tame, (c) the project was divisible, (d) the deliverables of the project were loosely coupled, (e) the project had technological fallback options, (f) the solution was mono-functional, and (g) the implementation happened incrementally. *Social complexity* proved to be manageable when (a) the dependence on user preferences was limited, (b) preferences and aims of all parties involved in the project were uniform and stable, (c) third parties did not hold blocking power, (d) the transformation time was short, and (e) the project could influence its social environment. The cases also showed that if projects focused more on the management of technical than on social complexity uncertainty in the project increased (Bruijn and Leijten 2008, p. 37).

A second line of inquiry in the extant project management literature is the contrast between size and complexity. Project size and duration have been used as a proxy for project complexity (McFarlan 1981, Zmud 1980, Sauer et al. 2007,

Jenkins et al. 1984, Wallace et al. 2004, Brooks 1975, Jiang and Klein 2000).

Studies demonstrated that increasing project size decreased the likelihood of project success (Standish 1999; Jones 2000).

This led to the common recommendation to keep projects short and small (e.g., Schmidt et al. 2001; Fortune and White 2006). More specific were the analyses by Cooke-Davies (2002) and Sauer et al. (2007). Sauer et al. found plummeting success rates for projects longer than 18 month, and recommend to keep projects under 12 months (Sauer et al. 2007). Similarly, Cooke-Davies (2002) recommended a cut-off of 3 years, ideally less than 1 year. Project duration has been linked to increasing complexity because longer projects have been found to be more exposed to quickly changing business strategies (Schmidt et al. 2001) and volatility of scope and the stability of requirements (Sauer et al. 2007). Additionally, longer project durations increased the social complexity through turnover in the senior management environment (Feeny et al. 1992).

The complexity of larger projects²⁰ has been linked to the complexity of the product and the complexity of the decision-making process associated with the project (Priemus and van Wee 2013, p. 1). Larger projects presented decisions that were more costly, complex, and that were in operation for a longer time (Priemus and van Wee 2013, p. 1). The decision making process of larger projects was complicated by the erratic and non-linear nature of megaprojects

²⁰ Commonly referred to as large-scale projects (Miller and Lessard 2001), megaprojects (e.g., Merrow 2011, Flyvbjerg et al. 2003) or major projects (Morris and Hough 1987)

(Leijten 2013) which is thought to reduce the predictability of the consequences of decisions made and thus cause frequent deviations from plans.

In sum, project complexity has been the dominant lens to analyze the project as a system and explain why project performance (markedly) deviates from expectations. The extant literature has studied the dimensionality of project complexity, whereas surveys relied mostly on more abstract approximations of complexity such as project size and duration. The next section will focus on the discussion of system-centric explanations in the IT project management literature.

2.4.1.3 IT Project Management Theories: Estimation Theory

In a seminal book, Glass (1997) documented 16 cases of failed IT projects from the IRS system to the Denver Airport Luggage system. Glass coined the term 'Software Runaways' to describe his cases. Glass pointed out: "Most of the runaway projects are (or were) huge. It is well known in the field that huge projects are much more troublesome than smaller ones." (Glass 1997, p.17). Glass (1997) found in his analysis of case studies that a multiplicity of causes resulted in the project failure. He noted that technological issues as well as management issues play a role. Glass cites that in his analysis "technology was just as often the cause as management" (Glass 1997:17). Particular technological challenges originated from innovation or the newness of the solutions, and high performance requirements.

Similarly, most studies of IT project performance were a-theoretical. The exception are studies, which were grounded in estimation theory and aimed at identifying and correcting the causes of inaccurate estimates of software development efforts, e.g., Lederer and Prasad (1995), Bergerer and St. Arnaud (1992), Kemerer (1987), Heemstra (1992).

Lederer and Prasad (1995) investigated a list of 24 causes of inaccurate estimates²¹. The results of their survey showed that the six most important causes of inaccurate estimates are: frequent requests for changes by users; users' lack of understanding of their own requirements; overlooked tasks; insufficient user-analyst communication and understanding; poor or imprecise problem definition; insufficient analysis when developing estimate. The authors argued that these unexpected responses point towards an insufficient user involvement in the estimation process as a key cause of inaccuracies.

²¹ The list is fairly broad: inability to anticipate skills of project team members; overlooked tasks; lack of an adequate methodology or guidelines for estimating; lack of coordination of systems development, technical services, operations, data administration, etc.; functions during development; lack of setting and review of standard durations for use in estimating; lack of historical data regarding past estimates and actual performance; changes in Information Systems Department personnel; lack of project control comparing estimates and actual performance; lack of diligence by systems analysts and programmers; poor or imprecise problem definition; insufficient time for testing; inability to tell where past estimates failed; insufficient analysis when developing estimate; frequent requests for changes by users; insufficient user-analyst communication and understanding; users' lack of data processing understanding; users' lack of understanding of their own requirements; performance reviews don't consider whether estimates were met; red tape; removal of padding from the estimate by management; reduction of project scope or quality to stay within estimate, resulting in extra work later; pressures from managers, users, or others to increase or reduce the estimate; lack of participation in estimating by the systems analysts and programmers who ultimately develop the system; and lack of careful examination of the estimate by Information Systems Department management.

Lederer and Prasad (1995) further analyzed the correlation of these factors with the level of inaccuracies. The five highest correlations (all statistically significant with $p < 0.01$) were: performance reviews don't consider whether estimates were met; lack of setting and review of standard durations for use in estimating; lack of project control comparing estimates and actual performance; lack of careful examination of the estimate by management; reduction of project scope or quality to stay within estimate, resulting in extra work later. The first four factors pointed towards technical explanations, i.e., lack of good models and good data and calibration to improve models and data. The fifth factor indicated a deliberate tinkering with the project forecast to fit certain targets. This insight was further strengthened by factor analysis which extracted four factors: estimation methodology, politics, user communication, and management control. However, a multiple regression analysis showed that only the factor management control had a statistically significant influence on cost overrun. The authors noted that all factors correlated significantly with estimation inaccuracies (except user communication).

Kemerer (1987) reviewed model-based forecasts and found that forecast inaccuracies were 772% for SLIM²² models, 601% for COCOMO²³, 102% for

²² SLIM (Software Lifecycle Management) developed by Putnam (1978) is an effort estimation model that depended on four parameters effort, development time, elapsed time, and state of technology. Because elapsed time was an independent variable in time the model was the first to reflect that effort was not uniformly distributed along the duration of the project.

²³ COCOMO (Constructive Cost Model) is a method developed to translate size estimates into cost estimates with the use of 15 cost drivers (reliability, database size, software complexity, execution time, memory constraints, hardware volatility, response time, quality analysts, experience with application, quality programmers, hardware experience, programming language experience, modern programming techniques, software tools, project duration

function points²⁴, and 85% for ESTI-MACS²⁵, with many projects being severely underestimated. Collopy (2007) discussed the findings and offered underestimation of development complexity as the key issue which explained the inaccuracies in these estimation examples. Bergerer and St. Arnaud (1992) found that the accuracy of software estimates depends on both the forecast, which is influenced by the method chosen, as well as the execution against the forecast. Heemstra (1992) explained inaccuracies with the complexity of the software cost estimation task. In the same vein, Heemstra found that lack of good data and good models was a key issue. For example, software estimations often lacked a stable scope model and lacked the ability to scale up efforts because of non-linearity in scale. Phan et al. (1998) identified two main causes for cost overruns: "Overly optimistic estimations and changes being made too frequently for estimation to be accurate" (Phan et al. 1998, p. 97). Kulk et al. (2009 and Kulk 2009) followed conventional estimation theory with controlling for project category, line of business, and executive center as project classification factors. Additionally, the study checked for generic risk factors: IT

constraints). COCOMO differentiates between three classes of software development organic (stable, small, less innovative), embedded (tightly constrained, highly volatile and innovative), and semi-detached (between organic and embedded software developments).

²⁴ Function points were initially developed by IBM to estimate the size of a software project. It aimed to replace the lines of code (LOC or KLOC) estimation methods by quantifying the number of functions the software is supposed to provide. Function points depend on the type of software (input, output, inquiry, interface, internal file) and 14 adjustment factors (data communication, distributed functions, performance, heavily used configuration, transaction rate, online data entry, end user efficiency, online update, complex processing, reusability, installation ease, operational ease, multiple sites, facilitate change) the software size was estimated. The function point estimate is then translated into effort, budget, and time estimates based on historical or benchmark productivity estimates.

²⁵ Estimacs is a black box model, in form of a software package, developed by Computer Associates. The software supported function point estimation and then uses 25 questions on the organisational complexity and the software complexity to estimate the effort required.

maturity (CMM), reporting unit size, project management tool, reporting quality of financial information, size of the development department, and the internal development staff size; as well as project specific risk factors, which were estimated cost, duration, project power, in-house/outsourced, and project category. A subsample of 79 projects provided data sufficiently homogeneous to model the impact of the factors on inaccuracies of estimates. Only two of the factors showed significant levels of association: percentage of internal development staff and percentage of development staff overall. Importantly, CMM did not turn out to be a significant factor in the sample studied.

In sum, the literature on IT project management that is grounded in estimation theory explained (markedly) deviations of actual from expected software development efforts through differences in forecasting methods, lack of good data and good models, social and psychological aspects of forecasting (organizational politics, user involvement, management control, staffing levels, optimism of forecasters), and project complexity (innovation and scale).

The next section will review the extant organizational, project management, and IT project management literature from the event-centric school of thought.

2.4.2 Event-Centric View

The event-centric view explains outliers as a consequence of (1) rare events, (2) surprises or (3) crises that happen upon an organization (e.g., March 1991, Christianson et al. 2008, Lampel et al. 2009, Rerup 2009, Starbuck 2009). The

four phases of a rare event illustrate the conceptualization (Roberts et al. 1994, Fink 1986):

- Pre-event phase/prodromal stage - the focus of the organization lies with preventive and planning activities
- Emergency phase/acute stage - the event hits the organization and the effects of the event are felt, organizations focus on rescue activities
- Intermediate phase/chronic stage - organizations focus on restoring normality
- Long-term phase/resolution stage - organizations focus on recovery and restoration efforts that take longer.

First, Christianson et al. (2008) defined *rare events* as events that occur outside the everyday experience of organization. Lampel et al. (2009) defined rare events as departures from the ordinary experience. Rare events presented themselves as interruptions that place novel demands on an organization (Barton and Sutcliffe 2009), which acted as an audit to an organization's existing responses repertoire and underlying organizational capabilities (Christianson et al. 2008). In other words, the occurrence of a rare event interrupts the organization and an insufficient response to the rare event then caused the outlier.

Second, based on the historically rich notion of military surprise, e.g., Trojan horse, Pearl Harbor, Yom Kippur war, events that happen upon organizations have been depicted as *strategic surprises*, which is an unanticipated strategic discontinuity (Ansoff 1975). Others defined surprises as events that had not been previously thought about or that had not been seriously considered (Eppel

2012, p. 882). In a similar vein, Watkins and Bazerman (2003) defined predictable surprises as disastrous events that organizations should have anticipated and prepared for. The causes of surprises were found in failures in planning. Conventional strategic planning often underestimated variability over time by simply extrapolating smoothly from the past (Ansoff 1975). In other cases surprises could have been forecasted if the organization had correctly identified and processed weak signals (Ansoff 1975). In contrast, Turner and Pidgeon (1997) dismissed the notion of surprises as a substantive conceptualization of rare events and disasters. Their key critique was that surprises were emotional responses to the cause not the cause of outliers in its own right (Turner and Pidgeon 1997). By focusing on the affect, the concept of surprise downplayed cognitive and behavioral responses. However, compared to rare events, defined as interruptions of operations, the concept of surprises introduced the aspect of strategic planning to the explanation of why outliers occurred.

Third, conventionally a *crisis* has been defined as an out-of-the-ordinary situation (Roux-Dufort 2007, p. 105). Traditionally major events such as natural disasters or man-made disasters, e.g., malicious rumors, plant explosions, and product recalls, have been portrayed as crises (Roux-Dufort 2007). Others define crises as an episode of a threat under uncertainty, which requires urgent action (Boin and 't Hart 2003) and as "low-probability, high-impact event that threatens the viability of the organization and is characterized by ambiguity of cause, effect, and means of resolution, as well as by a belief that decisions must

be made swiftly." (Pearson and Clair 1998, p. 60). Crises are also broadly defined as negative events (James et al. 2011, p. 458) or deviant events that disrupt an organization (James et al. 2011, p. 460). Crises were characterized by posing a significant threat, short decision time, and surprise (James et al. 2011, p. 460). In more general terms, crises were an abrupt change or shift from a normal situation to a crisis situation (Roux-Dufort 2007, p. 107). In a similar vein, and often interchangeably, the term disaster is used, however, with a slight difference. Typically disasters and catastrophes are defined quantitatively by their impact, e.g., an event that caused more than 20 deaths, or probability, e.g., a 100-year flood (Madsen 2009). Compared to the concepts of rare events and surprises, the concept of crises puts more emphasis on the high-impact of a force majeure event and the activities of recovery and restoration.

The concepts of rare events, surprises, and crises also showed considerable overlap, which is why this literature review combined these three perspectives under the event-centric view. In sum, the definitions of key concepts developed in the event-centric view emphasized that a special negative event occurs creating the need for a response by the organization.

The event-centric school of thought has explained the failure of an organization to adequately respond to a rare event through: classical cognition, sense-making, framing and decision-making, as well as organizational learning from

and through events. But to date, these theoretical explanations²⁶ have rarely been used to explain outliers in project performance. One exception, Miller and Lessard (2000), studied the role of rare events, which they called turbulences.

They found that:

“Projects experience difficulties, however, not so much because engineers cannot cope with technical complications or external effects, but because sponsors cannot rise to the managerial challenges of coping with unforeseen turbulence. [...] The longer the development time, the higher the likelihood that the project will be affected by emergent events. Turbulence can originate from two sources: exogenous events, occurring outside the control of management, and endogenous events, arising from within the project organization. [...] Turbulence, both exogenous and endogenous, is negatively linked to project performance.” (Miller and Lessard 2000, p. 22-23).

Similarly, Mallak et al. (1997) found that crises are inevitable in projects. Miller and Lessard observed that “Few projects sailed through without meeting turbulence: on average, projects met close to five unexpected events [...] some had up to twelve rocky events.” (Miller and Lessard 2000, p. 22).

Pich et al. (2002) have argued that all projects expect the occurrence and impact of uncertain events, which is at the core of any risk management activity.

Hallgren and Wilson (2008) observed that events in the form of mundane interruptions were dealt with within the existing project system. More disruptive events were dealt with through the formation of task forces.

However, the actual work of the task forces showed too few commonalities to

²⁶ For reasons of space and because project management views rare events as typical in projects the full discussion of organizational theories can be found in Appendix 8.1

allow for greater theory building (Hallgren and Willson 2008). Further studies documented that crises in projects offered opportunities of individual learning (Hallgren and Willson 2011) as well as organizational learning (Kazanjian et al. 2000). Overall, because rare events and crises were expected in projects and managed through routine risk management activities – these studies suggest that ineffective risk management and ineffective project sponsorship contribute to the occurrence of outliers, that is markedly deviating project performance.

In sum, explaining performance outcomes as a consequence of rare events has not been a strong tradition in the project management literature. The occurrence and impact of rare events has been demonstrated as being inevitable and has not drawn dedicated research attention. It has only been linked to shortcomings in a project's risk management and sponsor. The next section is going to review the process-centric school of thought that explains the antecedents of markedly deviating organizational performance.

2.4.3 Process-Centric School of Thought

Three main theories have been discussed in this school of thought (1) the sociological account of man-made disasters, (2) the cultural production that leads to the normalization of deviance, and (3) the escalation of commitment to a failing course of action in the face of negative feedback. This section now will review each of the theories in turn.

2.4.3.1 *Man-Made Disasters*

The process centric view was first suggested in the analysis of man-made disasters (Turner and Pidgeon 1997). Man-made disasters were defined as unintended, unpremeditated or unforeseen events, which move an unsafe, erratically behaving system beyond or out of control of its managers (Turner 2005, p. 215). In contrast to the technological perspective of the system-centric school of thought, man-made disasters were social in origin and unfold over time.

First, the theory contrasted accidents (technical) and disasters (social). Disasters "pose a significant disruption or collapse of the existing cultural beliefs and norms about hazards [causing accidents]" (Pidgeon 1997, p. 2). Turner (2005) cited as support for this hypothesis investigations of engineering disasters, such as collapses of bridges, which found that the underlying technical fault was known to the engineers but the actual failure was that of the social system to act upon this knowledge.

Second, disasters unfold over time, which has been conceptualized as the incubation period (Turner and Pidgeon 1997). The analysis of 84 accident reports demonstrated that, over long periods of time, events accumulated that were at odds with existing beliefs (Turner and Pidgeon 1997, p. 86). A chain of events, not a singular event, led to the disaster. Each of the events in the chain went unnoticed, which means that outliers were determined not by a single failure but by multiple failures in the system (Turner and Pidgeon 1997).

During the incubation period a multiplicity of minor causes, misperception, misunderstandings, and miscommunications accumulated (Turner and Pidgeon 1997). Misperceptions and misunderstandings arose from difficulties of handling complex sets of information and cultural lag which made individuals accept minor issues and or made them actively hide issues to prevent blame (Turner and Pidgeon 1997). Another cause was miscommunication. Man-made disasters were failures in communication because of cultural, political, and social pressures which prevented a redistribution of information, which could have ended the incubation period and prevented disaster (Turner and Pidgeon 1997).

In sum, Turner's theory is a "cultural process of defective reality testing and denial" (Pidgeon 1997, p. 3). The underlying causes of this cultural process centered on technological and administrative control (Turner 2005). Disasters incubated over a long period of time during which a multiplicity of minor causes, misperception of weak signals, misunderstandings, and miscommunications accumulated.

2.4.3.2 Normalization of Deviance

Vaughan (1996) re-framed the incubation period as a process of normalization of technical deviance in her analysis of the Challenger disaster. Similar to the sociological aspect of the incubation process, the normalization of deviance was a production of culture as a result of how the organization managed technology risks. Vaughan (1996, p. 154) identified five steps that explain how the

normalization of deviance slowly and incrementally enlarged the bounds of acceptable risks:

- (1) Signals of potential danger were noted by members of the organization.
- (2) Potential danger was acknowledged and triggered a risk assessment.
- (3) Risk assessment identified potential causes of the danger; mitigating actions were taken as a result of an investigation.
- (4) Risk is declared acceptable.
- (5) A successful flight evidenced to the organization that the risk was now under control.

The conventional understanding of the Challenger disaster at the time of the study was that the disaster was caused by a failure in engineering analysis, communication failure, and group think (United States Presidential Commission on the Space Shuttle Challenger Accident 1986). Vaughan's analysis disagreed with those findings. The case study demonstrated that the active production of culture led to a dysfunctional course of action.

Additionally, the analysis pointed to the importance of prior policy decisions, which resulted in a trickle-down of top management decisions via the organizational structure and organizational culture. Similarly, Vaughan (1997) showed the trickle-down of strategic change mandated by the top management. Her analysis found that budget pressures at NASA reoriented the strategy from exploration and research to operating a commercial space flight program. As a result of those pressures the shuttle program was declared operational. A

decision, which trickled down the organizational structure and changed the organizational culture so that decisions were re-framed as cost-benefit trade-off, which hid the engineering risks (Vaughan 1997). Vaughn also found a conflict between the engineering culture, on the one hand, which actively produced concerns and, on the other hand, the administrative culture, which produced bureaucratic accountability. In sum, the complex organizational culture, shaped by trickle-down policy decisions better explained the Challenger disaster than prevailing explanations (Vaughan 1996, Vaughan 1997).

The 2003 Columbia disaster led researchers to revisit Vaughan's (1996) findings. The analysis found that normalization of deviance also played a role in the case of the Columbia disaster: "the more frequently we see events, conditions, and limitations, the more we think of them as normal and simply accept them as a fact of life" (Starbuck and Farjoun 2005, p. xviii). Studies further found that organizational cultures had contributed to the accident. In particular culture manifested two-fold through (1) the language and (2) the role of time. First, Ocasio (2012) showed how the vocabulary of safety drifted prior to the Challenger disaster. A linguistic analysis showed that safety was ambiguously defined and seen in the organizational culture as an engineering constraint and not as an organizational goal (Ocasio 2012). Second, analyses found that the actual organizational goals were meeting tightly constraint and overambitious deadlines. Pressures to meet the narrow Columbia launch windows made time the central focus of the organization, leading to stress and constantly increasing pressures to meet deadlines. In terms of organizational goals time and safety

required trade-off decisions. The shift in the language of safety, tight time constraints, and increasing importance of meeting deadlines biased how the organization processed information in a way that safety risks were underestimated and schedule risks were overestimated (Starbuck and Farjoun 2005).

The process-centric view of disaster has also embraced arguments from the system-centric view. For instance, Turner & Pidgeon (1997), outlined that a key to negentropic processes in high-reliability organizations was organizational learning, particularly creating the opportunity to learn from early mistakes. Baumard & Starbuck (2005) analysis of organizational learning questioned this argument though. They found that small cases of failure are easily dismissed by organizations: small failures are labelled experiments and become ideological playgrounds where in the end corporate core beliefs are reinforced instead of being challenged. Their observation supports the threat-rigidity hypothesis (Staw et al. 1981): Under threat organizations resorted to rigidity, i.e. followed a learned and dominant response. The more intense the threat the higher the restriction of information and constriction of control that caused the rigidity (Staw et al. 1981). Baumard and Starbuck (2005) showed not only a rigid response and a tendency towards groupthink²⁷ but also an incremental change in the underlying core beliefs. Learning was most prohibited, when failure challenged core beliefs, such episodes were quickly dismissed as unique or

²⁷ Under threat groups reduced flexibility, sealed off new information, and controlled deviant responses.

idiosyncratic. Additionally, managers of small failures moved to other parts of the organization where their individual learning was irrelevant. Managers who questioned core beliefs in post-mortem analysis of failures were demoted (Baumard and Starbuck 2005, p. 209). In sum, they argued, that large failures led to less learning by individuals and the organization than smaller failures.

In sum, the process-centric view argued large failures accumulate over time. Vaughn (1997) and others (Turner and Pidgeon 1997), have shown that smaller failures did not go unnoticed, they were acknowledged by individuals yet they were not salient to the organization: either because issues were actively hidden or because issues were seen as solved or because issues are seen as mere complications at the time. In nearly all cases the issues, and their accumulation over time, were only revealed to the organization for what they were ex-post (Vaughan 1997; Baumard and Starbuck 2005, p. 294). Outliers are caused by conflicting goals and cultures. Threats to goals, in particular, led to rigid responses that inhibited learning and continually enlarge the level of accepted risk by the organization.

2.4.3.3 Escalation of Commitment

The theory of the escalation of commitment explains why organizations persevere with a failing course of action in the face of negative feedback (Staw 1981). This conceptualization of the process emphasizes the existence of negative feedback (Staw 1981). Thus escalation of commitment differs from man-made disaster and normalization of deviance, because here decision

makers were aware of negative information, for instance in Vaughan's concept of normalization of deviance, the increasing risk was not salient to the decision-makers²⁸. Escalation of commitment has been demonstrated in diverse fields such as finance, marketing, accounting, and information systems (Sleesman et al. 2012). Individual cases studied under this perspective include projects such as the Shoreham Nuclear Power Plant (Staw and Ross 1986), the failed IT project at the London Stock Exchange (Drummond 1997), or the Denver Airport Luggage System (Montealegre and Keil 1998).

Escalation of commitment has been explained as a result of biased information acquisition and processing (Staw and Ross 1986). For instance, Baumard and Starbuck (2005) observed incomplete issue reporting, where managers of multiple failed projects in a telecoms company actively omitted issues that questioned the rationale for the venture and issues they expected to diminish in the long run. Additionally, recipients of issue reports heavily discounted troubles (Baumard and Starbuck 2005, p. 293). The information that got reported included routine issues, e.g., lack of coordination and efficiency problems. Difficult to explain issues were left out of reports (Baumard and Starbuck 2005). Contrarily, Drummond (1997) found no evidence for biased information processing. The top executives at the London Stock Exchange not

²⁸ Consequently, escalation of commitment questions the efficacy of the concept of mindfulness in high reliability organizations, because escalation of commitment exists in the face of feedback about failure.

only receive troublesome news they also acknowledged and highlighted them in memos they received (Drummond 1997).

Other explanations of why escalation of commitment occurs are less contested (Keil et al. 2000; Sleesman et al. 2012):

- Self-justification theory, where a failing course of action is followed to justify the initial decision to embark on this course (Staw 1981)²⁹.
- Prospect theory, where a failing course of action is framed as a potential loss which induces risk seeking behavior (Kahneman and Tversky 1979, Tversky and Kahneman 1992)³⁰.
- Goal substitution effect, which explains how a goal such as economic profitability is replaced by a different goal such as quick completion (Garland and Conlon 1998).
- Principal agent theory, which predicts that agent have an incentive to stay on a failing course of action that is undesirable for the principal (Booth and Schulz 2004, Jensen and Meckling 1976)³¹.

²⁹ Self-justification theory builds on the need for consistency (McGuire 1960) and the theory of cognitive dissonance (Festinger 1962), which explains that decision makers persist with a course of action in order to be self-consistent with prior attitudes, judgements, and decisions. Self-justification is psychological and social (Whyte 1986). Self-justification as an explanation of escalation of commitment has been suggested by Staw and Ross (1987) and Brockner (1992).

³⁰ Prospect theory (Kahneman and Tversky 1979) and cumulative prospect theory (Tversky and Kahneman 1992) explain the violation of expected utility theory (Neuman and Morgenstern 1947) observed in decision under uncertainty. Prospect theory predicts that subjective probabilities are non-linear, which created a concave value function for gains and a convex function for losses. Thus differences in framing predict different risk appetites. If a prospect offers a gain decision makers are risk averse. If the prospect presents itself as a loss decision makers are risk seeking. In sum, prospect theory predicts that escalated projects are perceived as a loss which is why decision makers increasingly seek risks.

³¹ Agency theory (Jensen and Meckling 1976) describes situations when a principle instructed an agent to perform a service. Due to goal incongruence and information asymmetries a principle-agent relationship can result in unexpected outcomes.

- Approach avoidance theory, which leads to a goal completion effect (Keil and Truex 1995)³².
- Self-presentation theory, which leads to saving face as well as conformity to group pressures (Sleesman et al. 2012).

Further antecedents that were initially considered, but have not been studied widely, are:

- Whether problems are perceived as temporary,
- Whether error investigation is perceived as effective,
- Size of sunk costs and estimated costs for completion, and
- History of previous failures (Staw and Ross 1986).

A meta-analytic review of 166 empirical studies established support for each of the theoretical explanations, except self-presentation theory (Sleesman et al. 2012). Recently the view of escalation has been applied to other dysfunctions in organization outside decision-making such as the escalation of deception (Fleming and Zyglidopoulos 2008).

Furthermore, the extant literature presented evidence that antecedents leading to escalation of commitment even predated the execution of chosen course of action. In other words, factors contributing to the escalation of commitment can be found in the planning stage of projects (Merrow et al. 1981). For instance, escalation has been linked to the decision logic behind investments. Cooper et al. (1988) documented a company, which evaluated new technological ventures

³² Approach avoidance theory (Rubin and Brockner 1975) explains why decision makers fail to walk away from a failing course of action. Three factors contribute to decision makers' avoidance of abandonment: size of the goal, cost of withdrawal, and proximity of the goal.

through worst-case scenarios. However, the chosen worst cases were consistently over-optimistic (Cooper et al. 1988). Rogers and Shoemaker (1971) found that tools used to evaluate and plan new ventures, such as discounted cash flows and net present value evaluations, are prone to framing effects, overconfidence, and preference for confirming evidence. Dosi and Lovallo (1997) concluded that a company's assessment of its own technological and competitive abilities is likely biased, because of illusions of control, overconfidence, and planning with the inside view. Baumard and Starbuck (2005) found that higher expectations for a project are directly linked to a lower likelihood of managers questioning the fundamental assumptions behind alternatives even before embarking on a failing course of action (Baumard and Starbuck 2005, p. 293).

2.4.3.4 Process-Centric School of Thought in Project Management

Of the three process theories found in the process-centric school of thought only the escalation of commitment is discussed in the extant project management literature. Keil (1995) was the first study to empirically test whether the escalation of commitment occurred in IT projects. Consistent with general theory, the study found that social, organizational, project, and psychological factors are associated with escalation in IT projects. Additionally, Keil (1995) observed emotional attachment to the project, empire building, and loose management controls as contributing to project escalation. In a follow-up study Keil et al. (2000) unpacked the causes of the escalation of commitment and tested the influence of self-justification, prospect, approach avoidance, and

principal agent theory. The study found that constructs from all four theories statistically significantly explained project escalation, whereas proximity of the finish (completion effect) had the most accurate predictive power of identifying escalated projects.

Among the case studies of escalation of commitment in projects Drummond (1997) had the broadest theoretical base. The case study documented the technology-led change project at the heart of the 1986 Big Bang, which deregulated the London capital market. Importantly, Drummond reported also non-findings, where the case study failed to support explanations suggested by prior research. Escalation of commitment had been explained through the argument of complexity vs. complication. Prior studies argued that decisions were complex and not complicated if decisions lacked objective verification and decision criteria. However, Drummond failed to find evidence supporting the claim that lack of objective verification of decisions led to the failure of the Taurus project. In other instances the case study contradicted existing theory. Staw and Ross (1989) argued that biased information processing and acquisition might explain escalation of commitment. Biased information processing and commitment was thought to be driven by the need for consistency which leads to confirmation bias. According to theories of how information is processed, contradicting information is first considered but then discounted in its validity. Drummond, however, found evidence in minutes and briefings that disconfirming information was considered and debated in an unbiased way. Similarly, the case study failed to find evidence for paucity of

decisions, incremental drift (decision-less decisions), irrational behavior of decision makers, group think, and sunk cost effects.

Drummond found evidence for several theoretical explanations linked to the social psychology of decision making. The case study demonstrated ignorance of information. Ignorance was partly innocent and due to a lack of managerial attention to the project as a whole. In some cases the ignorance manifested through differences in attention to supporting information. Drummond also found evidence that ignorance was deliberately created. In particular project supporters and decision makers raised the cost of dissent. Equally, group think (Janis 1973) was at play. Drummond demonstrated that decision meetings suffered from exhaustion of attendees, false consensus, and a tendency to save the cozy atmosphere in which the capital markets operated. Lastly, the case study found evidence in line with prospect theory (Kahneman and Tversky 1979) namely that actors on the project consistently preferred large and uncertain future losses over small but certain immediate losses.

In sum, the Taurus case concluded that many decisions were based on organizational myths. Drummond argued that myths gained and lost legitimacy and thus acceptance in the organization over time. A snapshot in time is thus characterized by the dominant viewpoint of an organization. A further symptom of how the decision making process contributed to the escalation of Taurus was the particular use of rhetoric. The rationale behind the project subtly shifted from the project being important to the project being vital

- instead of aiming to achieve beneficial means for the organizations involved, continuing the project became a mean in its own right.

Second, Drummond offered structural explanations for the failure of Taurus.

The explanation builds on decision dilemma theory. Escalation of commitment was caused by structural unaccountability and actors maximizing individual pay-offs, which led to locally rational but systemically sub-optimal decision outcomes, in other words creating a tragedy of commons (Hardin 1968).

Similarly, Drummond observed that actors tried to solve problems by delegating them away. This led to a weakening of the project control structures. While complexity of the structure of responsibilities increased, the effectiveness of formal mechanisms of communication decreased, further weakening control over the project.

Decision dilemma invokes the idea that highly complex decisions contributed to the escalation of Taurus. On the one hand, the decision complexity was technical, e.g., many technological aspects of the front-to-back integration of the process to trade, price, and settle capital markets products were highly innovative. Thus decision makers operated near the boundaries of knowledge. Decision complexity also manifested through unforeseen consequences of actions taken and decisions made. Moreover, the organization lacked processing capabilities to make adequately informed decisions, e.g., it lacked analysts and subject matter experts (Drummond 1996).

On the other hand, the decision complexity was characterized by social factors. Decision complexity increased because power differentials suppressed negative feedback. Additionally, organizational conflicts and incongruence between powers and responsibilities increased the complexity of decision in the project.

In sum, Drummond (1996) argued that the escalation of the Taurus project could be best understood as a tragedy of the commons. On the surface the cause of why the project escalated were the complexity of the project, that decision makers operated on the boundary of available knowledge, and that the organizational structure failed to adequately communicate and control problem solving and decision making. At the root cause of these problems was that the project was built on a contradiction and that macro-level influences prevented the project from being stopped, because it was being built for the great and the good.

Montealegre and Keil (2000) took a slightly different approach to theorizing escalation. They focused on a process theory of de-escalation³³. Based on process theories of escalation they hypothesized that de-escalation, in other words abandoning the failing project, requires changes in top management, public announcements of limits of escalation ('the buck stops here') and minimum performance targets for the solution, availability of technical

³³ As noted by Conlon (1999), the Taurus case study (Drummond 1996) documented for the first time the power and inertia of institutional factors. Drummond reported that the incoming CEO of the London Stock Exchange wanted to stop the project but was not able to. It took more than a year, including an expert review of the project and changes in public communication to slowly shift the dominant viewpoint of the organisation to be able to stop the project.

alternatives, allowing the possibility of failure, separation of project evaluation and initiation, appeals to stakeholders, external pressures on the organization, unambiguous negative feedback, visibility of project cost, and efforts to de-institutionalize the project. The key finding, similar to Drummond (1996), was that the organization persisted for much longer than feedback had accumulated - pointing towards a high level of inertia and to a process of abandonment rather than a singular decision. In the Denver Airport case, the process of de-escalation began with a problem recognition after which an external review of the project was ordered. Then followed a phase of re-examination of the prior course of action, followed by a phase in which alternatives were sought out. This happened through the set-up of a task force that scouted alternatives and then got the green light to begin the construction of an alternative system. Lastly, the exit strategy was implemented which was finalized by adapting the contractual arrangements of the project (Montealegre and Keil 2000).

In the first phase - problem recognition - two key activities took place the recognition of negative feedback and a response to external pressure. In the following phase, of the re-examination of the prior course of action - the magnitude of the problem was clarified and the problem was redefined. The next phase - search for an alternative - involved identification and legitimization of alternative solutions as well as impression management. The final phase - implementing the exit strategy - involved appealing to the project stakeholder and de-institutionalizing the project. The latter was achieved through shifting the dominant viewpoint of the organization away from the

system being defining for the new airport (Montealegre and Keil 2000).

Similarly, Natovich (2003) used the theory of escalation of commitment to explain the escalation of commitment through the break-down of trust between the project organization and the outsourcer delivering the project de-escalation of project failure. Likewise, Yourdon (1996) discussed famous cases of the 'death march project'³⁴. His explanations pointed to a combination of optimism of team members, unexpected events and crises, and organizational politics, e.g. of overpromising towards the clients which itself was a result of competitive pressures.

In sum, the extant literature in the process-centric view focusses mostly on the escalation of projects to explain why outliers, i.e., observations of markedly deviant performance occurred. The next section will review a closely related field, that explained not extreme but average performance through the lens of success and failure factors in IT projects.

2.4.3.5 Success and Risk Factors of IT Projects

Success factors are defined as areas where results ensure successful competitive performance of an organization (Daniels 1961, Rockart 1979). Rubin and Seeling (1967) first applied the concept of success factors to project management. In project management theory, success factors were defined as factors that

³⁴ Which he defined as being at least 50% different from comparable projects (e.g., working under a compressed schedule of only 50% of comparable projects, with 50% less staff, 50% reduced budget) and having a 50% chance of failure.

improved the chance of successful project implementation (Pinto and Slevin 1987).

Pinto and Slevin (1987) identified 10 success factors through analyzing research and development projects: clearly defined project goals, top management support, detailed project planning, user involvement, suitable personnel, sufficient technical knowledge, client acceptance, monitoring and feedback, communication, and trouble-shooting. These factors were grouped in strategic and tactical factors. In subsequent analysis Pinto and Prescott (1988) found that the importance of success factors varied throughout the project life cycle and according to the success measure used. If external measures of success were employed strategic factors have a consistently higher importance than tactical factors throughout the project life cycle. Moreover, the factors communication, monitoring and feedback, and personnel held little association to project success in the longitudinal view (Pinto and Prescott 1988). The findings are at odds with the explanation suggested by the escalation literature.

Cooke-Davies (2002) critically tested Pinto and Slevin's work and found that schedule performance was influenced by risk management, project governance, and duration of the project. Cost performance was associated to scope control and performance measurement. Further success factors were learning from experience, project status communication, strategy alignment, and benefits management.

Two review articles summarized the more than 40 years of research in this field: Belassi and Tukel (1996) and Fortune and White (2004). Belassi and Tukel (1996) identified 7 conceptual studies and 9 empirical studies. In their analysis Belassi and Tukel (1996) grouped the success factors into subsystems of the project: those related to the project, related to the project manager and project team, related to the organization, and related to the environment. Success factors often take a system-centric view of explaining project performance, but Belassi and Tukel argued that the dynamics are more important to explain performance, for example, lack of top management support combined with inexperienced leadership might lead to failure, whereas individually these factors would not matter (Belassi and Tukel 1996).

Fortune and White (2004) reviewed 64 publications and again used a systems decomposition approach to draw attention to the interdependencies between success factors. They collated a list of success factors ranked by the success factor's count of studies: senior management support, clear realistic objectives, detailed up-to-date plans, good communication, user involvement, sufficiently qualified team, effective change management, competent project manager, strong business case, sufficient resources, good leadership, proven technology, realistic schedule, risk management, project sponsorship, effective monitoring and control, adequate budget, organizational culture, good performance by suppliers, project reviews, training, political stability, project management methodology, environmental influences, learning from past experiences, project size/complexity/duration.

Boehm (1991) followed the success factor school of thought but focused on identifying risk factors. He outlined the top 10 risks for IT projects³⁵ and identified the cost drivers behind these risks: personnel, requirements, reusability, and tools and environment.

In a similar vein, Addison and Vallabh (2002) identified key risks to IT projects³⁶. The analysis found that control, without further unpacking the related theory, mitigated seven of the ten identified risk factors. Belout and Gauvreau (2004) based their study on the resource-based view of the firm. They found that a key reason for project performance shortfalls is that they "have been managed as technical systems instead behavioral systems." (Belout and Gauvreau 2004, p. 2). Sauer and Cuthbertson (2003) cited as risk factors of cost overruns and schedule delays: lack of management commitment, confusion over the objectives, and a lack of user commitment. Further factors that were associated with the likelihood of project failure were complexity, operationalized as the links with other technology and business processes, and volatility within the organization, e.g., changing objectives, personnel turnover. The study was further unpacked in Sauer et al. (2007), which tested the effect of

³⁵ Personnel shortfalls, unrealistic schedules and budgets, developing the wrong functions and properties, developing the wrong user interface, gold-plating, continuing stream of requirements changes, shortfalls in externally furnished components, shortfalls in externally performed tasks, real-time performance shortfalls, and straining computer-science capabilities

³⁶ Unclear or misunderstood scope and objectives, unrealistic schedules and budgets, lack of senior commitment, failure to gain user involvement, inadequate knowledge/skills, lack of effective project management methodology, misunderstanding requirements, gold-plating, continuous requirement changes, developing the wrong software functions, subcontracting, resource usage and performance, introduction of new technology, and failure to manage end user expectations

size³⁷ and volatility³⁸ on project risk. The results showed that size (in particular effort in PD) and volatility (in particular changes in project management) increase the risk of IT project failure.

In sum, the literature on success and risk factors of IT projects strengthened the point that common deviations from expected IT project performance are associated with success and risk factors of project execution. In particular, social factors (e.g., staffing, management control, and organizational politics) and project complexity factors (e.g., size, duration) have been problematized in organizational and project management literatures. The next section will review the literature on IT project abandonment.

2.4.3.6 Factors Leading to IT Project Abandonment

Missed deadlines, exceeded budgets, dropped functionalities, reduced scope ought to be considered normal features of IT projects abandonment offered conservative definition of failure (Ewusi-Mensah and Przasnyski 1991, Sauer 1993). The review on project management studies of de-escalation of commitment already showed that project abandonment³⁹ is difficult and a long process.

³⁷ Operationalised through effort, duration, and team size

³⁸ Operationalised through governance volatility (changes in the project manager or sponsor) and target volatility (changes in schedule, budget, scope)

³⁹ It has been suspected that project abandonment is particularly common in IT projects. However, no surveys exist about the extent of the issue. Harris et al. (2003) reported that 1% of private finance infrastructure projects were abandoned mostly shortly after financing was closed and before actual construction started.

Ewusi-Mensah and Przasnyski (1991) hypothesized that three groups of reasons lead to abandonment: economic, organizational, and technological. Economic reasons implied that cost overruns and schedule delays have made the project economically unviable. Organizational reasons, invoked contingency theories, namely the fact that abandoned projects overlooked behavioral, political, and other organizational concerns. Technological reasons were mostly the risk of the developed technology being leap-frogged. In their analysis organizational reasons emerged as the most dominant in explaining abandonment decisions.

In an effort to replicate the findings, Oz and Sosik (2000) extracted five factors from 30 items elicited in a survey of reasons of IT project abandonment. The extracted factors were poorly communicated goals/deliverables, lack of corporate leadership, inadequate skills and means, poor project management, and deviation from timetable/budget. The authors used a structural equation model to test the relationships between factors. Of the five factors only inadequate skills and means showed a statistically significant association with deviations from timetable and budgets. Oz and Sosnik found that:

"Deviations from budget and timetable are often the major, or the only, reason for cancelling an IS development project. Our analysis could not support the hypothesis that such deviations from budget and timetable are related to poorly communicated goals and deliverables. In the long run, corporate leadership may foster an atmosphere in which only people equipped with the proper skills are put on a project team, and the team enjoys adequate hardware and software." (Oz and Sosnik 2000, p. 74).

Similarly, Sauer (1993) presented the Mandata case study, and five shorter cases as validation, to explain the failure of IT projects. A central argument of his analysis was that the processes through which an IT project is delivered are situated in a way that it was impossible for the organization to prevent project failure. The case analysis found that projects were abandoned only after they lost organizational support. Therefore the root causes for abandonment were found in the relationship between the IT project, the project supporters, and the IT system itself. These relationships were located in the context of exogenous constraints and contingencies, as well as endogenous constraints of the system and the technology. Those constraints were met with problem-solving activities. Problem-solving was flawed, which led to sub-optimal decisions or decision deferral. Project abandonment was ultimately a consequence of the interaction of flaws in solving problems and loss of internal and external support for the project. In short, the analysis made the important point that technological failure alone did not lead to project abandonment. The analysis also showed that support was not stable throughout the project. Finally, the analysis showed that some failures were inevitable due to project constraints. The Mandata case inevitably failed because of tightly constrained support, i.e., the project was only supported under the exact operating conditions without any options or room to solve problems (Sauer 1993).

Pan and Flynn (2003) further unpacked the role of organizational politics in IT project abandonment. They identified six factors associated with project politics that explain project abandonment: political distrust among stakeholders,

formation of opposing coalition, threats of retaliation, failure to maintain top management support, lack of political campaign, and political insensitivity.

In sum, the literature on IT project abandonment enlarged the perspective offered by project management studies of de-escalation of commitment, which showed that abandonment is a long process, which requires considerable organizational efforts to walk away from a project. The studies of IT project abandonment reviewed here showed that abandonment ultimately has been the result of three factors: economic (project became an unviable proposition), organizational (project and/or system lost political support), and technological (technical solution became outdated).

However, the field of success and risk factor research is contested. March and Sutton (1997) questioned whether a cross-sectional view of an organization can sufficiently establish the causal relationship that was often implied when success factor research has been translated into management recommendations. Other researchers have questioned whether stable success factors exist at all (e.g., Lampel and Shapira 1995) and whether the trade-off between academic rigor and practical relevance can be overcome (Kieser and Nicolai 2005).

2.4.4 Summary of the Literature Review

The first chapter of the thesis closed with posing two broad questions: How frequent are outliers, in the sense of markedly deviating observations, in IT project performance? Why do outliers in IT project performance occur?

The first part of the literature review found that, to date, no study had surveyed the occurrence of outliers in IT projects. The literature review drew on findings from studies of IT project performance in terms of cost overruns, schedule delays, and benefit shortfalls as well as studies of other IT (project) indicators. The literature review found that studies of IT project performance reported a high level of variation between results and within results. Moreover, the literature review revealed that, when reported, studies counted a high over incidence of outliers. In fact, as noted by Kulk (2009) and Kitchenham (1998), the over incidence of outliers questioned the normal distribution assumption. The literature review also demonstrated that exclusion of outliers has been the dominant analysis strategy, with few exceptions in PERT studies of task effort and duration. These analyses had replaced the normal distribution assumption with more extreme value distributions, which allowed the studies to incorporate outlier in their models. In sum, the first research question: *How frequent do outliers, in the sense of markedly deviating observations of performance, occur in IT projects?* can thus be translated into three hypothesis for further analysis. Under the thin-tailed assumption, e.g., that of a normal distribution, statistical outliers do not occur in IT project performance (H1.1). IT project performance is normally distributed (H1.2) and extreme value distributions offer a better alternative description of IT project performance (H1.3).

The second part of the literature review focused on the second research question: *Why did outliers, in the sense of markedly deviating observations of cost, schedule, benefits, and effort performance, occur in IT projects?* The literature review

identified three schools of thought and in each organizational, project management, and IT project management literatures were reviewed. Table 1 summarizes the factors potentially contributing to the occurrence of outliers, that is markedly deviant performance, in IT project performance. The review of the extant literature showed substantial theorizing within each of the three schools of thought but rarely across schools.

Table 1 Overview of Literature of Why Outliers Occur in IT Project Performance

Theory	Schools of thought		
	System-centric	Event-centric	Process-centric
Organizational	<i>Normal Accident Theory</i> : In interactively complex and tightly coupled systems failure is inevitable (Perrow 1981). <i>High reliability organizations</i> tighten and loose coupling (Roberts 1989, Weick and Sutcliffe 2007) and attend to weak signals (Weick 2006).	Failure is result of insufficient response to <i>rare events, surprises, or crises</i> . (Christianson et al. 2008, Ansoff 1975, Roux-Duffort 2007). Insufficient responses are the result of failures of sense-making, framing, and learning (Baumard and Starbuck 2005, Weick 2004, Rerup 2009, Christianson et al. 2008).	<i>Man-made disasters</i> incubate over a period of time. Misperception, misunderstanding and miscommunication of error accumulates (Turner and Pidgeon 1997). <i>Normalization of deviance</i> enlarged the accepted risks over time (Vaughan 1996) <i>Escalation of commitment</i> influenced by project, psychological, social, and structural factors (Staw and Ross 1986, Sleesman et al. 2012).
Project management	<i>Social and technical complexity</i> cause average overruns in megaprojects (McFarlan 1981, Zmud 1980, Whitty 2009, de Bruijn & Leijten 2008).	Management of <i>turbulences</i> is core of project risk management (Miller and Lessard 2000, Pich et al. 2002).	Self-justification, prospect, agency, and approach avoidance theory all explain project escalation (Keil et al. 2000).
IT project management	Inaccuracies of <i>effort and cost estimates</i> are linked to project complexity, social and technical factors (Addison and Vallabh 2002, Glass 1997, Boehm 1991, Phalp and Cox 2000, Heltewig 2008).	–	<i>IT project abandonment</i> linked to of lack of economic viability (Oz and Sosik 2000), loss of political support (Sauer 1993), and leapfrogged technology (Ewusi-Mensah and Przasnyski 1991). Average project outcomes determined

Size and duration are proxies for interactive complexity and volatility of scope and requirements (Jones 2000, Sauer et al. 2007, McFarlan 1981, Zmud 1980, Jenkins et al. 1984, Wallace et al. 2004, Brooks 1975).	by multitude of <i>success and risk factors</i> (Fortune and White 2002, Belassi and Tukul 1996, Pinto and Slevin 1987, Cooke-Davies 2002)

The *system-centric school of thought*'s key theoretical explanations of outliers are Normal Accident Theory (Perrow 1981, 1984) and the theory of high reliability organizations (Roberts 1989, Weick and Sutcliffe 2007, Weick 2006). In short, both theories have the premise that highly complex organizations, i.e., those characterized by high interactive complexity and tight coupling, have a higher propensity to fail. Failure can be prevented through technical designs, influencing organizational and human relationships, and attention to weak signals. In project management theories these ideas have been translated into the concepts of social and technical complexity (McFarlan 1981, Zmud 1980, Whitty and Maylor 2009, Bruijn and Leijten 2008). In the IT project management literature these concepts have been rooted in forecasting theory, where project complexities increase errors in forecast and thus lead to overruns. The notion of complexity has been operationalized through the proxies of size and duration (e.g. Jones 2000, Sauer et al. 2007, McFarlan 1981, Zmud 1980, Jenkins et al. 1984, Wallace et al. 2004a, Brooks 1975). Other studies have operationalized complexity through assessments of difficulty or manageability of social factors (e.g., Addison and Vallabh 2002, Glass 1997, Yourdon 1996, Boehm 1991) and

technical factors (e.g. Phalp and Cox 2000, Heltewig 2008, Collopy 2007, Bergeron and St. Arnaud 1992, Heemstra 1992).

The *event-centric school of thought*'s key theoretical explanations of outliers are theories of rare events and the management of catastrophes. The notion of rare events has been linked to interruptions of organizational routines (Christianson et al. 2008), strategic surprises (Ansoff 1975) and crises (Roux-Douffort 2007). In sum, these theories explain outliers as the effect of an insufficient organizational response to a rare event. Those insufficient organizational responses are the result of failure to make sense of a rare event, framing a rare event in a way that inhibits the needed response, and failure to learn from and through rare events (Baumard and Starbuck 2005, Weick 2004, Rerup 2009, Christianson et al. 2008). In project management literature the concept of rare events is translated into the notion of turbulences, which is the core of project risk management (Miller and Lessard 2000, 2001, 2008, Pich et al. 2002). This line of inquiry has failed to attract much research in project and IT project management fields.

The *process-centric school of thought*'s key theoretical explanations of why outliers occur are the theories of man-made disasters (Turner and Pidgeon 1997), normalization of deviance (Vaughan 1996), and the escalation of commitment (Ross and Staw 1986, Sleesman et al. 2012). In project management literature the escalation of commitment is the dominant theoretical lens, which has been particularly studied as project escalation (Keil et al. 2000). Other theoretical explanations of why IT projects have become outliers has been the study of

project abandonment which is rooted in the support theory, i.e., projects were abandoned because they lost economic support (Oz and Sosik 2000), political support (Sauer 1993), or technological support (Ewusi-Mensah and Przasnyski 1991). Lastly, the explanation of common overruns in IT project performance in the process-centric school of thought has often been atheoretical and focused on identifying success factors of project execution and delivery (Pinto and Slevin 1987, Cooke-Davies 2002, Fortune and White 2002, Bleassi and Tukul 1996).

Based on the extant literature the thesis selected a long list of variables (cf. Appendix 8.2), which was narrowed down through test interviews with two project managers and discussions with academic experts and practitioners.

The final list (cf. Table 2), includes as variables from the *system-centric school of thought* (H2): project size (H2.2), social complexity of the project (H2.2), technical complexity of the project (H2.3), complexity of the software solution (H2.4). Two hypotheses were added from an on-going research project: whether the project was planned using the inside-view (Kahneman and Lovallo 1993) or the outside view (H2.5.1) and a statement to evaluate the project in hindsight: 'If we had known the real costs of the project it would not have been built' (H5.1).

Because of reasons of interview time only one variable was included from the *event-centric school of thought* (H3): Ineffective risk management (H3.1).

Variables taken from the *process-centric school of thought* (H4) were: effectiveness of project management processes (H4.1), suitability of staffing of the project (H4.2), effectiveness of project governance (H4.3), support of the project (H4.4), and communication within the project (H4.5). The next chapter will describe the methodology of data collection and develop the analyses used in the thesis.

Table 2 Final Selection of Variables to Explain IT Project Outliers

VARIABLES	SCHOOLS OF THOUGHT		
	System-centric	Event-centric	Process-centric
	H2.1: Project size H2.1.1: Estimated cost H2.1.2: Planned project duration H2.1.3: Actual project budget H2.1.4: Actual project duration H2.2: Social complexity H2.2.1: Unclear business case, goals, success criteria H2.2.2: Lack of managerial influence over environment H2.2.3: Organizational resistance H2.2.4: Lack of user involvement H2.2.5: Political intent H2.2.6: Unrealistic and unreliable schedule H2.2.7: Project perceived as unique H2.3: Technical complexity H2.3.1: Volatile requirements H2.3.2: Innovation H2.3.3: Tight coupling H2.3.4: Wide and volatile scope H2.3.5: Unsuitable project management methods H2.3.6: Many unknowns in the design H2.3.7: Big bang implementation H2.4 Software complexity H2.4.1: Processing requirements H2.4.2: Reusability requirements H2.4.3: Performance requirements H2.4.4: Security requirements H2.5: Planning H2.5.1: Project has taken an inside view H5.1: "If we had known the real costs of the project it would not have been built"	H3.1: Ineffective risk management	H4.1: Project processes H4.1.1: Ineffective monitoring and control H4.1.2: No benefits management H4.1.3: Lack of detailed and up-to-date plans H4.1.4: Ineffective change management H4.2: Staffing H4.2.1: Unsuitable mix of internal and external resources H4.2.2: Lack of competence of project management team H4.2.3: Lack of qualification and motivation of project team H4.2.4: Lack of suitable staffing H4.2.5: Years of work experience of most senior project manager H4.3: Governance H4.3.1: Ineffective project governance H4.3.2: Ineffective project sponsor H4.4: Support of the project H4.4.1: Lack of senior management support H4.4.2: Lack of strong support by the system vendor H4.4.3: Weak contracting H4.4.4: Non-aligned project stakeholders H4.5: Communication H4.5.1: Non-transparent status communication H4.5.2: Lack of team communication

3 Methodology

This chapter will now discuss four methodological aspects (1) the survey instrument, with which data on the variables, identified at the end of chapter 2, were collected, (2) the data collection process, (3) the statistical methods developed and used to test the hypotheses, and (4) a description of the sample that was collected.

It is important to note, that the thesis methodologically improves upon existing research - first, the survey instrument, measures project performance from documentary evidence, i.e., cost and effort overruns, schedule delays, benefit shortfalls, in conjunction with explanatory variables from the management literature. This will provide a less biased view of project performance compared to self-reported surveys, as problematized by Moløkken-Østvold et al. (2004).

Secondly, all research studies need to make a trade-off between the number of observations and the richness of variables, which is even more aggravated when studying rare events (King and Zeng 2001). One research strategy is to sample on the dependent variable (for example, Keil 1995 used of two surveys one for escalated and one for non-escalated projects). To ensure that the results are representative, the thesis will not sample on the dependent variable.

Moreover, the study design combines the large-scale study of project performance with a more in-depth survey of explanatory variables identified in chapter 2.4.4.

Thirdly, as discussed in chapter 2.3.2 few academic studies have problematized the presence of outliers in the data or attempted to explain the occurrence of outliers. The thesis will improve on the conventional analysis strategy⁴⁰ of logistic regressions with introducing multilevel models, which are able to account for organizational differences between projects, an issue in prior studies problematized by Verhoef (2004).

Fourth, analyzing outliers with logistic regression models requires the researcher to classify the observations into two groups. For example Kulk (2009) classifies misestimated projects as more than 5% overrun or more than 2.5% underrun; Keil et al. (2000) left the decision to the respondent asking for troubled projects versus projects “which progressed smoothly enough that you never thought [they] should be discontinued or redirected” (Keil et al. 2000, p. 661). Similarly, studies into success and risk factors asked their respondents to state whether the project was a success or typically not used in a multi- (e.g., Verner et al. 2007, Shenhar et al. 2002, Pinto and Slevin 1988). The thesis improves upon these methods of inquiry by introducing a second, novel analysis, which directly compares the thickness of the tails of the distribution of IT project performance. Thus the analysis does not require the exogenous distinction whether an observation was a special or a common data point.

40 As the literature review showed antecedents of outliers have mostly been studied through case studies. Quantitative approaches have analyzed project escalation through logistic regression (Rai et al. 2003) also in combination with machine learning (Zhang et al. 2003). Logistic regression has been used to model antecedents of project abandonment (Kumar et al. 1996) and the factors of misestimation of projects (Kulk 2009)

3.1 The Survey Instrument

The survey instrument consisted of two parts (1) a form for the archival research to capture the performance data (a full definition of all technical terms is included in Appendix 8.4) and (2) Likert-scaled and anchored items to conduct a project post-mortem analysis.

The second part of the questionnaire, translated the variables identified in table 2 of chapter 2 into statements to be rated on a 10-point Likert scale, which was anchored at values 1, 4, 7, and 10. The anchors were defined using the preliminary results from the first two organizations and expert input to help participants compare their project to other projects.

Table 3 shows selected variables, namely the business case (H2.2.1), project plan (H4.1.3), and staffing (H4.2.4). The table illustrates the anchoring that was presented to informants. The full list of variables and their anchoring is included in Appendix 8.5.

Table 3 Selected survey items and anchoring (full table in Appendix 8.5)

Item	(1)	(4)	(7)	(10)
The goals, business case, and success criteria were clear.	No defined business case, goals nor success criteria	Project goals vaguely defined; only effort cost estimate no full business case	Project objectives stated w/o prioritization; rough business case exists	Clear, concise, prioritized goals and objectives; sound business case incl. benefits; clearly defined success measures
Was the project planned in detail and the plan updated regularly?	No reliable plan existed if one existed it was completely out-of-date	Project plan existed/was made in the beginning but rarely updated; not a useful tool to	Detailed project plan existed, was updated infrequently, helped in informing future work,	Cost, schedule, benefits/functionality planned in detail (incl. dependencies) and tracked regularly to inform future work

		inform future work	but could have been more efficient or effective	
Was the project lacking suitable staff?	Sufficient number of experienced and qualified personnel are assigned full-time to key tasks	Personnel is experienced and qualified with a few gaps; most work full-time or 80% for the project	Personnel mostly 50% on the project with key roles in other projects, several competency gaps and lack of experience	Personnel responsible for key tasks is not fully qualified, lacks experience, and is assigned only part-time to the project

3.2 Data Collection & Preparation

Data were collected between 2009 and 2013 from four different data sources.

This section first describes in detail each of the data sources. Then the section describes possible limitations of the sample.

The aim of the data collection was to build a large sample while ensuring high quality and consistency of the data. In particular the data collection aimed to improve on previous methodologies, which (a) used of aggregated data (e.g., Standish 1994-2010), (b) narrowly focused on software development (e.g., Jones 2000, Jørgensen 2004), and (c) used self-reported performance (e.g., Wallace et al. 2004ab, Sauer and Cuthbertson 2003). Additionally, the use of multiple data sources reduces bias and increases validity (Campbell and Fiske 1959, Jick 1979) and avoids common method bias (Chang et al. 2010, Podsakoff et al. 2003).

Data for this thesis was collected from four different sources:

1. Study of participating organizations,
2. Desk study of auditor reports,

3. Budget requests for investment titles to the US Office of Management and Budget (OMB), and
4. Project status reporting to the OMB.

3.2.1.1 Study Projects of Participating Organizations

First, data collection from participating organizations proved most time consuming. The effort needed for the archival research was a barrier to recruit participants.

The first approach at recruiting participants was made through collaborating with McKinsey's Business Technology Office. In total, 62 of their partners were briefed through several presentations of the research project. Partners were asked to invite their clients to participate. At the same time, organizations in the network of the BT Centre for Major Programme Management were approached (e.g., Advisory Board members, students in Executive Education, and participants of other research projects). The project was also presented at McKinsey & Company's annual CIO/COO/CTO conference to nearly 75 c-level executives with an IT and operations background from Europe and America.

In total, 367 organizations were contacted (127 via meetings and telephone calls, 240 via snail mail). Initially, 28 organizations agreed to participate but 8 of the 28 found that the organization had no archival data available. From the remaining 20 organizations data on 805 projects were collected. The low response rate of 5% might have introduced a conservative bias in the sample, because organizations with very low project management maturity were excluded from the sample. Moreover, upon closer inspection most data was of

unreliable quality, e.g., labelling forecasts as the initial baseline but being the last approved baseline, and data on only 169 projects could be used. Then, preliminary results were used to demonstrate the insights generated by the research. The results were presented at a wide range of events for IT executives. In December 2010, a printed copy of the preliminary report was mailed to all clients of McKinsey's Business Technology Office. It was followed-up with more snail mails and presentations after preliminary findings were published in the Harvard Business Review in September 2011. In total, these efforts resulted in reliable data on 1,365 projects from 34 organizations. The data collection was audited in September 2013 by a postdoctoral researcher for accuracy and quality.

The project post-mortem part of the questionnaire was administered to 220 projects of the sample. Initially the questionnaire contained 28 questions. Nine questions were added after preliminary analysis. In total 220 post-mortem surveys were filled, 168 of the 37-question version and 52 of the 28-question version.

Initially, data collection focused on large-scale projects. Participating organizations were asked to provide access to projects with at least five to eight million Euro budget. According to benchmarks (Jones 2000) this budget corresponds to 10,000 function points in size, which delineates the top 6% of the function point database. Data collection with the first two participants showed that the definition which projects are to be considered large-scale is highly

dependent on the organization. Subsequently the question, with which participants were approached, was changed to:

“The survey aims to benchmark your organization’s project performance against a database of the industry. To get a representative picture of your organization the survey would like to collect the last 20-30 finished projects.”

Dropping a reference to project size led to a sample that span a wide range of project sizes. No cut-offs in terms of project size were defined in either stream of data collection. This enabled the thesis to study the effects of scale. Moreover, it reduces bias in the results which can be introduced by defining cut-off values as demonstrated by Jørgensen et al. (2012). However, routine projects were not included in the sample and only 19 out of the total 4,307 projects lasted less than 30 days.

3.2.1.2 Desk Study

The second source of data was a desk study. Auditor reports published by the GAO, the NAO, and their counter-parts in Australia, Denmark, Sweden, Norway, Switzerland, Germany, France, and Austria⁴¹ were reviewed; as well as academic publications. Audits reports of IT projects were screened for budget, schedule, and benefits figures. In most cases only forecasted and actual costs could be obtained. Most data were collected from UK and the US where

⁴¹ These countries were selected because of language and because these have a central government auditor who reported on IT projects.

the audit offices regularly conduct project audits and issue recommendations for performance improvement⁴².

In most other countries, only projects of public interest tend to be audited. Often because these projects have sparked public interest and debate, leading to a potentially biased collection of particularly badly performing projects. A total of 289 projects were collected from these data sources. The data collection was audited by a postdoctoral researcher for quality and accuracy.

3.2.1.3 Study of US Federal Budget Requests

A third source of data was US federal budget requests. Every capital investment project under the 1996 Clinger-Cohen Act needs to submit a budget request form (named Exhibit 300 or E300) to the office of management and budget. Additionally, agencies are required to annually submit another budget request form (named Exhibit 53 or E53) for IT projects. The forms are used to request annual budgets from the OMB and the forms are used by the OMB to build the federal IT project portfolio. Each agency individually consults with the OMB to define which of their investment projects falls under the reporting requirements. The forms include a project description, as well as detailed actual and estimated cost. For every budget year a new request is submitted including updated actual and forecasted figures.

⁴² In most European countries, as well as in Australia government auditors focus on auditing government accounts, which often results in reviews of annualized budgets and thus lacking a complete forecast and/or baseline.

Currently, no legislation is in place that criminalizes misreported numbers. Nevertheless, data quality of the forms was deemed sufficient because each project budget request is signed by the responsible project manager, including full name, e-mail address, and telephone number, and recently made publicly available on the internet⁴³. During the data collection for this thesis the GAO audited the data quality of the IT dashboard three times (July 2010, March 2011, and November 2011). GAO found that by November 2011 the data quality had increased and that the audited project reported up-to-date and correct figures. The GAO's reports asked for further improvements of the verbal CIO recommendations and the traffic light status reported for each projects (GAO 2005, GAO 2010); these two data points were subsequently not included in this analysis.

The electronic IT dashboard only covers the budget years 2010 and onwards. It was decided to collect pre-2010 forms from websites and via freedom-of-information act requests from all 134 federal bureaus. In total 1,921 forms were collected. A standard mapping was scraped that translated the form fields into an Excel spreadsheet. This mapping allowed the automated data collection from 1,548 forms. The remaining 372 forms, mostly non-automatically readable scans, were manually processed by research assistants. In total this process

⁴³ Online availability of the data was politicized during the budget talks in 2012 (Byers 2012). Project performance was not updated while Congress discussed the White House budget in 2012, the tool went offline during the government shut-down between October 2013 and February 2014. When the dashboard re-launched access was limited in access to US government employees. In April 2014, after Congress passed the 2014 budget in March, access was made public again.

yielded 14,669 individual budget requests dating from 2003 to 2010. The data required extensive cleanup. First, the 14,699 records were manually coded item by item to exclude 1,526 requests which were double entries. These double entries mainly came from overlaps in paper documents from agencies and the electronic versions of 2010 budget requests from the Office of Management and Budget. The remaining 13,166 budget requests were hand coded with a unique identifier to allow a grouping of multi-year projects. This was done manually since identifiers on the requests, as well as project names, changed frequently. This resulted in 5,173 unique projects with 13,166 budget requests. Next, on-going projects were excluded. Percent complete was calculated based on actual spend divided by most recent baseline and 2,582 projects with less than 80% complete were excluded. This resulted in data on 2,555 finished projects. Then projects with less than one year duration were excluded because revised estimated and actual figures are only available for multi-year projects, i.e., when at least two budget requests are issued. This resulted in a final set of 1,180 projects. Lastly, projects were checked for common error patterns, e.g., unlikely budgets that might be due to reporting or processing errors. A further 151 projects were excluded because of concerns over data quality. This resulted in a final sample of 1,029 projects.

In 2011, the reporting was changed and the focus shifted from budget requests to progress reporting. Beginning with the requests for budget year 2012 investment titles were broken down into sub-projects and for budget year 2013 projects started to report real-time performance broken down by milestones,

deliverables, and contracts. In total 3,269 projects were reported via the federal IT dashboard. All projects include unique identifiers, formats of those identifiers changed year-on-year and required manual mapping to identify projects.

Overall, the analysis of the US federal IT portfolio from the IT dashboard and budget requests to the Office of Management and Budget collected 21,249 budget requests (E300s and E53s), dating back to 1985. Those requests represented 9,319 IT projects. 4,424 of these were at least 80% complete. 2,653 were documented through budget requests and 1,771 were reported through progress reports to the web dashboard.

However, interviews with US government CIOs raised considerable doubt about the data quality, despite the positive audit by the GAO. Only one person interviewed was willing to go on record and stated:

“...in every experience I have had in the PSP [Public Sector Projects], a sample of roughly 10, the data that goes to our Office of Management and Budget ... is highly suspect. There is quite a lot of papering over budget overruns, shifting money around halfway through a project, changing the scope without admitting to it, etc. That happens. It's pervasive throughout the US federal government, unfortunately...” (Interview with a consulting partner, July 2013).

Statistical analysis corroborates the informant's suspicion, that the introduction of the IT dashboard changed project outcomes. Before the web dashboard introduction the median cost overrun was 0%, with an IQR⁴⁴ of 0.568, after the

⁴⁴ Inter-quartile range, the width of the box in a boxplot, defined as the distance between the top (75th) and bottom (25th) quartile (Tukey 1977).

introduction the median cost overrun is 0% with an IQR of 0.017. This shows that the variability in the data decreased: the distance between the bottom and the top quartile changed from 57pp to 2pp. To test the significance of this change a non-parametric Wilcoxon test was applied. The null hypothesis is that both samples come from the same population. The test statistic is $W = 2480617$, which gives a $p\text{-value} = 0.0008$. Cost overruns post and prior the web dashboard shifted with overwhelmingly statistical significance. Thus the decision was made to exclude the 1,771 projects, which came from the new project reporting into the OMB's web dashboard. This resulted in final sample of 2,653 observations from the third source.

In sum, the data collection across all three sources aimed to establish a reliable and robust dataset. Multiple data sources were used to collect forecasted and actual cost, schedule, effort, and benefits from documentary evidence. The reliance on archival data increased the data collection effort, yet it avoided interviewee biases. However, reliance on archival data might have biased the sample because the method excludes organizations with a low planning maturity, i.e., organizations that do not keep project records. Thus the sample errs on the conservative side.

Additionally, because considerable efforts were required to collect the archival evidence, access negotiations proved difficult and were only successful where mandated by top management. Access negotiations with mid-level managers, e.g., heads of PMOs, never resulted in access to data. Again, this potentially led

to a conservative bias in the sample, because the research failed to get access to organizations where top management was not interested in the performance of IT projects. This might have excluded organizations with a lower maturity and therefore a worse track record of project outcomes.

Overall, the data collection methods have been chosen to provide a reliable and representative sample of project performance. The three data collection efforts resulted in a final sample of 4,307 IT projects. While particular attention was placed on only collecting high-quality data, the way data were collected potentially led to a conservative bias in the sample, because organizations without documentary evidence or top management attention to IT project performance are not part of the sample. In other words, the results might overrepresent IT project performance.

3.3 Statistical methods

The thesis will use two different approaches to analyzing the explanatory variables on the occurrence of outliers: (1) rare events multilevel regression and (2) comparing the thickness of the tails of the performance distribution. Table 4 compares the two methods.

Rare events logistic regression (King and Zeng 2001) allows to analyze determinant of rare events using explanatory scale variables. Project performance, the dependent variable, has a hierarchical structure that a multilevel model (Lindstrom and Bates 1990) can account for. Projects have a hierarchical structure, they typically part of a program and projects are carried out by an organization. Effects, such as learning from past projects, spill-over from projects in the same program, might occur and might need to be controlled for. However, a rare events logistic regression requires that the observations are classified into observations that are rare events and those that are common events.

The second method addresses this particular issue through directly comparing the distributions. If the distribution and its tails can be fitted using a maximum likelihood estimator two distributions can be compared by testing whether the distribution parameters statistically significantly differ⁴⁵. This approach does not require to specify upfront which observations are rare events and which are common events. However, the approach does not allow to control for the

⁴⁵ Handcock and Morris (1999) provided a non-parametric alternative.

hierarchical structure in the data. The biggest disadvantage of this analysis strategy is that it only allows for groupwise comparisons, e.g., comparing big and small projects instead of using the full range of project size, thus some information is lost.

Table 4 Overview of the Two Analyses Used to Identify Why Outliers Occurred in IT Project Performance

	Rare Events Multilevel Logistic Regression	Testing the Thickness of the Tails
Advantages	• Use scale independent variables • Deal with hierarchical structure of dependent variable	• Treat outliers as part of the tail • Computationally simple
Disadvantages	• Classify an observation as an outlier • (Sample on the dependent variable) • Rare events bias in ML estimator	• Group-wise comparison (convert scale variables into binary variables) • Large sample sizes because of uncertainty of ML estimator

3.3.1 Rare Events Multilevel Logistic Regression

The thesis formulates a multilevel logistic regression of the kind:

$$\ln \left[\frac{p_{ij}}{1 - p_{ij}} \right] = \beta_0 + \beta_1 x_{ij} + \dots + \beta_p x_{ij} + u_{ij} x_{ij} + u_{0j} + e_{ij}$$

Where i is the project subscript and j is the first level, e.g., program or organization subscript (cf. Gelman and Hill 2007).

The basic idea is to take a generalized linear model of the form $\ell(\eta) = X_i b$ and add a random effect similar to a linear mixed model $\ell(\eta) = X_i b + Z_i u$, with $u \sim N(0, \gamma)$. Some specifications do not add an error term (Gelman and Hill 2007).

The expression for the likelihood of a multilevel model is an integral over the first level space. Because of the first level the likelihood function includes the integral $\int f(y_{ij}|u_i, b, \gamma) f(u_i|Data) \pi(u_i)$. Thus the restricted maximum likelihood (REML) solution to exactly evaluate the likelihood cannot be applied. Instead an approximation is needed. Three types of approximations exist: penalized quasi-likelihood (PQL), Laplace method, and adaptive Gaussian quadrature (AGQ). This thesis employed the first two types PQL as implemented in `glmmPQL` of the `MASS` package in R (Veneables and Ripley 2002) and the Laplace method as implemented in `glmer` of the `lme4` package in R (Bates et al. 2014).

Rare events logistic regression can introduce three biases into the analysis: (1) small sample bias, (2) non iid bias, (3) rare events bias. First, King and Zeng (2001) studied the applicability of logistic regression when analyzing rare events. They found that the MLE estimator for the regression coefficients β is approximately unbiased in small samples (King and Zeng 2001, p. 140).

Second, a bigger potential source of bias is if the data collection samples on the dependent variable. In any study, but particularly in the study of rare events, a simple trade-off presents itself: more observations versus more variables. If the event rarely occurs a large number of events are needed to acquire sufficient observations; usually this comes at the expense of the quality of the variables (King and Zeng 2001, p. 143). Studies have addressed this issue by sampling on the dependent variable. For example, Keil et al. (2000) sent out two

questionnaire versions. One version asked the respondent about a non-escalated project, i.e., being on-budget and on-time, and the other version asked for an escalated project. Sampling on the dependent variable has two consequences. On the one hand, the frequency of events cannot be established from the data set. On the other hand, a sample drawn by sampling on the random variable violates the maximum likelihood assumption that the observations are independent and identically distributed (iid). Rare events logistic regression analyses offers two solutions for the bias introduced in MLE by sampling on the dependent variable: prior correction and weighting (King and Zeng 2001, p. 143). However, this thesis used random sampling and did not sample on the dependent variable to avoid this bias altogether.

Third, rare events bias has been reported, i.e., the systematic bias that logistic regressions underestimate the probability of rare events. In logistic regression the bias is minimal the closer the probability of an observation is to 0.5. The further the probability of a rare event is away from 0.5 the higher the underestimation bias. King and Zeng (2001) approximated the rare events bias with

$$E(\hat{b}_0 - b_0) \gg \frac{\bar{p} - 0.5}{n\bar{p}(1 - \bar{p})}, \text{ where } \bar{p} = \frac{1}{n} \sum_{i=1}^n p_i,$$

where π_i is the probability that the dependent variable takes on the value 1 for all observations $i \dots n$. The formula also demonstrates that the bias is

conservative for rare events, i.e., events that are per definition $\pi_i \ll 0.5$. Lastly the formula shows that the bias decreases with increasing sample size at $1/n$.

King and Zheng (2001) further found through simulations that the bias directly affects the intercept and other coefficients primarily as a consequence of the effect on the intercept.

In sum, multiple sources for potential biases exist when using logistic regression with maximum likelihood estimators (MLE). The first bias is MLE's small sample bias, which King and Zheng (2001) found did not introduce large errors in logistic regression. The second bias, comes from sampling on the dependent variable, which is a common strategy to balance the trade-off between collecting more observations or collecting better variables. Sampling on the dependent variable, limits not only the ability to forecast the frequency of rare events but also violates the MLE assumption that variables are independently identically distributed (iid.). This thesis did use random sampling at the expense of requiring more effort to avoid the bias. The third bias comes from using logistic regressions where the average probability of the rare event occurring is much lower than 0.5.

3.3.2 Fitting and Testing the Thickness of the Tails

Outliers are a subjective post-data manifestation. In other words, outliers are outliers only in relation to an assumption made about the data. Outliers have four distinguishing features:

1. The outlier is a random manifestation, rather than a deterministic feature;

2. The outlier can only engender surprise relative to some (at least subjective or implicit) initial (working) model for the whole sample;
3. Assessment of discordancy is relative to the initial model: we need an alternative model in whose favor we reject the initial model;
4. The alternative model may be an inherent one: under which the data now appear as a homogeneous sample, e.g., Cauchy rather than Normal (Barnett 1978, p. 245).

The basic idea of testing the thickness of the tails is based on the idea of formulating an alternative model. Thus the idea is to avoid the need to make an assumption that delineates common and rare observations. Pisarenko and Sornette (2003) argued classifying observations between common and rare observations makes the assumption that outliers are from a different population of uncertainty. Alternatively, if outliers can be smoothly extrapolated from smaller observations, they do not form a fundamentally different category in terms of uncertainty. Thus observation that markedly deviate from the rest of the sample can be understood as being a valuable signal instead of noise that is to be treated differently (Mandelbrot 1963, Steele and Huber 2004).

The thesis will show in chapter 4.4.2 that, among the common extreme value distributions (Gumbel 1958/2004), power laws are the best fit of the data.

Lampel et al. (2009) first suggested that rare events in organizations might follow power laws: "What is rare under a normal distribution may therefore not be considered rare under a power-law distribution." (Lampel et al. 2009, p. 836).

Power laws are a family of extreme value distributions. Power laws are also called Pareto distributions (Newman 2005)⁴⁶.

Power laws were initially discovered in geography describing city sizes (Auerbach 1913, Zipf 1949) and first used in Economics to study wealth concentration (Pareto 1964) and distribution of firm sizes (Simon and Bonini 1958). Power laws have been established for a wide array of social and natural phenomena. Newman (2005) lists among others word frequency, citations of academic papers, visits to a website, or the magnitude of earthquakes. A distribution follows a power law if the tail above a cut-off value x_0 follows the form: $p(x|\alpha) = Ax^{-\alpha}$, where x is the observed overrun, A is a normalization constant, and α is the exponent.

The power law exponent can be estimated using a Maximum Likelihood estimator (see Appendix 8.6 for the full derivation of formulae). Given are n real-valued data x_i with $i = 1, 2, \dots, n$, which are assumed to be independent and identically distributed (i.i.d.). To fit a power law $p(x|\alpha) = Ax^{-\alpha}$ the tail is normalized to unity so that $1 = A \int_{x_0}^{\infty} x^{-\alpha} \partial x$. Solving for A gives $A = \frac{\alpha-1}{x_0^{1-\alpha}}$, which results in $p(x|\alpha) = \frac{\alpha-1}{x_0} \left(\frac{x}{x_0}\right)^{-\alpha}$.

⁴⁶ It is important to note that power laws and Pareto distributions differ in their specific estimators. Pareto distributions assume the probability density function $f(x) = \alpha x_0^\alpha x^{-(\alpha+1)}$ and Power laws assume the probability density function $f(x) = Ax^{-\alpha}$ for $x > x_0$. As a result the estimated alpha differs by the value of 1.

Putting the power law probability density function into the equivalence for the likelihood gives $\mathcal{L}(\alpha|x) \equiv \left(\frac{\alpha-1}{x_0}\right)^n \cdot \prod_{i=1}^n \left(\frac{x_i}{x_0}\right)^{-\alpha}$. Maximizing the log-likelihood results in the estimator $\hat{\alpha} = 1 + n \left[\sum_{i=1}^n \ln\left(\frac{x_i}{x_0}\right) \right]^{-1}$, where n is the sample size.

The uncertainty of the maximum likelihood estimator is approximately normally distributed with $\hat{\alpha} \sim N\left(\alpha, \frac{(\alpha-1)^2}{n}\right)$ for large n . Furthermore the moments of the distribution are defined as $\langle m \rangle = \frac{\alpha-1}{\alpha-1-m} x_0^m$, where m is the m^{th} moment. Importantly, the formula shows that not all higher order moments are defined. If alpha is smaller than 1 none of the higher order moments is finite. If alpha is between 2 and 3 only the first order moment, the mean is finite.

Variance, the second order moment, and all higher order moments (e.g., Kurtosis and Skew) are infinite. If alpha falls between 3 and 4 only the first two moments, mean and variance, are finite. Mandelbrot (1997) distinguishes seven states of randomness. Three states are applicable: slow randomness as described by the lognormal distribution, wild randomness where variance is infinite, and extreme randomness where the mean is infinite.

The uncertainty in the maximum likelihood estimator can be used for testing whether two power law tails are statistically significantly different. Two alternatives can be used for this test: (1) z-test and (2) likelihood ratio test.

The *z-test* compares the estimated exponents α_1 and α_2 of two power law tails.

The *z-value* tests whether both sub-groups are drawn from the same population because they have the same shape parameter. The test uses the

limiting properties of maximum likelihood estimators for large samples. One limiting property is that maximum likelihood estimators approximate a normal distribution with variance of 1 over Fisher Information (I), i.e., $\hat{\theta} \xrightarrow{D} N\left(\theta, \frac{1}{I(\theta)}\right)$, which the thesis used earlier to derive $\hat{\alpha} \sim N\left(\alpha, \frac{(\alpha-1)^2}{n}\right)$. Thus under $H_0: \alpha_1 = \alpha_2$, the difference between the estimators is normally distributed around 0 with

$$\hat{\alpha}_1 - \hat{\alpha}_2 \sim N\left(0, \frac{(\alpha_1 - 1)^2}{n_1} + \frac{(\alpha_2 - 1)^2}{n_2}\right)$$

Where $\hat{\alpha}_1$ is the estimated power law exponent of the first distribution fitted to a sample with size n_1 observations and $\hat{\alpha}_2$ is the estimated power law exponent of the second distribution fitted to the sample of size n_2 . Standardizing the difference between the maximum likelihood estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ to $N(0,1)$ gives

$$\frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\sqrt{\frac{(\hat{\alpha}_1 - 1)^2}{n_1} + \frac{(\hat{\alpha}_2 - 1)^2}{n_2}}} \sim N(0,1)$$

Where $\hat{\alpha} = 1 + n \left[\sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) \right]^{-1}$ for $x \geq x_0 > 0$. This results in a test statistic z , which asymptotically follows the standard normal distribution:

$$z = \frac{\left(1 + n_1 \left[\sum_{i=1}^{n_1} \ln\left(\frac{x_{1i}}{x_0}\right) \right]^{-1}\right) - \left(1 + n_2 \left[\sum_{i=1}^{n_2} \ln\left(\frac{x_{2i}}{x_0}\right) \right]^{-1}\right)}{\sqrt{\frac{n_1}{\left[\sum_{i=1}^{n_1} \ln\left(\frac{x_{1i}}{x_0}\right) \right]^2} + \frac{n_2}{\left[\sum_{i=1}^{n_2} \ln\left(\frac{x_{2i}}{x_0}\right) \right]^2}}} \sim N(0,1)$$

An alternative hypothesis test is the *likelihood-ratio test*. The likelihood-ratio test compares nested models. The test of the null hypothesis $H_0: \alpha_1 = \alpha_2$, translates into the question whether a model that splits the observations into two subgroups and estimates two tail parameter $\hat{\alpha}_1$ and $\hat{\alpha}_2$ provides a statistically significantly better fit than the model of the pooled sample and only one parameter $\hat{\alpha}$. A model with more parameters, in this case the two shape parameters instead of one, will always fit better. Hence the question is crucial whether the model fit is significantly better given the additional complexity of the two models. Thus if H_0 was true the sample should not be split into groups. In this thesis the null model is the pooled model and the alternative is the two subgroups model.

The likelihood ratio test (Vuong 1989) answers this question. It defines a test statistic $D = 2 \ln \frac{\mathcal{L}(\text{alternative model})}{\mathcal{L}(\text{null model})}$, which for the two sample case is

$$\begin{aligned}
 D &= 2(\ln \mathcal{L}_{\text{alternative}} - \ln \mathcal{L}_{\text{null}}) \\
 &= 2 \left(\ln n_1 \cdot \ln(\hat{\alpha}_1 - 1) - n_1 \ln x_0 - \hat{\alpha}_1 \sum_{i=1}^{n_1} \ln \left(\frac{x_{1i}}{x_0} \right) \right) \\
 &\quad + 2 \left(\ln n_2 \cdot \ln(\hat{\alpha}_2 - 1) - n_2 \ln x_0 - \hat{\alpha}_2 \sum_{i=1}^{n_2} \ln \left(\frac{x_{2i}}{x_0} \right) \right) \\
 &\quad - 2 \left(\ln(n_1 + n_2) \cdot \ln(\hat{\alpha} - 1) - (n_1 + n_2) \ln x_0 - \hat{\alpha} \sum_{i=1}^n \ln \left(\frac{x_i}{x_0} \right) \right)
 \end{aligned}$$

D is approximately χ^2 distributed (Wilks 1938, p. 62) with degrees of freedom of $df(\text{alternative model}) - df(\text{null model})$ where $df()$ is the number of free parameters. In the case of two subgroups the additional shape parameter estimated in the alternative model means that $df=2$.

The literature review identified multiple variables that potentially vary with the thickness of the tails. Simultaneously conducting independent statistical tests can lead to false positives or type 1 errors, i.e., the rejection of null hypothesis that should not be rejected. The results need to be corrected for the false discovery rate or the less conservative family-wise error rate. In statistical inference the probability of making a type 1 error is the significance level alpha, which per long standing convention is set at 5%. When simultaneously running 42 independent tests with a significance level of 0.05, there is an 88% chance⁴⁷ that at least one falsely rejects the null hypothesis. P-values from simultaneous tests thus need to be adjusted so that probability of at least one false positive finding is less than 5%.

Three approaches are commonly used: (1) Bonferroni adjustment, (2) Holm adjustment, and (3) Benjamini Hochberg procedure. Bonferroni and Holm correct for both the false discovery rate and the family-wise error rate, Benjamini Hochberg corrects for the family-wise error rate.

⁴⁷ Probability of at least one false positive = $1-(1-\alpha)^N$

First, Bonferroni adjustment is computationally simple. Bonferroni adjusts the p-value for the family-wise error rate by multiplying the number of comparisons N to the unadjusted p-value, i.e., $\tilde{p}_i = \min\{Np_i, 1\}$.

Second, Holm's procedure (Holm 1979) improves the Bonferroni adjustment by first ordering all p values so that $p_1 \leq p_2 \leq p_3 \leq \dots \leq p_n$ rejecting all hypothesis if $p_j \leq \frac{\alpha}{n-j+1}$, where α is the level of significance. Thus the first hypothesis is using a Bonferroni adjustment, the second adjusts with $N-1$, the third with $N-2$ and so forth. This gives the Holm adjusted p-value $\tilde{p}_i = \max_{j \leq i} \{\min\{(N - j + 1)p_i, 1\}\}$.

Benjamini and Hochberg (1995) showed that the Bonferroni adjustment controls for the false discovery rate. If statistical tests are independent then a weaker control is sufficient, the so called family-wise error rate (for a comparison of all three adjustment procedures see Williams et al. (1999)). The Benjamini-Hochberg procedure is a step-up procedure to identify the true null hypotheses (m_0) out of a total of m simultaneously tested null hypotheses. Testing the null hypotheses $H_1, H_2, H_3, \dots, H_n$ results in n p-values $p_1, p_2, p_3, \dots, p_n$, which are ordered $p_{(1)} \leq p_{(2)} \leq p_{(3)} \leq \dots \leq p_{(n)}$, so that H_i corresponds to the p-value p_i . Given a level of significance q^* , e.g., following long standing convention set at 0.05, let k be the largest i for which $p_{(i)} \leq \frac{i}{m} q^*$. Then reject all $H_{(i)}$ for $i=1, 2, \dots, k$. In other words, if m hypotheses are tested if and only if the m^{th} largest p-value is less then the critical test value $p=0.05$ all hypotheses are assumed true.

In sum, this section discussed the statistical analyses of the data where outliers are present and the object of interest. Two methods were discussed. Prior studies use rare events logistic regression, which this thesis improves upon by introducing a multilevel model to account for the hierarchical structure in the dependent variable, in this case whether an observation is an outlier or a common observation using Tukey's rule. The disadvantage of this method is that it requires a classification of the observations. A second analysis was developed which directly compares the parameters describing the distribution of the IT project performance. The analysis, thus, does not need a classification of observations into outliers and common observations. However, it is restricted to group-wise comparisons, which loses some of the information contained in the explanatory variables.

3.4 The Sample

This section will give a descriptive overview of the final sample.

3.4.1 *Sample Size*

The sample consists of 4,307 projects. Each of the 4,307 projects belongs to one of 190 organizations. An organization is typically the owner of the portfolio that the studied project is part of. Organizations in the public sector are typically known as government bureaus and agencies. In the private sector these are typically independent business units and subsidiaries. The 190 organizations in this sample can be grouped into 84 government department or private sector companies. On average the data contain 28 projects per organization (median 7), and 64 projects per company (median 5).

3.4.2 *Frequency of Project and System Types*

The project descriptions were used to classify the type of each project. Most projects in the sample were system implementation (89.9%) followed by IT infrastructure (5.4%), IT architecture (0.5%), communication technology (2.7%), and others (1.4%), which includes IT post-merger management, and IT outsourcing. System implementation projects were classified by software type. Implemented systems were supply chain management systems (26.9%), ERP⁴⁸ systems (19.7%), HR management systems (9.5%), management information systems (4.9%), production scheduling and management systems (3.2%),

48 ERP = Enterprise Resource Planning, the name refers to accounting and management systems, in the public sector also commonly referred to as financial management or budget management systems

document and content management systems (2.8%), unknown/classified systems (4.8%) and a wide variety of other system types (28.2%).

3.4.3 Size and Duration of IT Projects in the Sample

Actual and estimated project cost were converted into 2010 USD using local country GDP deflators and average annual exchange rates at the time of estimate. The average estimated cost is USD 24.6 million (in 2010 terms), the median estimated cost is USD 1.2 million, the lower quartile is USD 0.2 million, the upper quartile is USD 8.7 million. The actual project costs were on average USD 21.3 million (in 2010 terms), with a median cost of USD 1.2 million, lower quartile at USD 0.2 million and upper quartile at USD 8.7 million. In sum, the total actual monetary value of the sample is approximately USD 85.0 billion (in 2010 terms), the total estimated value is USD 86.9 billion (in 2010 terms).

The average estimated duration is 280 days, median 182 days. The lower quartile is 113 days, the top quartile is 323 days. The average actual duration of the projects is 1,202 days, median is 1,096 days, lower quartile is 376 days, and the top quartile 1,826 days.

Table 5 shows that the average project size, 24.6 million in this sample, is larger than the projects in previous studies, except for Patanakul (2014). The average duration, 280 days in this sample, is slightly shorter than Little (2006), which was the one study giving the project size in duration.

Table 5 Project Sizes in Previous Studies

Author	Year	Sample size	Average project size (USD millions)
Jenkins et al.	1984	72	0.13
Kulk et al.	2009	165	2.2
McKeen	1983	32	0.04
Mitchell and Zmud	1999	42	3.8
Patanakul	2014	16	594
Standish	2002-2010	80,000	3-5
Bergeron and St. Arnaud	1992	89	1251 PD*
Moløkken-Østfold et al.	2004	36	400 PD*
Little	2006	106	329 days

* Verhoef (2002) used day rates of 500-4,000 USD per PD with an average of 1,800; this sample the median day rate is 1,300 USD per PD in 2010 terms.

Note: None of the studies reported the base year of the cost.

The two dimensions time and cost are moderately but not strongly correlated.

The correlation of estimated duration with estimated cost is $r = 0.457$, with a 95 percent confidence interval of the correlation coefficient of $[0.389, 0.521]$, the correlation is highly significantly different from zero, with $t = 12.0461$ at $df = 549$ with $p < 0.01$. The correlation indicates that planners couple cost and schedule; to the point where at least 20% of the estimated cost variation can be explained by variation in estimated duration and vice versa. On the other hand, the correlation between actual cost and actual duration is only $r = 0.053$ with a 95 percent confidence interval of $[0.021, 0.084]$ with $t = 3.225$, $df = 3723$, p -value < 0.01 . This indicates that only 0.28% of the variance of actual cost are explained by actual duration and vice versa. The difference between the two correlation coefficients is striking and an indication of the trade-offs between cost and time made by projects.

3.4.4 *Heterogeneity in the Sample*

Lastly, the question arises whether the data are too heterogeneous to be pooled as it was the case in Kulk (2009) and Kulk et al. (2009), who did not pool data across organizations. Five characteristics have been suggested as potential sources of heterogeneity: (1) programs, (2) geography, (3) organizations, (4) project types, and (5) sources of data. The question can be approached through the question of independency, in other words are observations independent and identically distributed? And that of reliability, i.e., Is an aggregate measure a reliable indicator for individual observations⁴⁹?

First, most projects in the dataset were part of a program. Only 833 projects were not associated to a program. Each program consists of multiple projects. The program level potentially introduces a dependency between projects, because projects in a program share monetary and non-monetary resources, aim for the same program goals, and share parts of their governance structures. Thus projects within a program might show a level of non-independence.

Second, studies of infrastructure project performance and geography have shown that the country in which the project was implemented might cause heterogeneity in the data (Cantarelli et al. 2012). This could be due to different

⁴⁹ The classic example is that explaining the performance of students through school, class, and student characteristics. Student performance is associated with school performance. The performance of individual students is not independent of their school or their class. Analyses need to control for class and school differences to make comparisons at the student level. At the same time student performance is a reliable indicator of class and school performance. That is to compare schools a random sample of students and classes adequately reflects the differences between schools.

regulatory environments, shared economic resources, macro-economic trends, and – in the case of the public sector – also shared monetary budgets.

Third, projects within an organization might be similarly non-independent.

Each project in the sample belonged to one specific organization that owned the project. Nearly all organizations in the sample owned multiple projects. Similar to the dependencies between projects in a program potential non-independence between projects in an organization comes from shared resources, shared governance structures, standardized processes, and shared organizational goals. Moreover, over time learning from earlier projects might influence later projects.

Fourth, project types are a potential characteristic of non-independence. Project types have been discussed in the IT literature (e.g., Stahlknecht and Hasenkamp 1997) and in the project management literature (e.g., Shenhar and Dvir 1996) in relation to the notion of domains. An example of a domain are ERP projects.

Potential dependencies between projects in a similar domain might come from shared knowledge, shared routines, shared experts, shared regulation and standards. For the ERP example there could potentially be dependencies between ERP projects because of shared regulatory environments, shared accounting standards, the use of the same experts and system integrators, the limited number of available standard software packages, and emulation of best practice processes.

Fifth, as documented earlier in this chapter multiple sources were used to collect the data. Thus heterogeneity in the data might come from the data collection itself.

Bliese (2000) suggested the use of two intraclass correlation coefficients (ICC1 and ICC2) to analyze potential dependencies in hierarchal data. The first intraclass correlation coefficient (ICC1) is defined as $\frac{\tau_{00}}{\tau_{00} + \sigma^2}$, where τ_{00} is the between group variance and σ^2 is the within group variance. ICC2 is defined as $\frac{\tau_{00}}{\tau_{00} + \frac{\sigma^2}{n_j}}$, where n_j is the size of group j (Bliese 2000). The intraclass correlation coefficients are estimated using a multilevel linear model with project cost performance as the dependent variable and the organization, project type, program, geography, and data source as the five different candidates for the first level. In the multilevel linear model τ_{00} is the variance of the intercept of the data and σ^2 is the residual variance of the model. ICC estimates range between 0 and 1. The value can be interpreted as the degree to which individual observations are associated with the group level and vice versa (Kenny and Judd 1986). Ozkaya et al. (2013) defined as a rule of thumb that an ICC of 0.10 is enough to warrant further examination.

Table 6 Intraclass Correlation Coefficients of Potential Hierarchical Dependencies in IT Project Cost Performance

	Program	Country	Organization	Project type	Source
ICC1	0.017	0.0012	0.122	0.020	0.0335
ICC2	0.018	0.0449	0.441	0.472	0.7305

Table 6 shows the result of the analysis. The data show low ICC1 and ICC2 for the *program level*. The low ICCs are a result of little between-group variance

being shared with within-group variance. The findings have two implications. On the one hand, the lack of intraclass correlation and thus shared variance between the project level and the program level indicate that the cost performance of one individual project depends very little on the overall program cost performance. On the other hand, low ICC1 and ICC2 also imply that program cost performance is not well characterized by the cost performance of individual projects in the program. Interpreting directions of influence through the lens of reliability of performance measurements shows that observing a singular project is not a reliable indicator for program cost performance and, equally, program cost performance is not a reliable indicator for project cost performance. Interpreting the results through the lens of independence, the analysis shows very low levels of dependence between the projects grouped under the umbrella of the same program.

Second, ICC1 and ICC2 are again nearly zero for the *country level*. Again, it indicates that very little between-groups variance is shared within within-group variance. Same as with programs, this finding implies that project cost performance depends very little on country. The finding also implies that country cost performance, if one was to calculate such an indicator, is not a reliable measurement for individual project cost performance. Project cost performance is also not a reliable measurement for country cost performance.

Third, the results of the analysis of *organizations* differ from the analyses of programs and countries. Neither, ICC1 nor ICC2, are as close to zero as they

were for the program variable. Applying the rule of thumb suggests that with $ICC1$ and $ICC2 > 0.1$ sufficient variance is shared between and within groups that justifies the use of an organizational level in the model. The findings show that project cost performance varies with organizational project cost performance. The results, through the lens of reliability, show that the cost performance of projects is a reliable indicator of an organization's cost performance; while at the same time an organization's cost performance is a reliable indicator of project cost performance. The results, viewed through the lens of independence, show a certain level of dependence, in the form of shared variance, within the project portfolios of organizations. In sum, because both ICCs are greater than 0.1 the organization suggests itself as a first level and an important covariate when analyzing IT project cost performance.

Lastly, the analyses of *project types* and *data sources* show unusual results. The results show in both cases a small $ICC1$ and a large $ICC2$, which Bliese (2000) argued is a rare pattern in real-world data. This pattern indicates a considerable amount of shared variance between groups but with very weak relation to individual observations. $ICC1$ of 0.020 means that a maximum of 2 percent of the variance of IT project cost performance can be accounted for by the project type. Similarly $ICC1$ of 0.0335 means that a maximum of 3 percent of the variance of IT project cost performance can be accounted for by the source of the data. In other words only 2 percent of the variance are shared among group members of a specific type and only 3 percent of the variance are shared by projects from the same source. Consequently, not accounting for project type or

data source would not introduce any strong biases in measuring IT project cost performance.

However, ICC2s are not close to zero. The structure of a low ICC1 and a high ICC2 characterizes emergent phenomena because ICC1 and ICC2 are related through group size (Bliese 2000), in other words low ICC1 and high ICC2 are indicative of group-level phenomena that are not evident in lower-level data. This interpretation has two immediate implications. First, viewing the results of the analysis through the lens of the reliability of measurement instruments, individual projects are not reliable indicators to identify the cost performance of project types. For measurement purposes large sample and group sizes are needed to characterize the aggregated cost performance, for instance to compare the typical ERP project compared to the typical MIS project. Second, the difference of cost performance between project types and data sources, is nearly undetectable when comparing individual projects. The data, in terms of independence, indicate that cost performance of one individual project is nearly completely independent from the project type or the data source. Yet, when looking at large groups of projects differences in cost performance emerge between different project types and data sources.

The analysis of cost performance by project type also provides evidence to how cost performance might be understood. When aggregating variables to higher levels compilation and composition processes are at play (Bliese 2000).

Compilation processes reflect the theoretical notion of discontinuity, while

composition processes reflect the notion of isomorphism (Bliese 2000, p. 366). The concept of composition depicts lower-level data as qualitatively different from higher-level data, the classic example is measuring gender on the lower level to form an aggregated group homogeneity measure on the higher level.

Compilation, on the other hand, depicts lower-level data as similar to each other and similar to higher level constructs (Bliese 2000, pp. 366-367). The explanatory process is isomorphism. Cost performance of projects is, by definition of this multilevel analysis, not qualitatively different from the cost performance of programs, organizations or project types. Hence, the observed differences between project types might be the outcome of a fuzzy composition process (Bliese 2000, pp. 369-376). Fuzzy composition processes are non-isomorphic or only partially isomorphic. Thus aggregated measurements are different while, at the same time, related to individual observations. The partial isomorphism results in 'subtle' group level phenomena (Bliese 2000). In this context, 'subtle' means that individual project cost performance lacks the shared unit property, i.e., cost performance by type. Extending this observation, the results imply that the group levels, project type and data sources, measure a different cost performance from the individual-level observation of project cost performance while retaining a nomological link (Bliese 2000) between the two levels.

In terms of reliability, the results show that individual project cost performance is an unreliable indicator of cost performance of project types and data sources. Equally, cost performance of project types and sources of data are unreliable

indicators for cost performance of individual projects. In terms of independence between observations, the data indicate a lack of isomorphism that would, if present, have created more shared variation between the individual observations. As result of the fuzzy, partially isomorphic composition process, the group level of project type performance can be conceptualized of as a separate construct nearly independent from the individual observations of project cost performance.

3.4.5 Summary of the Sample Statistics

The descriptive statistics showed that the sample collected is larger than most of the previous samples collected. The sample spans a wide range of organizations, project and system types. The projects in the sample are larger, in terms of budget, than those in previous studies. Lastly, this raised the question about data heterogeneity the analyses of the intraclass correlation coefficient using a multilevel model of IT project cost performance showed that the project cost performance is independent of program cost performance and geography and vice versa. The analysis also showed that organization is a covariate that needs special attention and ideally needs to be controlled for. Finally the analysis showed a somewhat more complex pattern for project types and data sources. Here the findings indicated low dependence and reliability between the cost performance of individual projects and higher-level variables, but with large sample sizes emergent group-level differences that can be detected which indicates the presence of weak isomorphic processes, if any, within project types and data sources.

3.5 Summary of the Methods Chapter

This chapter documented the steps this thesis has taken to collect and prepare the data. The chapter first documented how the variables were translated into a questionnaire with an anchored 10-point scale. The chapter then documented how the data were collected and prepared for data analysis. Next the chapter derived the statistical analyses needed to fit the tail of the distribution and test the hypotheses. Lastly, the chapter described the sample and analyzed the issue of heterogeneity in the sample, which concluded that of the potential covariates and levels the organization is the one level that needs to be controlled for when testing the hypotheses in the next chapter. After this preparatory work, chapter 4, will use the sample to answer the first research question: How frequent are outliers, that is markedly deviant observations, in IT project performance? Before chapter 5 will test the variables that potentially explain why outliers occurred in IT project performance.

4 Findings: How Frequent are Outliers in IT Project Performance?

4.1 Outline of the Chapter

Chapter 1 illustrated the problem of the adverse impact of outliers, in the sense of markedly deviant observations of IT project performance. Chapter 1 closed with posing two research questions: (1) How frequent are outliers, in the sense of markedly deviating observations, in IT project performance? (2) Why did outliers occur in IT project performance? This chapter presents the findings for the first research question.

The literature review, chapter 2, found that no prior studies exist that answered the question of how frequent outliers are. The literature review drew on related fields of literature, namely the study of common cost overruns, schedule delays, and benefits/ functionality/ quality shortfalls and studies of distributions of other IT project indicators. The literature review showed that cost overruns, schedule delays, and benefits shortfalls were common, and that the large variation of reported project performance within and between studies might indicate the presence of outliers. With regards to IT projects a normal distribution is commonly assumed, only PERT studies of effort and schedule distribution of tasks have previously assumed fatter tails, using e.g., lognormal, Weibull, Gamma distributions, but not with extreme value distributions, e.g. generalized Pareto distributions. The literature review derived three research hypotheses:

- H1: Under the assumption of thin tails, markedly deviant observations of IT project performance did not occur.
- H2: Performance of IT projects is normally distributed.
- H3: IT project performance data fit power law tails, an extreme value distribution.

This chapter will now analyze each of the first three hypotheses in turn.

4.2 Hypothesis 1: Under the Assumption of Thin Tails, Markedly Deviant Observations of IT Project Performance did not Occur

This section will test the first hypothesis along the IT project performance dimensions of cost, schedule, benefits, and effort.

4.2.1 Cost Performance

The 4,307 IT projects in the sample had an average cost overrun of +106.7%. In other words, on average the projects more than doubled their estimate. At the same time the sample median was $\pm 0\%$. The bottom quartile was -11.1%, the top quartile +35.8%. Thus the interquartile range was 46.9 percentage points.

Following Tukey's convention that the whiskers span four interquartile ranges, outliers are observations that are further than 1.5 interquartile ranges away from the box in a boxplot (Tukey 1977). Observations that fell below -81.4% cost overrun were outliers in the left tail. Observations that fell above +106.1% were outliers in the right tail. It is unusual that the sample average falls beyond the whiskers and is a potential outlier.

This gives a clear indication that skew in the sample was considerable - to the point where the arithmetic mean is highly influenced by observations in the tail. The sample standard deviation for cost performance was 957.6 percentage points.

The robust measures, median and median absolute deviation give a better indication of how the performance is spread. For example, the standard deviation is calculated as the square root of the average squared distance of all

observations from the mean. Thus large observations have a large impact since their distance is squared. A robust measure of variation is the median absolute deviation, which is the median of the deviation of each observation from the sample median. The sample median was $\pm 0\%$ with a median absolute deviation⁵⁰ of 27.5 percentage points, the unscaled median absolute deviation was 18.5 percentage points. In other words, half the observations were further than 18.5 percentage points away from the median of the sample and half the observations fell within 18.5 percentage points above and below the median.

The non-parametric Wilcoxon can tests the center of the sample distribution by ranking all observations by their distance from $\pm 0\%$ cost overrun – by ranking the observations the influence of large outliers is limited at the cost of losing information. The test rejects with overwhelming significance the null hypothesis that the sample centers symmetrically on $\pm 0\%$ cost overrun ($W = 3,184,822$, $p < 0.001$).

In terms of outliers, 59 observations in the left tail had cost underruns smaller than -81.4%. The rate of outliers in the left tail is 1.6% of all observations. In the right tail 465 observations fell above +106.1%. This corresponds to 12.7% of all observations.

In a thin-tailed distribution, such as the standard normal, the frequency expected of outliers is maximum 0.7% or in each tail 0.35%, if the distribution is

⁵⁰ The median absolute deviation is the median (M) of the differences between observations and the sample median $MAD = 1.486 * M(|M(x) - x_i|)$. The constant 1.486 scales the MAD so that for the standard normal distribution $N(0,1)$ MAD equates the standard deviation.

symmetric. The number of outliers above and below a cut-off point follows a binomial distribution. A thin-tailed distribution would expect 26 observations beyond the whiskers of the boxplot, the data show 524 observations outside this range. The Binomial distribution for large n estimates that the 95% confidence interval of the rate of outliers is between 13.2% and 15.4%. This is statistically significantly higher than the expected rate of 0.7% under the thin tail assumption (Poisson test, $p < 0.001$).⁵¹

4.2.2 *Schedule Performance*

The IT projects in the sample had an average schedule delay of +38.9%. The median schedule delay was $\pm 0\%$. The bottom quartile was also $\pm 0\%$, which indicates a very high concentration of observations at $\pm 0\%$ with limited observations in the left tail. The top quartile was +36.5%. Thus the interquartile range was 36.5 percentage points.

Again, outliers were defined as observations more than 1.5 interquartile ranges away from the box in the boxplot. Outliers were observations that fell below -

⁵¹ Three tests are commonly used to test the rate of rare events, all tests have slightly different biases in constructing the confidence interval. (1) *Poisson test*: The number of outliers is binomially distributed, which is approximately a Poisson distribution with $\lambda = n * p$, where p is the rate of expected occurrences. A Poisson test with $\lambda = 26$ estimates the 95% confidence interval of outliers between 13.1% and 15.5%. (2) *Rate ratio test* is a uniformly most power test on the ratio of rates of outliers of two Poisson counts. Here the first count (the null count) is the theoretical rate of 0.7%. The test assumes the null hypothesis that the ratio of the two rates is 1, i.e., that 0.7% over 14.3% is equal to 1. The 95% confidence interval of the rate ratio translates into an interval of the rate of outliers between 9.6% and 22.1%. (3) *Fisher's exact test for count data*, has the null hypothesis that the odds ratio of expected to observed rate is 1. The 95% confidence interval of the odds ratio translates into a probability of outliers between 10.0% and 20.3%. The p -values for all tests were consistently highly significant, with $p < 0.001$.

54.8% or above +91.3% schedule delay. The sample mean for schedule overruns fell outside the box but within the whiskers of the boxplot.

The median absolute deviation was 8.6 percentage points (5.8 percentage points unscaled). In other words half of the observations fell within a range of 5.8 percentage points above and below the median schedule delay. The sample standard deviation was 126.0 percentage points. This again indicates considerable skew in the data due to large schedule overruns. The large difference between the scaled median absolute deviation and the standard deviation questioned the usefulness of standard deviation as a measure of variation. The non-parametric Wilcoxon test with the null hypothesis that the location of the sample distribution is $\pm 0\%$ is statistically significant and can reject the null hypothesis ($W = 137,366$, $p < 0.001$). This gives further evidence that the sample is skewed towards overruns.

In terms of outliers, 27 observations fell below the cut-off point of -54.8% schedule delay and 122 observations fell above the cut-off point of +91.3% schedule delay. This corresponds to 15.4% of the observations (2.8% in the left tail and 12.6% in the right tail). The binomial distribution estimates the 95% confidence interval around the observed occurrence of outliers between 13.2% and 17.8%. The 15.4% rate of observed outliers is statistically significantly higher than the assumption of thin tails suggests where the expected rate of outliers is maximum 0.7% (exact tests $p < 0.001$).

4.2.3 Effort Performance

The IT projects in the sample had an average effort overrun of +118.0%. The median effort overrun was +25.4%, the bottom quartile was $\pm 0\%$, the top quartile +110.1%. Similar to the schedule overrun the mean falls beyond the box in the boxplot, i.e., above the top quartile but within the whiskers of the boxplot. Consequently, the interquartile range is 110.1 percentage points.

Effort outliers can only occur in the right tail above a cut-off of +275.3%. The lower bound of the effort overrun distribution is less than 1.5 interquartile ranges away from the bottom quartile. Outliers in the left tail would fall below -165.2%, which is per definition an impossible outcome.

Sample standard deviation is 535.7 percentage points. Standard deviation is a very large given the sample mean and the location of median and quartiles, again this shows that mean and standard deviation are influenced by large effort overruns. The median absolute deviation was 54.3 percentage points (the unscaled median absolute deviation was 36.6 percentage points), i.e., half the observations in the sample fell within 36.6 percentage points around the median.

The non-parametric Wilcoxon test was able to reject the null hypothesis that the location of the sample distribution is $\pm 0\%$ effort overrun ($W = 8388$, $p < 0.001$). In the sample 10.1% of the observations were classified as outliers, i.e., falling into the right tail beyond the whiskers of the distribution in a boxplot. Given the small sample size for effort estimates the uncertainty around the observed

number of outliers is greater than for schedule and cost overruns. The observed rate of 10.1% has a 95% confidence interval based on the binomial distribution between 5.9% and 15.9%. The rate of observed outliers is statistically significantly higher than expected (exact tests all $p < 0.001$).

4.2.4 Benefits Performance

The IT projects in the sample had an average benefits shortfall of -24.2%. For benefits shortfalls a positive value of actual-forecasted/forecasted indicates that more benefits were realized than planned, which is a positive outcome, a negative outcome indicates that less benefits were realized than planned, which is a negative outcome. Thus the direction of the scale of benefits shortfalls is opposite that of cost, schedule, and effort performance. The median benefit shortfall was $\pm 0\%$. The top quartile was $\pm 0\%$, which indicates that benefits performance is skewed towards benefits shortfalls. The bottom quartile was -48.5%. Thus the interquartile range was 48.5 percentage points. The standard deviation is 64.9 percentage points and the median absolute deviation is zero percentage points, which shows that the data are very lumpy on the median, which is unsurprising given the small sample size. The non-parametric Wilcoxon test rejected the null hypothesis that the location of the sample distribution is $\pm 0\%$ benefits shortfalls ($W = 86$, $p < 0.001$).

Applying Tukey's definition, outliers are observations that have a shortfall of less than -121.3% or more than +72.8%. As discussed earlier the value of -121.3% is counterintuitive. The following example illustrates how the benefits

shortfall can be less than -100%: A project aimed to create benefits from saving IT cost by replacing legacy systems with new systems that are cheaper to operate, e.g., saving 5 million per annum by reducing annual running costs from 15 to 10 million. Benefits shortfalls exceed -100% if the project underestimated the true costs of operation of the new system to the extent that the project implemented a solution that costs more than the system it replaced, e.g., the new system costs 17 million to operate instead of the estimated 10. In this example, the operations cost estimate has been off by 7 million. In the cost-benefit analysis the change in operations cost reduced the benefits from 5 million to -2 million, which is an $\text{actual/estimate} = -0.4$ and as $(\text{actual-estimate})/\text{estimate} = -140\%$, which exceeds -100%.

Both outcomes in the left tail (where projects that tried to achieve positive benefits actually had negative benefits) and in the right tail (where more positive benefits were achieved than forecasted) were observed. 2 outliers in the left tail fall below the cut-off value and 3 observations fall above the cut-off value in the right tail. This gives an outlier rate of 6.3% in the sample. The Poisson test rejected the null hypothesis that the rate is the same as the thin-tailed expectation of 0.7% ($p < 0.001$). The Binomial distribution identifies the 95% confidence interval of outliers between 2.1% and 14.2%.

4.2.5 *Summary of Findings of Hypothesis 1*

Table 7 summarizes the findings for the four dimensions of cost, schedule, effort, and benefits performance. In all four dimensions the tests rejected the

null hypothesis that markedly deviating observations, or outliers, do not occur more frequently than in the thin-tailed standard normal distribution, where the expected rate of outliers is maximum of 0.7%. Moreover, the location parameters showed that because of this high rate of outliers the standard deviation and the mean are inadequate measures to characterize IT project performance. More robust measures, i.e., median, quartiles, interquartile range, and median absolute deviation give a better overview of the data. These measures and the boxplots demonstrate that considerable skew is present in the data, with a lumpiness around being on-budget, on-time, on-effort, and on-benefits (cf. figures 1 and 2).

Table 7 Summary of Project Performance Distribution

	Cost	Schedule	Effort	Benefits
1st quartile	0.889	1.000	1.000	0.515
Median	1.000	1.000	1.254	1.000
Mean	2.067	1.389	2.180	0.758
3rd quartile	1.358	1.365	2.101	1.000
n	3672	965	158	79
IQR	0.469	0.365	1.101	0.485
Cut-off outlier left tail <=	0.186	0.453	-0.652	-0.213
Cut-off outlier right tail >=	2.061	1.913	3.753	1.728
SD	9.576	1.260	5.357	0.649
MAD ⁵²	0.275	0.086	0.543	0.000
MAD (unscaled)	0.185	0.058	0.366	0.000
No. of outliers left tail	59	27	0	2
No. of common observations	3148	816	142	74
No. of outliers right tail	465	122	16	3
Outliers left tail	1.6%	2.8%	0.0%	2.5%
Outliers right tail	12.7%	12.6%	10.1%	3.8%
Outliers both tails	14.3%	15.4%	10.1%	6.3%

⁵² MAD is calculated using the standard formula in R, which incl. a normalization constant of 1.4826 so that the E(MAD)=SD in normally distributed data.

Wilcoxon test; H_0 : True
location = 0

W	3184822	137366	8388	86
p	< 0.001	< 0.001	< 0.001	< 0.001
Over incidence left tail	359%	699%	-100%	623%
Over incidence right tail	3518%	3512%	2793%	985%
Over incidence both tails	1939%	2106%	1347%	804%
Outliers	524	149	16	5
Expected	26	7	1	1
p-value (Poisson)	< 0.001	< 0.001	< 0.001	0.004
Confidence interval of frequency of outliers	[0.132, 0.154]	[0.132, 0.179]	[0.059, 0.159]	[0.021, 0.142]

Figure 1 Boxplot Project Performance (actual/estimate, LN Scale)

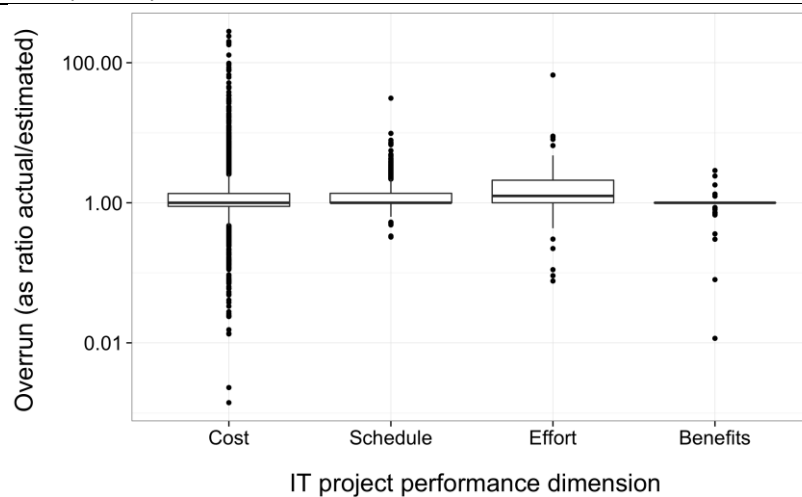
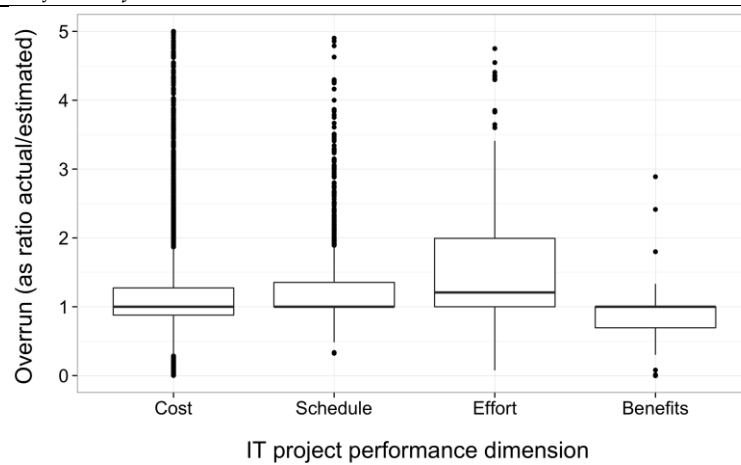


Figure 2 Boxplot Project Performance (Linear Scale, truncated at $0 \leq \text{actual/estimate} \leq 5$)



4.3 Hypothesis 2: Performance of IT projects is not Normally Distributed

Table 8 Anderson-Darling Test for Normality

	Cost	Schedule	Effort	Benefits
Test statistic A	INF	INF	INF	8.685
p-value	< 0.001	< 0.001	< 0.001	< 0.001

Table 8 shows the results of the test for normality. The test rejects the null hypothesis that the data is normally distributed with an overwhelming level of statistical significance $p < 0.001$.

So far, this chapter rejected the first hypothesis that outliers, that is markedly deviant observations of IT project performance, do not occur more frequently than in a thin-tailed normal distribution. The second hypothesis was also rejected, namely that cost overruns, schedule delays, effort overruns, and benefits shortfalls can be approximated by a normal distribution. In sum, the rejection of the first two research hypotheses has established that the observation of outliers is not a chance event and that it refutes the assumption of normality. Now, the third research hypothesis questions whether power laws, an extreme value distribution, provide an appropriate new working assumption to describe the data.

4.4 Hypothesis 3: IT Project Performance Data Fit Power Law Tails, an Extreme Value Distribution

The test of the third research hypothesis - the tail of the data can be fit by a power law - requires five steps:

1. Check with QQ-plots whether the cumulative distribution function follows an extreme value distribution
2. Visual fit on a log-log binned histogram
3. Estimate parameters using maximum likelihood estimator (cf. chapter 3.3.2)
4. Perform goodness of fit tests on the estimated distribution.

4.4.1 Visual Fit to Common Extreme Value Distributions

Figures 3-5 show the visual fit of the tail of the cost, schedule, and effort performance distribution. Benefits performance excluded due to insufficient sample size, only 27 observations had a benefits shortfall of more than 0%, i.e., actual/estimated benefits < 1.

Figures 3-5 depict QQ-plots. QQ-plots show the quantiles of a theoretical assumed distribution on the x-axis and the empirical quantiles of the sample on the y-axis. If the QQ-plot approximates a 45 degree straight line the family of distributions is a candidate for fitting the data. The QQ-plots in the figures 3-5 depict the families of distributions, which have been used to model IT project indicators before (cf. chapter 2.3.2). The QQ-plots show consistently for all three dimensions, cost, schedule, and effort performance, that the data deviate from a *Normal distribution* and *Student's distribution*.

Distributions with fatter tails, that have been studied before were Poisson, Lognormal, Weibull, Beta, and Gamma⁵³. The QQ-plots show that *lognormal*, *Poisson*, *Weibull*, *Beta*, and *Gamma* are better representations of the data than the normal distribution, i.e., the curve in the QQ-plots is closer to the diagonal. The QQ-plots for schedule and effort performance show, that *Beta* and *Gamma* distributions fit the data with the exception of the very end of the tail of the distribution.

The *Generalized Extreme Value* distribution shows a similar pattern: The distribution might fit the data in the tail but is still under representing the true thickness of the tails. *Frechet*, *Cauchy*, and *Levy* are all distributions with very extreme values, which shows in the QQ-plots as a fit of the extreme ends but very condensed middle ranges.

Lastly, The QQ-plots depict three power laws: *Pareto distribution*, *Zipf-Mandelbrot distribution*, and *Finite Zipf-Mandelbrot distribution*. The *Pareto distribution* is consistently the best representation for all three dimensions of cost, schedule, and effort for the whole right tail of each distribution.

⁵³ The Erlang distribution is a special case of the Gamma distribution

Figure 3 QQ-plots Project Cost Performance

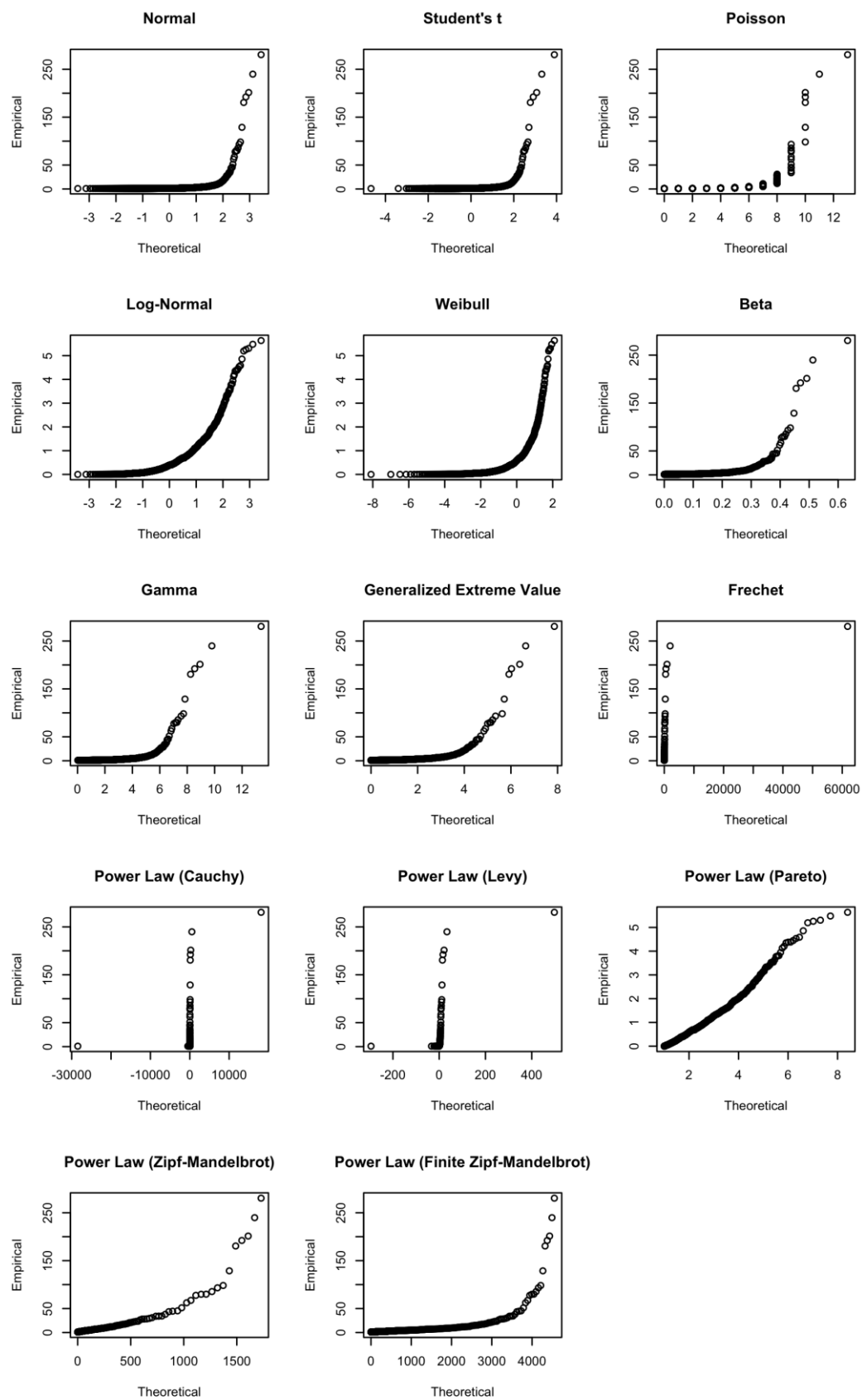


Figure 4 QQ-plots Project Schedule Performance

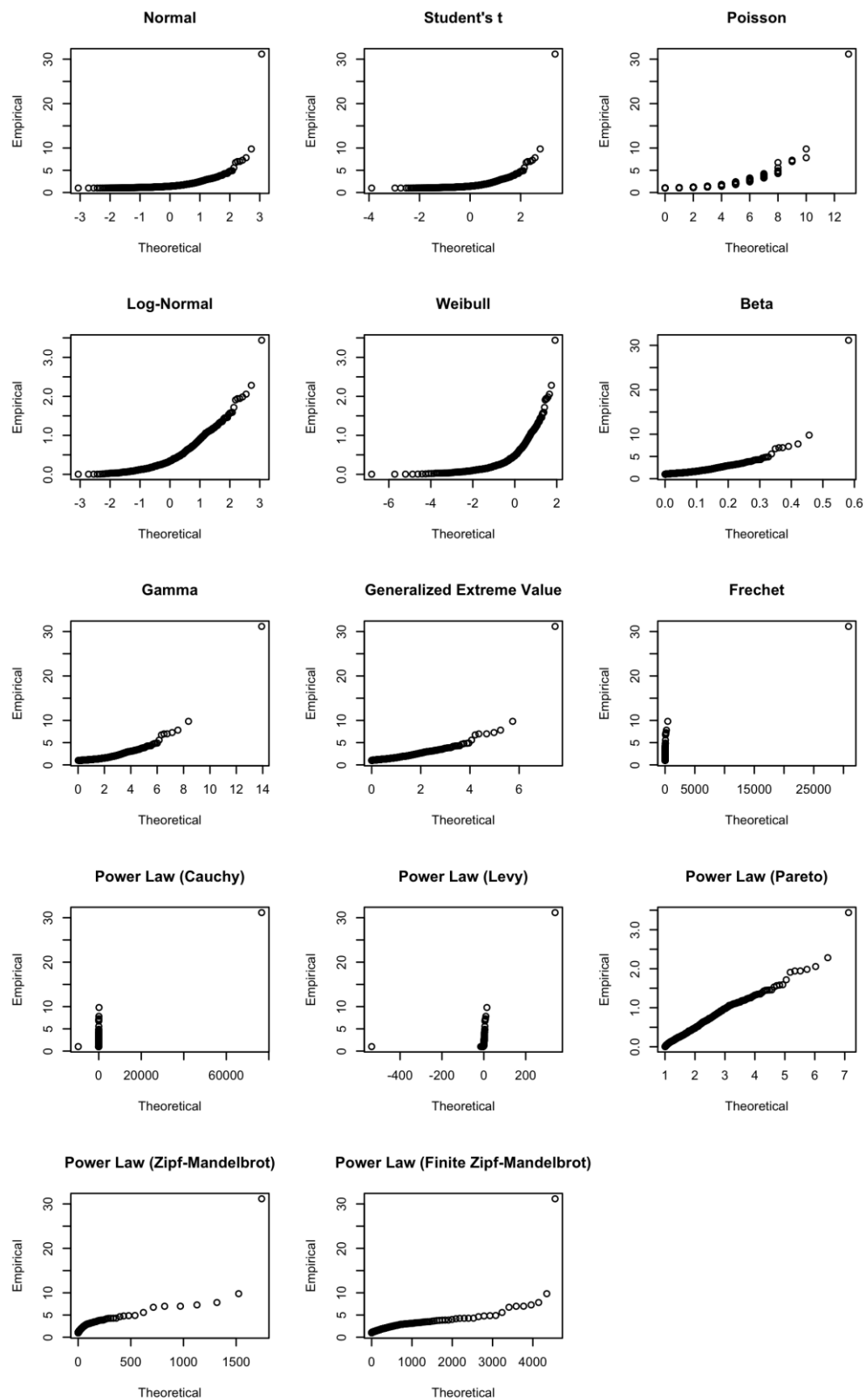
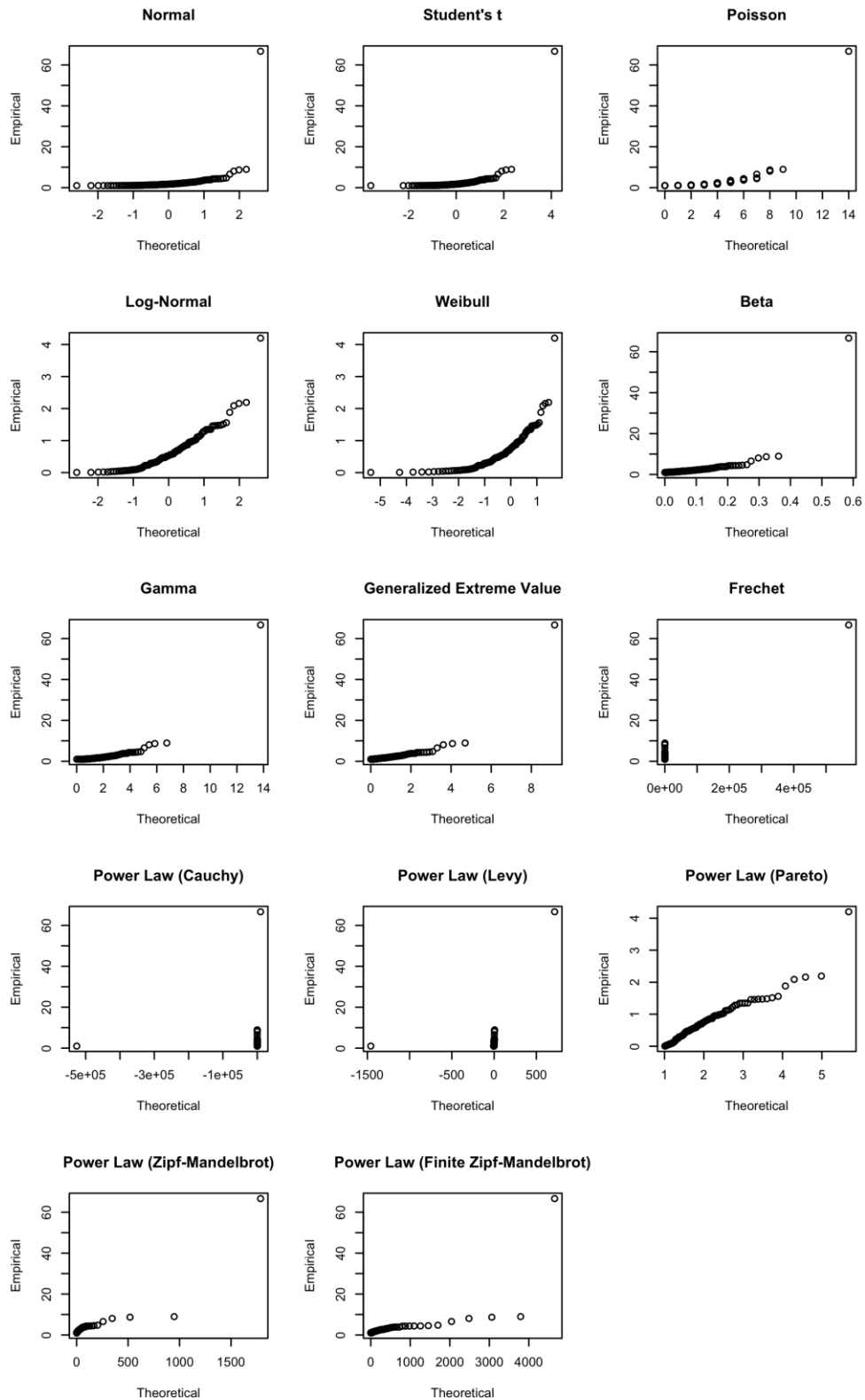


Figure 5 QQ-plots Project Effort Performance



4.4.2 Power Law Fit of the Tails

Figures 6-8 depict different resolutions of ln-ln histograms⁵⁴ for cost, schedule and effort performance. The x-axis shows the midpoints of the binned and ln-transformed actual/estimated cost, schedule, and effort. The y-axis shows the ln-transformed frequency count. Each graph shows the histogram with a different number of breaks, i.e. different widths of the bins. In each graph the number of breaks to create the bins is decreasing. Binning removes noise from the data, at the same time, it limits the autocorrelation of the data points that is introduced by sorting them.

The *cost performance* data show that the tail resembles a straight line on the ln-ln-transformed histogram (cf. figure 6). The fit line has a slope of -2.03⁵⁵. The binned histogram also indicate that alpha decays further towards the end of the tail.

Schedule performance (cf. figure 7) and *effort performance* (cf. figure 8) also depict a straight line but behave differently. The slope of the fitted line in the binned ln-ln histogram is -1.6 for schedule performance and -0.9 for effort performance. The data show that the fit is not accurately depicting the end of the tail (cf. Sornette and Ouillon 2012 argued that this is to be expected), which contains more noise than the cost performance data.

⁵⁴ In R notation log is the natural logarithm ln

⁵⁵ Goldstein et al. (2004) show that the fit on a ln-ln binned histogram is biased.

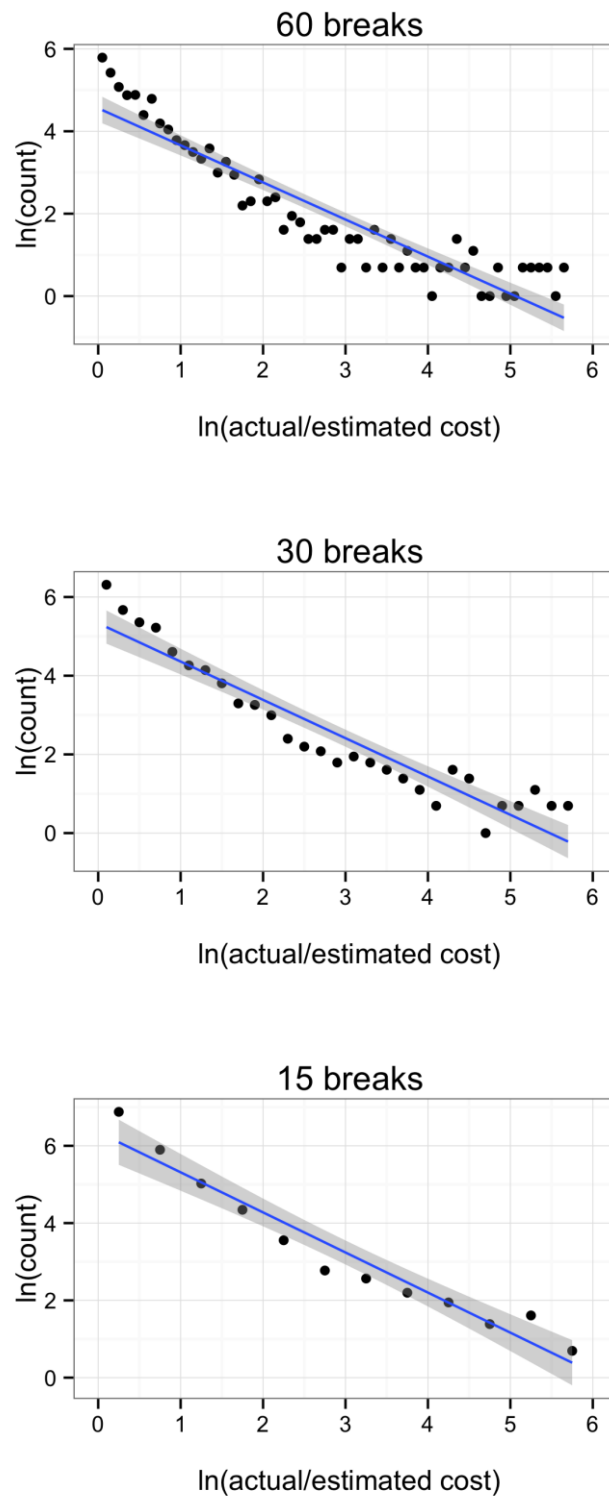
Figure 6 Binned LN-LN Histogram Project Cost Performance

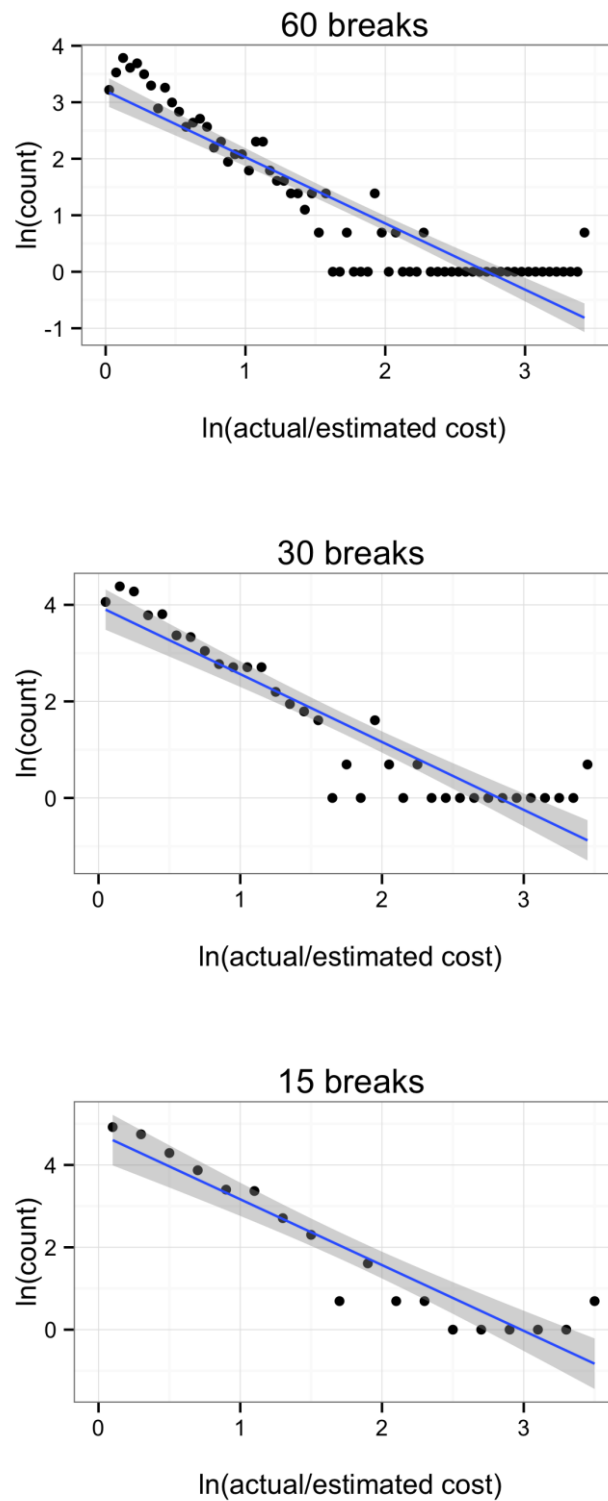
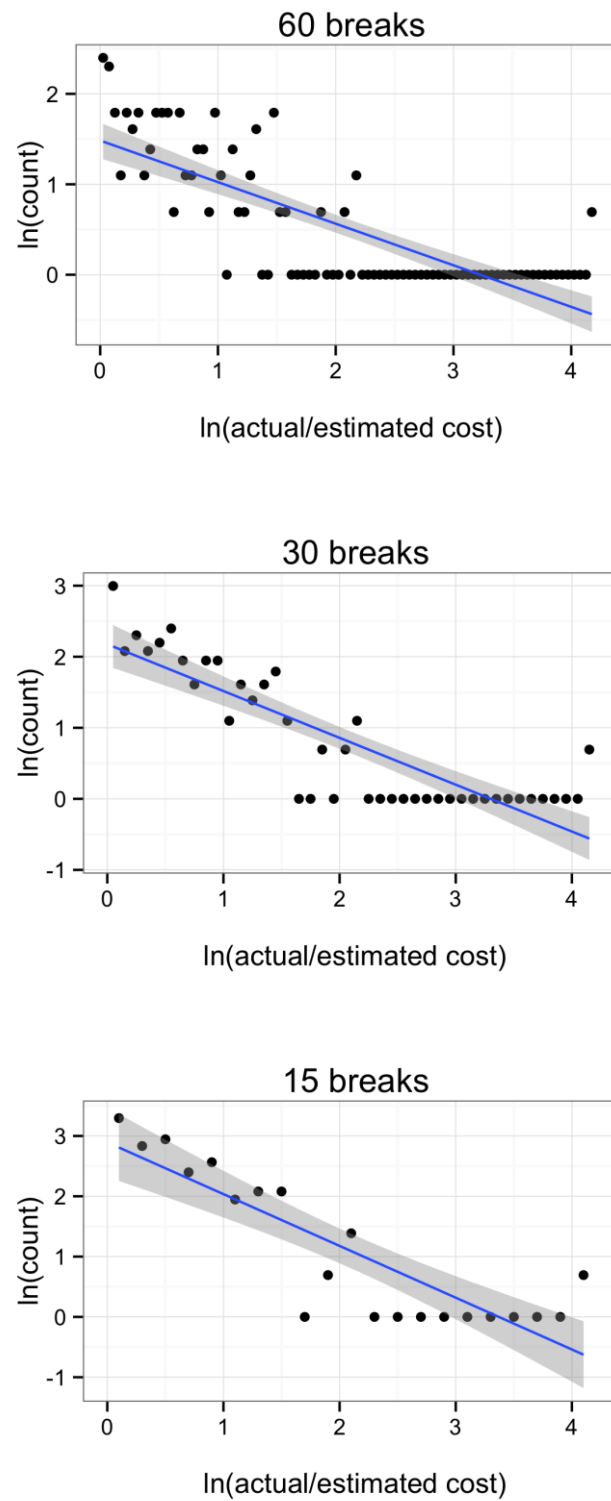
Figure 7 Binned LN-LN Histogram Schedule Performance

Figure 8 Binned LN-LN Histograms Project Effort Performance

The graphs in figures 6-8 reveal a regularity in the observations that become visible on a log-log histogram. The pattern illustrates that very large overruns can be forecasted from the occurrence of smaller overruns. The figures show transformed data using the natural logarithm, which has base $e \approx 2.7$. Thus the fit line of -2 in the cost overrun data, for example, shows that a cost overrun 3-times higher is about 7 times less frequent ($e^{-2} \approx 0.135 \approx 1/7$).

Fitting the tail of the distribution to a power law of the form $p(x|\alpha) = Ax^{-\alpha}$ with the maximum likelihood protocol (c.f. chapter 3.3.2, Newman 2005, Stouffer et al. 2005, Goldstein et al. 2004, Clauset et al. 2009) is less biased than the OLS fit to log-log histograms. Fitting a power law tail to the data follows a two-step approach first the model is fitted based on the minimum value $x_0 = 1$, which is the median in each dimension. Then the procedure iterates through all x in the dataset to minimize the Kolmogorov-Smirnov distance ($D = \max_{x \geq x_0} |S(x) - P(x)|$) between the empirical distribution $S(x)$ and the theoretical estimated distribution $P(x)$. Table 9 shows that the initial maximum likelihood estimate of alpha of the cost performance is 2.6. After minimizing D the tail is fitted to all $x \geq x_0 = 1.087$, which leads to a new estimated alpha of cost performance of 2.5. The estimated alpha of schedule performance is 3.1 before and after minimizing the distance between the empirical and the theoretical distribution; with a final supported tail of $x \geq x_0 = 1.091$. The estimated alpha after minimization of effort performance is 2.7 for $x \geq x_0 = 1.562$. Figures 10-12 graphically show the fitted distributions on a log-log plot where the x-axis

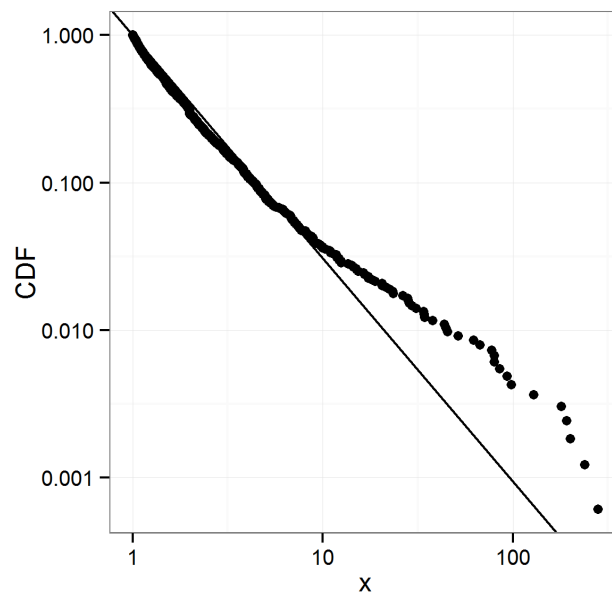
represents the overruns and the y-axis the inverse cumulative distribution function, which is defined as the probability of an observation larger than x ($Pr(X > x)$). Thicker tails, i.e., power laws with a smaller α , show as flatter lines.

The estimators indicate that a cost overrun double in size is about 6-times less likely, schedule overrun double in size is about 8-times less likely, and effort overrun is about 7-times less likely⁵⁶.

Table 9 Maximum Likelihood Power Law Fit

	Initial fit			Adjusted x_0 to minimize D		
	x_0	alpha	D	x_0	alpha	D
Cost	1.000	2.616	0.066	1.087	2.514	0.043
Schedule	1.004	3.089	0.066	1.091	3.089	0.035
Effort	1.006	2.481	0.068	1.562	2.685	0.058

Figure 9 Power Law Fit Cost Performance (Log-Log Plot of Inverse Cumulative Distribution Function (CDF) for cost overruns $x \geq x_0 = 1.087$, $a = 2.5$)



⁵⁶ Probability $p(x|\alpha) = Ax^{-\alpha}$, $2^{-2.5} \approx 0.18 \approx 1/6$, $2^{-3.1} \approx 0.12 \approx 1/8$, $2^{-2.7} \approx 0.15 \approx 1/7$

Figure 10 Power Law Fit Schedule Performance (Log-Log Plot of Inverse Cumulative Distribution Function (CDF) for schedule overruns $x \geq x_0 = 1.091$, $a = 3.1$)

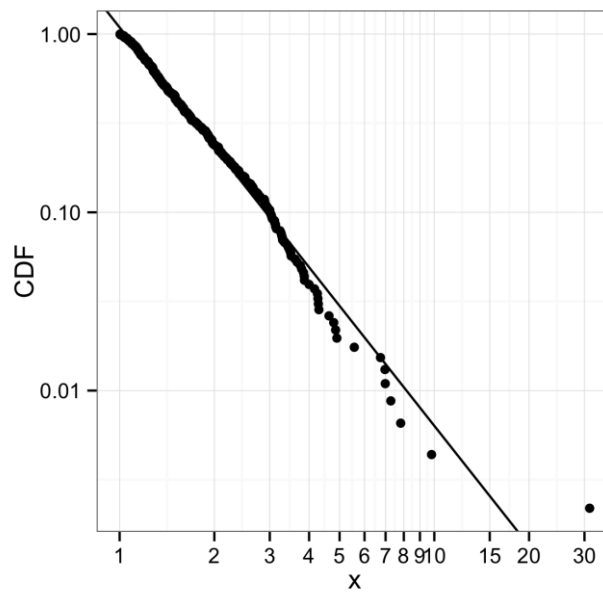
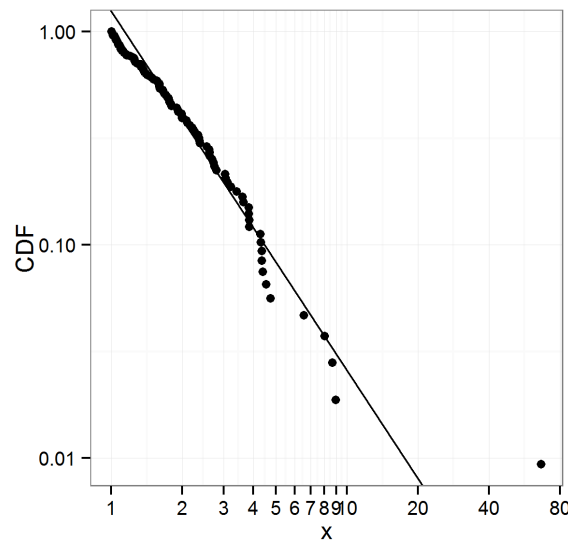


Figure 11 Power Law Fit Effort Performance (Log-Log Plot of Inverse Cumulative Distribution Function (CDF) for effort overruns $x \geq x_0 = 1.562$, $a = 2.7$)



To understand the uncertainty around the values of x_0 and α a non-parametric bootstrapping is performed. This is done by drawing a uniformly random sample of the same length as the sample size from the original data. Given a large number of draws, here 5,000, the uncertainty in the estimates of x_0

and alpha can be approximated. Figure 12 shows the results for the estimator of the cost performance distribution. The x_0 estimator has a cumulative mean of 1.903 with a cumulative standard deviation of 1.348. The alpha estimator has a cumulative mean 2.353 with a cumulative standard deviation of 0.193. As Figure 13 shows, the exponent and the cut-off value are negatively correlated. The bootstrap confirms the earlier impression from binned histograms that alpha is decreasing towards the end of the tail. That means if the cut-off value for the tail would increase the estimate for the tail would increasingly be thicker.

The bootstrap for the schedule performance leads to similar results. Figure 14 shows the cumulative average estimate and the 95% confidence interval of x_0 and alpha as well as their standard deviations. After 5,000 simulations the cumulative average x_0 was 1.119 and the cumulative average alpha was 3.244, with a standard deviation of x_0 of 0.146 and a standard deviation of alpha of 0.144. Figure 15 shows that the estimated alpha is positively correlated with the cut-off value of the tail, which is the opposite direction compared to the project cost performance.

Lastly, the bootstrap of the effort performance (cf. figure 16) shows that the cumulative average of x_0 is 1.343 with a standard deviation of 0.376. The cumulative average of alpha is 2.655 with a standard deviation of 0.537. The maximum likelihood estimate of parameter alpha shows a positive correlation

with the cut-off point in the tail (cf. figure 17), thus a cut-off chosen further out in the tail leads to higher estimates of alpha.

Figure 12 Uncertainty of the Fit of the Cost Performance Distribution

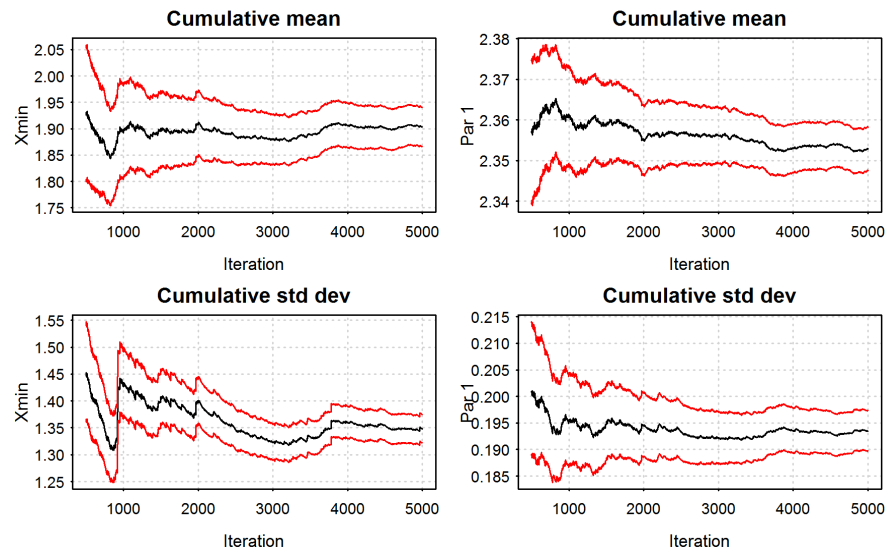


Figure 13 Uncertainty of alpha and x_0 of Cost Performance Distribution

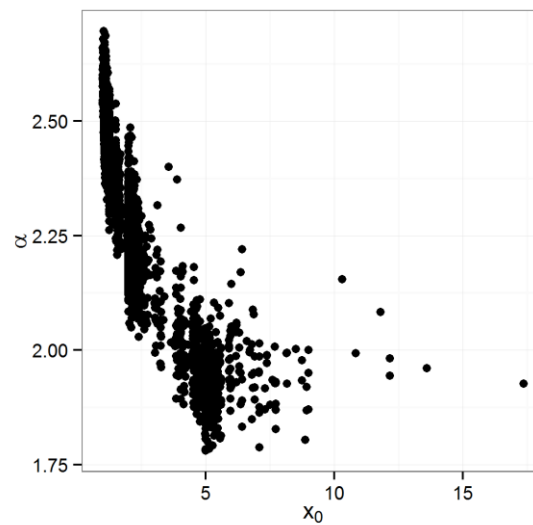


Figure 14 Uncertainty of the Fit of the Schedule Performance Distribution

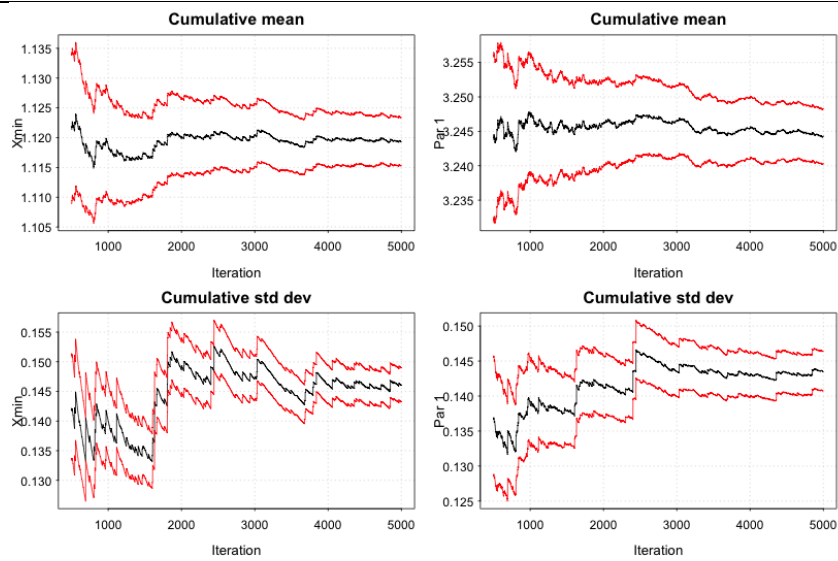
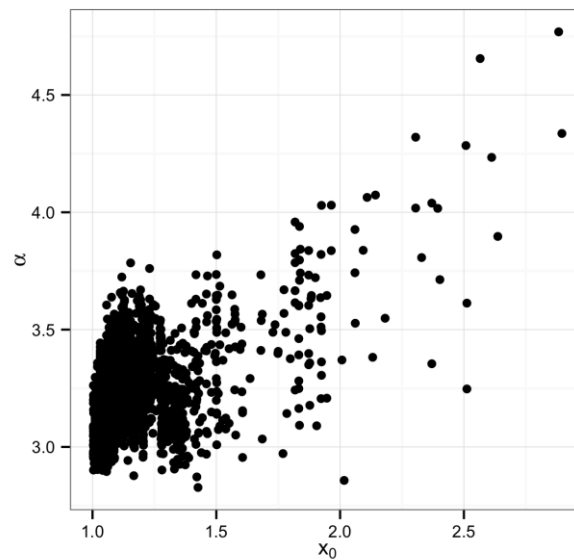
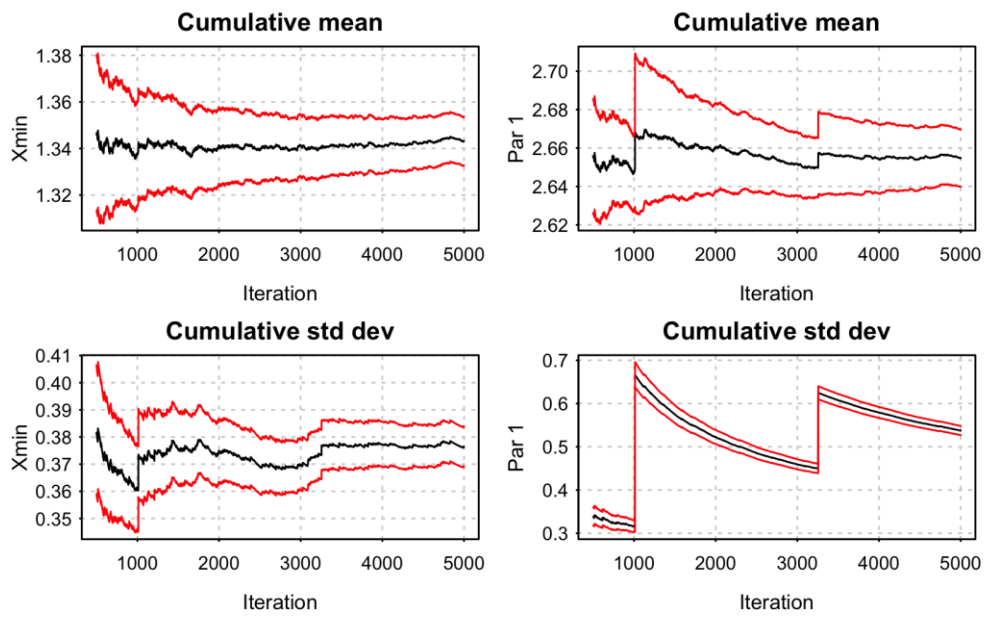
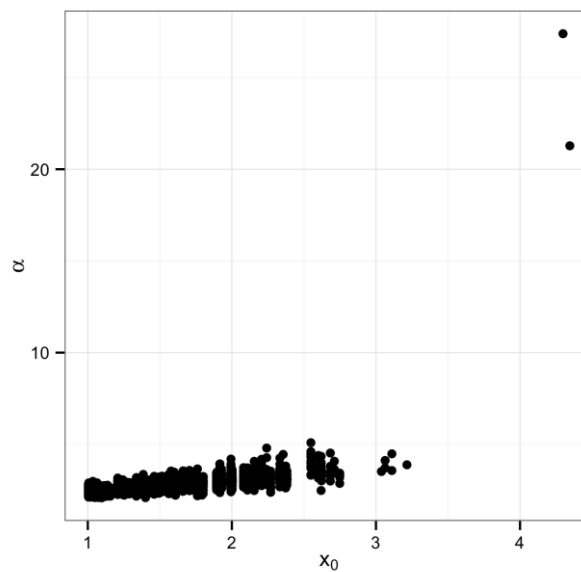
Figure 15 Uncertainty of α and x_0 of Schedule Performance Distribution

Figure 16 Uncertainty of the Fit of the Effort Performance Distribution

Figure 17 Uncertainty of α and x_0 of Effort Performance Distribution

4.4.3 Goodness of Fit Test

Maximum likelihood estimators have no statistical power, thus the estimator can be fitted to any data in the tail of a distribution without an indication of the goodness of fit. Moreover, the analysis derived the conclusion that it is best

described by a power law from the data itself. However, the goodness of fit can be established through bootstrapping (Clauset et al. 2009).

The null hypothesis of the goodness of fit test is:

H_0 : Data are generated from a power law distribution.

If the p-value of the bootstrap procedure is reasonably small, the null hypothesis can be rejected and an alternative distribution might be more suitable to fit the data. Figures 18-20 show the goodness of fit tests for cost, schedule, and effort performance. As before the non-parametric bootstrap uniformly draws 5,000 samples with replacement. The final cumulative averages for the p-values were highly non-significant (cost $p > 0.999$, schedule $p > 0.999$, effort $p = 0.998$). Thus the null hypothesis cannot be rejected that the data are generated from a power law distribution.

Figure 18 Goodness of Fit Test Project Cost Performance

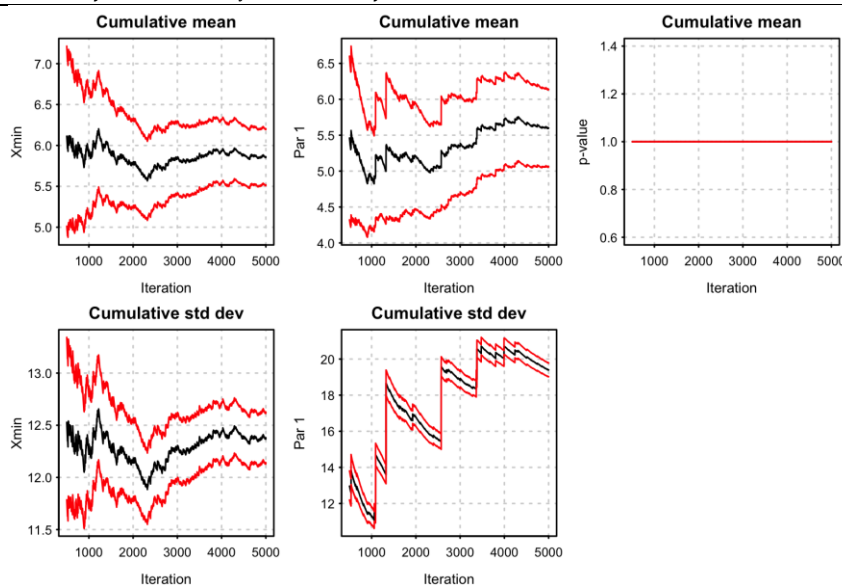


Figure 19 Goodness of Fit Test Project Schedule Performance

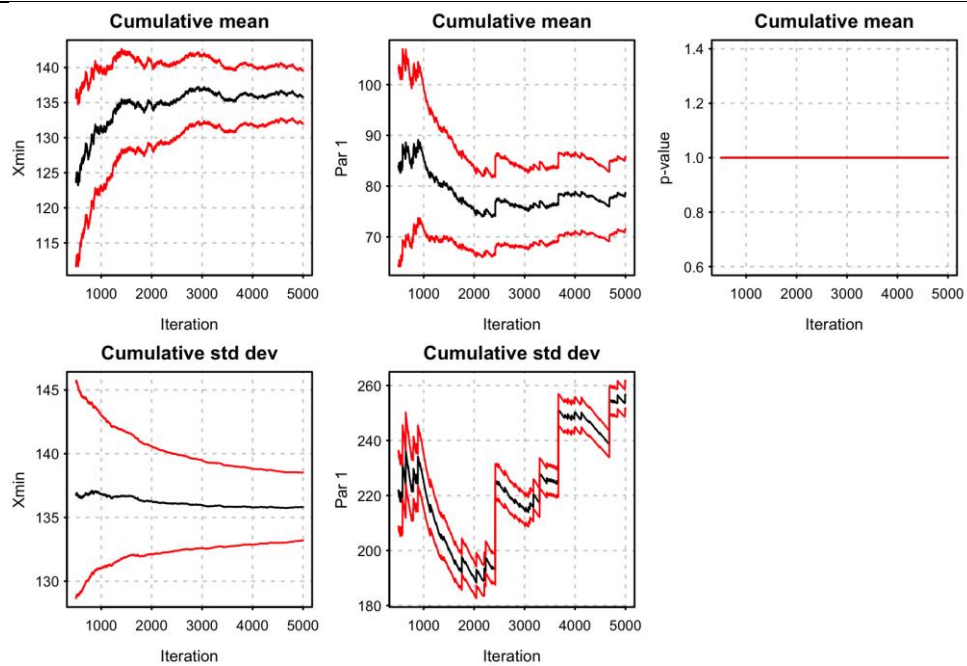
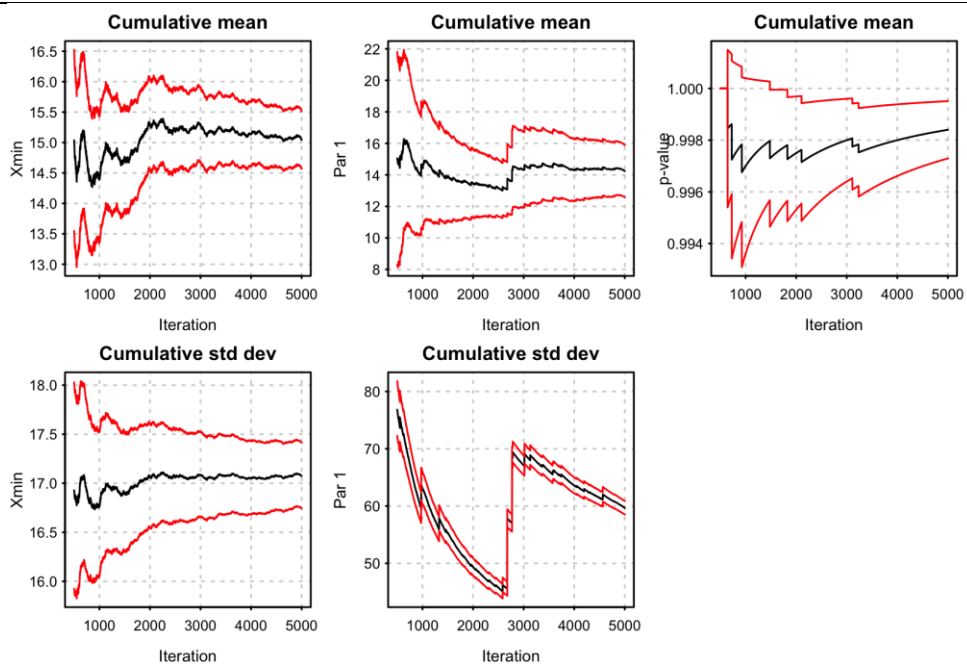


Figure 20 Goodness of Fit Test Project Effort Performance



4.4.4 Higher Order Moments

The beginning of the chapter showed that empirical means and standard deviations are strongly influenced by large overruns in the tails. The higher order moments of a power law tail are defined as:

$$\langle x^m \rangle = \frac{\alpha - 1}{\alpha - 1 - m} x_0^m$$

Most importantly, the higher order moments above $m \geq \alpha - 1$ are infinite. Table 10 lists the higher order moments for the fitted tails of the distributions. For cost, alpha is 2.5, which means that only the first order moment is defined, all other moments (variance, skewness, kurtosis) are infinite. According to Mandelbrot (1993) distributions whose variance is infinite are also called heavy tailed. The first order moment, the mean, is 3.2 or +220.0%. In other words, if a project overruns its budget it more than triples the forecasted cost. The mean schedule overrun is 2.1 or +109.3%, with a variance of 28.058. For schedule performance skewness and kurtosis are infinite. This is characteristic of a fat tailed distribution according to Mandelbrot (1993). The mean effort overrun is 3.8 or +284.1%, again all other higher moments (variance, skewness, kurtosis) are infinite, which makes it a heavy tail distribution.

Table 10 Higher Order Moments of the Cost, Schedule, and Effort Performance Tails

		Cost	Schedule	Effort
Fitted power law tail	x_0	1.087	1.091	1.562
	α	2.514	3.089	2.685
Moments	Mean	3.200	2.093	3.841
	Variance	∞	28.058	∞
	Skewness	∞	∞	∞
	Kurtosis	∞	∞	∞

The key consequence of the infinite moments is illustrated in figures 21 and 22 using synthetic data. For a range of sample sizes from $n=1$ to 10,000 the figures show the sample mean and variance for the same randomly generated sample. The figures illustrate how slowly the sample mean and variance converge to the population mean and population variance with increasing sample sizes, which is what makes them ill-defined. In other words, because large overruns frequently occur the noise in the data drowns out the signal, even though theoretically the weak law of large numbers still holds true.

As the figures also illustrate, the estimation of mean and variance from empirical data requires very large samples. The problem is so severe that to estimate the population variance for cost and effort overruns an infinite sample would be needed (Mandelbrot 1993), which of course is impossible. To put it differently, the noise in the data drowns out the signal to the extent that with practically obtainable sample sizes correct estimation of the population moments is highly questionable.

Figure 21 illustrates a *heavy-tailed power law* with exponent 2.5, which is the fit for the cost performance data. Even though the mean is defined it converges very slowly to the population mean. This slow convergence illustrates the weak law of large numbers. On the other hand, the variance, which is infinite, continuously grows with the sample size. To correctly estimate the population variance an infinite sample would be needed.

Figure 22 illustrates a *fat-tailed power law* with exponent 3.1, which is the fit for the schedule performance data. Here the tail is less fat than for cost performance, which is why the mean converges faster and the population variance is defined, even though it converges very slowly.

Figure 21 The sample mean and standard deviation for a heavy-tailed power law distribution with $\alpha=2.5$ for different sample sizes, for each sample size variance and mean are the estimates from the same synthetic data

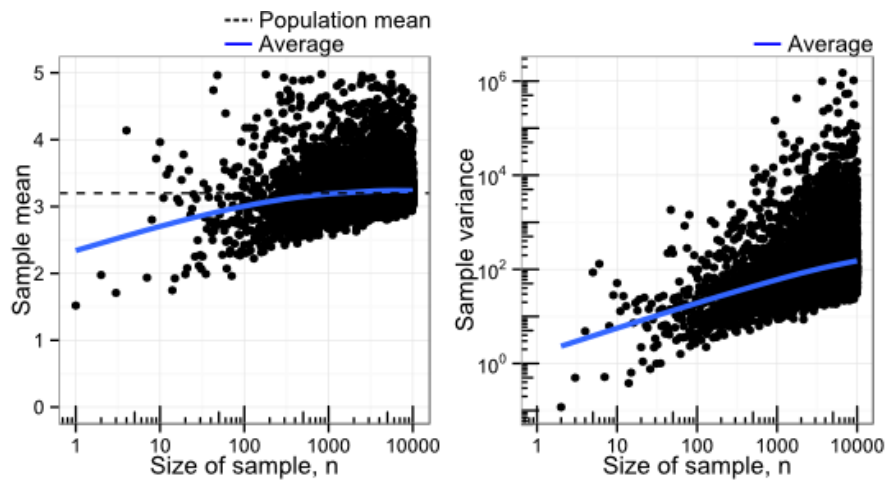
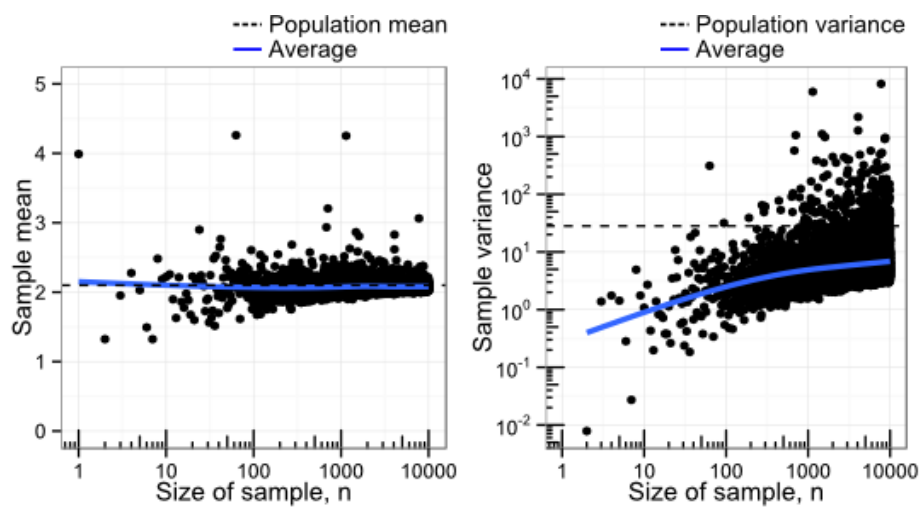


Figure 22 The sample mean and standard deviation for a fat-tailed power law distribution with $\alpha=3.1$ for different sample sizes, for each sample size variance and mean are the estimates from the same synthetic data



4.4.5 *Alternative Explanations of the Power Law Tails*

The fitted models showed that the distributions of cost, schedule, and effort performance have fat tails. However, three alternative explanations exist, why these fat tails could be a spurious result: (1) fat tails are the result of a mixture of log normal distributions, (2) fat tailed distributions are the result of a mixture of normal distributions with the same mean and different variances, and (3) fat tails are the results of aggregating observations over non-stationary parameters. The next sections will test whether each alternative explanation holds true. If the fat tails are a spurious result from aggregating observations they will disappear if the data is disaggregated. For example the “Power Law of Learning” (Anderson 2000) describes that the observation of how fast humans carry out a learned activity follows a power law. The key implication is that learning slows down over time. However, Murre and Chessa (2011) showed that this result is most likely an artefact of averaging the learning rates of different persons in a study. For an individual learning does not slow down, it follows an exponential distribution, but when aggregated it creates a power law. Consequently if the data were studied subject by subject the power law tail would disappear.

4.4.5.1 *Spurious Fat Tails as a Mixture of Lognormal Distribution*

A common alternative explanation of fat tail behavior⁵⁷ is that power law tails are better fitted to a lognormal distribution. Fat tails are thought to result from

⁵⁷ This alternative explanation goes back to the dispute around the paper by Barabasi (2005). The paper, published in *Nature*, found power law tails in response times to emails. The study later became the basis of a popular book (Barabasi 2010). The analysis was found to be flawed

multiplicative errors over time. Following the log rule $\log(a*b)=\log(a)+\log(b)$, multiplicative error could also arise from aggregating lognormal data, which is not fat tailed, in the Mandelbrotian sense, because all higher order moments are defined for the lognormal distribution. The Vuong test (Vuong 1989), a version of the likelihood ratio test, is used to test this alternative explanation. The Vuong test establishes whether the power law model performs significantly better than the lognormal model. The null hypothesis is that both theoretical distributions are indistinguishable, i.e., that both distributions are equally distant from the true distribution. The test statistic D is the ratio of the log-likelihoods of the data between the power law model and the lognormal model (chapter 3.3.2 used the same test to develop a strategy for inferences about power law tail behavior).

For the project cost performance models the test statistics was $D = 7.29$, which gave two-sided p -value of $p < 0.001$. This allowed to reject the null hypothesis that both fitted distributions are indistinguishable. The positive sign of the test statistic R indicated that the first model, which in this case the power law model, is closer to the true distribution of the data with overwhelming statistical significance.

For the project schedule performance models the test statistic $D = 1.74$ gave a p -value of 0.082, which indicates that the power law distribution fits better than

(Stouffer et al. 2005) to the effect that the phenomenon could better be explained with lognormal distributions.

the lognormal model and the test is nearly significant. Again the test indicated, this time with weaker statistical evidence, that project performance followed a power law and not a lognormal distribution.

As before project effort performance was problematic because of the small sample size of $n = 107$ for observations where $x > x_0 = 1$. The Vuong test of the power law model over the lognormal model gave a test statistic of $D = 1.38$ and a two-sided p-value of $p = 0.167$. Again the test statistics favored the power law model. The p-value, however, indicated that there is only weak statistical evidence that either model is further away from the true distribution.

4.4.5.2 Fat Tails as a Spurious Result of Mixtures with Same Means and Different Variances

Second, Fama (1965, p. 55) describes the alternative explanation of fat tails:

«Perhaps the most popular approach to explaining long-tailed distributions has been to hypothesize that the distribution ... is actually a mixture of several normal distributions with possibly the same mean, but substantially different variances. There are, of course, many possible variants of this line of attack, and little can be done to test them unless the investigator is prepared to specify some details of the mechanism instead of merely talking vaguely of 'contamination'. »

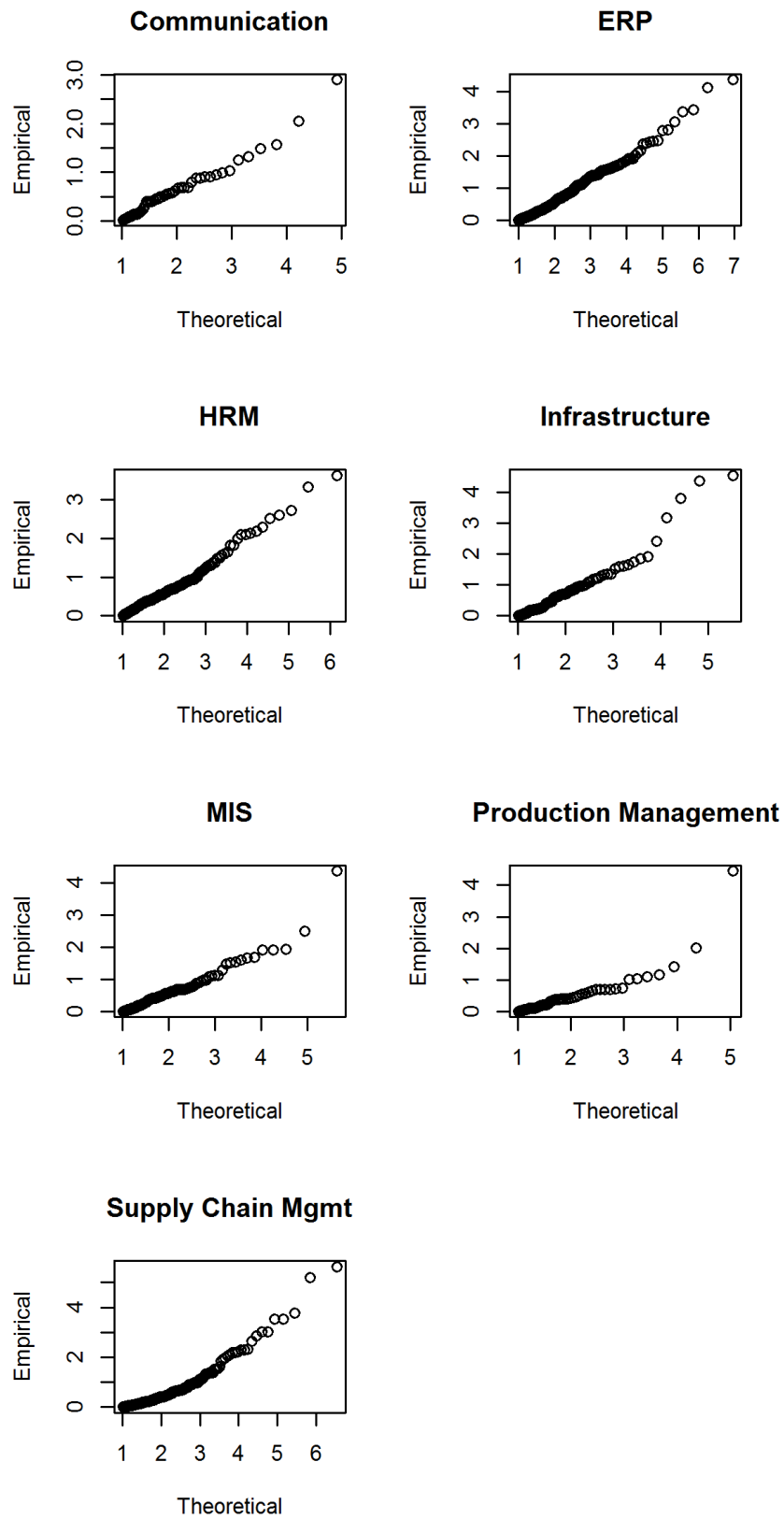
In other words, fat tails in a distribution can be the spurious symptom from mixing thin-tailed normal distributions with same means and different variances.

In this dataset an alternative explanation to the rejection of the third research hypothesis could come from inappropriately mixing different project types,

which could artificially create fat tails. In other words, the fat tails could simply be the result of mixing distribution with similar average performance but highly different variances. Figure 23 tested this potential alternative explanation. The figure shows the QQ-plots for power law tails for seven different project types, namely communication technology, ERP, HRM, IT infrastructure, MIS, production management systems, and SCM projects. All seven subsamples showed a straight line on the QQ-plot, which indicates that power laws were a suitable fit for project cost overruns. In other words, all three subsamples showed the same fat tail behavior.

Appendix 8.7 replicates this analysis with variables that could potentially explain how non-homogenous observations were mixed which led to spurious power law patterns at the aggregate level. The sample was split by project type, data source, country, year of the decision to build the project. Bootstrapping the goodness of fit⁵⁸ shows that the null hypothesis that the subsample does follow a power law cannot be rejected with the exception of cost performance in projects started in 2009. Thus the fat tails observed are most likely not an artifact of mixing the underlying performances of different project types.

⁵⁸ In this analysis with 1,000 uniformly drawn samples due to the long computing time to conduct each analysis

Figure 23 Power Law QQ-plots for Different Project Types

4.4.5.3 Fat Tails as a Spurious Result of Aggregating over Non-stationary Parameters

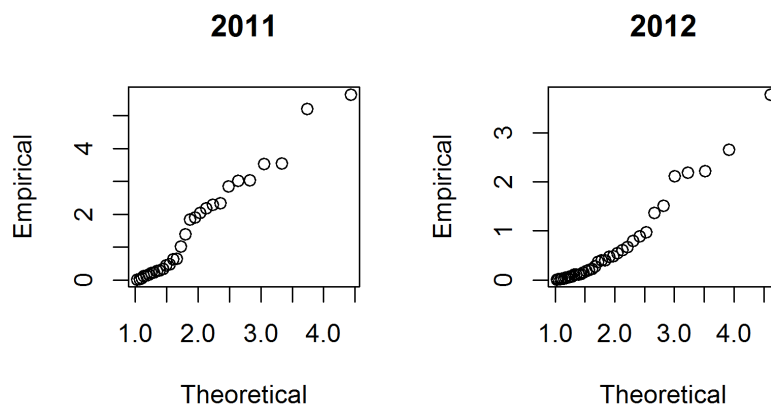
Third, fat tails also spuriously emerge if an outcome is measured over non-stationary parameters. For instance in the analysis of stock market prices fat tails appear if the underlying risk of the company changes over time. In this context, fat tails could be the spurious result of changes to the way an organization delivers or plans projects, e.g., introduction of lessons learned schemes, agile project methodology, and new incentive structures. In these three examples underlying parameters that influence project performance might change, similar to the stock market example where the underlying risk of investing in a company changes. In other words, organizational changes could have caused changes in the average project performance and/or the variance structure of project performances. In turn, aggregating project outcomes across those changes could artificially result in fat tails.

To check the problem of non-stationary underlying parameters, the one participating organization was selected where it was known that a major organizational change program tried to improve IT project performance. At the end of 2011, this organization re-organized the internal IT department. The company rolled out a «lean IT» program, which among other things introduced performance measurement, standardized project planning and execution, and refocused the performance-based pay for its project managers. The organizational changes came into effect in 2012. If the observed pattern in project performance was the result of aggregating over a non-stationary

underlying parameter, which influences project performance, then comparing the distributions of project performance before and after should show that the fat tails disappeared.

In 2011, the organization experienced a median cost overrun of +13%, with an IQR of 3.751. In 2012, the median cost overrun was reduced to -0.7%, with an IQR of 0.512. The data show that not only the location shifted but also the variation around it. A non-parametric Wilcoxon test, presented some statistical evidence that a change had occurred. The location shift is nearly significant at $p = 0.061$, with $W = 1589$.

Figure 24 QQ-plots Before and After Organizational Change



To check whether the aggregation of project outcomes across changing parameters has led to the observation of fat tails; the 2011 and 2012 Q-Q plots were checked. The 2011 and 2012 Q-Q plots for power law tails (cf. figure 24) showed that, within each year, the project performance of this organization showed fat tails. Despite the shift in location, i.e., the median reducing from +13% to -0.7%, and despite reducing the variability in project outcomes from

IQR = 3.751 to IQR = 0.512. The distribution of the cost performance of IT projects in 2011 and 2012 both can be suitably fitted to power laws. Table 11 shows the fitted power law parameters and the goodness of fit tests. The results shows that the project cost performance in 2011 had a fatter tail than the project cost performance in 2012. Nevertheless, in both subsamples the null hypothesis that the cost performance follows a power law cannot be rejected with $p_{2011} = 0.944$ and $p_{2012} = 0.864$.

Table 11 Goodness of Fit Test Power Law Tails in the Organizational Change Example

	2011	2012
x₀	6.27	1.12
Alpha	1.77	2.14
KS-Distance	0.083	0.099
Goodness of fit p-value	0.944	0.864

In sum, the analysis was based on the one participating organization where it was known that a major organizational change program was implemented with the aim to improve project performance. The organization changed multiple underlying variables that influence project performance. The results of which were visible in an increase in the power law exponent which describes the tail of the project cost performance. In other words, the changes implement reduced the thickness of the tail. Most importantly, the analysis failed to show that fat tails spuriously emerged from aggregating performance across non-stationary underlying parameters.

4.5 Summary of Findings

In sum, this chapter focused on the first research question: How frequent are outliers, in the sense of markedly deviations, in IT project performance? The sample showed that 14.3% of the projects were cost performance outliers, 15.4% were schedule performance outliers, 10.1% were effort performance outliers, and 6.3% were benefits performance outliers. Outliers frequently occurred and very high overruns are not impossible.

The first hypothesis, that outliers do not occur more often than in thin-tailed normal distributions was rejected for cost, schedule, effort, and benefits performance. The second hypothesis, that the cost, schedule, effort, and benefits performance are normally distributed was also rejected. The third hypothesis then tested whether power laws offer an alternative explanation. First, QQ-plots indicated that power laws are the most suitable family of distributions to fit the data. The assumption was further supported by the analysis of binned histograms. Finally, the maximum likelihood estimator was used to estimate the shape parameter and cut-off value for the tail in each dimension of project performance. Goodness of fit were adequate.

Lastly, the higher order moments were derived to approximate the population mean and population variance. Higher order moments were ill-defined: cost and effort performance have infinite variance; schedule performance has infinite kurtosis and skewness. In Mandelbrotian terms cost and effort performance are heavy tailed, schedule performance is fat tailed. The analysis

illustrated that the outliers, that is markedly deviant observations, frequently occur. To the effect that the noise drowns out the signal: the frequent occurrence of large overruns leads to very slow convergence of sample means and variance to the population mean and variance. Moreover, for the heavy-tailed distributions of cost and effort performance the variance increases with increasing sample sizes without converging. The finding further supports the argument that the variance between and within studies in the extant literature is a result of frequently occurring outliers.

Finally, three alternative explanations were tested, which create spurious fat tails out of thin-tailed distributions. The first alternative explanation is that of mistaking a lognormal distribution for a power law distribution. This explanation could not reject the null-hypothesis that the tails fit power laws. The second alternative explanation is the aggregation of distributions with same means but different variance structures. The analysis demonstrated that the characteristic fat tails exist in every subsample. Again, this explanation could not reject the null-hypothesis that the tails fit power laws. The third alternative explanation is the aggregation of observations over non-stationary parameters. IT project performance at one private sector company was compared before and after a major reorganization of the IT department. While the distribution shifted, i.e. the project performance improved, fat tails did not disappear. Again, this alternative explanation could not reject the null-hypothesis that the tails fit power laws.

In sum, the analysis in this chapter concluded that performance outliers occur frequently and consistently in IT projects and it is concluded that this conclusion is highly robust. Chapter 5 now turns towards the second research question: Why do the outliers, i.e., markedly deviating observations, occur in IT project performance? Then, chapter 6 will discuss the implications of the findings of this and the next chapter.

5 Findings: Why did Outliers Occur?

The previous chapter analyzed the first of the two research questions: How frequent are outliers, in the sense of markedly deviating observations, in IT project performance? The data showed that outliers occur frequently and consistently in IT projects. This chapter now turns to the second research question: Why do outliers occur in IT project performance?

5.1 Determinants of Cost Outliers in IT Projects

In this section the determinants of outliers of IT project cost performance will be tested. The literature review, chapter 2, identified potential explanatory variables from the extant research on common overruns and on outliers. Hence, the analysis will first analyze the common overruns in this sample in an attempt to replicate prior results, to test whether the sample differs from prior studies and if so, how the sample might differ. Common overruns in this context are defined as the center of the distribution – as opposed to the tails.

Then, the analysis will test the explanations of the tails. Chapters 3.3.1 and 3.3.2 described the two analyses that will be used: multilevel rare events logistic regression, where the rare event in the analysis is the occurrence of outliers, and the comparison of the thickness of the tail⁵⁹.

⁵⁹ Analyses of power law behavior have eliminated theoretical explanations based on suggested generating mechanisms. In other words, whether abstract models suggested by theories generate outcomes that fit power law distributions. Explanations from all three schools of thought can be translated into models that generate power laws (cf. appendix 8.12). Thus, due to the finding alone that IT project performance is best described by power laws none of the theoretical explanations can be rejected.

5.1.1 *Determinants of Common Cost Overruns in IT Projects*

To explain the common cost overruns, the test split the sample in two subsamples along the median of the explanatory variables, e.g., into big and small projects. This analysis was chosen because the large number of explanatory variables. Regression modelling, as typically common, overfits the data. Moreover, the explanatory variables are correlated (cf. the full correlation matrix is included in Appendix 8.6) a factor analysis did not result in a meaningful reduction of the variables (see Appendix 8.7).

Table 12 shows the results of the test of determinants of common cost overruns. The first two columns give the abbreviated title of the explanatory variable, then the table contains the median at which the sample was split and the interquartile range (IQR) as a measure of variation in the explanatory variable. The next six columns compare the subsamples, with their respective sample sizes, median overruns, and IQRs of the overruns. For many variables the split resulted in unequal subsamples because of ties at the median⁶⁰. The last three columns give the test result of the non-parametric Wilcoxon test. The null hypothesis of the test is that there is no difference in the ranks of cost overruns between the two subsamples. A non-parametric ranking was chosen to deal with the large number of outliers in the cost overrun data (cf. chapter 4.1). The

⁶⁰ To a small extend also because of missing values in the dependent variable, although the problem is more prominent in the analysis of common schedule and effort overruns because those data contain more missing values.

very last column gives the adjusted p*-value, using the Benjamini-Hochberg procedure (cf. chapter 3.3.2).

Table 12 Non-Parametric Wilcoxon Test for Location Shift in Cost Overrun Distribution

Abbreviated description	Median	IQR	N ₁	N ₂	M ₁	M ₂	IQR ₁	IQR ₂	W	p-value	adjusted p*-value
H2.1.1 Estimated cost	1.2	8.4	1745	1739	1.00	1.00	0.364	0.626	1404360	0.000 ***	0.003 **
H2.1.2 Estimated duration	181	203	329	217	1.00	1.00	0.305	0.325	38956	0.071 ..	0.228
H2.1.3 Actual cost	0.8	6.7	1954	1496	1.00	1.00	0.551	0.483	1634861	0.000 ***	0.000 ***
H2.1.4 Actual duration	1096	1461	1862	1413	1.00	1.00	0.447	0.598	1261851	0.045 *	0.189 .
H2.2.1 Clear goals, business case	8.0	2.3	63	105	1.00	1.03	0.179	0.443	3019	0.344	0.525
H2.2.2 Environmental influence	8.0	3.0	83	84	1.01	1.00	0.350	0.343	3559	0.816	0.859
H2.2.3 No resistance	3.0	4.0	81	87	1.00	1.00	0.475	0.264	3674	0.634	0.796
H2.2.4 User involvement	8.0	2.0	55	97	1.00	1.07	0.145	0.424	2299	0.158	0.414
H2.2.5 No political intent	3.0	3.0	59	59	1.00	1.00	0.557	0.301	1930	0.309	0.520
H2.2.6 Realistic, reliable schedule	8.0	3.0	86	82	1.00	1.07	0.175	0.501	2806	0.022 *	0.117 .
H2.2.7 Uniqueness	5.0	4.0	59	108	1.00	1.00	0.522	0.298	3701	0.085 ..	0.254
H2.3.1 Stable requirements	5.0	4.0	75	93	1.10	1.00	0.492	0.185	4073	0.062 ..	0.220
H2.3.2 Innovation	7.0	3.0	63	104	1.01	1.00	0.243	0.429	3520	0.422	0.611
H2.3.3 Decoupled deliverables	7.0	3.0	91	77	1.00	1.03	0.358	0.324	3151	0.262	0.484
H2.3.4 Minimal & stable scope	5.0	4.0	40	49	1.00	1.11	0.264	0.645	855	0.303	0.520
H2.3.5 Effective PM method	8.0	2.0	44	122	1.00	1.01	0.380	0.370	2573	0.684	0.799
H2.3.6 Few unknowns in the design	5.0	4.0	75	93	1.09	1.00	0.487	0.248	4188	0.026 *	0.119 .
H2.3.7 Incremental approach	6.0	6.0	77	90	1.00	1.01	0.310	0.389	3564	0.753	0.855
H2.4.1 Low processing requirements	7.0	3.0	76	90	1.00	1.00	0.357	0.306	3365	0.858	0.865
H2.4.2 Low reusability requirements	8.0	4.0	49	118	1.00	1.01	0.506	0.269	2533	0.209	0.448
H2.4.3 Low performance requirements	8.0	3.0	75	92	1.00	1.00	0.257	0.368	3313	0.659	0.796
H2.4.4 Low security requirements	8.0	4.0	90	77	1.00	1.05	0.201	0.492	3118	0.265	0.484
H2.5.1 Outside view	5.0	4.8	41	76	1.00	1.00	0.240	0.442	1509	0.781	0.859
H3.1 Effective risk management	8.0	2.0	43	123	1.00	1.04	0.173	0.409	2000	0.018 *	0.116 .
H4.1.1 Monitor & control	8.0	2.0	69	98	1.00	1.00	0.334	0.356	3247	0.663	0.796
H4.1.2 Benefits measured	0.0	0.0	146	352	1.00	1.00	0.425	0.473	248254	0.466	0.631
H4.1.3 Detailed, up-to-date plan	8.0	2.0	36	52	1.00	1.00	0.591	0.441	916	0.865	0.865
H4.1.4 Effective change management	8.0	1.3	26	62	0.98	1.00	0.342	0.515	701	0.339	0.525
H4.2.1 Sustainable resource mix	8.0	2.0	16	68	1.08	1.00	0.435	0.541	627	0.350	0.525
H4.2.2 Competent project managers	9.0	2.0	61	105	1.00	1.01	0.231	0.424	3034	0.572	0.751
H4.2.3 Qualified & motivated team	9.0	2.0	52	116	1.00	1.01	0.327	0.385	2618	0.172	0.426
H4.2.4 Suitably staffed	5.0	5.0	97	68	1.00	1.02	0.220	0.535	3067	0.444	0.622
H4.2.5 Years of experience PM	10.0	7.0	52	60	1.06	1.00	0.484	0.293	1976	0.015 *	0.116 .
H4.3.1 Effective governance	8.0	2.0	56	112	1.00	1.07	0.090	0.414	2441	0.019 *	0.116 .
H4.3.2 Effective sponsor	8.0	2.0	68	100	1.00	1.04	0.272	0.390	2952	0.148	0.414
H4.4.1 Senior management support	8.5	3.0	82	86	1.00	1.01	0.269	0.390	3125	0.203	0.448
H4.4.2 Well performing supplier	7.0	1.0	81	84	1.00	1.04	0.321	0.356	2663	0.016 *	0.116 .
H4.4.3 Robust contracting	9.0	3.0	30	132	1.00	1.02	0.106	0.396	1549	0.063 ..	0.220
H4.4.4 Aligned stakeholders	9.0	3.0	53	113	1.00	1.03	0.169	0.412	2663	0.251	0.484
H4.5.1 Transparent status comm.	8.0	2.0	41	47	1.00	1.01	0.640	0.422	936	0.818	0.859
H4.5.2 Good team communication	7.0	3.0	62	106	1.00	1.01	0.374	0.313	2907	0.213	0.448
H5.1 If true costs were known...	1.0	2.0	51	65	1.20	1.00	0.653	0.178	2318	0.000 ***	0.003 **

Note: M₁ is the median of the subsample where variable > median, M₂ is the median of the subsample where the variable ≤ median, N₁, N₂ and IQR₁, IQR₂ are subsample sizes and IQRs of the subsamples respectively.

W is the non-parametric Wilcoxon test statistic, the adjusted p-value uses the Benjamini-Hochberg procedure, significance levels are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1, . < 0.2

Table 12 shows that, after controlling for the false discovery rate, the cost overruns of only three subsamples vary with statistical significance of $p^* < 0.05$. Six further variables were significant at $p^* < 0.12$. The three variables that are significant at the 0.05 level are estimated cost, actual cost, and the variable ‘if the true cost were known...’.

First, the variable *estimated cost* (H2.1.1) reflects the hypothesis that the common cost overruns do not vary with the estimated cost expressed in 2010 USD millions. The median estimated cost is 1.204 million, with an IQR of 8.4 million. Splitting the total sample along the median, with ties, creates two subsamples. The first subsample includes 1,745 projects that were larger than the median estimated cost of USD 1.204 million. The second subsample includes 1,739 projects smaller or equal to the median estimated cost of USD 1.204 million. In both subsamples the median cost actual/estimated cost is 1.00. However the IQRs of the cost overruns differ: the IQR of the larger projects is 0.364, the IQR in the subsample of the smaller projects is 0.626. In other words, the variability of cost overruns is smaller in the large projects, where the width of the box in the boxplot is only half the size of the box in the boxplot of the smaller projects. Subsequently, the non-parametric Wilcoxon test rejects the null hypothesis that the true location in both subsamples is the same with high significance of $p < 0.001$, which adjusted for the false discovery rate of testing multiple hypotheses is $p^* = 0.003$.

Second, the variable *actual cost* (2.1.3) is also statistically significant. The median actual cost were USD 0.8 million, in 2012 terms, with an IQR of USD 6.7 million. Splitting the sample at the median results in two subsamples of 1,954 projects with an actual cost > USD 0.8 million and 1,496 projects with actual cost ≤ USD 0.8 million. The median cost overrun in both samples is actual/estimated cost = 1.00. The IQRs differ. In the sample of the projects > USD 0.8 million the IQR was 55.1 percentage points cost overrun. In the sample of projects ≤ USD 0.8 million the IQR was 48.3 percentage points. The difference between both subsamples is highly significant $p^* < 0.001$.

Third, the perceived project outcome, measured with the variable *“If we had known the real costs of the project it would not have been built”* varies significantly with the cost overruns. The median answer to the question was 1.0. Splitting the sample at the median separates the projects that answered “1 = No regrets; none or very limited cost overruns or delays” from all other projects. The median cost overrun for projects that answered >1 was +20% (IQR = 65.3%). The median cost overrun for the projects that answered with 1 was ±0% (IQR = 17.8%). The difference is highly statistically significant with $p < 0.001$ and an adjusted $p^* = 0.003$.

In sum, among all hypotheses the test was only able to reject the null hypotheses with high statistical significance:

H2.1.1: Estimated costs do not vary with IT project cost overruns.

H2.1.3: Actual project cost do not vary with IT project cost overruns.

H5.1: The variable 'If the true cost were known...' does not vary with cost overruns.

The finding that these three hypothesis are significant is unsurprising. All three variables are strongly correlated with the variable of cost overruns, which is calculated as actual/estimated cost. Estimated cost should be negatively correlated to cost overruns. While actual cost should be positively correlated. Indeed, the correlation coefficient between $\ln(\text{estimated cost})$ and $\ln(\text{cost overrun})$ is $r = -0.08$ ($p < 0.001$), the correlation between actual cost and $\ln(\text{cost overrun})$ is $r = 0.17$ ($p < 0.001$). However, figures 25 and 26, which compare the log-log plot with the boxplots, show that the difference in the distribution is due to the difference in variation between small and large projects.

Additionally, that the variable 'if we had known the true cost...' is correlated to cost overruns is also unsurprising. However, it relates the findings back to the question of project failure and success. As the literature review argued, a cost overrun does not necessarily equate project failure. On the one hand, this argument is supported by the data: even projects that disagreed with the statement experienced cost overruns (see figure 27). On the other hand projects that agreed more with the statement had higher cost overruns and vice versa.

Figure 25 Correlation between estimated project size and cost overruns (blue line indicates the median at which the two groups are split in the boxplots)

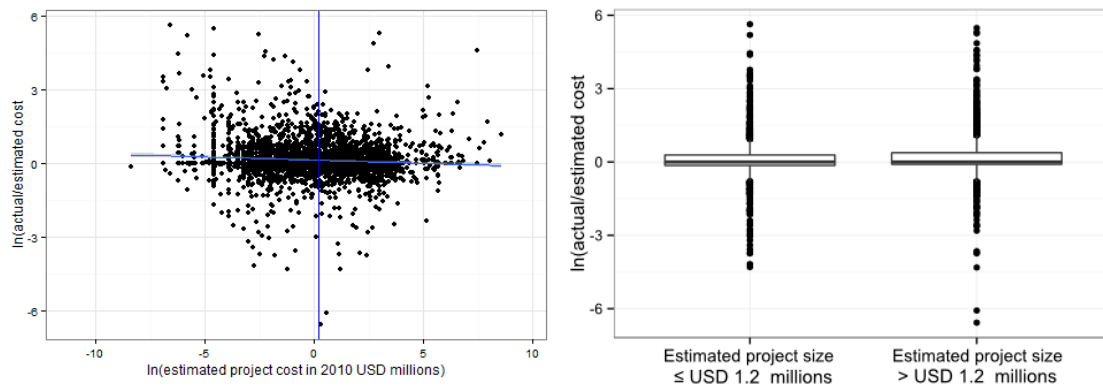


Figure 26 Correlation between actual project size and cost overruns (blue line indicates the median at which the two groups are split in the boxplots)

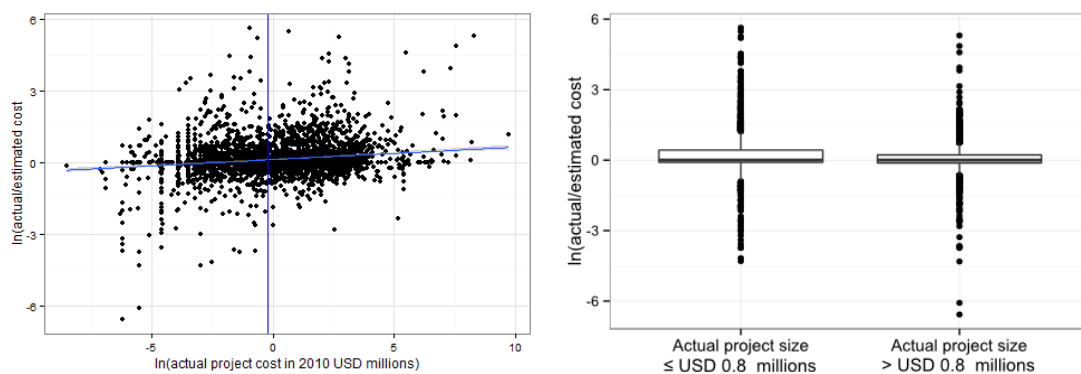
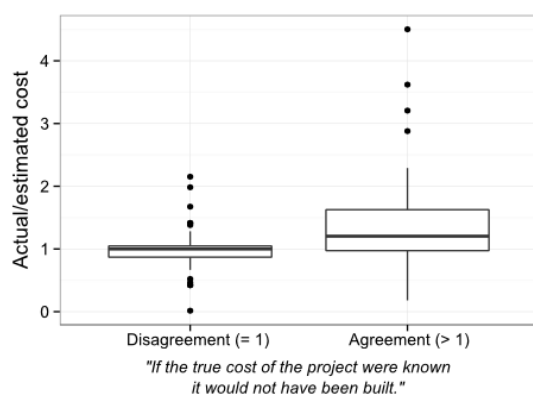


Figure 27 Cost Overruns in Relation to 'If the True Cost of the Project were Known...'



In addition to the three variables of estimated cost, actual cost, and 'if we had know the true cost...', there is weak statistical evidence to reject seven further null hypotheses. The non-parametric Wilcoxon test rejected the hypotheses at a level of significance of $\alpha = 0.05$; but controlling for the false discovery rate of multiple tests, the hypotheses can only be rejected at a significance level of $\alpha < 0.20$. The seven hypotheses were:

- H2.1.4: Actual project durations do not vary with IT project cost overruns.
- H2.2.6: Perceptions of whether the project followed a realistic and reliable schedule do not vary with IT project cost overruns.
- H2.3.6: Perceptions of whether the project had a large number of unknowns in the solution design do not vary with IT project cost overruns.
- H3.1: Perceptions of whether the project had an effective risk management do not vary with IT project cost overruns.
- H4.2.5: IT project cost overruns do not vary with the number of years of experience of the most senior project manager.
- H4.3.1: Perception of whether the project had an effective governance do not vary with IT project cost overruns.
- H4.4.2: Perceptions of whether the project was supported by a well performing supplier do not vary with IT project cost overruns.

Lastly, several null hypotheses were rejected by the non-parametric Wilcoxon test, but after adjusting for the false discovery rate of multiple tests, insufficient statistical evidence remains to reject the null hypotheses. Those null hypotheses tested the effect of the planned project duration, involvement of users, perceived uniqueness of the project, stability of requirements, qualification and

motivation of the project team, effectiveness of the project sponsor, and robustness of contracting.

In sum, the first test shows that only few of the explanatory variables suggested by the literature explained differences in cost overruns. The two steps taken by this thesis to improve upon previous studies, particularly those from the project management literature, might explain the difference. First, the thesis uses the actual and estimated cost baselines from documentary evidence to calculate cost overruns instead of using possibly biased self-reported survey responses. Second, the analysis controls for the false discovery rate, a step which was not taken or reported by prior studies.

5.1.2 Determinants of the Thickness of the Upper Power Law Tail of the IT Project Cost Performance Distribution

Same as in the previous test, the sample was split at the median of the determinant to create two subsamples and the thickness of the tails of the cost overruns between the two subsamples was compared. The likelihood ratio test tested whether splitting the sample better explains the cost overruns than the pooled sample. In other words it tested the null hypothesis that the two exponents that describe the tail are the same. The likelihood ratio test rejected with very high statistical significance the two null hypotheses (cf. table 13):

H2.1.1: The power law tails of IT project cost performance do not differ
between projects with different estimated cost.

H2.1.4: The power law tails of IT project cost performance do not differ between projects with different actual durations.

Table 13 Test for Thickness of the Upper Tail of the Cost Overrun Distribution

Abbreviated description	Median	IQR	N ₁	N ₂	Alpha ₁	Alpha ₂	LR	p-value	adj. p-value
H2.1.1 Estimated cost	1.2	8.4	768	815	2.997	2.320	57.190	0.000***	0.000***
H2.1.2 Estimated duration	181	203	148	87	3.048	2.564	3.558	0.169.	0.699
H2.1.3 Actual cost	0.8	6.7	971	603	2.557	2.622	0.530	0.767	0.999
H2.1.4 Actual duration	1096	1461	855	638	2.803	2.337	27.909	0.000***	0.000***
H2.2.1 Clear goals, business case	8.0	2.3	25	58	4.953	3.876	1.595	0.450	0.999
H2.2.2 Environmental influence	8.0	3.0	44	39	4.454	3.837	0.755	0.686	0.999
H2.2.3 No resistance	3.0	4.0	39	44	3.551	4.929	3.638	0.162.	0.699
H2.2.4 User involvement	8.0	2.0	17	56	3.282	4.107	1.216	0.544	0.999
H2.2.5 No political intent	3.0	3.0	26	26	2.880	4.967	6.557	0.038*	0.396
H2.2.6 Realistic, reliable schedule	8.0	3.0	35	48	5.466	3.573	5.588	0.061..	0.514
H2.2.7 Uniqueness	5.0	4.0	30	52	3.192	5.206	7.920	0.019*	0.267
H2.3.1 Stable requirements	5.0	4.0	44	39	3.718	4.785	2.113	0.348	0.973
H2.3.2 Innovation	7.0	3.0	34	49	4.732	3.820	1.460	0.482	0.999
H2.3.3 Decoupled deliverables	7.0	3.0	39	44	3.906	4.367	0.424	0.809	0.999
H2.3.4 Minimal & stable scope	5.0	4.0	15	27	3.416	3.177	0.093	0.954	0.999
H2.3.5 Effective PM method	8.0	2.0	20	63	3.716	4.294	0.548	0.760	0.999
H2.3.6 Few unknowns in the design	5.0	4.0	46	37	3.751	4.789	1.954	0.376	0.988
H2.3.7 Incremental approach	6.0	6.0	37	46	3.749	4.530	1.213	0.545	0.999
H2.4.1 Low processing requirements	7.0	3.0	39	43	4.521	3.872	0.797	0.671	0.999
H2.4.2 Low reusability requirements	8.0	4.0	19	63	3.202	4.616	3.647	0.161.	0.699
H2.4.3 Low performance requirements	8.0	3.0	38	44	4.853	3.718	2.312	0.315	0.973
H2.4.4 Low security requirements	8.0	4.0	39	43	4.531	3.865	0.837	0.658	0.999
H2.5.1 Outside view	5.0	4.8	16	36	3.757	3.469	0.123	0.940	0.999
H3.1 Effective risk management	8.0	2.0	13	70	4.678	4.050	0.348	0.840	0.999
H4.1.1 Monitor & control	8.0	2.0	34	49	4.363	3.992	0.256	0.880	0.999
H4.1.2 Benefits measured	0.0	0.0	661	577	3.115	2.600	3.937	0.140.	0.699
H4.1.3 Detailed, up-to-date plan	8.0	2.0	17	25	2.977	3.497	0.505	0.777	0.999
H4.1.4 Effective change management	8.0	1.3	11	31	3.501	3.181	0.135	0.935	0.999
H4.2.1 Sustainable resource mix	8.0	2.0	9	31	2.945	3.359	0.242	0.886	0.999
H4.2.2 Competent project managers	9.0	2.0	28	54	5.000	3.781	2.204	0.332	0.973
H4.2.3 Qualified & motivated team	9.0	2.0	23	60	4.809	3.934	1.027	0.598	0.999
H4.2.4 Suitably staffed	5.0	5.0	46	35	4.815	3.497	3.396	0.183.	0.699
H4.2.5 Years of experience PM	10.0	7.0	29	20	3.392	3.584	0.065	0.968	0.999
H4.3.1 Effective governance	8.0	2.0	17	66	4.765	4.004	0.624	0.732	0.999
H4.3.2 Effective sponsor	8.0	2.0	24	59	3.543	4.460	1.581	0.454	0.999
H4.4.1 Senior management support	8.5	3.0	36	47	4.186	4.094	0.016	0.992	0.999
H4.4.2 Well performing supplier	7.0	1.0	31	51	4.196	4.168	0.001	0.999	0.999
H4.4.3 Robust contracting	9.0	3.0	8	73	5.339	4.053	0.771	0.680	0.999
H4.4.4 Aligned stakeholders	9.0	3.0	21	61	4.921	3.890	1.306	0.521	0.999
H4.5.1 Transparent status comm.	8.0	2.0	18	24	2.892	3.638	1.039	0.595	0.999
H4.5.2 Good team communication	7.0	3.0	26	57	3.390	4.651	3.151	0.207	0.724
H5.1 If true costs were known...	1.0	2.0	34	18	3.159	4.881	3.505	0.173.	0.699

Note: Alpha₁ is the estimated power law exponent alpha of the subsample where variable > median, Alpha₂ is the estimated power law exponent of the subsample where the variable ≤ median, N₁ and N₂ are the subsample sizes respectively.

LR is the Likelihood Ratio test statistic, the adjusted p-value uses the Benjamini-Hochberg procedure, significance levels are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1, . < 0.2

First, the median *estimated cost* was USD 1.2 million (expressed in 2010 USD terms) with an IQR of USD 8.4 million. Splitting the sample at the median resulted in two subsamples of 1,745 IT projects above the median and 1,739 IT projects less or equal to the median. The power law was fitted to the tails of the distributions of both subsamples above a cut-off $x_0 = 1.0$.

This resulted in 768 valid observations in the tail of the cost performance distribution of the projects larger than the median project budget and 815 valid observations in the tail of the cost performance distribution of the projects smaller than the median. The likelihood ratio test for nested hypotheses requires the same cut-off value for the hypotheses to be nested, hence the tail was estimated simply using the maximum likelihood estimator without the second step of searching through all possible cut-off values x_0 to optimize the fit.

The estimated power law tail coefficient for the subsample of the larger projects was $\alpha = 2.997$. The estimated power law tail coefficient for the subsample of the smaller projects was $\alpha = 2.320$. The Likelihood Ratio was $D = 57.190$. The null hypothesis “the model that split the sample into two subsamples fits the data better than a model with the combined sample” was rejected with overwhelming statistical significance of $p < 0.001$. Controlling for the false discovery rate, the adjusted p^* -value also rejected the null hypothesis with overwhelmingly statistical significance of $p^* < 0.001$. In other words, the exponents of the two samples are statistically significantly different, where the

subsample of the smaller projects ($\leq$ USD 1.2 million) had a thicker tail ($\alpha = 2.3$) than the subsample of larger projects ($\alpha = 3.0$). This is to say that outliers, that is markedly deviant observations of IT cost performance, are more frequent in projects smaller or equal to USD 1.2 million than in larger projects above USD 1.2 million.

Second, the median *actual duration* was 1,096 days, i.e., 3 years and 1 day. The IQR was 1461 days, i.e., 4 years and 1 day. Splitting the sample at the median actual duration results in a subsample of 1,862 projects longer than the median and a subsample of 1,413 project shorter than the median actual duration.

Again, the power law was fitted only to the tails of the cost performance distributions above the cut-off value of $x_0 = 1$. This resulted in 855 valid observations in the tail of the subsample of the projects longer than the median actual duration. 638 valid observations were available in the tail of the cost performance of the projects shorter than the median actual duration.

The estimated power law exponent in the subsample of projects longer than the median actual duration was $\alpha = 2.803$. The estimated power law exponent in the subsample of projects shorter or equal to the median actual duration was $\alpha = 2.337$. The Likelihood Ratio was $D = 27.909$, which is highly statistically significant at $p < 0.001$. Controlling for the false discovery rate of multiple tests, the null hypothesis that the power law exponents of the two distributions of IT project cost performance differ can be rejected with high statistical significance of $p^* < 0.001$.

In other words, the tails of the distribution of the cost overruns of the IT projects are statistically significantly different. Shorter projects (≤ 3 years) have a thicker tail ($\alpha = 2.3$) than longer projects (> 3 years, $\alpha = 2.8$).

Nine other variables provided weak statistical evidence from the Likelihood Ratio tests, but could not be rejected with the adjusted p^* -value controlling for the false discovery rate of multiple tests. Those nine variables were:

H2.1.2: The power law tails of IT project cost performance do not differ between projects with different estimated durations – shorter projects (≤ 181 days) had a thicker tail ($\alpha = 2.6$) than longer projects ($\alpha = 3.0$).

H2.2.3: The power law tails of IT project cost performance do not differ between projects with different perceived levels of organizational resistance – projects which encountered more resistance had a thicker tail ($\alpha = 3.6$) than projects with fewer resistance ($\alpha = 4.9$).

H2.2.5: The power law tails of IT project cost performance do not differ between projects with different perceived levels of political intent behind plans and budgets – projects with high levels of political intent influencing the project plans and budgets had a thicker tail ($\alpha = 3.6$) than projects with less political intent ($\alpha = 5.0$).

H2.2.6: The power law tails of IT project cost performance do not differ between projects with different perceived levels of realism and reliability in their schedules – projects that had a high level of realism in their project plans had a thinner tail ($\alpha = 5.5$) than projects with less realistic and reliable schedules ($\alpha = 3.6$).

H2.2.7: The power law tails of IT project cost performance do not differ between projects with different perceived levels of uniqueness – projects that were perceived as being more unique had a thicker tail ($\alpha = 3.2$) than projects that were seen as less unique ($\alpha = 5.2$).

H2.4.2: The power law tails of IT project cost performance do not differ between projects with different reusability requirements – projects that had low reusability requirements had a thicker tail ($\alpha = 3.2$) than projects with more required reusability ($\alpha = 4.6$).

H4.1.2: The power law tails of IT project cost performance do not differ between projects that measured benefits and those who did not – projects that measured benefits had a thicker tail ($\alpha = 2.6$) than projects who did not measure benefits ($\alpha = 3.1$).

H4.2.5: The power law tails of IT project cost performance do not differ between projects with different perceived levels of suitable staffing – projects that were more suitably staffed had a thinner tail ($\alpha = 4.8$) than projects that were not suitably staffed ($\alpha = 3.5$).

H5.1: The power law tails of IT project cost performance do not differ between projects with different agreements to the statement 'if the true cost had been known it would not have been built' – projects that agreed more with the statement had a thicker tail ($\alpha = 3.2$) than projects that agreed less ($\alpha = 4.9$).

In sum, two determinants had strong enough statistical evidence to reject the null hypothesis, i.e., where the thickness of the cost overrun tails differs if the sample is split into two subsamples along the median of the explanatory variables. Those two explanatory variables were (1) the estimated cost, where larger project budgets reduce the thickness of the tails of the cost performance distribution. And (2) the actual duration, where longer projects showed a statistically significant reduction in the thickness of the tail.

For further nine variables the tests found weak evidence, which could not be rejected once controlled for the false discovery rate. The nine factors were: estimated duration, organizational resistance, political intent in plans, realistic and reliable schedule, uniqueness of the project, reusability requirements, whether benefits were measured, suitable staffing, and the statement 'if the true cost were known...'.

It is noteworthy, that the variables reusability requirements and benefits measurement showed the opposite direction than the literature suggests. Lower reusability requirements and managing benefits are expected to make a project less risky, however the analysis showed that those have in fact thicker tails, i.e.,

they are more risky because outliers are more frequent in projects with more required reusability and in projects that measured benefits.

Lastly, the statement ‘if the true cost of the project were known it would not have been built’ emphasizes that the variables were measured post-mortem, which should be taken into account when interpreting the results of this analysis. The variable ‘if the true cost were known...’ also links the thickness of the tails back to the question of success and failure of projects. The thesis argued in the literature review that a cost overrun per se does not constitute a failed project. This analysis, however, has shown that thinner tails, i.e., where outliers were more frequent, are more likely to agree that the project should not have been built.

5.1.3 Explanatory Multilevel Regression Models

Two multilevel regression models were constructed from the potential explanatory determinants of common cost overruns and the thickness of the tail of cost overruns of IT projects. The first model used the log-transformed actual/estimated cost as dependent variable, the second model used the binary classification as potential outlier⁶¹ as the dependent variable. Both models used the organization as a first level, also called a random factor, in their models.

⁶¹ Following Tukey’s definition of being more than 1.5 IQRs away from the box in the boxplot.

5.1.3.1 Multilevel Model of Common Cost Overruns

Before building the model a power analysis was conducted⁶². The variable ‘Years of experience of the most senior project manager’ had the weakest statistical power. Data was only available for 11 organizations of the 21 organizations that had filled out the questionnaire. For 21 groups, a power level of 0.8, and a moderate effect size 17 observations are needed in each group. For the same power level, number of organizations, and a small effect size 101 observations would be needed per organization. Given that most organizations provided data for 10-20 projects, the regression analysis provided a conservative test of the hypothesized relationships.

Table 14 Correlations and Sample Sizes of Fixed Effects in Model to Explain Common Cost Overruns

	LN(est. cost)	Realistic and reliable schedule	Few unknowns in the design	Effective risk mgmt.	Experience project manager	Effective governance	Senior mgmt. support
LN(estimated cost)		181	181	178	120	181	181
Realistic and reliable schedule	0.01		220	217	145	220	220
Few unknowns in the design	0.28***	-0.19**		217	145	220	220
Effective risk management	0.05	0.52***	-0.10		142	217	217
Experience project manager	0.21*	-0.02	0.01	0.07		145	145
Effective governance	0.05	0.52***	-0.23***	0.53***	0.01		220
Well performing supplier	0.08	0.15*	-0.04	0.19**	0.00	0.47***	

Note: Upper half sample n, lower half correlations with significances are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1,

. < 0.2

Table 15 Fixed Effects Model to Explain Common Cost Overruns (LN transformed)

Estimate	Std. Error	df	t-value	p-value	F-value	p-value

62 Following Cohen (1988) using the R package `power`

<i>Intercept</i>	0.491	0.484	82.870	1.015	0.313	-	-
<i>LN(estimated cost)</i>	0.006	0.039	49.290	0.150	0.881	0.023	0.881
<i>Realistic and reliable schedule</i>	-0.011	0.039	82.590	-0.278	0.781	0.078	0.781
<i>Few unknowns in the design</i>	-0.021	0.027	78.010	-0.770	0.444	0.593	0.444
<i>Effective risk management</i>	-0.046	0.035	80.130	-1.310	0.194	1.716	0.194
<i>Experience project manager</i>	0.030	0.012	80.800	2.467	0.016	6.084	0.016
<i>Effective governance</i>	0.047	0.048	76.050	0.971	0.335	0.943	0.334
<i>Well performing supplier</i>	-0.069	0.048	80.440	-1.420	0.160	2.017	0.159

Note: t-values calculated using Wald test; F-values calculated using Satterthwaite approximation of degrees of freedom

The multilevel model results are shown in table 14 and table 15. Table 14 shows that several significant correlations exist between the independent variables in the model. Thus the model will have considerable interaction effects.

The first level variance is 0.075, with a standard deviation of 0.284. This shows a relatively small effect of the first level but a large variation. The standard deviation is directly comparable with the fixed effects of the model (cf. table 15), which shows that the random effect is strong compared to the determinants of cost overruns. The Wald test estimates the p-value of first level ($X^2 = 4.84$, $p = 0.0278$). These results indicates that the level holds explanatory value and needs to be kept in the model.

Table 15 shows the fixed effects model. The interpretation of a p-value differs from that of OLS models: here a coefficient's p-value is an approximation to the p-value of a model that is fitted without this particular coefficient. The results show that the number of years of experience of the most senior project manager are a statistically significant determinant of the average cost overrun. Supplier

performance and effective risk management offer some weak statistical evidence of influencing the average cost overrun.

In other words, all other things being equal the number of years of experience of the most senior project manager is the most significant predictor of the average cost overrun. A project manager with ten years of work experience has on average a 19 percentage points higher cost overrun than a project manager with 5 year work experience. One potential explanation is the significant correlation with estimated cost, more experienced project managers manage larger projects. The effect can be controlled for estimated cost by entering estimated cost as a covariate before explaining the average cost overruns with the years of experience of the most senior project manager. In the resulting multilevel model the effect of the experience is non-significant ($p = 0.275$), though still positive: every year of work experience adds 1.6 percentage points of cost overruns above and beyond the effects of estimated cost and organizational differences. The variable has also significant correlations with other variables not in the final model, which indicates that it is a proxy of other risk and success factors. The years of experience of the most senior project manager correlate significantly with the less scope volatility ($r = -0.24$, $p = 0.010$), more detailed and up-to-date plans ($r = 0.21$, $p = 0.020$), more competent project management team ($r = 0.20$, $p = 0.020$), more qualified and motivated project team ($r = 0.19$, $p = 0.020$), use of a more effective project management method ($r = 0.19$, $p = 0.020$), and lower performance requirements ($r = 0.20$, $p = 0.040$).

5.1.3.2 Multilevel Model of Cost Outliers

For the prediction of cost outliers a multilevel model was fitted to explain the dependent variable, which classified outliers according to Tukey's rule of being at least 1.5 IQRs away from the box in the boxplot. The organization of the project was entered as a first level, i.e., the random effect, in the model.

Table 16 Fixed Effects Multilevel Logistic Regression for Cost Outliers

	Estimate	Std. Error	z-value	p-value	F-value
Intercept	-9.440105	8.624191	-1.095	0.274	-
LN(Estimated cost)	0.302164	0.456767	0.662	0.508	0.1255
LN(Estimated duration)	1.718645	1.467624	1.171	0.242	0.3212
Organizational resistance	-0.152483	0.384041	-0.397	0.691	0.0417
Political intent	0.003024	0.377991	0.008	0.994	0.9985
Realistic and reliable schedule	-0.354183	0.426995	-0.830	0.407	0.0236
Uniqueness	0.320409	0.280637	1.142	0.254	1.6916
Reusability requirements	0.142200	0.277022	0.513	0.608	0.0957
Measurement of benefits	2.330946	2.269770	1.027	0.304	1.0980
Suitably staffed	-0.523330	0.608082	-0.861	0.389	0.0001
Good team communication	-0.544039	0.674393	-0.807	0.420	0.0000
If true cost were known...	0.498179	0.406889	1.224	0.221	1.4991

Table 16 shows that the model with all potential determinants of outliers. None of the variables entered is statistically significant. Table 17 shows the correlations between the variables in the analysis. Strong and significant multicollinearity is present in the data, which justifies a more parsimonious model through excluding insignificant effects. Moreover the variance and the standard deviation of the random effect are 0. That is that in this model the random effects model does not explain any variance and can be excluded from the model as well.

Table 17 Correlation between Fixed Effects

	LN (P. cost)	LN (P. duration)	Org. resistance	Political intent	Realistic schedule	Unique- ness	Reusability req.	Measur. benefits	Suitably staffed	Team comm.
LN(Planned cost)		570	181	130	181	180	180	3553	177	181
LN(Estimated duration)	0.42***		167	117	167	166	166	1102	166	167
Organizational resistance	0.19**	0.09		168	220	219	219	220	216	220
Political intent	0.06	0.06	0.33***		168	168	168	168	165	168
Realistic and reliable schedule	0.01	0	-0.13*	-0.19**		219	219	220	216	220
Uniqueness	0.09	-0.02	0.14*	0.27***	-0.17**		219	219	215	219
Reusability requirements	0.04	-0.02	0.14*	0.1	0.06	-0.02		219	215	219
Measurement of benefits	0.01	0.03	0.14*	0.19*	0.09.	0.08	0.01		216	220
Suitably staffed	0.28***	0.14..	0.2***	0.12.	0.2***	-0.03	-0.02	0.18**		216
Good team communication	-0.34***	-0.12.	-0.02	0.1.	-0.25***	0.12 ..	0.05	-0.25***	-0.45***	

Note: Upper half sample n, lower half correlations with significances are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1,

. < 0.2

Table 18 shows the most parsimonious model, which does not include a random effect. It shows that planned project duration, perceived uniqueness, and whether benefits are measured provide weak statistical evidence of influencing the odds of a project being classified as an outlier.

Estimated duration was log-transformed to correct for outliers and skew for the analysis. Thus the resulting effect is not linear. For example, increasing the estimated duration from one year to two years increases the probability of a cost outlier occurring from 14% to 30% all other things being equal.

Similarly, perceived uniqueness is measured on a scale 1-10. An increase in the perceived uniqueness of the project by 1 unit increased the probability of an outlier occurring from 14% to 20%.

Lastly, the dummy variable whether benefits were measured provides a counterintuitive result given that the extant literature on projects expects projects that measure their benefits to lower their risk. In this sample, when benefits are not measured, however, the probability of a cost outlier occurring dropped from 14% to 2%. While the variable benefits measured has the lowest significance, excluding the variable shows that it is a suppressor⁶³ variable in this model.

Table 18 Logistic Regression for Cost Outliers

	Estimate	Std. Error	z-value	p-value
Intercept	-14.0207	5.4564	-2.570	0.0102
LN(Estimated duration)	1.3833	0.7868	1.758	0.0787
Uniqueness	0.3921	0.2292	1.711	0.0871
Benefits measured	2.1131	1.2930	1.634	0.1022

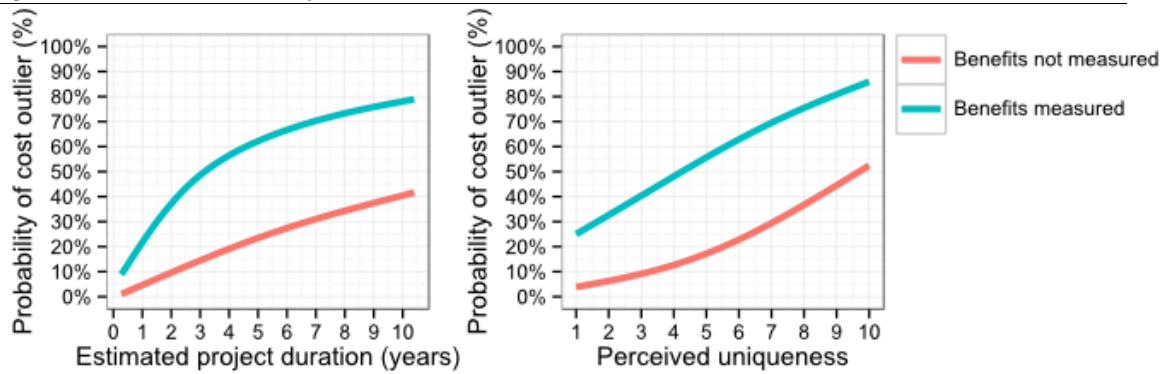
Figure 28 shows the average predicted probability of a project being a cost outlier over the range of values in all three variables. The variable whether benefits were measured or not is binary, thus it is depicted as two separate lines. However, when interpreting the results one should bear in mind that the variable is non-significant and included as a suppressor because it reduces the error of prediction. In other words, the distance between the two lines is not statistically significantly different from 0.

⁶³ The dummy variable whether benefits were measured has only a weak relationship with the dependent variable, which are the log-odds of an outlier occurring. It also does not significantly correlate with the other two variables (cf. table 17). However, if the dummy variable whether benefits were measured is excluded from the model the significance of the other effects decreases. This is most likely because the dummy variable reduces the error term in the regression analysis, which improves the model. The reduction in the error term is also called suppression.

In both cases the probability of a project being a cost outlier increases with increasing duration and uniqueness. A two-year project has on average a 10% probability of being a cost outlier if its benefits are not measured, while a four-year project, whose benefits are not measured, has a 19% probability of being a cost outlier. Again, it is important to note that the difference between projects which measured benefits and which did not measure benefits is non-significant, i.e., the effect is not significantly different from zero. Notwithstanding, if benefits are measured, the two-year project has on average a 37% probability of being a cost outlier and a four-year project 57% probability of being a cost outlier. A possible explanation of the effect of benefits management might be that the project is managed through different trade-offs. If a project is managed against the three goals of cost, schedule, and benefits the project might be more likely to incur cost overruns in order to realize those benefits.

Similarly, a project that is perceived as ranking 4 = "Project had a few unique aspects to it; but not fundamentally different from other projects" has a 12% probability of being a cost outlier if its benefits are not measured and a 48% probability if its benefits are measured. A project that is perceived as ranking 7 = "Project was unique to parts of the team and users but had an experienced vendor or project leaders; overall difficult to compare to other projects" has on average a 29% probability of being a cost outlier if its benefits are not measured and on average a 69% probability of being a cost outlier if its benefits are measured.

Figure 28 Prediction Cost Performance Outliers



A final multilevel logistic regression model is fitted solely with estimated duration (H2.1.1) and estimated cost (H2.1.2). As the data collection showed, these are often the sole decision criteria available at the green light decision. Thus a model of the risk of an IT project only based on these two criteria can be useful to understand the risk of a project at its inception. Moreover, the model benefits from a larger sample size of $n = 526$, where all three variables were reliably available. The random effect is significant again, with a variance of 1.364 and a standard deviation of 1.168. Again comparing the standard deviation of the random effect to the estimate of the fixed effects shows that a considerable amount of variation is included in the first level of the data.

Table 19 Multilevel Logistic Regression for Cost Outliers

	Estimate	Std. Error	z-value	p-value
Intercept	-13.5576	2.08426	-6.505	< 0.001
LN(Estimated duration)	1.63750	0.30460	5.376	< 0.001
LN(Estimated cost)	-2.36955	0.48410	-4.895	< 0.001
Interaction effect	0.33188	0.07357	4.511	< 0.001

Table 19 shows the results of the model. All coefficients are highly statistically significant, including the interaction effect between them. The direction effect of the planned project cost is negative, i.e., it reduces the log odds by -2.4 per one

unit increase in the log-transformed estimated cost. In other words, the probability of an outlier decreases by 7% for every additional USD 3 million estimated cost⁶⁴. However, the direct effect is overshadowed by the interaction effect: increasing estimated project cost has a direct effect that decreases the probability of cost outliers, but the negative direct effect of estimated cost is increasingly moderated by the interaction effect, which points in the opposite direction. Moreover, interpretation is complicated by the log transformation of both variables estimated cost and estimated duration.

Figure 29 helps interpreting the effects. The figure charts the average predicted probability of the project being classified as a cost outlier depending on the estimated cost and the estimated duration, while varying the other variable. In other words, the curve in the graph is the moving average of the predicted probabilities at a given point of the variable for the range of all other factors.

The graph shows that a three year project has on average⁶⁵ a 9% probability of being a cost outlier and a six year project has on average a 50% probability of being a cost outlier. A small project with USD 0.1 million estimated cost has on average a 20% probability of being a cost outlier, a USD 2 million project has on average a 26% probability of being a cost outlier, and a USD 100 million project has on average a 38% probability of being a cost outlier. This shows that although estimated cost have a negative direct effect on the probability of

⁶⁴ $\text{LN}(3) = 1.098612$

⁶⁵ i.e. this is the average of all predictions for a range of estimated cost.

outliers occurring the effect is dominated by the direct effect of project duration and the interaction effect.

Figure 29 Predicted Average Probability of a Project Being a Cost Outlier

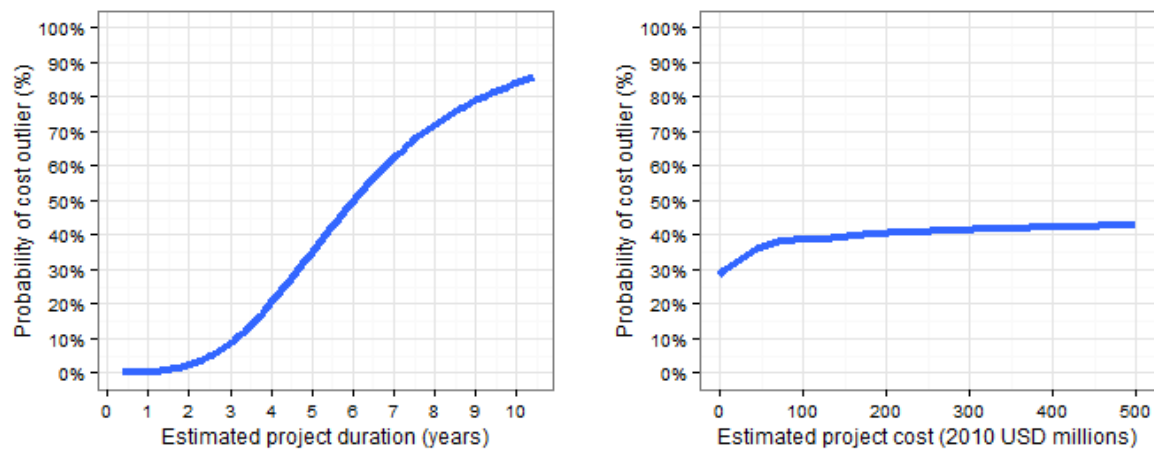
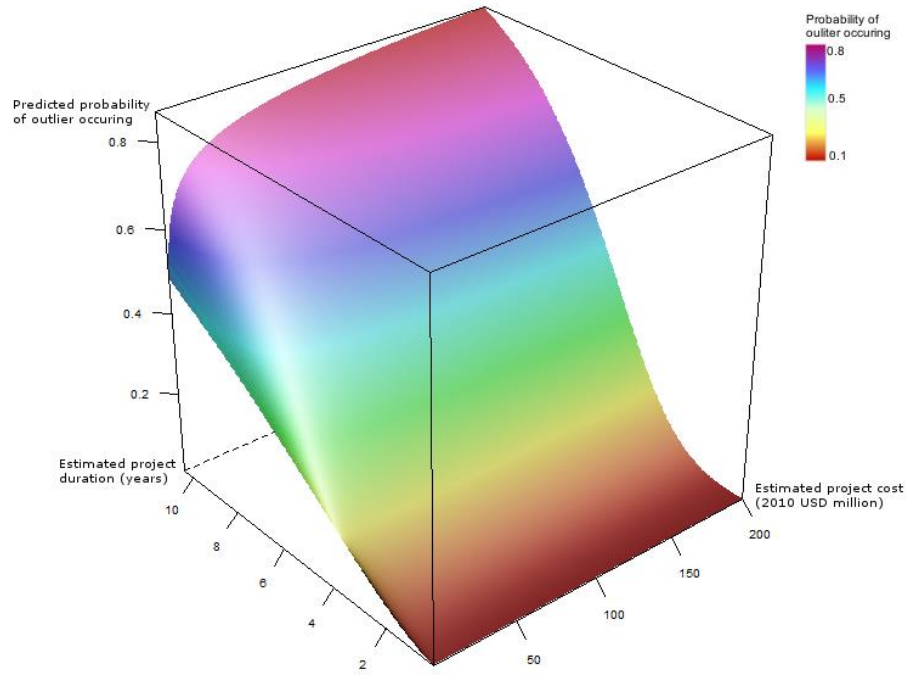


Figure 30 helps to further compare the strength of both effects. The graphs shows the full three dimensional shape for the range of values of estimated cost and duration found in the data and not only the cross section at the average of the unobserved dimension. The shape shows that the direct effect of the estimated duration and the interaction effect are much stronger than the direct effect of the estimated project budget.

Figure 30 Predicted Probability of the Occurrence of Outliers for Given Project Cost and Schedule Estimates



Lastly, the question arises of how biased these rare events logistic regression models are because they predict outliers, that is the occurrence of markedly deviant observations, where the reasonable expectation is that the rare event occurs less frequently than 0.5. King and Zeng (2001) approximated the rare events bias with

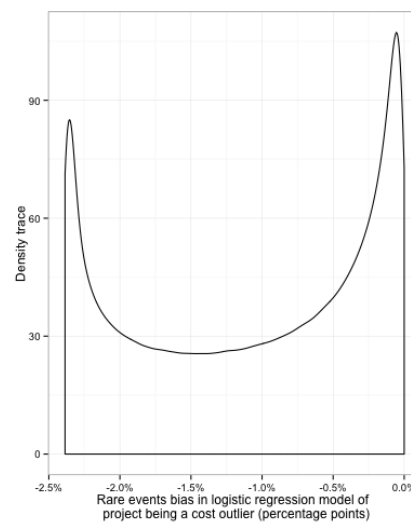
$$E(\hat{b}_0 - b_0) \gg \frac{\bar{p} - 0.5}{n\bar{p}(1 - \bar{p})}, \text{ with } \bar{p} = \frac{1}{n} \sum_{i=1}^n p_i,$$

where π_i is the probability that the dependent variable takes on the value 1.

Given the sample size the rare events bias in the explanatory model with estimated duration, uniqueness, and benefits measured is -0.0955. In the more parsimonious model with estimated cost and duration the bias is -0.0258, mostly because of the increased sample size. In effect, the bias is leading to an average underestimation of the probability of a project being a cost outlier by 1 percentage point. The bias ranges (cf. figure 31) up to 2.4 percentage points,

which is much smaller than the average predicted effect size. It shows that the bias is limited, because outliers are not very rare events, and that the prediction subsequently errs on the conservative side slightly underestimating the occurrence of outliers.

Figure 31 Rare Events Bias in Logistic Regression Explaining Cost Outliers



In sum: Why did cost outliers, that is markedly deviant observations of the cost performance of IT projects, occur? The comparison of the tail of the distribution showed that estimated cost and actual duration changed the thickness of the tails. Other variables with weaker statistical evidence, that changed the thickness of the tails, were estimated duration, organizational resistance, political intent in plans and budgets, realism and reliability of the schedule, perceived uniqueness of the project, reusability requirements, benefits measurement, and staffing. Where reusability and benefits measurement showed somewhat unexpected results compared to the extant literature, namely that less required reusability and the measurement of benefits increased

the risk of outliers. The latter effect is a non-significant suppressor in the rare events logistic regression explaining outliers and should only be considered within the model of uniqueness and estimated duration. The multilevel model showed that estimated duration and perceived uniqueness of the project best explain the occurrence of cost outliers, i.e., markedly deviant observations. In terms, of prediction even a very parsimonious model of only estimated duration and estimated cost shows that project duration is the most influential factor behind the occurrence of outliers in the cost performance of IT projects.

5.2 Determinants of Schedule Performance in IT Projects

This section follows the same analysis steps as the previous section. The section will first test the determinants of common schedule overruns in IT projects, which provides a comparison to the extant literature from which most of the explanatory variables were adapted. The section will then test which determinants influence the thickness of the tail of the schedule performance distribution. Finally, multilevel models will be fitted to explain both common schedule overruns and schedule outliers while controlling for the differences in schedule performance between organizations.

5.2.1 Determinants of Common Schedule Overruns in IT Projects

Table 20 shows the results of the non-parametric test for a statistically significant shift in the location of the schedule overrun distribution. As before the sample was split at the median of the potential determinant to test whether the schedule performance significantly differs between the two subsamples.

Table 20 Non-Parametric Wilcoxon Test for Location Shift in Schedule Overrun Distribution

Abbreviated description	Median	IQR	N ₁	N ₂	M ₁	M ₂	IQR ₁	IQR ₂	W	p-value	adj. p*-value
H2.1.1 Estimated cost	1.2	8.4	108	351	1.251	1.053	0.652	0.423	21453	0.034 *	0.128 .
H2.1.2 Estimated duration	181	203	458	501	1.036	1.000	0.400	0.351	119380	0.257	0.539
H2.1.3 Actual cost	0.8	6.7	148	791	1.082	1.000	0.553	0.328	65166	0.022 *	0.128 .
H2.1.4 Actual duration	1096	1461	30	926	1.382	1.000	0.532	0.347	18657	0.001 ***	0.009 **
H2.2.1 Clear goals, business case	8.0	2.3	47	62	1.000	1.219	0.303	0.590	11940	0.103 .	0.282
H2.2.2 Environmental influence	8.0	3.0	58	50	1.047	1.268	0.326	0.600	10990	0.028 *	0.128 .
H2.2.3 No resistance	3.0	4.0	52	57	1.134	1.091	0.553	0.348	16060	0.449	0.664
H2.2.4 User involvement	8.0	2.0	35	60	1.037	1.125	0.303	0.506	9620	0.492	0.688
H2.2.5 No political intent	3.0	3.0	32	32	1.138	1.127	0.528	0.341	5300	0.816	0.816
H2.2.6 Realistic, reliable schedule	8.0	3.0	65	44	1.000	1.319	0.273	0.588	9070	0.001 **	0.009 **
H2.2.7 Uniqueness	5.0	4.0	26	82	1.055	1.127	0.326	0.500	10170	0.721	0.774
H2.3.1 Stable requirements	5.0	4.0	41	68	1.304	1.050	0.552	0.344	16440	0.114 .	0.282
H2.3.2 Innovation	7.0	3.0	45	64	1.031	1.143	0.389	0.501	11970	0.130 .	0.304
H2.3.3 Decoupled deliverables	7.0	3.0	65	44	1.125	1.089	0.500	0.446	14860	0.728	0.774
H2.3.4 Minimal & stable scope	5.0	4.0	25	19	1.204	1.306	1.224	0.399	2190	0.661	0.771
H2.3.5 Effective PM method	8.0	2.0	23	85	1.161	1.087	0.445	0.500	10220	0.738	0.774
H2.3.6 Few unknowns in the design	5.0	4.0	49	60	1.303	1.082	0.600	0.314	16870	0.181 .	0.400
H2.3.7 Incremental approach	6.0	6.0	45	64	1.034	1.171	0.303	0.569	10920	0.030 *	0.128 .

H2.4.1 Low processing requirements	7.0	3.0	48	59	1.089	1.129	0.425	0.464	1329	0.582	0.741
H2.4.2 Low reusability requirements	8.0	4.0	31	77	1.143	1.091	0.619	0.375	1320	0.385	0.646
H2.4.3 Low performance requirements	8.0	3.0	51	57	1.143	1.077	0.501	0.363	1597	0.372	0.646
H2.4.4 Low security requirements	8.0	4.0	68	40	1.070	1.231	0.344	0.590	1109	0.105	0.282
H2.5.1 Outside view	5.0	4.8	26	38	1.000	1.304	0.190	0.568	320	0.015	* 0.108
H3.1 Effective risk management	8.0	2.0	25	83	1.037	1.125	0.363	0.500	999	0.779	0.798
H4.1.1 Monitor & control	8.0	2.0	49	60	1.077	1.190	0.303	0.585	1125	0.033	* 0.128
H4.1.2 Benefits measured	0.0	0.0	44	92	1.296	1.000	1.153	0.348	2599	0.001	*** 0.009
H4.1.3 Detailed, up-to-date plan	8.0	2.0	19	25	1.129	1.348	0.432	1.247	165	0.085	.. 0.282
H4.1.4 Effective change management	8.0	1.3	10	34	1.183	1.305	0.347	0.622	141	0.424	0.660
H4.2.1 Sustainable resource mix	8.0	2.0	8	33	1.171	1.304	0.180	0.558	120	0.705	0.774
H4.2.2 Competent project managers	9.0	2.0	43	66	1.125	1.106	0.500	0.484	1502	0.604	0.746
H4.2.3 Qualified & motivated team	9.0	2.0	36	73	1.000	1.200	0.203	0.686	813	0.001	** 0.009
H4.2.4 Suitably staffed	5.0	5.0	73	35	1.087	1.180	0.500	0.530	1166	0.458	0.664
H4.2.5 Years of experience PM	10.0	7.0	24	34	1.304	1.000	0.468	0.409	507	0.113	.. 0.282
H4.3.1 Effective governance	8.0	2.0	38	71	1.000	1.143	0.391	0.510	1214	0.386	0.646
H4.3.2 Effective sponsor	8.0	2.0	46	63	1.110	1.125	0.383	0.579	1378	0.659	0.771
H4.4.1 Senior management support	8.5	3.0	57	52	1.037	1.143	0.348	0.553	1214	0.100	.. 0.282
H4.4.2 Well performing supplier	7.0	1.0	52	55	1.072	1.125	0.392	0.502	1325	0.509	0.690
H4.4.3 Robust contracting	9.0	3.0	24	81	1.000	1.125	0.226	0.502	843	0.318	0.636
H4.4.4 Aligned stakeholders	9.0	3.0	37	70	1.037	1.136	0.363	0.514	1209	0.568	0.741
H4.5.1 Transparent status comm.	8.0	2.0	19	25	1.161	1.306	0.498	1.172	202	0.400	0.646
H4.5.2 Good team communication	7.0	3.0	30	79	1.192	1.077	0.500	0.464	1322	0.348	0.646
H5.1 If true costs were known...	1.0	2.0	28	35	1.510	1.000	1.715	0.145	831	0.000	*** 0.000

Note: M_1 is the median of the subsample where variable > median, M_2 is the median of the subsample where the variable $\leq$ median, N_1 , N_2 and IQR_1 , IQR_2 are subsample sizes and IQRs of the subsamples respectively.

W is the non-parametric Wilcoxon test statistic, the adjusted p-value uses the Benjamini-Hochberg procedure, significance levels are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1, . < 0.2

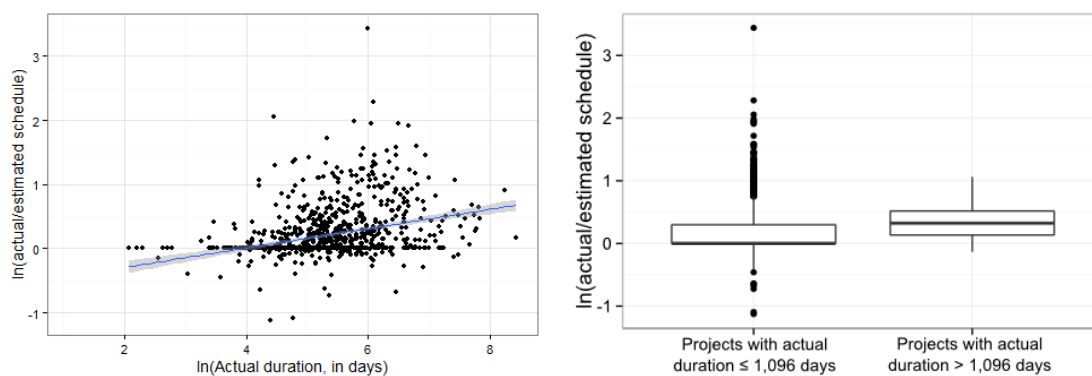
Five hypotheses were rejected with high statistical significance after controlling for the false discovery rate of multiple independent statistical tests:

- H2.1.4: Actual project durations do not vary with IT project schedule overruns.
- H2.2.6: Perceptions of whether the project followed a realistic and reliable schedule do not vary with IT project schedule overruns.
- H4.1.2: Whether the project measured benefits or not does not vary with IT project schedule overruns.
- H4.2.3: Perceptions of the qualification and motivation of the project team do not vary with IT project schedule overruns.

H5.1: Perceptions of “if the true were known the project would not have been built” do not vary with IT project schedule overruns.

The result that *actual duration* (H2.1.4) is related to the average schedule overrun is not surprising, because schedule overrun is defined as actual/estimated schedule. The correlation coefficient between $\ln(\text{actual schedule})$ and $\ln(\text{schedule overrun})$ is $r = 0.32$ ($p < 0.001$). Figure 32 shows the correlation and the distributions for projects split at the median in comparison.

Figure 32 Correlation between Actual Duration and Schedule Overruns



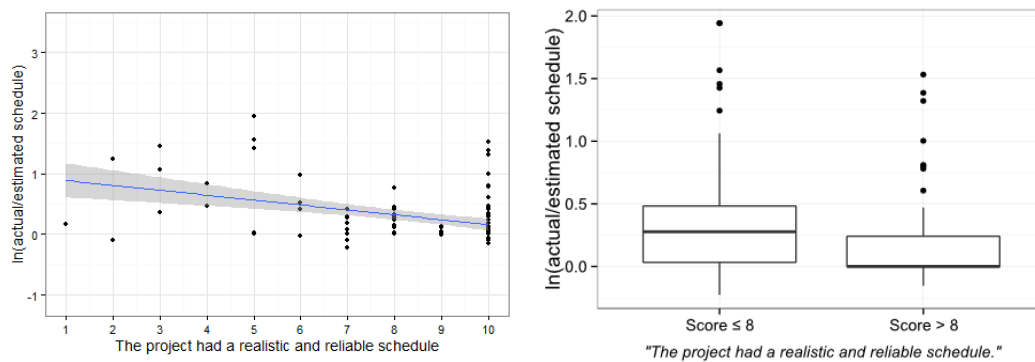
The variable ‘*if the true cost were known...*’ (H5.1) is again significant. If the organization did not completely disagree with the statement the median schedule overrun was +51% in the other projects the median was 0%. This illustrates that even though a schedule overrun does not necessarily constitute project success or failure those projects with an unfavorable outcome had significantly higher schedule overruns.

To interpret the variable *realistic and reliable schedule* (H2.2.6) also needs to take into account that the data comes from post-mortem analyses. The occurrence of

schedule overruns questions whether the project schedule was realistic.

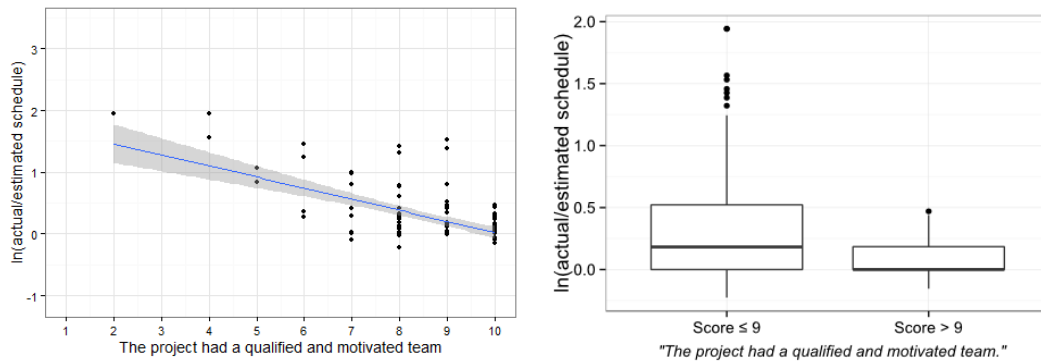
Unsurprisingly, projects that found that the schedule was less reliable and realistic (scoring ≤ 8 on a ten-point scale) had a median schedule overrun of +32%, whereas projects with a more reliable and realistic schedule (scoring > 8) had a median schedule overrun of +0%. However, as figure 33 shows projects experienced schedule overruns and were still perceived as having reliable and realistic schedules.

Figure 33 Schedule Overruns and Realistic and Reliable Schedules



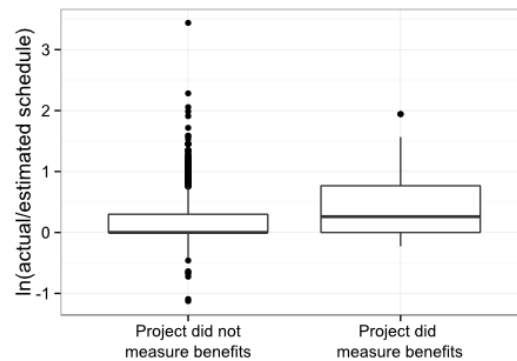
The variable *qualification and the motivation of the team* (H4.2.3) shows the expected result that with a more qualified and motivated team the average schedule overrun decreased (cf. figure 34). The figure also illustrates that the variation decreased.

Figure 34 Schedule Overruns and Qualification and Motivation of the Team



The dummy variable whether the project *measured benefits* (H.4.1.2) is again giving unexpected results (cf. chapter 5.1.1). The median schedule overrun of project that measured benefits was +29.6% and of those who did not it was 0%. The extant literature on project management, however, suggests that benefits management leads to a reduction in project risk. The data show (cf. figure 35) that the projects that measured benefits had not only a significantly higher median but also more variation. One possible explanation of this finding is that if projects are managed against quantified cost, schedule, and benefits goals then different trade-offs are made. In other words, when benefits are measured throughout the project, projects are more likely to incur delays, possibly to achieve those benefits.

Figure 35 Schedule Overrun and Projects that Measured Benefits



Moreover, the data provide weak statistical evidence that the project schedule performance differs depending on estimated cost, actual project cost, whether the project was able to influence the project environment, delivering the project in an incremental approach, whether the project was able to compare itself to other projects – i.e. taking an outside view, and whether the project was effectively monitored and controlled. These further six hypothesis were rejected by the unadjusted non-parametric Wilcoxon test, but only provides a level of significance $\alpha < 0.2$ once the results are penalized for conducting multiple tests.

In sum, the analyses show that only few of the explanatory variables suggested by the extant management literature were able to explain variation in common schedule overruns. As argued before, the two steps taken by this thesis to improve upon previous studies, namely the use of documented and not self-reported performance and controlling for the false discovery rate, might explain the phenomenon. Additionally, the available sample size of schedule overruns was smaller than the available sample of cost overruns. The analysis now turns to explaining the thickness of the tails.

5.2.2 Determinants of the Thickness of the Upper Power Law Tail of the IT Project Schedule Performance

Following the same steps as chapter 5.1.2 took to analyze the cost performance, the sample was split along the median of a potential explanatory variable. For each subsample the upper tail of the distribution of schedule overruns was fitted to a power law, i.e., the actual/estimated duration for all observations above $x_0 > 1$. A likelihood ratio test was then used to compare both fits and determine, whether the model with the two subsamples explains the distribution significantly better than a model with the pooled sample. Table 21 shows the results.

Table 21 Test for Thickness of the Upper Tail of the Schedule Performance Distribution

Abbreviated description	Median	IQR	N ₁	N ₂	Alpha ₁	Alpha ₂	LR	p-value	adj. p*-value
H2.1.1 Estimated cost	1.2	8.4	68	182	2.966	3.098	0.188	0.910	1.000
H2.1.2 Estimated duration	181	203	240	214	3.428	2.785	9.594	0.008 **	0.116 .
H2.1.3 Actual cost	0.8	6.7	83	363	3.083	3.078	0.000	1.000	1.000
H2.1.4 Actual duration	1096	1461	24	428	3.280	3.071	0.183	0.912	1.000
H2.2.1 Clear goals, business case	8.0	2.3	23	42	3.545	2.777	1.647	0.439	1.000
H2.2.2 Environmental influence	8.0	3.0	30	34	3.519	2.687	2.249	0.325	1.000
H2.2.3 No resistance	3.0	4.0	32	33	2.793	3.226	0.676	0.713	1.000
H2.2.4 User involvement	8.0	2.0	19	37	4.093	2.807	3.078	0.215	1.000
H2.2.5 No political intent	3.0	3.0	18	20	2.482	3.240	1.413	0.493	1.000
H2.2.6 Realistic, reliable schedule	8.0	3.0	31	34	3.452	2.698	1.930	0.381	1.000
H2.2.7 Uniqueness	5.0	4.0	16	48	3.638	2.857	1.249	0.535	1.000
H2.3.1 Stable requirements	5.0	4.0	29	36	2.986	2.992	0.000	1.000	1.000
H2.3.2 Innovation	7.0	3.0	23	42	3.429	2.810	1.112	0.574	1.000
H2.3.3 Decoupled deliverables	7.0	3.0	39	26	2.976	3.010	0.004	0.998	1.000
H2.3.4 Minimal & stable scope	5.0	4.0	17	16	2.417	3.535	2.404	0.301	1.000
H2.3.5 Effective PM method	8.0	2.0	15	49	3.632	2.872	1.126	0.570	1.000
H2.3.6 Few unknowns in the design	5.0	4.0	31	34	2.668	3.414	1.973	0.373	1.000
H2.3.7 Incremental approach	6.0	6.0	24	41	4.011	2.660	4.530	0.104 .	0.872
H2.4.1 Low processing requirements	7.0	3.0	28	35	3.509	2.738	1.838	0.399	1.000
H2.4.2 Low reusability requirements	8.0	4.0	19	45	2.750	3.138	0.488	0.784	1.000
H2.4.3 Low performance requirements	8.0	3.0	31	33	2.999	3.012	0.001	1.000	1.000
H2.4.4 Low security requirements	8.0	4.0	37	27	3.311	2.698	1.334	0.513	1.000
H2.5.1 Outside view	5.0	4.8	10	28	3.050	2.729	0.183	0.913	1.000
H3.1 Effective risk management	8.0	2.0	14	50	3.679	2.876	1.165	0.559	1.000
H4.1.1 Monitor & control	8.0	2.0	27	38	4.105	2.585	6.073	0.048 *	0.504
H4.1.2 Benefits measured	0.0	0.0	30	427	2.516	3.133	3.181	0.204	1.000
H4.1.3 Detailed, up-to-date plan	8.0	2.0	13	20	3.983	2.433	3.497	0.174 .	1.000

<i>H4.1.4 Effective change management</i>	8.0	1.3	7	26	3.490	2.677	0.706	0.703	1.000
<i>H4.2.1 Sustainable resource mix</i>	8.0	2.0	7	24	3.728	2.846	0.692	0.708	1.000
<i>H4.2.2 Competent project managers</i>	9.0	2.0	25	40	2.945	3.018	0.019	0.991	1.000
<i>H4.2.3 Qualified & motivated team</i>	9.0	2.0	17	48	5.834	2.646	11.027	0.004**	0.116.
<i>H4.2.4 Suitably staffed</i>	5.0	5.0	41	24	3.120	2.800	0.363	0.834	1.000
<i>H4.2.5 Years of experience PM</i>	10.0	7.0	19	16	3.183	2.467	1.202	0.548	1.000
<i>H4.3.1 Effective governance</i>	8.0	2.0	18	47	3.051	2.967	0.020	0.990	1.000
<i>H4.3.2 Effective sponsor</i>	8.0	2.0	25	40	3.491	2.767	1.572	0.456	1.000
<i>H4.4.1 Senior management support</i>	8.5	3.0	30	35	3.318	2.774	1.017	0.602	1.000
<i>H4.4.2 Well performing supplier</i>	7.0	1.0	28	36	3.098	3.071	0.002	0.999	1.000
<i>H4.4.3 Robust contracting</i>	9.0	3.0	10	52	2.917	3.129	0.084	0.959	1.000
<i>H4.4.4 Aligned stakeholders</i>	9.0	3.0	20	45	3.293	2.879	0.478	0.788	1.000
<i>H4.5.1 Transparent status comm.</i>	8.0	2.0	14	19	3.375	2.530	1.329	0.515	1.000
<i>H4.5.2 Good team communication</i>	7.0	3.0	22	43	3.536	2.792	1.509	0.470	1.000
<i>H5.1 If true costs were known...</i>	1.0	2.0	24	13	2.352	5.779	10.272	0.006**	0.116.

Note: Alpha₁ is the estimated power law exponent alpha of the subsample where variable > median, Alpha₂ is the estimated power law exponent of the subsample where the variable ≤ median, N₁ and N₂ are the subsample sizes respectively.

LR is the Likelihood Ratio test statistic, the adjusted p-value uses the Benjamini-Hochberg procedure, significance levels are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1, . < 0.2

Before penalizing the results for conducting multiple independent statistical tests, the likelihood ratio test is able to reject four null hypotheses, after adjusting the p*-values only three hypotheses are rejected with weak statistical evidence ($p^* < 0.2$):

H2.1.2: The power law tails of IT project schedule performance do not differ between projects with different estimated durations – projects with a duration ≤ 181 days had a thicker schedule overrun tail ($\alpha = 2.8$) than projects > 181 days ($\alpha = 3.4$).

H4.2.3: The power law tails of IT project schedule performance do not differ between projects with different perceived levels of project team qualification and motivation – projects with a less than highly qualified and motivated team (those scoring ≤ 9 on a 10-point scale) had a thicker schedule overrun tail ($\alpha = 2.6$) than projects with a highly qualified and motivated team (those scoring = 10 had $\alpha = 5.8$).

H2.1.2: The power law tails of IT project cost performance do not differ between projects that differ in their perceptions whether “if the true costs were known the project would not have been built” – projects that fully disagreed with the statement (scoring = 1 on a 10-point scale) had a thicker tail ($\alpha = 2.4$) than projects not disagreeing with the statement ($\alpha = 5.8$).

Additionally, the variable ‘effectiveness of monitoring and controlling the projects’ (H4.1.1), could only be rejected before adjusting for the false discovery rate and thus might be a false a positive. Projects that had less effective monitoring and controlling (scoring ≤ 8 on a 10-point scale) had thicker tails ($\alpha = 2.6$) than those with more effective monitoring and controlling (> 8), where $\alpha = 4.1$.

The results for the variables *motivation and qualification of the team* (H4.2.3) and effectiveness of *monitoring and controlling the project* (H4.1.1, which might be a false positive) were expected, given the extant literature on project management. Higher scores in both variables lead to thinner schedule overrun tails and thus reduced project risks, i.e., fewer schedule outliers. Similarly, the variable ‘*if the true cost of the project were known...*’ (H2.1.2) is indicative of a project failure. Schedule overruns, as argued before, do not automatically imply that a project failed. The results show, however, that projects who agreed with the statement had thicker tails, i.e., schedule outliers occurred more frequently. The finding can also be interpreted in the opposite direction of cause and effect: the thicker the tails of schedule performance, and thus the higher the probability and impact of outliers, the more likely that the project agreed with

the statement ‘if the true cost of the project were known it would not have been built.’

Lastly, the variable *estimated duration* (H2.1.2) is noteworthy because shorter projects had thicker tails and thus higher risks. On the one hand, schedule overruns are expected to be inversely correlated to estimated durations, because schedule overruns are calculated as actual/estimated schedule. On the other hand, the finding might also indicate that longer estimated durations include more buffer, which reduces schedule overruns.

5.2.3 *Exploratory Multilevel Regression Models*

5.2.3.1 *Multilevel Model of Common Schedule Overruns*

The explanatory multilevel model includes the organization as a random effect. The random effect has an intercept of 0.004 (standard deviation = 0.063) and a residual of 0.136 (standard deviation of 0.368). This shows that little variability is located in the first level of the model compared to the fixed effect strengths. However, the level should be kept in the model.

Table 22 shows the fixed effects. Four independent variables have a statistically significant effects on the log-transformed schedule performance. The average schedule overrun decreases with a more *qualified and motivated team*. Schedule overruns are on average 14 percentage points lower for a unit increase in the perceived ranking of the team on a scale of 1-10 ($b = -0.152$, $p = 0.001$). Similarly a unit increase in the perceived effectiveness of the monitoring and controlling

of the project, for example through a PMO, reduced the schedule overrun on average by 5 percentage points ($b = -0.055$, $p = 0.045$).

Furthermore, weak statistical evidence shows the influence of *actual duration* ($b = 0.144$, $p = 0.073$) – actual duration can, of course, only be an explanatory not a predictive factor. Moreover, actual duration is a confounded variable, because it is included in the dependent variable, which is actual /estimated project schedule. Hence a positive relationship between those two variables is expected. However, the variable acts as a proxy for other variables in the model (cf. table 23) and other variables of the analysis (cf. appendix 8.8 for the full correlation matrix). The longer the project was the higher the planned and actual project cost and the less likely that the benefits were measured on this project. Moreover, longer projects tended to be more innovative, faced more organizational resistance, and had higher security requirements.

Lastly, a unit increase in the perception of whether the project was able to *influence its environment* to be successful leads to increase of the average schedule overrun by 4 percentage points ($b = 0.036$, $p = 0.082$). This is surprising given the unexpected direction of the effect, namely that more influence increases the average schedule risk. This result could again be explained through its correlation with other variables (cf. table 23 and appendix 8.8). Influence over the project environment was rated higher if the higher the planned and actual cost. Moreover, projects that had more influence over the project environment also had higher processing and performance requirements

and a less effective team communication. On the positive side, given the discussion in the extant literature, projects with more influence over the project environment had a more effective monitoring and controlling, and more effective contracting.

Table 22 Fixed Effects of the Multilevel Model to Explain Common Schedule Overruns (LN Transformed)

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>p-value</i>	<i>F-value</i>	<i>p-value</i>
<i>Intercept</i>	0.779	0.579	46.63	1.345	0.185	-	-
<i>LN(Estimated cost)</i>	-0.045	0.092	47.52	-0.487	0.629	0.237	0.629
<i>LN(Actual cost)</i>	0.017	0.084	47.86	0.205	0.838	0.042	0.838
<i>LN(Actual duration)</i>	0.144	0.077	24.51	1.874	0.073	3.510	0.073
<i>Influence over project environment</i>	0.036	0.020	47.25	1.775	0.082	3.152	0.082
<i>Realistic and reliable schedule</i>	-0.018	0.034	31.31	-0.514	0.611	0.264	0.611
<i>Incremental development</i>	-0.001	0.020	47.62	-0.029	0.977	0.001	0.977
<i>Outside view</i>	0.013	0.023	47.57	0.57	0.571	0.325	0.571
<i>Effective monitor & control</i>	-0.055	0.027	47.75	-2.057	0.045	4.231	0.045
<i>Benefits measured</i>	0.200	0.124	10.64	1.613	0.136	2.603	0.136
<i>Qualified & motivated team</i>	-0.152	0.044	45.53	-3.425	0.001	11.733	0.001
<i>If the true cost were known...</i>	0.028	0.023	26.97	1.225	0.231	1.500	0.231

Note: t-values calculated using Wald test; F-values calculated using Satterthwaite approximation of degrees of freedom

Table 23 Correlation between Fixed Effects

	<i>LN (Est. cost)</i>	<i>LN (Actual cost)</i>	<i>LN (Actual dur.)</i>	<i>Env. Influence</i>	<i>Realistic schedule</i>	<i>Increm. Dev.</i>	<i>Outside view</i>	<i>Monitor, control</i>	<i>Benefits measured</i>	<i>Qual. Team</i>	<i>If the true</i>
<i>LN(Estimated cost)</i>		3467	3298	179	181	180	128	179	3553	181	128
<i>LN(Actual cost)</i>	0.95 ***		3747	193	195	195	143	194	4005	195	143
<i>LN(Actual duration)</i>	0.37 ***	0.46 ***		186	188	187	140	187	3842	188	140
<i>Influence over project environment</i>	0.18 *	0.19 **	-0.01		218	217	165	216	218	218	165
<i>Realistic and reliable schedule</i>	0.01	0.02	0.01	0.07		219	166	218	220	220	166
<i>Incremental development</i>	0.15 *	0.12 ..	0.11 .	-0.12 ..	-0.02		166	218	219	219	166
<i>Outside view</i>	0.04	0.04	0.01	0.1 .	0.17 *	0.04		165	166	166	165

<i>Effective monitor & control</i>	0.16	0.15 *	0.08	0.17 *	0.23 ***	0.01	-0.11 .		218	218	165
<i>Benefits measured</i>	0.01	0.02	-0.15 ***	0.04	0.09 .	-0.29 ***	-0.22 **	0.05		220	166
<i>Qualified & motivated team</i>	0.09	0.07	0.09	0.09	0.41 ***	0.14 *	0.14 ..	0.31 ***	-0.10 .		166
<i>If the true cost were known...</i>	0.07	0.13 .	0.06	0.00	-0.39 ***	-0.08	0.00	-0.03	0.17 *	-0.26 ***	128

Note: Upper half sample n, lower half correlations with significances are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1,

. < 0.2

Stepwise exclusion of model terms led to a more parsimonious model that explains the determinants of common schedule overruns (cf. table 24). The multilevel model retains its random effect of the organization (var = 0.009, std. dev. = 0.092) and residual (var = 0.114, std. dev. = 0.338). The remaining factors in the model are the log-transformed actual project duration, the perceived influence over the project environment, perceived effectiveness of monitoring and controlling the project, whether the benefits were measured, and the perceived qualification and motivation of the project team. Effect strengths and directions are very close to the previous model, which indicates that neither suppressor nor confounding variables were excluded from the model.

Table 24 Parsimonious Explanatory Fixed Effects Model

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>p-value</i>	<i>F-value</i>	<i>p-value</i>
<i>Intercept</i>	0.848	0.369	55.41	2.297	0.025	-	-
<i>LN(Actual duration)</i>	0.153	0.048	86.23	3.206	0.002	10.2779	0.002
<i>Influence over project environment</i>	0.027	0.015	95.83	1.846	0.068	3.4075	0.068
<i>Effective monitor & control</i>	-0.055	0.019	99.06	-2.804	0.006	7.8645	0.006
<i>Benefits measured</i>	0.192	0.090	16.96	2.134	0.048	4.5538	0.048
<i>Qualified & motivated team</i>	-0.154	0.027	95.11	-5.777	<0.001	33.3689	<0.001

Note: t-values calculated using Wald test; F-values calculated using Satterthwaite approximation of degrees of freedom

5.2.3.2 Multilevel Model of Schedule Outliers

Next, the fitted multilevel model is fitted, where the dependent variable is the odds of an outlier occurring, which were classified following Tukey's rule. The multilevel model includes organization as a random effect (var = 5.314, std. dev. = 2.305). The random effect accounts for considerable variation compared to the effect strength (cf. table 25). The effects in the final model⁶⁶ are the log-transformed estimated duration, the perceived effectiveness of monitor and controlling, and the perceived qualification and motivation of the project team.

Table 25 Multilevel Fixed Effects of Model for Schedule Outliers

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>	<i>F-value</i>
<i>Intercept</i>	24.259	7.193	3.372	0.001	24.259
<i>LN(Estimated duration)</i>	-1.817	0.730	-2.488	0.013	-1.817
<i>Effective monitor & control</i>	-0.638	0.242	-2.636	0.008	-0.638
<i>Qualified & motivated team</i>	-1.518	0.464	-3.269	0.001	-1.518

Note: t-values calculated using Wald test; F-values calculated using Satterthwaite approximation of degrees of freedom

All three effects are negative. That is they reduce the log odds of a project being a schedule outlier for a unit increase in the explanatory variable. Figure 36 depicts the predicted average probability of a schedule outlier for a the range of values of the three explanatory variables in the model.

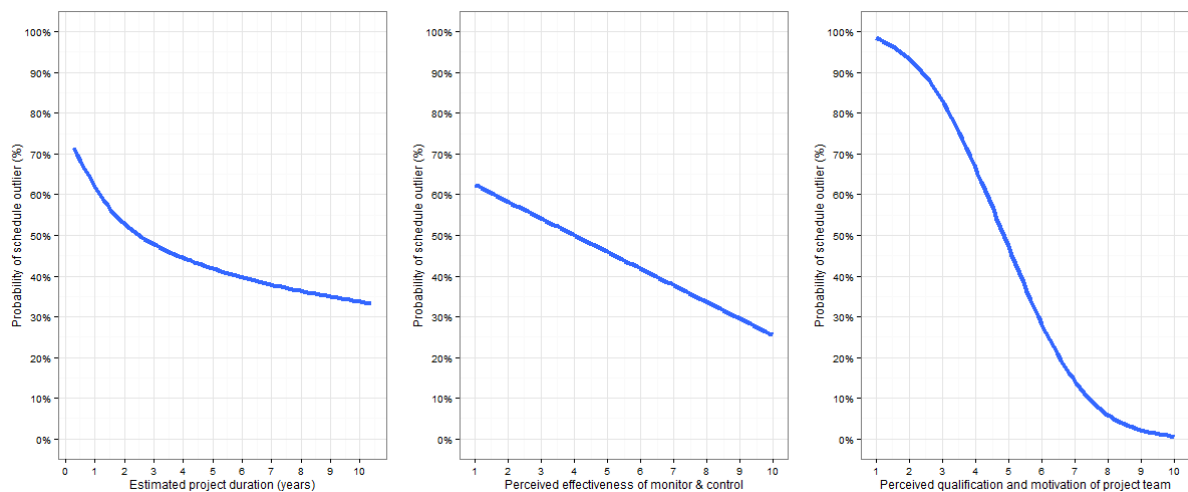
Figure 36 shows that the effect of longer *estimated duration* slowly decays: a one year project has on average a 63% chance of being a schedule outlier, a two year project has on average 54% chance of being a schedule outlier, and a nine year

⁶⁶ The only variable that was identified by comparing the thickness of the tails that was excluded from the model was "If we had known the true cost the project would not have been built" (p = 0.2645). Excluding the variable from the model did not impact the other estimates.

project has on average a 35% chance of being a schedule outlier. As argued when comparing the tails of the schedule overrun distribution, the variables estimated duration and schedule overrun are expected to be inversely correlated because schedule overruns are defined as actual/estimated schedule. Moreover, longer estimated durations could signify buffer in project plans, which allow to catch-up slippage. However, the graph shows that his relationship is not linear.

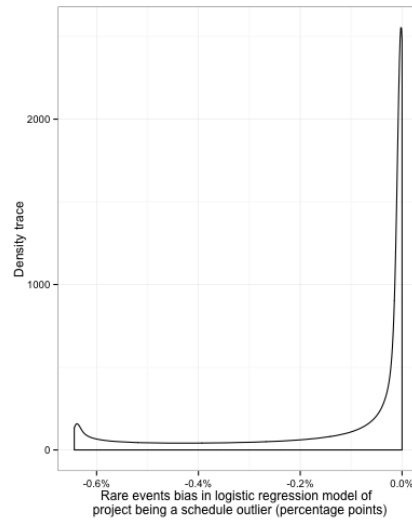
The effect of the perceived effectiveness of the *monitoring and controlling* of the project is more linear. A project rated 1 = "Project had no control over spending, resource consumption, and progress" has on average a 63% probability of being a schedule outlier, whereas a project rated 7 = "Information sufficient to support decision-making" has on average a 38% probability of being a schedule outlier.

The effect of the perceived *qualification and motivation of the project team* was the strongest. A project that is rated 1 = "Team mostly inexperienced, lacking qualification, and/or not overly motivated" has an average a 98% probability of being a schedule outlier, while a project that is rated 7 = "Motivated, qualified and experienced team members on the critical tasks, with the occasional gap in the wider team" has on average a 14% chance of being a schedule outlier.

Figure 36 Predictions of Average Probability of Schedule Outlier

Lastly, the question arises of how biased these rare events logistic regression models are because they predict an event, that is the occurrence of a markedly deviant observations, where the reasonable expectation is that the rare event occurs less frequently than 0.5. Given the sample size the rare events bias in the first explanatory model is estimated to be -0.0258, in the more parsimonious final model the bias is -0.006. The average bias is very small with 0.1 percentage points and ranges with maximum of approximately -0.6 percentage points (cf. figure 37), which is much smaller than the average predicted effect size. It shows that the bias is limited, because outliers are not very rare events, and that the prediction errs on the conservative side minimally underestimating the occurrence of schedule outliers.

Figure 37 Rare Events Bias in Logistic Regression Explaining Cost Outliers



In sum, the comparison of the thickness of the tails and the multilevel models showed that the project team's motivation and qualification and the effectiveness of monitoring and controlling are explanatory factors of why outliers, that is markedly deviating schedule overruns occurred.

5.3 Summary of Findings Why Outliers Occurred

This chapter analyzed why outliers, that is markedly deviant observations, occurred in the cost and schedule performance⁶⁷ of IT projects. Several models were fitted. Five factors explained the occurrence of outliers, both in the thickness of the tail as well as the odds of an outlier occurring: estimated cost, estimated duration, perceived uniqueness of the project, effectiveness of monitoring and controlling, and qualification and motivation of the project team. Table 26 on the next page shows the summary of the findings of the effects. The results can be contrasted two-fold: (1) the comparison of explanations of overruns versus outliers, and (2) the comparison of cost versus schedule outliers.

First, the variables that explain common cost and schedule overruns in this sample do not equally explain why outliers occurred. For example, estimated cost and duration explained the occurrence of outliers but not cost overruns⁶⁸.

Similarly, estimated duration explained schedule outliers, but not schedule overruns. Actual duration explained cost overruns, an effect due to the confounding of the two variables, but actual cost did not explain outliers.

Further variables that explained cost overruns but not cost outliers were effectiveness of risk management, years of experience of the most senior project

⁶⁷ The analysis of effort estimates did not yield any further findings and can be found in appendix 8.10.

⁶⁸ The pairwise comparison of small and large projects by estimated and actual cost was significant while the median cost overruns were the same in both samples, which indicated that the differences in skew led to the rejection of the hypothesis

manager, and the performance of suppliers. In the same vein, environmental influence of the project, realism and reliability of the schedule, and whether the project measured the benefits explained schedule overruns but not outliers.

Table 26 Summary of Findings

Schools of thought		Cost		Schedule		Effort	
	Abbreviated description	Overrun	Outlier	Overrun	Outlier	Overrun	Outlier
System-centric	H2.1.1 Estimated cost	--	***†	(-)			
	H2.1.2 Estimated duration	(-)	** / ***†		***		
	H2.1.3 Actual cost	--		(-)			
	H2.1.4 Actual duration	(-)	--	***			
	H2.2.1 Clear goals, business case			(-)		(-)	(-)
	H2.2.2 Environmental influence			*		(-)	
	H2.2.3 No resistance		(-)				*
	H2.2.4 User involvement	(-)					
	H2.2.5 No political intent		(-)				
	H2.2.6 Realistic, reliable schedule	(-)	(-)	--		(-)	
	H2.2.7 Uniqueness	(-)	**				(-)
	H2.3.1 Stable requirements	(-)		(-)		(-)	
	H2.3.2 Innovation			(-)			
	H2.3.3 Decoupled deliverables						
	H2.3.4 Minimal & stable scope						
	H2.3.5 Effective PM method					(-)	
	H2.3.6 Few unknowns in the design	(-)		(-)			
	H2.3.7 Incremental approach			(-)	(-)		(-)
	H2.4.1 Low processing requirements		(-)			(-)	
	H2.4.2 Low reusability requirements						
	H2.4.3 Low performance requirements						
	H2.4.4 Low security requirements			(-)			
	H2.5.1 Outside view			(-)			
Event-centric	H3.1 Effective risk management	*					(-)
Process-centric	H4.1.1 Monitor & control			***	***	(-)	(-)
	H4.1.2 Benefits measured		*	**		(-)	
	H4.1.3 Detailed, up-to-date plan			(-)	(-)		
	H4.1.4 Effective change management						
	H4.2.1 Sustainable resource mix					(-)	
	H4.2.2 Competent project managers					(-)	(-)
	H4.2.3 Qualified & motivated team			***	***		
	H4.2.4 Suitably staffed		(-)				(-)
	H4.2.5 Years of experience PM	**		(-)		(-)	(-)
	H4.3.1 Effective governance	(-)					
	H4.3.2 Effective sponsor	(-)					
	H4.4.1 Senior management support			(-)		(-)	
	H4.4.2 Well performing supplier	*					
	H4.4.3 Robust contracting	(-)					
	H4.4.4 Aligned stakeholders					(-)	
	H4.5.1 Transparent status comm.					(-)	
	H4.5.2 Good team communication					(-)	

Effort analysis is included in appendix 8.10, regression results: *** $p < 0.001$, ** $p < 0.05$, * $p < 0.2$; pairwise comparisons: -- $p < 0.05$, not significant in regression models, (-) false positive, i.e., effect only in pairwise comparisons before adjusting for false discovery rate, † denotes the fitted predictive model that only included estimated duration and cost

Second, cost and schedule outliers can be explained both with estimated schedule⁶⁹. Estimated cost and perceived uniqueness of the project only explained cost outliers. While schedule outliers were associated with the perceived effectiveness of monitoring and controlling and the perceived qualification and motivation of the project team.

In sum, Figures 38 and 39 compare the five variables that explained why cost and schedule outlier occurred. Figure 38 compares the thickness of the tails and figure 39 explains the odds of outliers occurring, that is markedly deviant observations of project performance, through multilevel rare events regression with the organization as random effect. The results of both analyses are consistent, except for the variable effectiveness of monitoring and controlling the project.

The comparison of the thickness (fig. 38) of the tails shows that projects with lower *estimated cost* and *estimated duration* had thicker tails. The analysis of the odds of the observing outliers (fig. 39) shows that the risk initially increased with increasing estimated cost and then stays flat. Moreover, it shows to opposite effects that with increasing estimated duration the probability of a cost outlier increased and the probability of schedule outliers occurring decreased.

The higher the *perceived uniqueness* of the project the thicker were the cost overrun tails, however it did have a significant effect on the schedule overrun

⁶⁹ Note: The attempted fit of a multilevel model predicting the probability of a project being a schedule outlier or an effort outlier based solely on the variables of estimated cost and estimated duration did not result in meaningful models.

tails. The regression analysis showed that the probability of the occurrence of a cost outlier increased with increasing uniqueness, while the probability of schedule outliers slightly decreased – however the effect on the probability of schedule outliers was not significant.

The *effectiveness of monitoring and controlling* had opposite effects. More effective monitoring and control increased the probability of cost outliers but decreased the probability of schedule outliers in the regression analysis. The schedule overrun tail was thicker for less effectively controlled projects. The difference between the tails was not statistically significant.

Finally, the *qualification and motivation of the project team* showed that better teams decreased the thickness of the schedule and cost overrun tails, but the effect was not significant for the cost overrun tail. The regression analysis came to the same result and indicated that both the probability of cost and schedule outliers occurring decreased with more qualified and motivated teams.

Lastly, the approximation of the rare events bias (cf. figure 40) in the multilevel regression models showed that the results err on the conservative side. A predictive model based on the above results would underestimate the occurrence of outliers by a very small degree compared to the effect sizes.

The next chapter will discuss the theoretical and practical implications of the findings of both research questions: How frequent are outliers, in the sense of

markedly deviating observations of IT project performance? And what factors explained the occurrence of outliers in IT project performance?

Figure 38 Comparison of Thickness of Power Law Tails for 5 Selected Variables

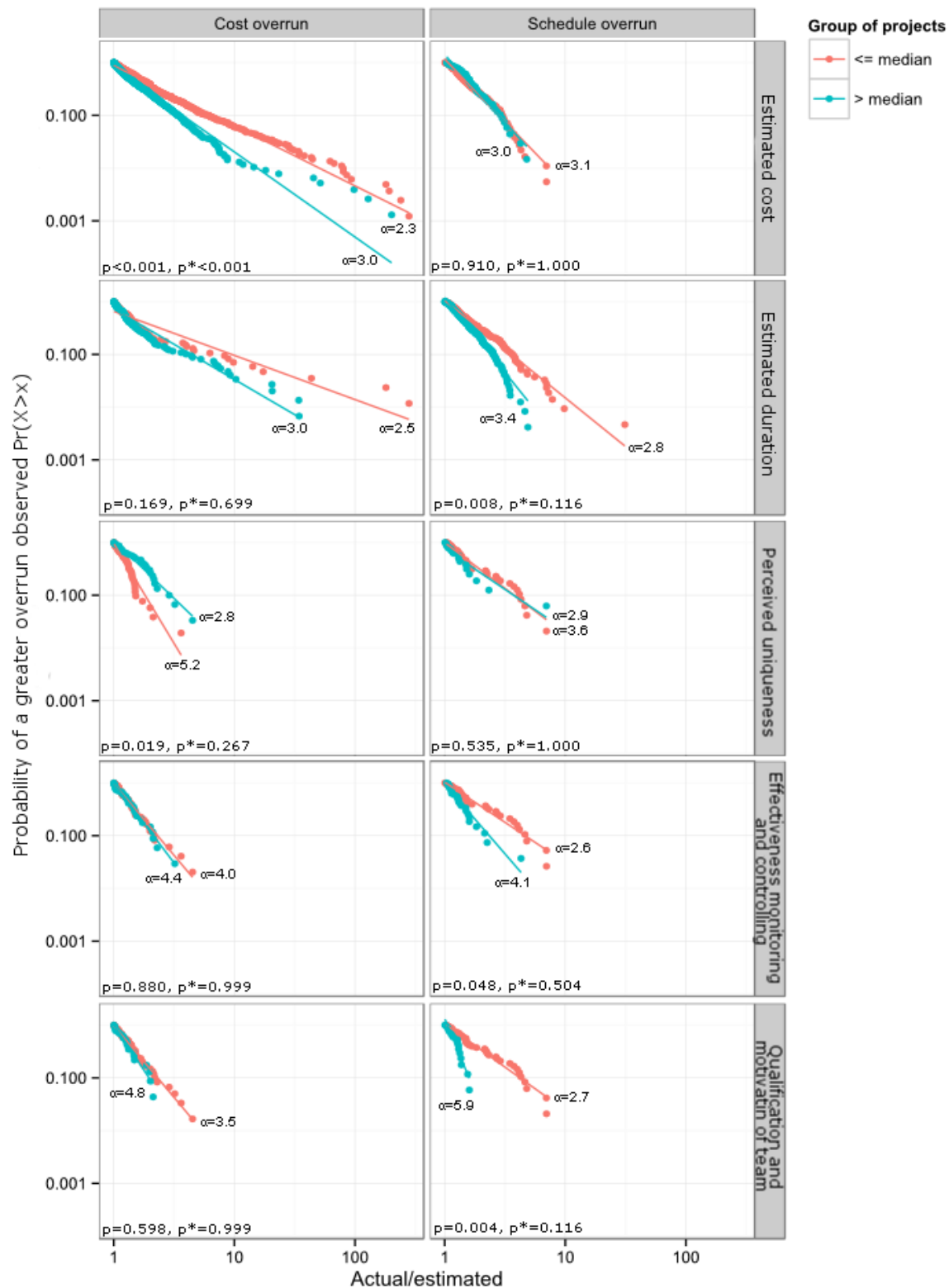


Figure 39 Main Effects of Logistic Regression to Explain Occurrence of Outliers in IT Projects

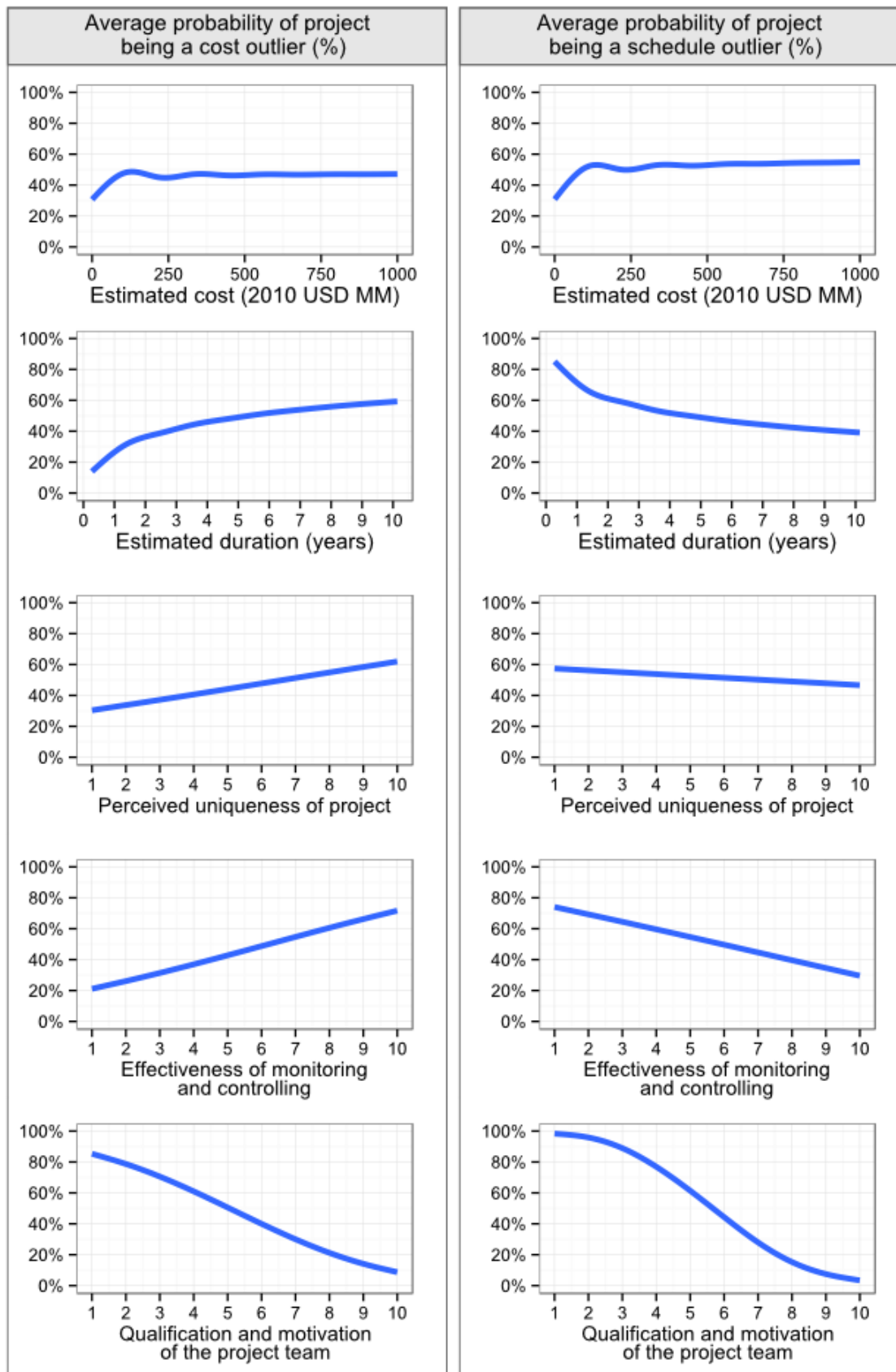
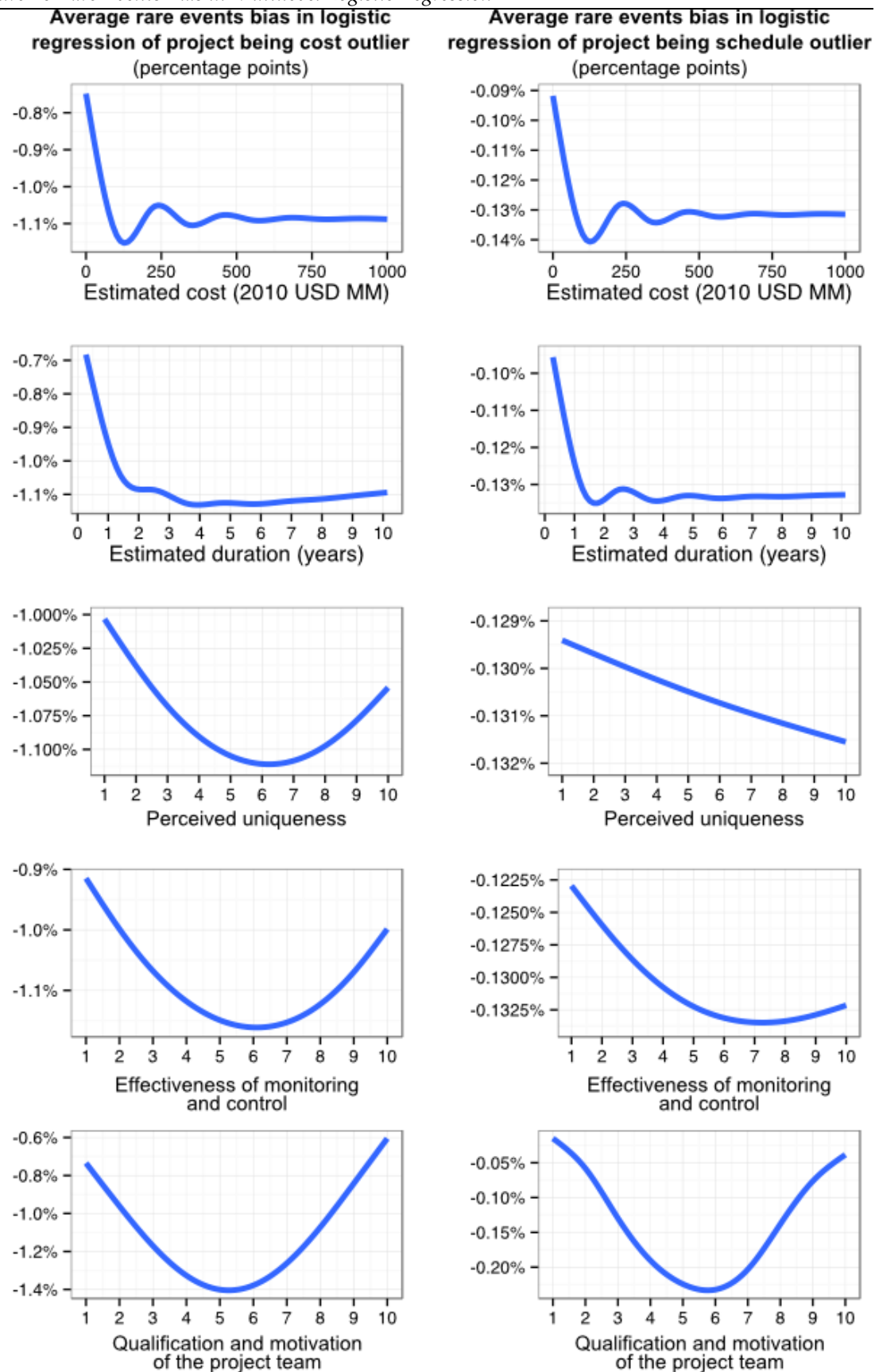


Figure 40 Rare Events Bias in Multilevel Logistic Regression



6 Discussion and Implications of Findings

This chapter will discuss the implications of the findings of the two research questions in turn:

I: How frequent are outliers in the performance of IT projects?

II: Why do outliers occur in the performance of IT projects?

6.1 How Frequent are Outliers in the Performance of IT Projects?

The analysis found that outliers, i.e. markedly deviant observations of IT project performance did occur more frequent than suggested by the assumption of thin tails. Moreover, IT project performance is not normally distributed. Instead, the tails of cost, effort and schedule overruns can be described by power laws. In sum, it can be concluded from these findings that (1) outliers in IT project performance occur frequently and consistently and (2) outliers are not from a different universe of uncertainty. These conclusions were found to be highly robust and ruled out alternative explanations of the power law tails, e.g., mixture of distributions, hierarchy of exponentials, and aggregation over non-stationary variables. The conclusion has implications for two areas, which will be discussed in the next section (1) reasearching IT project performance and (2) IT project forecasting and risk modelling.

6.1.1 *Implications of Power Law Tails for Researching Project Performance*

The findings have two implications for researching IT project performance. First, the analysis concluded that outliers are not from a different universe of uncertainty than common observations in a sample. The finding implies that

the frequently used strategy of rejecting outliers is not suitable to deal with outliers when analyzing project performance. To paraphrase Steele and Huber (2004) the thesis showed that:

Markedly deviant observations of IT project performance are not unusual and should not be thrown out. On the contrary, outliers of IT project performance might hold the clue to data behaviors that are not revealed by the rest of the information.

Moreover, the same implications can be drawn for other fields with fat-tailed behavior. Verhoef (2004) and Kitchenham (1998) have shown that similar distributions are present in other variables used to research IT projects, e.g., distribution of project sizes in a portfolio, software development productivity. Outliers were also present in the study of other project types, e.g., hydroelectric dams (Ansar et al. 2014) and urban rail projects (Flyvbjerg 2007). Moreover, other areas in operations management have suggested fat-tailed behaviors, for example, inventory and demand management (Tadikamalla 1979), operational risks in banking and insurances (Embrechts et al. 2003), the occurrence of cascading failures in networked systems (e.g., blackouts in the electric grid Zaretski et al. 2009), and human productivity (e.g., surgical procedure times Strum et al. 2000). In sum, the thesis has demonstrated how outliers can be incorporated in the analysis of determinants of these types of distributions.

Second, the findings imply an explanation of why prior studies of IT project performance lacked consensus and showed a large amount of variation between and within studies, as discussed in chapter 2.3.1.

On the one hand, power law tails imply that reported mean and variance should be taken with a grain of salt. While mean and variance can be calculated for every empirical sample, describing observations in this way makes assumptions about the nature of uncertainty behind the observed phenomenon. Most importantly, that the mean and standard deviation are not skewed by outliers. In the language of statisticians: reporting mean and variance assumes that the central limit theorem holds true – the average of a random variables converges to the true population mean with large samples. The strong law of large numbers assumes that variance and mean are both finite. This does not hold true for power law tails, i.e., higher order moments only slowly converge, if at all.

The higher order moments for power law tails are defined by the formula $\langle m \rangle = \frac{\alpha-1}{\alpha-1-m} x_0^m$. The median is defined as $\langle x_{1/2} \rangle = 2^{1/(\alpha-1)} x_0$. Table 27 shows the calculated mean and variance for the power law tails of cost, schedule, and effort overruns. The expectation value of the observations in the tail is that if costs overrun the budget the expected cost overrun is +220% (the median cost overrun in the tail is +72%), if projects overrun their schedule the expected schedule overrun is +97%, and if projects overrun their effort estimates the expected effort overrun is +284%.

Table 27 Power Law Exponents and Higher Order Moments of the Key Performance Dimensions of IT Projects

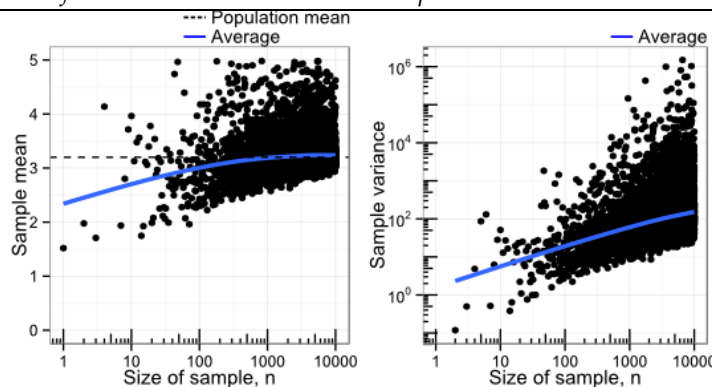
Dimension	x_0	alpha	Mean	Variance	Median
Cost performance (in actual/estimated cost)	1.09	2.5	3.20	∞	1.72

Schedule performance (in actual/estimated duration)	1.09	3.2	1.97	11.33	1.49
Effort performance (in actual/estimated effort in PD)	1.56	2.7	3.84	∞	2.36

While, these first order moments are limited in their application, because they only describe the tail, they demonstrate a more important issue: In the case of cost and effort performance variance is infinite, which implies that the average is not well defined and explains why the findings of extant studies failed to reach consensus and cannot serve as a benchmark.

Infinite variance implies that the mean converges slowly. Figure 41 demonstrates with synthetic data that this convergence process for the mean is slow because variance increases with increasing sample sizes. The synthetic cost overrun data suggests, for example, that sample sizes $> 1,000$ observations are needed to reach convergence.

Figure 41 Convergence of the Mean in Cost Overrun Samples



In other words, the slow converge is due to the occurrence outliers, that is markedly deviant observations of IT project performance. In power law tails the spacing of observed average maximum value follows a power law itself. The observed average maximum is the average maximum observed in samples of

size n . The value of the observed average maximum overrun x_{max} is related to the sample size through the formula:

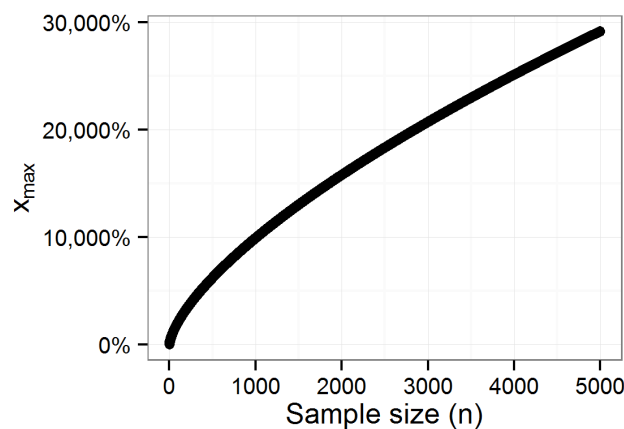
$$\langle x_{max} \rangle \sim n^{\frac{1}{\alpha-1}}$$

For the cost overruns of IT projects the value is

$$\langle x_{max} \rangle \sim n^{2/3}$$

Figure 42 depicts this relationship: the observed average maximum values grow with larger sample sizes. The figure also illustrates that the average maximum cost overrun has no upper bound. In other words, very large cost overruns are found if the horizon of observation is long or big enough. This implies that a small samples will lead to bias that underrepresents the risk of cost overruns. The power law might also explain why the larger n studies (Eveleens and Verhoef 2009, Standish 1994-2004) reported higher average cost and schedule overruns, because means are influenced by the maximum observed value.

Figure 42 Observed Average Maximum Cost Overrun (x_{max}) Depending on the Sample Size (n)



In sum, the thesis draws three implications for the research of IT project performance. (1) Outliers should not be excluded from data sets, they are an important part of the phenomenon. (2) The mean and standard deviation are influenced by outliers, which leads to slow convergence of true values, if they exist at all. (3) Large n studies are needed to adequately characterize the phenomenon. These implications might also apply to other areas of quantitative research in project and operations management, which were not part of the empirical basis of this thesis.

6.1.2 Implications of Power Law Tails for IT Project Forecasting and Risk Modelling Practice

The power law tails of the performance of IT projects have implications for the forecasting and modelling of risk. For example, the findings showed that tail of the distribution of cost performance of IT projects, expressed as actual/estimated cost, can be fitted to a power law of the form

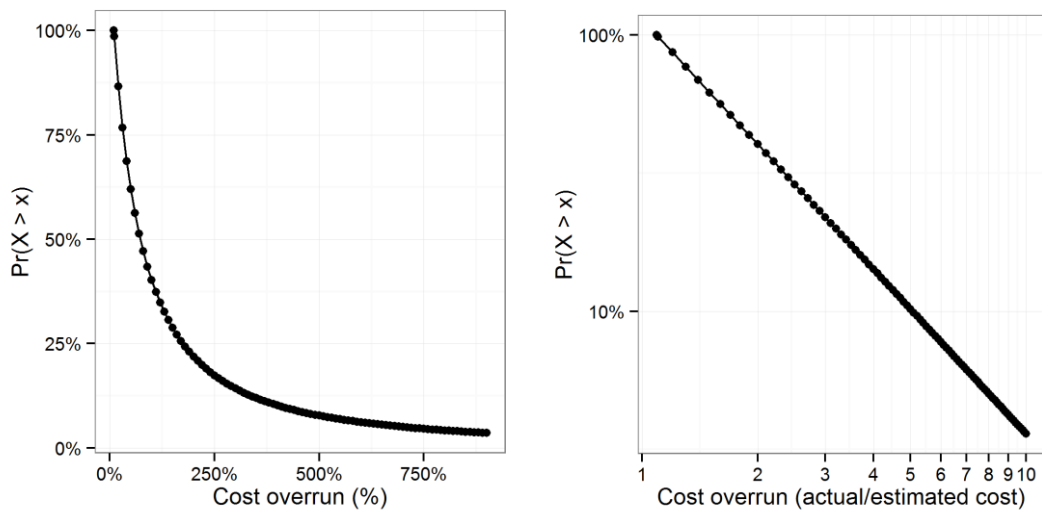
$$p(x| \propto) = \left(\frac{1.5}{1.09}\right) \left(\frac{x}{1.09}\right)^{-2.5}$$

Schedule overruns follow a power law of the form $p(x| \propto) = \left(\frac{2.2}{1.09}\right) \left(\frac{x}{1.09}\right)^{-3.2}$ and effort overruns follow a power law $p(x| \propto) = \left(\frac{1.6}{1.56}\right) \left(\frac{x}{1.56}\right)^{-2.7}$. The inverse cumulative distribution function of the cost overrun tail, that is the probability of observing an overrun larger than x is

$$\Pr(X > x) = \left(\frac{x}{1.09}\right)^{-1.5}$$

Visualizing the cumulative distribution shows the typical quick decay and plotted on a double logarithmic diagram shows the signature straight line (cf. fig. 43). The cumulative distribution has six implications for the cost risk analysis.

Figure 43 Inverse Cumulative Distribution of Cost Overruns $\Pr(X > x)$ on Linear-Linear and Log-Log Plot



First, very large overruns happen infrequent but are not impossible.

Second, not all higher order moments converge. Empirically measured moments depend on the sample size and might incorrectly suggest that risk is bounded.

Third, the probability of a cost overrun in the tail, that is above than the cut-off value $x_0 = 1.09$, which is equivalent to being nearly on budget, decays but much slower than a thin-tailed Gaussian distribution assumes. For example, if a project has a cost overrun there is a 62% probability of the cost overrun being

greater than +50%, a 40% probability of the cost overrun being greater than +100%, and a 22% probability of the cost overrun being greater than 200%.

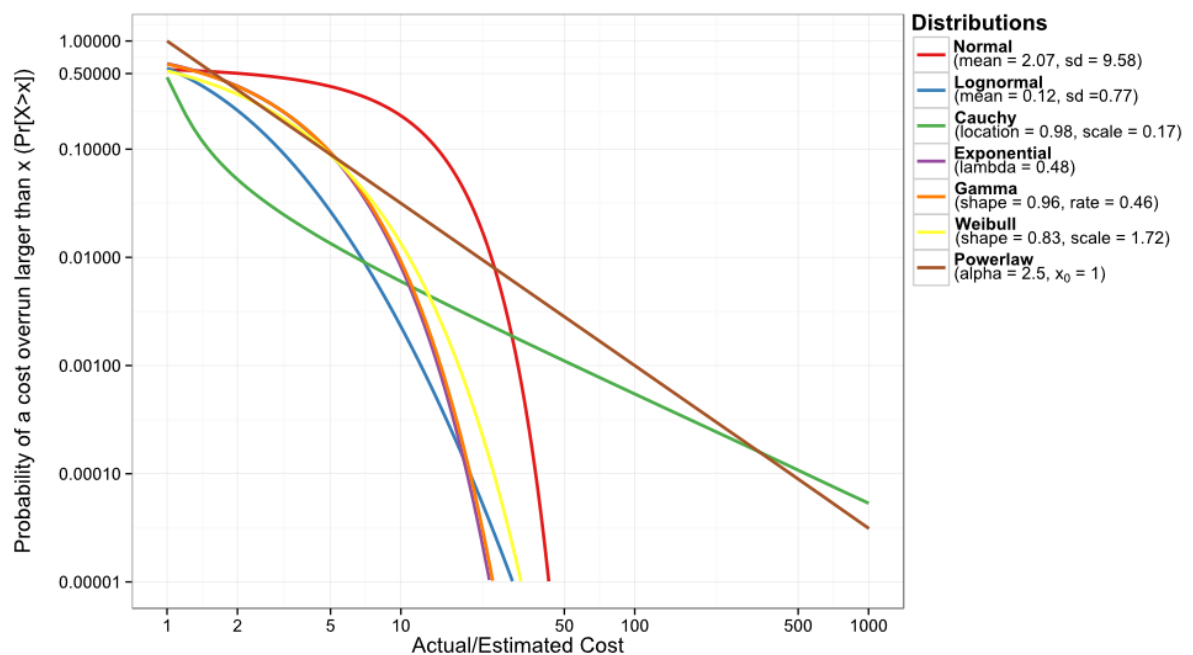
Additionally, the observed average maximum $\langle x_{max} \rangle \sim n^{\frac{1}{\alpha-1}}$, which was discussed above, also allows to forecast the extreme downside risk of IT projects. This application is similar to risk forecasts of catastrophic events, e.g., earthquakes or floods. The observed average maximum of cost overruns implies, that a +10,000% cost overrun is equivalent to an event occurring in one in a thousand projects. A +2,054% cost overruns is an event occurring in one in a hundred projects and a +364% cost overrun is an event occurring in one in ten projects.

Fifth, the results presented evidence that IT project performance can be fitted to power law tails. The implications for risk modelling can be illustrated by comparing the different types of distributions used in risk models. Figure 44 illustrates the distributions if the cost overrun data are fitted. The figure shows that models with a finite variance, i.e., normal, lognormal, Gamma, Weibull, or exponential models, underrepresent the occurrence of large overruns. For example, the probability of observing a cost overrun larger +900% (actual/estimated cost = 10) has a probability of about 0.002, that is 1 in 500 observations when using a lognormal model, whereas the power law distribution describes the probability a cost overrun of +900% to be approximately 0.03 or 1 in 30 observations.

Underrepresenting large overruns is more severe the larger the overrun.

Models assuming finite variance predict that overruns with actual/estimated cost > 50 are highly unlikely, i.e., less than 1 in 10,000. The power law, however, describes them to occur with a probability of 1 in 300.

Figure 44 Comparison of the Tails of Commonly Used Distributions in Risk Modelling



Note: the probability of a cost overrun larger than x has been cut off at 0.00001

Sixth, when modelling the risk of IT projects a common practitioners' concern is that relative cost overruns are not an adequate expression of the true risk for the organization. A +100% cost overrun on a USD 1 million project is less threatening to firm survival than a +100% cost overrun on a USD 100 million project. In other words, the nominal overruns are a more important risk than the relative overrun. Table 28 compares the nominal and relative overruns. Nominal overruns are here expressed in 2010 USD millions for cost and in calendar days for schedule.

Table 28 Power Law Tails of Key IT Project Risks

Dimension	x_0	alpha	Mean	Variance	Median
Cost performance (in actual/estimated cost)	1.09	2.5	3.20	∞	1.72
Nominal cost overrun (in 2010 USD millions)	2.72	1.6	∞	∞	8.28
Schedule performance (in actual/estimated duration)	1.09	3.2	1.97	11.33	1.49
Nominal delay (in calendar days)	266	3.2	491.16	851047	365.50

Table 28 shows that the implication for the understanding the true risk of IT projects are even more pronounced when studying nominal cost overruns. The tail of the nominal cost overrun fits a power law with $\alpha = 1.6$ for all $x > 2.72$, that is projects more than 3-times over budget. The tail is fatter than the tail of cost overrun expressed as actual/estimated cost. While the median nominal cost overrun is USD 8.3 million, the mean nominal cost overrun is infinite.

In other words, the tails of the distribution of nominal cost overruns in USD terms are so fat that the distribution does not have a mean (or any other higher order moment). This property can be illustrated with the top heaviness of the distribution. For example, concentration of wealth in a country is typically shown to be top heavy by demonstrating that 80% of the total wealth is held by 20% of the population. Similarly one could ask the question: How many dollars of overruns are incurred by the top 20% projects⁷⁰?

Because the power law exponent α is less than two the answer to this question is: all of the nominal cost overruns – if the number of observations is

⁷⁰ In statistical terms which share of cost is incurred by projects that exceed a cost overrun of x is given by $W(x) = \frac{\int_x^\infty x' p(x) dx'}{\int_{x_0}^\infty x' p(x) dx'}$

large there will be one project that had a higher nominal overrun than the rest of the projects combined. To paraphrase Mandelbrot (1997): Not only the top 50% or top 20%, but even the worst 1% of projects account for essentially all USD overruns. Put differently, the nominal cost overruns expressed in absolute USD terms show that the monetary damage is infinitely top heavy.

In sum, the findings imply that the true risk of IT projects are dominated by the fat tails of the distribution. In the most extreme case, of nominal cost overruns in USD terms, the expected risk is infinite.

Lastly, one implication is noteworthy for forecasting theory. Not all distributions observed here follow a power law tail. Comparing the actual and estimated cost shows a surprising pattern. Estimated cost hardly follow a power law only the very end of the distribution above USD 1.8bn can be fitted to a power law. In fact, the Vuong test suggests that the distribution has a lognormal tail with the parameters $\mu = 0.24$ and $\sigma = 2.51$ for all $x > 0.01$. The log normal distribution has finite higher order moments. The mean is $e^{\mu+\sigma^2/2} = e^{3.39} = 29.7$ and the variance is $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2} = 478,476$. While the variance is very large for all practical purposes it is finite. At the same time the tail of the actual cost distribution can be fit by a power law with $\alpha = 1.9$ for all $x > 143.93$ (cf. table 29).

Table 29 Power Law Tails of Estimated and Actual Cost of IT Projects

Dimension	x_0	alpha	Mean	Variance	Median
Estimated cost (in 2010 USD millions)	1769	2.3	8108	∞	3041.86
Actual cost (in 2010 USD millions)	144	1.9	∞	∞	320.34

The pattern in table 29 is noteworthy because it illustrates the effect of randomness. One common generating mechanism of power law tails is multiplicative error. For example, bottom-up project cost estimates are often derived as the sum of effort in person days needed multiplied with cost per person day. However, the multiplication of errors in both numbers does not necessarily result in a power law distribution because it might create distributions with non-universal exponents (cf. the proof in Bouchaud 2001, Gabaix 2009). A true power law requires additional randomness, either via a lower reflective boundary or a discontinuity of some form, e.g., uniformly random noise (cf. the discussions in Farmer 2004, Newman 2006). The empirical findings fit this explanation: bottom-up planning displays fat tail behavior in the distribution of estimated cost because of multiplicative errors in a cascade of tasks. The actual cost, in comparison, have fatter tails due to additional discontinuities. This pattern has an important implication for (IT) project forecasting and planning: planning inaccuracies might be explained due to discontinuity, e.g., random noise, introduced during project execution. Thus models need to adequately represent these discontinuities if forecasts are to reflect the true risk of projects. In other words, using the terms coined by Mandelbrot (1997), the uncertainty during project execution turns a project

planned for mild or slow randomness into an outcome of extreme randomness⁷¹.

In practical terms, the analysis refutes with overwhelming statistical evidence that IT project performance approximates a Gaussian normal distribution. The data show that other distributions with bounded uncertainty, e.g., Gamma, Weibull, lognormal, also underrepresent the risk of projects. Subsequently, the results have implications for application areas where the normal distribution is a frequently made assumption: (1) in the cone of uncertainty, (2) project portfolio analysis, (3) real options valuation, (4) earned value management, and (5) software analytics.

First, the *cone of uncertainty* (McConnell 1996) is a concept often invoked to discuss the reduction of technical estimation errors through the lifecycle of a project. The cone of uncertainty describes a symmetric conical shape of how deviations (expressed as standard errors, standard deviation, or variance) of forecasts converge towards the actual cost as the project progresses and more information is known to improve the accuracy of forecasts. One conclusion drawn from the cone of uncertainty is that if more and better information were available earlier accuracy of forecasts would improve. The analysis of forecasts at the decision to build has shown that cost and effort forecasts have infinite variance and schedule forecasts have a very large and slowly converging

⁷¹ Mandelbrot (1997) defines the log normal distribution as slow randomness with finite and localized moments. Wild randomness is characterized by infinite second order moments. Extreme randomness is characterized by infinite higher order moments.

variance, which depends on the size of the underlying data set. Eveleens and Verhoef (2009) attempted to fit lognormal distributions to the cone of uncertainty and found that only some of their samples fit adequately. They further argued that due to organizational biases no general distribution might exist. This analysis, has excluded biased forecasts, moreover it showed that the analysis of the general underlying distribution is important, since, as these findings imply, the higher order moments might be infinite which can only be concluded from the general but not the empirical distribution of the data. Moreover, Eveleens and Verhoef (2009) suggest to improve the concept of the cone of uncertainty by using DeMarco's estimation quality factor⁷², which is a time-weighted average of the forecast inaccuracy over the course of a project. This analysis has identified that the unboundedness of the risk of overruns is the key to adequately model risks. Both the cone of uncertainty and the use of a median estimation quality factor to compare different organizations do not adequately represent the risk of IT project performance overruns.

Second, *project portfolio analysis* (McFarlan 1981) suggests that companies should balance the risk of projects: Portfolios loaded with high risk projects threaten the operations of companies. Extending the idea other authors suggested that a diversification strategy can protect a company from disruptions due to failing projects (e.g., Marchewka and Keil 1995). The results of this analysis, however,

⁷² $EQF = \frac{\int_0^{t_a} a \, dt}{\int_0^{t_a} |a - e(t)| \, dt}$, where a is the actual value, $e(t)$ the estimated value at discrete points in time $t=0, \dots, t_a$ if only one estimate is made the $EQF = 1/\text{overrun}$.

showed that the downside risk of individual projects is very large, in some cases (e.g., actual outturn cost, average maximum overrun) even infinite, which questioned whether risks can be mitigated through a portfolio approach.

Third, in *risk modelling* practice, infinite variance implies that option pricing of IT projects (as suggested by e.g., Leslie and Michaels 1997, Tiwana et al. 2006, Huchzermeier and Loch 2001, Benaroch and Kaufman 1999) is problematic because the Black-Scholes price for a call option (the right to buy the next project) converges to zero⁷³.

Fourth, *earned value management* is a common practice to forecast the total cost and schedule of a project at completion (Fleming and Koppelman 2000). In forecasting practice, the budget at completion is estimated as the sum of actual cost at the time of forecasting plus the estimated budget and the earned value adjusted by a performance index. The assumption behind the model is that the performance index of the project is stable. Recent improvements to the earned value method (e.g., Narbaev and De Marco 2014, Caron et al. 2013) focus on estimating budgets and schedules at completion for non-stable performance indices. However, even these improvements assume that the forecasting error is bounded and not, as this analysis has shown, infinite.

⁷³ $C(S, t) = d_1 - \sigma\sqrt{T-t}$ where C is the value of the call option, $T-t$ is the time to maturity, σ the volatility of the asset, and $d_1 = \frac{1}{\sigma\sqrt{T-t}} \left[\ln\left(\frac{S}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t) \right]$, where S is the spot price, K the strike price, and r the risk free rate of return.

Fifth, *software analytics*, that is data-driven predictions of defects, effort, and duration of software development processes are challenged. For example, MacDonnell and Shepperd (2003) compared different estimation techniques on a skewed data set and found that least squares regression gave the most accurate results but the analysis did not include robust methods. Raffo and Kellner (2000) used simulation studies to predict defects, effort, and duration of a software development task. Despite using 'real-world data' results of the simulation were normally distributed. The prediction model tends to be highly dependent on the underlying data, which

“... is frequently heteroskedastic, and may contain a number of outlier observations. As a result we may see a prediction technique perform well on a subset of the data, but then perform very poorly on the remaining data points.” (MacDonnell and Shepperd 2003, p. 92).

To compare the often incongruous results of multiple estimation methods measures of the spread of predictions (the mean magnitude relative error MMRE) and peakedness (kurtosis or the number of predictions within 25% of the actual) are used (Kitchenham et al. 2001). Hassan et al. (2013) argued that future developments of software analytics and predictions should broaden the focus beyond developer tasks, such as coding and bug fixing. The analysis in this thesis showed which analytical problems arise once the focus broadens to capture the full IT project performance from the point of view of the client. Namely, that the challenge of outliers in data sets becomes even more pressing.

Moreover, case based reasoning has been suggested as an alternative to improve prediction accuracy (Finnie et al. 1997) due to its similarity to expert based judgments, which prevail in the industry (Moløkken and Jorgensen 2003). However, this analysis has shown that samples converge very slowly to the population mean and standard deviation, which implies that if case-based reasoning is to be used to predict project performance large very case databases are necessary.

In sum, the findings showed that very large overruns happen infrequently but are not impossible. Moreover, the findings showed that IT projects, that suffered from large overruns, are not special. Overruns are not idiosyncratic chance events from a different universe of uncertainty. They show regular and stable patterns. These findings have three key implications.

First, the findings implied for research that outliers need to be incorporated into analysis instead of being rejected. The scale-free nature of power law behavior suggests that valid explanations of large overruns can be derived from the study of smaller overruns. Moreover, the power law tails clearly show that the empirical sample mean and variance are problematic measures to characterize and compare the outcome of IT projects.

Second, the findings showed that distributions converge slowly. Large sample sizes are needed to understand the nature of the risk of IT projects. For some dimensions, e.g., nominal cost overruns, samples even need to be infinite – which for practical purposes is unachievable.

Second, the findings imply that outliers carry a large risk for IT projects. The performance distributions are top heavy. Outliers lead to infinite variances and in some dimensions, e.g., nominal USD cost overruns, also to infinite means. Risk models need to reflect the larger than anticipated amount of risk in IT projects. Models can adequately reflect the risk by explicitly assuming distributions that have infinite variance. The implications have shown that models used in practice to conceptualize the risk of IT projects, e.g., earned value management, IT project portfolio analysis, the cone of uncertainty, underrepresent the true risk of IT projects because they are based on thin-tailed distributions. Building on the argument that “only risk that is underappreciated and unmanaged has the power to surprise” (Tiwana and Keil 2004, p. 77), it can be deduced that a flawed risk model that underrepresents outliers, which are the true risk of IT projects, is adding more risk to an It project than the risk of overruns itself.

6.2 Theoretical Implications: Why did Outliers Occur in IT Project Performance?

The analysis in chapter 5 attempted to explain the odds of the occurrence of outliers and the thickness of the overrun tails with explanations suggested by the extant management literature. Only five variables had a statistically significant effect. Table 30 summarizes the significant effects. How does the extant literature, reviewed in chapter 2, hold up to these results?

Table 30 Summary of Results of Multilevel Logistic Regression Models: Why do Outliers Occur?

Schools of thought	Abbreviated description	Cost outlier	Schedule outlier
System-centric	H2.1.1 Estimated cost	***	
	H2.1.2 Estimated duration	*/***	***
	H2.2.7 Uniqueness	*	
Process-centric	H4.1.1 Monitor & control		***
	H4.2.3 Qualified & motivated team		***

Note: In all logistic regression models the organization was entered as a first level

*** overwhelming significance ($p < 0.001$), ** very significant ($0.001 \leq p < 0.01$), * significant ($0.01 \leq p < 0.05$),

Before turning to theoretical implications, the findings have implications for research practice. Chapter 5 argued that the low number of might be due to changes in the methodology, which aimed to improve upon prior studies. This thesis put additional rigor on the measurement of the dependent variable.

Outliers were defined based on documented cost, schedule, and effort performance instead of survey methods, which may have been affected by self-selection (Moløkken-Østfold et al. 2004). Moreover, access to data depended on top management support and availability of documents, which most likely introduced a conservative bias in the data.

Second, additional rigor was put on the analyses, making the results conservative. Findings were penalized for the false discovery rate. For example, when attempting to replicate the explanations of common cost overruns 14 variables out of 41 were found to statistically significant at $p < 0.1$. After adjusting for the false discovery rate only 3 variables were still statistically significant at $p < 0.1$. This illustrates that findings from multiple tests need to be controlled for the false discovery rate even if the results then err on the conservative side. This applies not only to success factor research using multiple comparisons (e.g., Pedersen 2014) but also to meta-analytical approaches, e.g., the approach to theory building through meta-analysis Belassi and Tukel (1996) or Ram and Corkingdale (2014).

The findings have theoretical implications. Chapter 2 began with a review of organizational theory before reviewing more domain specific literatures of project management and IT project management. The review showed that most studies were based on one school of thought and little theorizing has taken place across the system-centric, event-centric, and process-centric view. No study has directly compared the effects in a single model⁷⁴. This thesis

⁷⁴ Closest in approach are the studies of Keil et al. (2000) which constructed 4 separate logistic regression models for the constructs of each of their theories, which all came from the process-centric view, but which also included project size. Another large study is Slesman et al. (2012) who conducted a meta-analysis but only of the process-centric view.

One exception are the IT project management theories of success and risk factors. While most studies were a-theoretical they simultaneously portrayed success factors as a design choice in the beginning of the project and as a effectiveness of execution variable. The thesis grouped them in the process-centric school of thought, based on meta-analytical reviews of the field (e.g., Fortune and White 2006), which highlighted the process character and the dynamic interplay between the success and risk factors.

attempted a first comparison of all three schools of thought through including explanatory variables of all three theoretical schools of thought in one analysis. At the same time, the thesis took the point of view of the project sponsor, i.e., comparing full project estimates, covering a broad range of project types, and focusing on the risk of the investment decision at full business case stage.

The analyses identified five factors: estimated cost, estimated duration, perceived uniqueness, effectiveness of monitor and control, and motivation and qualification of the team. The findings implied that the system-centric and the process-centric schools of thought offer theoretical explanations of why outliers occur. The event-centric school of thought is not only the weakest line of inquiry in the extant project management literature but the related construct (effectiveness of risk management) did not provide strong explanatory value.

In sum, this analysis showed, that similar conclusions can be made to explain outliers: project design and project execution matter. The occurrence of outliers is linked to elements of *project design*. Theoretically these aspects are rooted in systems-centric theories of projects. In prior studies those are often conceptualized as dimensions of complexity or the manageability of complexity, e.g., Cooke-Davis et al. (2007), Jolivet and Navarre (1996), Alderman and Ivory (2007), Bruijn & Leijten (2008). The IT literature equated complexity of the project then with the difficulty of executing it, e.g., Dumke and Abran (2011), Baccarini (1996). However, the system-centric theories of high reliability organizations and normal accident theories clearly point out

that these are design choices made by decision makers. In other words, the findings of this analysis demonstrated that project design (in particular size, duration, and the uniqueness of the project) explained outliers above and beyond the execution. Project design is imposing more than just constraints to the project that need to be required to be matched by effective tactics in project execution.

On the other hand, the occurrence of outliers is linked to aspects of the *project execution*. In particular, the two aspects of project processes (through the variable effectiveness of monitoring and controlling) and human aspects (through the variable qualification and motivation of the project team) provided significant explanations above and beyond variables linked to project design. These variables were suggested by success and risk factor analyses of projects (e.g., Fortune and White 2002, Belassi and Tukel 1996, Pinto and Slevin 1987, Cooke-Davies 2002). Some of the findings, in particular that better monitored and controlled projects increase the odds of cost outliers occurring, might be indicative of escalation of commitment behavior (Staw and Ross 1986, Sleesman et al. 2012, Keil et al. 2000). Even though the analysis was limited to a cross-sectional and post-mortem analysis that is limited in testing the longitudinal explanations suggested by the theories of escalation.

6.2.1 Implications for System-Centric Theories

The findings of this thesis imply that *estimated cost and estimated duration* are a strong explanation of the occurrence of outliers. The data showed that long

projects are more likely to become outliers. Projects with larger budgets are less likely to become cost outliers, however the effect is dominated by the stronger effect of project duration.

Most theories of so-called project complexity predicted that bigger projects suffer from increased interactive complexity (e.g., Simon 1962) in IT projects the data show that this is not the case. This study questions the findings of e.g., Jones (2000), Sauer et al. (2007), McFarlan (1981), Zmud (1980), Jenkins et al. (1984), Wallace et al. (2004), Brooks (1975). Instead this study showed that scaling in project in size is not linked to outliers, scaling in project in time is. In practical terms, concurrent software engineering might be less problematic than managing long time horizons in IT projects.

The findings thus imply a more complicated relationship of size and performance. The importance of time supports ideas of volatility suggested by Sauer et al. (2007), Kulk (2009), which argues that risk of IT projects is linked to time through the concept of volatility, that is changes to requirements, scope and personnel turn-over. In the similar vein, Miller and Lessard (2000) argued that longer projects increase their exposure to turbulences, which disrupt the project. However, the analysis in this thesis found that larger and longer projects faced less scope volatility⁷⁵.

⁷⁵ Requirements volatility was positively correlated with estimated cost ($r=0.14$) and estimated duration ($r=0.05$) but the correlations were not significant.

Moreover, the estimated cost and estimated duration were also a proxy for other variables (cf. the correlation matrix in appendix 8.8). Larger and longer projects faced more organizational resistance, were more innovative, had more unknowns in the design, tighter coupling of deliverables, had higher processing, security, and performance requirements, had less effective communication in the project team, and had a less sustainable mix of internal and external resources. Larger and longer projects also had more influence over the project environment, less scope volatility, more experienced project managers, and better staffing. In addition, longer projects had more qualified and motivated teams. Some of these factors – in particular the higher degree of innovation, higher processing, security, and performance requirements – can also be interpreted in the opposite direction of cause and effect, i.e., these factors might explain why the projects had higher estimated cost and longer schedules to begin with. Nevertheless, these findings demonstrate not only the challenges of scaling up IT projects, but also imply that the concept and role of time in project requires further research beyond the idea of turbulences and volatility of scope and requirements.

The third design factor, explaining the occurrence of outliers, was the *perceived uniqueness of the project*. The finding supports the argument that social factors influence project outcomes made for example by McFarlan 1981, Zmud 1980, Whitty 2009, de Bruijn and Leijten 2008, Addison and Vallabh 2002, Glass 1997, Yourdon 1996, Boehm 1991. Nonetheless, the importance of the concept of uniqueness is a novel finding. The higher the perceived uniqueness the higher

the probability of a project being a cost outlier. The extant literature often makes the assertion that projects are unique undertakings (e.g., Baccarini 2004, Chapman 1998, Conroy and Soltan, 1998; Czuchry and Yasin, 2003, Gareis 1991, Jaafari 2007, Perminova et al. 2008, PMI 2004). Others have argued that projects are not necessarily unique (e.g., Ekstedt et al. 1993, Engwall 2003, Lundin and Soderholm 1998, Hellstrom and Wikstrom 2005). Few studies found that uniqueness contributes to the risk of IT projects (e.g., by Hellstrom and Wikstrom 2005, Artto 2001)⁷⁶. Based on these findings Williams and Samset (2010) have demonstrated that managers can undertake activities, such as searching for similar projects outside the organization, to reduce the uniqueness of a project. The findings of this analysis underline the importance of these activities.

In the extant literature, perceived uniqueness has been found to inhibit learning (Newell et al. 2006) and is linked to the lack of experience of key project team members (Engwall 2003). The perceived uniqueness of projects in this sample was also a proxy variable for other explanatory variables (cf. correlation matrix Appendix 8.8). Projects being perceived as more unique had more political intent behind plans, and less realistic and reliable schedules, they faced higher requirements volatility, had more unknowns in the design, and higher processing requirements. More unique projects were also less effective in their

⁷⁶ Note: the data in this analysis do not show a distinction of the type big and unique vs. small and repetitive projects. The correlation between size and uniqueness $r = 0.01$ $p = 0.92$ and correlation between duration and uniqueness $r = -0.04$, $p = 0.62$ are both not significant.

monitoring and controlling, had a less sustainable mix of internal and external resources, and worse performing suppliers. Overall, more unique projects were found to be more likely to be non-viable after the fact, as expressed through the statement “If the true cost were known the project would not have been built.”

The findings support the argument that uniqueness of IT projects is not an exogenous project constraint but is a design parameter that can and ought to be influenced by the project. The concept demands further academic and managerial attention.

In sum, the system-centric organizational literature argued that highly complex systems inevitably fail. The analysis suggests that, indeed, increased complexity in the sense of technical and social complexity, as suggested by the project management and IT project literatures, contribute to the occurrence of outliers in IT projects.

6.2.2 *Implications for Process-Centric and Event-Centric Theories*

The findings demonstrated that the two variables effectiveness of monitoring and controlling and the qualification and the motivation of the project team contributed to the occurrence of outliers. The findings support the argument that process and human aspects of project execution explain the occurrence of outliers.

The findings showed that *effectiveness of monitoring and controlling* the project explained the occurrence of outliers. This finding supports arguments made by the literature on success and risk factors (e.g., Fortune and White 2002, Belassi

and Tukul 1996, Pinto and Slevin 1987, Cooke-Davies 2002). The effectiveness of monitoring and controlling was also related to other variables (cf. Appendix 8.8 for the full correlation matrix). More effective controlling and monitoring was associated with more robust business cases, more environmental influence of the project, more user involvement, more reliable and robust schedules, less perceived uniqueness, less requirements volatility, better suitability of the project management method chosen, more effective risk management, more detailed and regularly updated project plans, more effective change management, more competent project managers, better motivated and more qualified project teams, more suitable staffing, better governance, more support from the project sponsor, better aligned stakeholders, and better communication of the true status of the project. On the other hand, better monitoring and control was also associated with risk factors: more innovation, and higher performance and security requirements. This demonstrates that the variable is a proxy for other success and risk factors discussed in the literature. Thus the results support previous findings made by the project management and IT project management literatures. However, the findings showed that more effective monitoring and control was associated with a higher probability of cost outliers occurring in the sample. A possible explanation can be made from how the correlations between the explanatory variables differ, if only projects that were cost outliers are analyzed. In this subset more effective monitoring and control is linked to tighter coupled deliverables ($r=0.87$, $p < 0.01$), higher security requirements ($r=0.72$, $p < 0.05$). Factors which are

described in the extant literature as increasing the risk of the project. However, the effectiveness of monitoring and control was also associated with more effective risk management ($r=0.85$, $p<0.01$), more detailed and up-to-date plans ($r=0.88$, $p < 0.01$), more suitable project management methods ($r=0.95$, $p<0.001$), more effective change management ($r=0.97$, $p<0.001$), more competent project managers ($r=0.91$, $p<0.001$), more effective governance ($r=0.68$, $p<0.05$), more effective project sponsor ($r=0.89$, $p<0.01$), and more transparent status communication ($r=0.97$, $p<0.001$). Taking into account that correlations do not imply a direction of the effect and that cost overruns do not imply project success or failure, these findings can be interpreted in the light of the literature on project escalation. The positive association between effectiveness of monitoring and control and support by senior management even in cost outliers might suggest escalated commitment to the projects (Staw 1981). A more effective project organization, signified by the proxy effectiveness of monitoring and control, might have been the result of sustained investment into the escalating project. Research into the determinants of escalation (cf. the meta-analysis in Sleesman et al. 2012) has focused on repeated decisions to continue to support of dysfunctional projects. Further longitudinal research is needed to unpack the relationship between design and process factors to fully explain these findings for example by tracing the changes of the project organization as characterized through its design and process variables over time of the escalation. In other words, the findings in this analysis highlight that further

research is needed to broaden theories of escalation beyond the focus of decision-making and integrating alternative theories.

Additionally, the effectiveness of risk management held some, albeit weak, explanatory value. The literature review located the concept in the event-centric view, which has a long tradition in organizational research. The core concept of event-centric theories is that an external event disrupts an organization, which in turn needs to respond to the disruption. Project management theory, however, has argued that foreseeing and dealing with turbulences is the core of risk management, an integral part of project management (Miller and Lessard 2002, Pich et al. 2002). The role of external events has found little reflection in project management theory. The findings of this thesis partially support this view. The analyses linked the effectiveness of risk management to cost overruns and very weakly to the occurrence of effort outliers. Overall the findings raise the question why external events and the internalization of external events through the process of risk management explained cost overruns but then failed to explain cost and schedule outliers. Further research is needed to be able to integrate organizational theories in this field into project management theories.

In sum, the findings offer support for process-centric and system-centric explanation of outliers. The findings demonstrated that neither school of thought fully explains in it of itself outliers in IT projects. Outliers were the result of an interplay of design and execution variables. More complex projects do not inevitably fail, at the same time that effective execution of success factors

does not prevent outliers from happening. The findings showed that estimated size, and even more so estimated duration and uniqueness explained why outliers occurred in the sample. The findings also showed that effectiveness of monitoring and controlling the project, and the qualification and motivation of team explained outliers. Each variable was associated with other factors. The findings also showed that the dynamics of escalation need to be integrated into these theoretical explanations. Moreover, the findings highlighted that more research is needed to not only integrate these two theoretical schools of thought but to also strengthen event-centric explanations, which, in the literature, is limited to discussions of risk management.

6.3 Practical Implications: Principles of Project Design and Project Execution

What, if anything, can be done about outliers in IT projects? The findings outline the challenges organizations face to prevent outliers and thus reduce the risk associated with investments in IT projects.

First, the single biggest design challenge to prevent outliers is to keep projects short. The analysis showed that projects do not have to be small but short to reduce the occurrence of outliers. At the same time the analysis has indicated that artificially crushing project schedules might lead to more schedule outliers.

Second, the analysis showed that the perceived uniqueness of projects explained why outliers occurred in the sample. If a project is seen as being unique, limited learning takes place from other projects, project plans tend to be developed bottom-up without a reality check through comparing itself and its plans to other projects.

A large body of prior practical and academic literature discusses the themes of size and uniqueness under the notion of complexity. The term complexity often adds confusion rather than being helpful. The findings of this analysis, however, imply where complexity could be best reduced. Casting the findings in this light, project complexity can be reduced in order to prevent outliers by reducing project durations, social and technical complexity, which were found to be linked to the perceived uniqueness of projects.

Third, the findings hold practical implications for the execution of IT projects.

One challenge points at project processes. The findings implied that monitor and control processes are an important challenge, while at the same time being a proxy for other project processes, such as change management, continuous planning, stakeholder management, and decision-making and communication processes.

Forth, the findings identified that the qualification and motivation of the project team is a key challenge to prevent outliers. The qualification and motivation of the project team in the projects in this sample was also linked to the project manager (experience and competency), more effective project processes (monitoring and controlling, change management, decision making, status communication), and a more effective project organization (mix internal/external resources, sponsor, senior management, stakeholders, and suppliers).

Finally and most importantly, the findings have implications for the practical forecasting and estimation of project risk. They showed that most practically used methods to model risk, e.g., through Monte-Carlo simulations or earned value management, underestimate the thickness of tails and thus under represent the risk of IT projects. The findings showed that the error in estimates is very large and even infinite in some forecasting dimensions. In similar vein, practical tools used to support decision making, e.g., sensitivity analyses and

real options valuation, currently lack the ability to address the issue of fat tails that is the unboundedness of downside risk and forecasting errors.

In sum, the findings implied that attempts at understanding and characterizing the performance of IT projects need to incorporate outliers. Benchmarks, models, and forecasts that exclude or ignore the thickness of tails over-represent the performance of IT projects and under-represent the risk of IT projects. Nevertheless, there is something that can be done about outliers. On the design side, the challenges are the sizing of projects as well as designing the social and technical elements of projects. On the execution side, the challenges are sound process fundamentals and staffing the project team.

6.4 Summary

In sum, the findings showed that very large overruns happen infrequently but are neither unlikely nor completely random. Moreover explaining the occurrence of outliers with concepts from management literature resulted in first explanations, albeit some only with weak statistical support.

First, the findings implied for *research methodology* that outliers need to be incorporated into analysis instead of being rejected. Second, concerning analysis of why outliers occurred the findings implied that rigor of analyses needs to be increased. On the one hand organizational differences need to be methodologically addressed. On the other hand, studies with large number of independent statistical tests need to take into account the false discovery rate. Additionally, many of the explanatory variables in studies are confounded and high levels of multicollinearity exist – a problem that several but not all studies have acknowledged in the past. Lastly, the power law nature of project performance suggest that mean and variance converge only slowly, if at all, which questions the use of small n studies to characterize IT project performance.

Second, the findings implied that explanations of why outliers occurred need to incorporate multiple theoretical dimensions and the dynamical interplay between those explanations. The extant literature offers rich explanations of system-centric theories (such as technical and social complexity of projects), event-centric theories (such as the management of turbulences), and process-

centric theories (such as effectiveness of managing success factors, or preventing escalation of decision-making). The findings showed that a combination of theories is best equipped to characterize and understand the performance of IT projects, but also that further research is needed.

7 Summary

The objective of this thesis has been to develop an understanding of business-embedded IT projects with markedly different performance, so called outliers. To do so, the thesis has drawn on the existing literature, provided new empirical evidence, and made methodological improvements to answer the questions: How frequently are outliers in IT project performance? Why did outliers in IT project performance occur?

7.1 How Frequent are Outliers?

Chapter 2 first reviewed empirical studies of IT project performance. The literature review documented a wide range of reported project performances in the three dimensions of cost, schedule, and benefits. Performance in these studies was expressed as the comparison of actual outturn cost, go-live dates, actual effort, and realized benefits with initial plans. The studies documented a wide variation not only within but also between studies.

One potential explanation of the large amount of variation in reported performance has been deliberate bias in forecasting (Kulk 2009). This thesis suggested outliers in the data as another explanation of variation. Only few studies have reported on the frequency of outliers. Outliers occurred between 33% (cost performance in Kulk 2009), 11.5% (actual duration in Yoon et al. 2007), 8% (actual and estimated cost in Verhoef 2005), 5% (software development productivity in Banker and Kauffman 1991; cost per function point in Verhoef 2005), 2% (schedule performance in Mitchell and Zmud 1999),

1.5% (actual duration in Verhoef 2005), to 0.2% (monetary value of schedule variation in Grant et al. 2006). The studies evidenced that outliers, even if unreported, are a recurring phenomenon.

Nearly all studies chose to deal with outliers through exclusion. This is the dominant strategy in research, with the exception of some scheduling studies, which used the lognormal distribution and other fat tailed distributions to incorporate outliers in models.

The empirical study that is presented in this thesis was thus formed with the objective of studying and understanding how IT projects perform, in particular, to advance the understanding of outliers and management factors contributing to outliers, i.e., markedly deviating observations of performance in IT projects. A quantitative survey based on decision documents was selected as the primary approach to avoid biases from self-reported survey tools that most prior studies in this field used. Wherever possible the data collection and analysis erred on the conservative side.

When studying rare events the trade-off in research designs is to balance the number of observations with the quality of the variables. Data were collected on 4,307 projects from 190 organizations. The data came from multiple sources. Project data were directly collected on 1,365 projects from 34 participating organizations. A desk study collected data on 289 projects from published academic studies and auditor reports. Finally data on 2,653 IT projects from the federal US government were collected. 1,029 through FOIA data for budget

requests from each of the 134 agencies and 1,624 from submissions to the Office of Management and Budget's IT project dashboard. For a subset of 220 projects a post-mortem was conducted.

Following long standing convention, the performance of the projects was calculated as actual/estimated for the four dimensions of cost, schedule, effort, and benefits. The data showed that the average cost overrun was +107% (median $\pm 0\%$), the average schedule overrun +39% (median $\pm 0\%$), effort overrun +118% (median +25%), and benefits shortfall was -24% (median $\pm 0\%$). The figures are consistent with the documented large amount of variation reported in the extant literature.

Outliers were defined following Tukey (1975) as being more than 1.5 interquartile ranges away from the box in the boxplot. 14% of the observations were an outlier of cost performance, 15% a schedule performance outlier, 10% an effort performance outlier, and 6% a benefits performance outlier. The analysis was able to overwhelmingly reject the first hypothesis, that outliers did not occur more frequent than under the assumption of thin tails (Poisson tests with $p < 0.004$).

The analysis was also able to overwhelmingly reject the second hypothesis that cost, schedule, effort, and benefits performance are normally distributed (Anderson-Darling test with $p < 0.001$).

Taken together the first and the second hypothesis provided evidence that outliers occur frequently and systematically to the extent that the commonly made assumption of normality is strongly refuted.

The third research hypothesis aimed to replace the assumption of normality with a new assumption about the general distribution in the data adequately reflecting the uncertainty of IT project performance. The extant literature on empirical data with large variation suggest that IT project performance might be best explained with extreme value distributions. Power laws (Pareto distribution) indicated a suitable fit. Power laws have been commonly reported, and misreported, in social sciences. For example the distribution of city sizes, links to a webpage, citation, scientific output, and the distribution of wealth follow power laws in their tails (Mitzenmacher 2004, Clauset et al. 2009). The analysis followed the protocol outlined by Clauset et al. (2009), Stouffer et al. (2005). It established that a power law, of the form $p(x|\alpha) = \frac{\alpha-1}{x_0} \left(\frac{x}{x_0}\right)^{-\alpha}$ for $x > x_0$, can be fit to the upper tail of the distribution and that this power law has the exponent $\alpha = 2.5$ for cost performance, 3.1 for schedule performance, and 2.7 for effort performance. Bootstrapping 5,000 simulations estimates the exponent alpha is on average 2.4 with a standard deviation of 0.2 for cost performance, alpha = 3.2 with s.d. = 0.1 for schedule performance, and alpha = 2.7 with s.d. = 0.5 for effort performance. In all three cases the hypothesis that the tail of the distribution can be fit by a power law could not be rejected ($p \geq 0.998$). In Mandelbrotian terms IT project performance is fat tailed and in some

dimensions (e.g., cost and effort) heavy tailed. As a result commonly reported measures to characterize IT project performance such as mean and variance converge very slowly if they exist at all.

Lastly, the analysis tested three commonly suggested alternative explanations that would indicate that the observed fat tails are spurious. The analysis was not able to establish that the fat tails emerge, either because of a mixture of lognormal distributions, of mixtures with similar means and different variances, nor because of aggregation across a non-stationary parameter.

Given these findings, the original characterization of the typical performance of IT projects and the resultant risk needs to be reformulated. The findings have direct implications for the understanding of the risk of IT project, methods of IT studying IT projects, and IT project forecasting and modeling of risks.

First, the results evidenced that outliers are not unique, idiosyncratic events. Large overruns are infrequent but not impossible. Observations of markedly deviant performance ought to be expected.

Second, outliers are not from a different universe of randomness than smaller overrun - thus they ought not to be excluded from analyses. On the positive side, outliers are not completely random either, they follow patterns of power laws. Power laws offer a pathway to study markedly deviant observations even with smaller sample sizes. The scale invariance and universality properties of power laws suggest that the same generating mechanisms cause the observed

performance along the full spectrum of observed outcomes. In other words, large overruns smoothly extrapolate from smaller overruns.

Third, the upper tails of the performance distributions are best described through power laws. Power law tails are (infinitely) top heavy. This implies that the largest observation strongly influences the center of the distribution. The performance of IT project has a large variance (e.g., for schedule performance and the nominal delay), in some dimensions the variance is infinite (e.g., cost performance) and in some dimensions (e.g., nominal cost overruns in USD terms, actual outturn cost) also the mean is infinite. This has direct implications to the modeling of risk, which in most state-of-the-art models is severely underestimated.

In sum, outliers occur frequently and regularly. In the conceptualization of Mandelbrot (1997) the field of IT projects forecasting is a field of non-Gaussian randomness, where outliers matter more than common observations. Some dimensions of IT project performance show, in Mandelbrot's terms, slow randomness, some show wild randomness (infinite variance), and some even extreme randomness (infinite means).

7.2 Why do outliers occur?

To inform the understanding of outliers, insights gleaned from a variety of theoretical explanations. The thesis grouped the extant explanations in three schools of thought: system-centric, event-centric, and process-centric theories of outliers. The literatures offered rich theorizing within schools of thought but rarely across. To address the question of why outliers occurred the thesis reviewed the organizational theories, project management theories, and IT project management theories in each of the schools of thought.

The *system-centric school of thought*'s key theoretical explanations of outliers are Normal Accident Theory (Perrow 1981) and the theory of high reliability organizations (Roberts 1989, Weick and Sutcliffe 2007, Weick 2006). In short, both theories have the premise that highly complex organizations, i.e., those characterized by high interactive complexity and tight coupling, have a higher propensity to fail. Failure can be prevented through technical designs, influencing organizational and human relationships, and attention to weak signals. In project management theories these ideas have been translated into the concepts of social and technical complexity (McFarlan 1981, Zmud 1980, Whitty 2009, de Bruijn and Leijten 2008). In the IT project management literature these concepts have been rooted in forecasting theory. The notion of complexity has been operationalized through the proxies of size and duration (e.g. Jones 2000, Sauer et al. 2007, McFarlan 1981, Zmud 1980, Jenkins et al. 1984, Wallace et al. 2004, Brooks 1975). Other studies have operationalized

complexity through assessments of difficulty or manageability of social factors (e.g., Addison and Vallabh 2002, Glass 1997, Youdon 1996, Boehm 1991) and technical factors (e.g. Phalp and Cox 2000, Heltewig 2008, Collopy 2007, Berger and St. Arnaud 1992, Heemstra 1992).

The *event-centric school of thought*'s key theoretical explanations of outliers are theories of rare events and the management of catastrophes. The notion of rare events has been linked to interruptions of organizational routines (Christianson et al. 2008), strategic surprises (Ansoff 1975) and crises (Roux-Douffort 2007). In sum, these theories explain outliers as the effect of an insufficient organizational response to a rare event. Those insufficient organizational responses are the result of failure to make sense of a rare event, framing a rare event in a way that inhibits the needed response, and failure to learn from and through rare events (Baumard and Starbuck 2005, Weick 2004, Rerup 2009, Christianson et al. 2008). In project management literature the concept of rare events is translated into the notion of turbulences, which is the core of project risk management (Miller and Lessard 2000, Pich et al. 2002). This line of inquiry has failed to attract much research in project and IT project management fields.

The *process-centric school of thought*'s key theoretical explanations of why outliers occur are the theories of man-made disasters (Turner and Pidgeon 1997), normalization of deviance (Vaughan 1996), and the escalation of commitment (Staw and Ross 1986, Sleesman et al. 2012). In project management literature the escalation of commitment is the dominant theoretical lens, which has been

particularly studied as project escalation (Keil et al. 2000). In IT projects the process-centric school of thought has often been atheoretical and focused on identifying success factors of project execution and delivery (Pinto and Slevin 1987, Cooke-Davies 2002, Fortune and White 2002, Bleassi and Tukel 1996). Further theoretical explanations of why IT projects have become outliers has been the study of project abandonment which is rooted in the support theory, i.e., projects were abandoned because they lost economic support (Oz and Sosik 2000), political support (Sauer 1993), or technological support (Ewusi-Mensah and Przasnyski 1991).

The empirical findings led to two conclusions based on (1) existence of power law tails, (2) tests to explain the thickness of tails and odds of a project being an outlier.

First, finding that the tails of the performance distribution of IT projects can be fit to a power law does not directly rule out any of the three theoretical propositions. System-centric theories would suggest models such as self-organized criticality and over optimization that have been shown to produce performance outcomes that fit power laws. Event-centric theories would suggest models such clustered volatility from herding and copycat effects or random exogenous news processes to generate performance outcomes that fit power law tails. Process-centric theories would suggest models such as random

proportional growth⁷⁷, models of multiplicative error with random noise, or rational bubble models to generate a performance outcome that fits a power law.

Second, the analysis combined two statistical analyses. It directly compared the thickness of the tails of the distribution between groups based on the asymptotic efficiency of maximum likelihood estimators and Vuong's test for model selection. This analysis is novel and has never been attempted before. It has the advantage of testing the full range of the tail making use of the scale-invariance and universality properties of power laws. However, the estimators suffer from small sample bias and requires pairwise comparisons of potentially correlated factors. To overcome the disadvantages the tail comparison was supplemented by multilevel rare events logistic regression. The more established method offers the advantage of having more developed methods to deal with hierarchically structured data, multicollinearity, and the ability to approximate rare events bias of the estimators. However, logistic regression requires an arbitrary classification of outcomes into outliers and common observations. The analysis used Tukey's (1977) rule of thumb to identify outliers, namely that outliers are observations more than 1.5 interquartile ranges away from the box in a boxplot. This thesis combined both analyses to suggest a first step towards new ways of studying markedly deviant observations.

⁷⁷ Also known as Zipf's law, Gibrat's law, Yule processes, Mathew's effect, and preferential attachment.

The two analyses reached similar conclusions across the dimensions of cost and schedule performance. The sample size proved insufficient to build explanatory models for benefits and effort performance.

Five variables statistically significantly varied with the occurrence of cost or schedule outliers: estimated cost, estimated duration, perceived uniqueness of the project, effectiveness of the monitoring and controlling of the project, and motivation and qualification of the project team. Figure 45 shows how the thickness of the performance tails varies with each of the variables. Figure 46 shows how the predicted average risk of a project being an outlier varies with each of the variables.

Figure 45 Comparison of Thickness of Power Law Tails for 5 Selected Variables

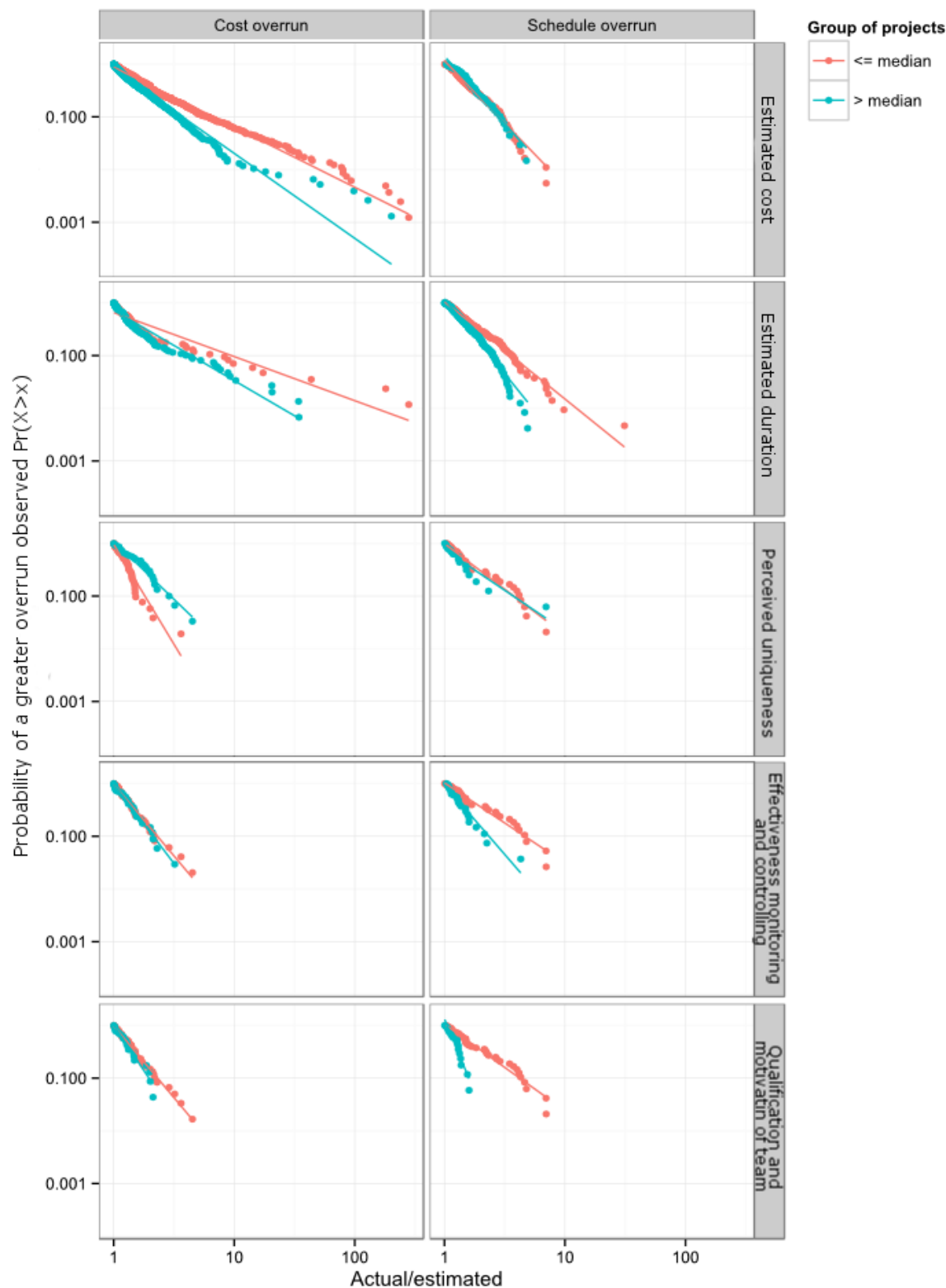
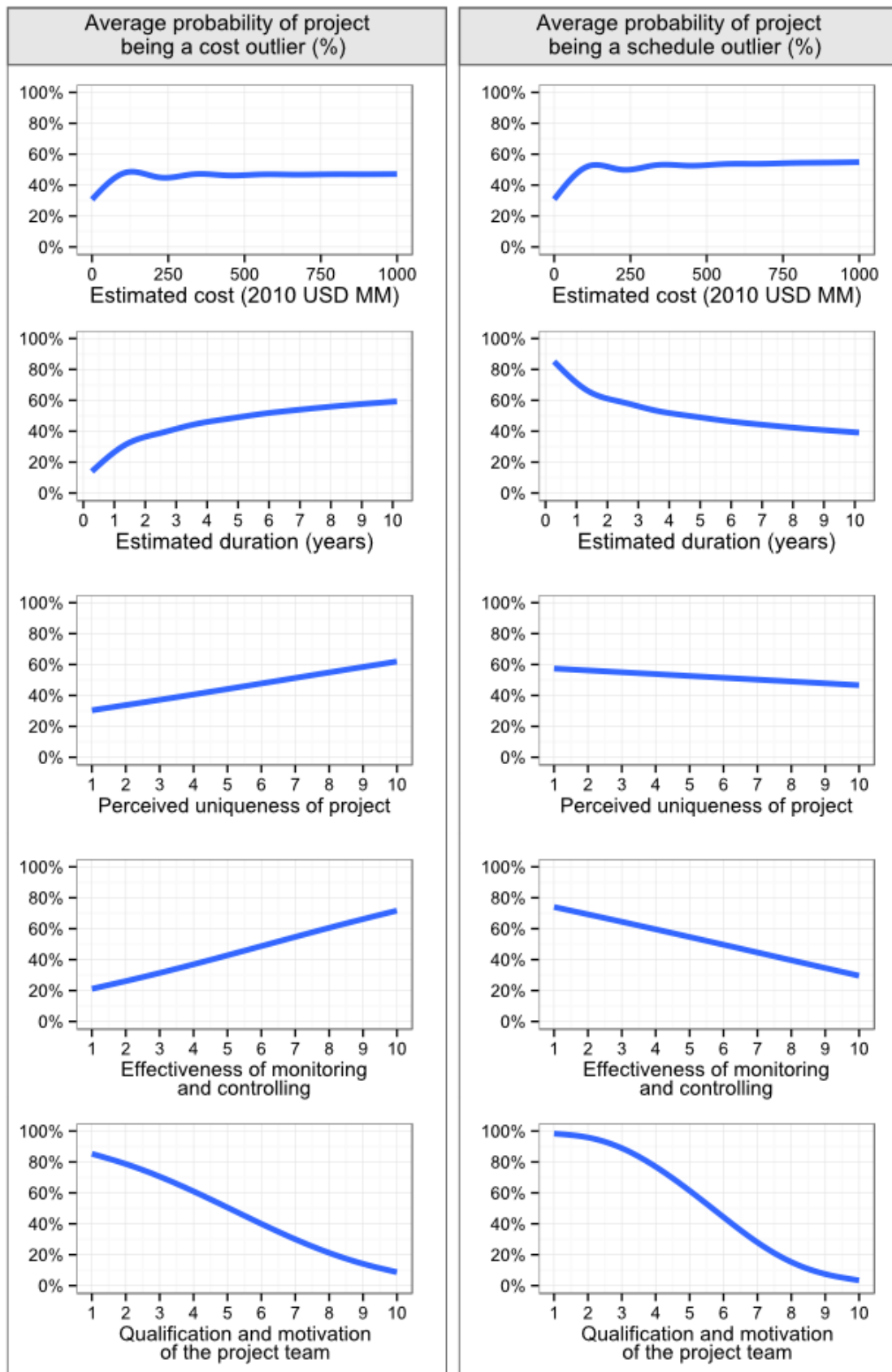


Figure 46 Main Effects of Logistic Regression to Explain Average Probability of Outliers in IT Projects



First, the *estimated project size*, in this sample measured in 2010 USD, contributed a small effect. Smaller projects have a slightly thicker tail of cost performance (fig. 45) but when controlling for all other factors the effect was mostly stable (fig. 46) but led to an increased average probability of a project being an outlier when scaling small up to 100 million.

Second, *estimated duration* showed that for both cost and schedule performance the tails of shorter projects were slightly thicker (fig. 45). Once all other factors were accounted for the duration of the project has an effect above and beyond other explanatory variables. The average probability of a project being a cost outlier increased and the probability of a schedule outlier decreased, most likely because of a more aggressive and therefore less realistic schedules when timeframes are constrained to be short (fig. 46). When modelled on their own, the risk reducing effect of increased estimated cost are moderated by an interaction effect showing that the risk increasing effect of estimated duration dominates the odds of outliers occurring.

Third, increased *perceived uniqueness of the project* led to thicker tails for cost and schedule performance (fig. 45). Similarly, as uniqueness increased the average probability of a project being a cost outlier increased above and beyond all other variables. The probability of the project being a schedule outlier slightly decreased, however this trend was not significant (fig. 46).

Fourth, projects with less *effective project monitoring and controlling* showed thicker tails for schedule performance and slightly thicker tails for cost

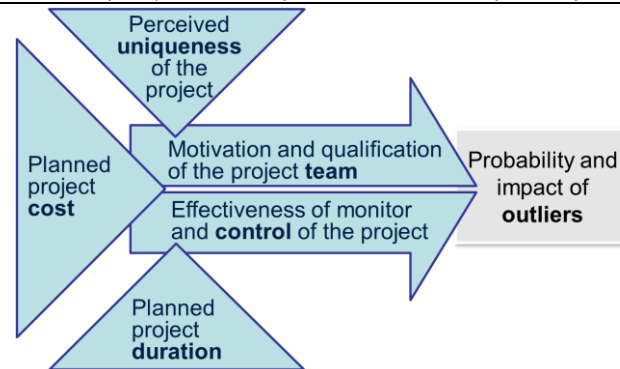
performance (fig. 45). Controlling all other variables the effect becomes more pronounced in the regression model. The less effective the control was the higher the average probability of a project being a schedule outlier. However, the trend was the opposite for cost outlier (fig. 46). Here the projects with the more effective control had a higher probability of being a cost outlier. This finding can be interpreted in the light of the literature on escalation of commitments to a failing course of action, i.e., the well documented escalation of projects suggests that once projects incur overruns decision makers continue to support the project even increasing their efforts.

Fifth, the *motivation and qualification of the project team* showed consistent effects in both dimensions and analyses. Projects with a less motivated and qualified team showed a thicker tail of both cost and schedule performance (fig. 45). Similarly the predicted probability of the project being either a cost or schedule outlier decreased with increasing motivation and qualification of the team (fig. 46).

Given these findings, the original framework of the three premises of system-centric, event-centric, and process-centric explanations should be reframed. Summarizing the empirical evidence a new framework (fig. 47) emerges. The five factors associated with the project being either a cost or schedule outlier consist of two groups. On the one hand, the factors of estimated cost, estimated duration, and perceived uniqueness of the project. On the other hand the

factors of qualification and motivation of the project team and effectiveness of monitor and control.

Figure 47 Restated Framework of Explanations of Outliers in IT Project Performance



The factors estimated cost and duration, and uniqueness of the team were adapted from theoretical explanations of the system-centric school of thought. Here, they are often characterized as organizational design parameters.

The factors qualification and motivation of the project team and effectiveness of monitoring and controlling the project were borrowed from theoretical explanations of the process-centric school of thought. These are often portrayed as process success and risk factors during project delivery, i.e., factors that are unknowable a priori and can only be judged a posteriori.

By furthering this understanding, this research showed how factors of organizational design of the IT project, as well as, factors of the delivery of the IT project contributed to projects being outliers. This thesis has thus made a strong argument for more cross-theoretical analysis and research.

7.3 Ideas for future research

The concepts and findings presented in this thesis are intended to provide motivations for more cross-theoretical research and practical ideas for researching the role, importance, and causes of outliers in IT project performance. A conceptual framework has been presented, with initial evidence from a large-scale survey showing that the framework shows potential merit.

In closing, it is clear that one of the avenues for future research would be to strengthen the research into the role of outliers in projects.

The thesis has shown that multi-theoretical explanations explain IT project performance. More research is needed to integrate the dominant perspectives of design-focused systems-centric theories, turbulence-focused event-centric theories, and execution-focused process-centric theories. The discussion of escalation of commitment dominates the field. The thesis has shown that other theoretical explanations can also contribute to understanding outliers.

However, the thesis was, due to research design constraints, not in a position to fully explore the dynamic interplay of those theories along the time it takes for projects to be executed. The particular focus of project dynamics is underrepresented in both theory as well as empirical evidence as it relates to project management. In short, the concept of time demands further academic attention and research.

The analysis identified the perceived uniqueness of IT projects as a factor contributing to the occurrence of outliers. Uniqueness is a contested topic. Some authors have argued that projects are per definition unique undertakings, while others have demonstrated detrimental effects through inhibition of individual and organizational learning and that uniqueness indicates lack of experience of project managers. Moreover, case studies have documented how project management can take active steps at reducing the uniqueness of projects. The results of this analysis underline the need for more research in this area.

Moreover, the thesis has shown how outliers are a large influence on the center of the distribution of IT project performance. This research should thus encourage studies in other areas of organizational performance. The thesis showed that due to the specific characteristics of the performance distribution large n samples are needed for research. The failure to reject some explanatory variables in this analysis might be linked to the limited statistical power of smaller subsamples and the conservative approach this thesis took to data collection and analyses. More empirical research in this and other areas is encouraged. Moreover, retesting existing data sets and creating more large n datasets in this and other areas would allow to validate the approach and methodology of comparing the thickness of tails as well as using rare events logistic regression.

Additionally, the implication of the findings for risk management, project finance, forecasting, and financial modeling need to be further developed. The

problem of IT projects having infinite variance and in some aspects infinite means requires a rethinking of not only commonly used analytical tools such as NPV calculations but also more advanced models of projects such as real option valuation.

Moreover, the findings implied that decision support tools used on IT projects and project portfolios require further development. Tools such as earned-value management, case-based and predictive forecasting, and simulation models of IT projects require research to be able to incorporate and adequately reflect the large downside risk of IT projects that this thesis documented.

With respect to the specific findings of the study, initial support has been provided for the antecedents of estimated cost, estimated duration, perceived uniqueness of the project, effectiveness of monitoring and controlling, and qualification and motivation of the project team in explaining outliers of IT project performance. This could be studied in a limitless number of contexts outside IT projects, from megaprojects in other areas to permanent organizations. The explanatory framework ought to be verified through additional, replicative studies conceived by others with a view to studying the field. Moreover, parts of the extant literature suggest concepts that were not included in this study, such as risk culture, organizational mindfulness, sense-making, and learning. Future research integrate these concepts from qualitative case studies in quantitative studies.

8 Appendices

8.1 Event-Centric Explanation of Why Outliers Occur

This section reviews the explanations suggested by theories of the event-centric school of thought. Four perspectives are reviewed: (1) cognitive perspective, (2) sense making perspective, (3) framing perspective, and (4) organizational learning perspective.

8.1.1 *Cognitive Perspective*

The cognitive tradition explained that events trigger fundamental causal attributions - specifically through the notions of 'victims', 'control', and 'impact' (Turner and Pidgeon 1997). The notions of 'victims'⁷⁸, 'loss of control' or 'out-of-control' all externally attributed the cause of the event. External meant that the cause of the markedly deviant observation was outside the control of the individual or the organization (Turner and Pidgeon 1997). Attribution also guided how organizations reacted to events. For instance Coombs and Holladay (1996) showed that the organizational response differed between (1) unintentional and internal causes, e.g., accidents, (2) intentional and internal causes, e.g., transgressions, (3), unintentional and external causes, e.g., faux pas, and (4) intentional and external causes, e.g., terrorism (Coombs and Holladay

⁷⁸ The notion of the victims is problematic in conventional studies of disasters. Turner & Pidgeon (1997) found that most accounts of disasters were first-hand accounts by victims. They argued that this posed an important challenge in terms of understanding and investigating the antecedents and consequences of disasters. Victims' memories and journalistic accounts of disasters were highly emotional, and relied on memory and perceptions made under great stress (Turner and Pidgeon 1997). Moreover, the analysis of man-made disasters demonstrated that the antecedents of disaster were found in the socio-technical system, with an emphasis on the social side of the system. Contrarily accounts by victims and journalists tended to point to technical causes or simply human error of an individual operator.

1996). This showed that errors in attribution are likely to influence choices of organizational responses to rare events.

Attribution theory (Weiner 2008) also explains spontaneous causal thinking of managers in organizations after rare events. Big failures required big causes. The locus of large failures tended to be perceived as external, controllability was perceived low, and the underlying theory of cause and effect was perceived as being unstable. Unstable cause and effect of failures was attributed by managers to exceptional circumstances, historic conditions, or societal changes (Baumard and Starbuck 2005). Baumard and Starbuck further described that managers attributed failures to lack of commitment to the venture, weak control, and high turn-over which resulted in a lack of ownership and dispersed stakeholders. Failures had, according to the managers, unique, idiosyncratic, unforeseeable, and external causes (Baumard and Starbuck 2005). As Baumard and Starbuck observe: "such explanations [idiosyncratic external events] surface after every failure, but they never appear after successes." (Baumard and Starbuck 2005, p. 295-296). The observations of spontaneous causal cognition by managers after rare events questioned whether learning from failure is at all possible. Regardless, the cognitive perspective demonstrated the importance of organizational and individual cognition and how markedly deviant performance outcomes can result from attribution errors, erroneous causal thinking, and imperfectly chosen responses to events.

8.1.2 *Sense Making Perspective*

Sense making stood in contrast to the cognitive causal analysis. "Sense making involves the ongoing retrospective development of plausible images that rationalize what people are doing." (Weick et al. 2005, page 409). Sense making has been defined as a retrospective process that combines action and interpretation. Thus sense making stood in contrast to the notion of decision making, which focused attention to influencing and evaluating choice (Weick et al. 2005, page 409). Additionally sense making has not solely been seen as a cognitive process. Sense making also had strong behavioral aspects due to the immediate need for action triggered by the event. Sense making combined the retrospective hindsight of 'what did just happen?' with the uncertain foresight of 'what now?' (Weick et al. 2005, page 412). Sense making thus avoided "the jaw of the fundamental attribution error" (Snook 2000, p. 206) that came from attributing the genesis of a failed decision to a decision maker: Who was wrong or who made a flawed evaluation? Instead, sense making focused on contextual factors, which drove organizational behavior (Snook 2000).

Studies of sense making after events contributed answers to two important research questions: "How does something come to be an event for organizational members? ... What does an event mean?" (Weick et al. 2005, p. 410).

First, from process perspective sense making started with flux, the chaotic multitude of potential antecedents and consequences of organizational life.

Some of the sense impressions of the flux stood out: they were noticed and bracketed as non-normal by members of the organization, in other words, some sense impressions became an events while others remained a non-event through labelling the sense impressions events. In more general sense, sense making is about labelling and thus stabilizing the flux of existence (Snook 2000, Weick 2005). Categorization and labelling are important processes linked to individual memorization and cognition (e.g., Barsalou et al. 1998; Barsalou 2008), individual reasoning and learning (e.g. Kahneman 2003), as well as organizational memorization and cognition (Douglas 1986). Labels and their underlying categories changed in their meaning through contextualization and influences from social systems. Labels and category systems also changed in their structure through shifts in salience, i.e., some instances moved from the periphery to the core of a category (Weick 2004). In sum, sense making is a process of organizing through communication (Weick 2004).

Second, explanations of meaning of events have drawn on the attention-based view of the firm (Ocasio 1997). Rare events redirected attention to specific features of an event and the organization (Levinthal and Rerup 2006; Rerup 2009). As an outcome of rare events organizations re-focused attention of individuals to new features of their environment and of their organization, e.g., attending to weak cues from lower levels (Rerup 2009)⁷⁹. Moreover, the sense

⁷⁹ The concept of changes of organizational attention to weak cues on lower levels has been linked to the practice of mindfulness (Levinthal and Rerup 2006). Mindfulness in turn has been studied in the system-centric view as part of the culture of high-reliability organizations (Weick and Sutcliffe 2006; Weick and Sutcliffe 2007).

making tradition of the event-centric perspective found, based on the attention-based view of the firm, that organizations did not simply observe, make sense, and react to rare events but they actively enacted events, in other words that organizations actively created the meaning of events. Decision makers enacted the event through redirecting attention. They defined what features of an event were salient and on the mind of the organization and what features were not salient to the organization (Weick and Sutcliffe 2006). At the same time enactment and sense-making also led to sense giving of rare events. Sense giving is characterized by defining a vision of an event and exerting influence to spread this vision throughout the organization (Gioia and Chittipeddi 1991). The sense making perspective offers the explanation how outliers, in the sense of markedly deviating performance outcomes, are the result of sense making processes after an event hit an organization. Sense making labelled and categorized the flux of organizational life into events and non-events. Sense making also influenced the salience of on contextual factors, which influenced organizational behavior. Thus preventing outliers depends on identifying and labelling the negative event, and understanding and enacting the meaning of the event and its criticality for the organization. Sense making also redirected attention in organizations and thus influenced which characteristics of the event, the organization, and the individuals are salient and acted upon.

8.1.3 Framing Perspective

This section will first review why frames matter and then review different definitions of frames in organizational and individual decision-making.

The literature on crisis management pointed out that successful crisis handling and leadership comprised two important aspects: cognitive framing and the subsequent emotional response (James et al. 2011). Framing is the cognitive concept similar to sense making discussed in the previous section. Framing strongly influences decision-making in organizations during times of crises (Nutt 1990; Nutt and Wilson 2010). An organization recognizes through framing that the organizational environment is not purely exogenous (Kaplan 2011) and therefore framing influenced how the organizational environment was brought into the decision-making process (Nutt 1990). Organizational responses to events were mediated by managers' interpretations of the environment (Kaplan 2011). The literature argued that frames were either motivated by bounded rationality (Simon 1991), in other words as a tool to simplify complexity, or by uncertainty (Knight 1921), in other words when organizations needed to make 'tough decisions', i.e., decisions under high ambiguity, high uncertainty, or conflict (Nutt 1990). When uncertainty in a decision situation is high politicized contest of frames occurred in organizations (Kaplan 2008). Thus frames were not only an ex-post justification of action through leadership, sense-giving, and power but also an ex-ante part of the political process that led to decision-making in organizations (Kaplan 2008).

Frames are defined as persistent knowledge structures representing knowledge elements and the structure between the elements (Walsh 1995). Similar terms used are schema, theories-in-use, cognitive maps, scripts, and mental models (Bartunek 1987; Kaplan 2011). Frames were commonly defined following

Goffman (1974) for instance by Kaplan 2008; Kaplan 2011; Schwenk 1988; Nutt 1990; James et al. 2011. In Goffman's definition frames were the "structure of experience, individuals have at any moment of their social lives" (Goffman 1974, p. 13) or more cryptically

"I assume that definitions of a situation are built up in accordance with principles of organization which govern events - at least social ones - and our subjective involvement in them; frame is the word I use to refer to such of these basic elements as I am able to identify." (Goffman 1974, p. 10-11).

On the other hand, frames in their original definition referred to experiential knowledge of a situation (Goffman 1974), which made Goffman's frames singular, monadic, frozen forms, without the concept of interaction (Denzin and Keller 1981). Modern convention of frame analysis (Ghaziani and Ventresca 2005) rooted frames in cultural sociology where frames were the underlying structures or organizing principles for symbols and idea elements to form packages of meaning (Creed et al. 2002). Frames directed attention by defining what is relevant and what is not (Creed et al. 2002).

Schemas, on the other hand, were often portrayed as process and outcome at the same time. They represented and structured the salient knowledge of an individual and could be shared within a collective (Elsbach et al. 2005).

Schemas objectified an abstract concept by delineating what it is and what it is not; as well as by sense making of what it does; and by deconstructing the relationships between elements and its environment, i.e., how and why it works (Elsbach et al. 2005). Schemas were most often used in top-down, theory-driven

approaches to process information and decision-making in organizations (Louis 1991). In other words, the concept of the schema added a process explanation to the concept of frames. The next section will now review how frames and schemas explain how events cause outliers in organizational performance. Two consequences of frames and schemas are important: (1) the link of frames to situated cognition in organizations and (2) the influence of biases and affective responses on the decision making process in organizations.

First, the concepts of frames and schemas were recursively entwined with the notion of situated cognition (Elsbach et al. 2005). Situated cognition was defined as temporally bounded interactions of individuals or collectives engaged in specific cognitive processes embedded in specific organizational context at particular points in time. Key cognitive organizational processes are decision making, problem interpretation, and learning (Elsbach et al. 2005). Situated cognition was both a process and a temporarily bounded outcome (Elsbach et al. 2005). It described the interpretation, i.e., sense making and sense giving, processes that led to a comprehension, understanding, explanation and attribution of a situation (Weick et al. 2005, Starbuck and Milliken 1988) as well as a temporary understanding of the situation. Situated cognition also extrapolated possible courses of action and predicts their outcomes (Starbuck and Milliken 1988, Elsbach et al. 2005). Frames, in the sense of knowledge or meaning structures were key to situated cognition. Frames are not concepts, which solely exist 'inside the head' of individuals and collectives of an organization (Elsbach et al. 2005). For instance, McNamara and Bromiley (1997)

showed that contextual factors mediate individual decision making on loan approvals. Thus situated cognition could be understood as the interaction between context and frames (Elsbach et al. 2005). The relationship between situated cognition and frames was recursive: situational context shaped the frames used in the decision situation and at the same time frames shaped how the situation was perceived (Elsbach et al. 2005, Brown and Long 1992).

Second, organizational framing demonstrated the same biases and heuristics that have been found in individual cognition. For instance organizations used different schemas depending on the frame of a problem. Dutton and Jackson (1987) showed that, consistent with Prospect Theory (Kahneman and Tversky 1979, Tversky and Kahneman 1992) the applied frames differed whether a problem was framed as a threat/loss or as an opportunity. For instance events that were framed as a crisis or threat implied a likely loss, minimal control, and thus motivated risk seeking. Conversely, framing issues as an opportunity implied a likely gain, high level of control, and motivated risk aversion (James et al. 2011, p. 476). At the same time frames also elicited a different emotional response. Threats were associated with negative emotions, such as anger, if the event was external, or guilt, if the event was internal, or anxiety, if the event was framed as an act of nature. Opportunities, on the other hand, were associated with positive emotions (James et al. 2011, p. 478). In either case emotion further drove behavior (James et al. 2011, p. 478). For instance negative emotions have been shown to reduce the chance of re-framing an issue as an opportunity (Brockner and James 2008).

However, frames might be less useful in novel situations (Louis 1991, Walsh 1995) or under high uncertainty (Boland et al. 2001). Frames dominate perception processes and therefore have a stronger impacts on outcomes of routine decisions and actions (Boland et al. 2001, Heracleous and Barrett 2001, Carson et al. 2003).

Lagadec (1993) showed that sense making was triggered when frames, which were used in cognition, broke-down. A rare event led to the collapse of meaning because the boundary frames were temporarily disrupted. Boundary frames enabled individuals to differentiate between things, i.e., to bracket and categorize. The disintegration of boundaries left actors confused. Weick (1993) argued that disintegration triggered processes of sense making which led to a reconfiguration and rebuilding of frames. Thus from this viewpoint a rare event, crisis, disaster or strategic surprise was more the result of an experience, i.e., the confusion following the collapse of meaning and sense making, rather than the result of an event; which implies that the rare event or crisis had not been a thing in itself, but existed only through the meaning given and enacted by individuals (Roux-Dufort 2007, p. 110).

In sum, the research has shown that frames, as structures of knowledge or meaning, are closely linked to situated cognition in organizations. In ambiguous and uncertain situations, e.g., after rare events, surprises, or crises, organizations used frames to make decisions about what actions to take. Frames used in situated cognition were the outcome of a politicized contest of

competing frames. Research also showed how the choice of frames influenced how decisions were made and how individuals reacted emotionally to rare events, which in turn influenced how decisions were made.

8.1.4 Organizational Learning Perspective

This section will review the literature on organizational learning from and through rare events by focusing on three questions: (1) Can organizations learn from rare events at all? (2) What is the organizational learning from rare events? (3) What are the consequences?

First, the dominant type of learning from and through rare events was trial-and-error learning (Woodhouse 1988). Because rare events have per definition a low frequency learning has been shown to take place if additional trials were generated by organizations, for example through learning from near-misses or near-accidents (Sagan 2004). Barton and Sutcliffe (2009) argued that rare events need to be internalized to improve the safety behavior of organizations. One such way to replace external rare events with an internal rare events was to voice concerns and identify disparate perspectives within the organization (Barton and Sutcliffe 2009). Equally, organizational theorists have argued that organizations have learned from "samples of one or fewer" (March et al. 1991, p. 465). Observations, for example of first-time ventures such as an acquisition, showed that organizations learned even before the activity was finished and most engaged in a similar activity before the first one was finished and performance feedback became available. March et al. (1991) used the notion of

'richly experiencing history' to explain how learning happened in these cases.

To experience history richly, the situation was broken down into a chain of smaller cases (March et al. 1991). In other words, aspects of how individual phases or stages of the event were managed and how very specific short-term decisions were made created ample opportunity for organizational learning.

Second, learning in the event-centric school of thought is commonly defined as activities to reduce the likelihood of future crises (Madsen 2009). Research has demonstrated that rare events triggered audits of organizational response repertoires and led to changes of these response repertoires as an outcome of organizational learning through and from rare events (Christianson et al. 2008).

In conventional studies of crisis, the response repertoires were sets of organizational behaviors in terms of crisis preparedness, crisis response, and crisis recovery operations with a firmly established norm of what should happen during crises (Comfort 2005).

Third, what are the consequences of learning from rare events? Christianson et al. (2008) demonstrated how learning from rare events changed organizational routines at four sites of learning. First, immediately after the rare events members of the organization focused their response to the event not on action but on interpretation. Members of the organization actively made sense of the event through processing ambiguous and uncertain information. Second, rare events changed the relation between members and organizations. Rare events questioned what individual members contributed to the organization, how

their contributions built the collective and thus how they represented themselves in the collective, and whether those contributions were no longer needed by the organization. Third, to deal with the event, the organization rebuilt and replaced inadequate structures and routines, as well as acquired lacking capabilities. Fourth, as an outcome of rare events, the organization redirected its identity. The survival of a rare event redefined the distinctiveness of the organization (Christianson et al. 2008). As a consequence it can be hypothesized that outliers are the result of the failure to learn, which not only comes in the form of failing to prevent future negative events but also in the form of failure to learn by failing to change and adapt because of the rare event.

However, the extreme criticality of rare events, crises, disasters, and surprises isolates individual instances and in turn hinders theorizing (Roux-Dufort 2007, p. 105). Crises were mostly defined after the effect (Roux-Dufort 2007, p. 107). The extreme impacts of rare events often portrayed them as unique, unprecedented, or even uncategorizable (Christianson et al. 2008, p. 846). As such

"rare events by definition are not equivalent; they can vary along a number of dimensions, including magnitude, frequency, valence (positive or negative), and cause (internal versus external)" (Christianson et al. 2008, p. 857).

A further critique is that rare events have been treated as "accidental manifestations of underlying organizational processes", i.e., sense-making, framing, situated cognition, and learning (Lampel et al. 2009, p. 835). To

address the challenge that theorizing of idiosyncratic events is not possible theorists of the system-centric view have argued that events should not be framed as unique, but as exaggerations of events that are routinely encountered at a smaller scale (Christianson et al. 2008)⁸⁰.

"A common misunderstanding about crises – understood here as epochs of profound uncertainty and urgent challenges to the problem-solving capacities of the socio-political order in which they occur – is that they are all unique" (Brändström et al. 2004, p. 191).

In sum, the event-centric school explains the occurrence of outliers as a consequence of rare events, disasters, crises, or surprises happening upon an organization. While each concept carries a slightly different connotation the concepts share the idea that a special negative event occurred, which created the need for a response by the organization. Key organizational processes that are part of the response are: classical cognition, sense making, situated cognition and framing, as well as organizational learning from and through events. Outliers are then a consequence of attribution error and wrong choices of an organizational response to the event (classical cognition), erroneously labelling critical events as non-events (sense making), failing to understand and enact the meaning and the criticality of the rare event to the organization (sense making), drawing attention to the wrong aspects of the event and the organization (sense giving), political and cognitive biases in situated cognition

⁸⁰ The argument has also been made by system-centric theorists about the concept of mindfulness, e.g., Levinthal and Rerup (2006), Weick and Sutcliffe (2006).

(framing), failure to learn from prior rare events (learning), and failure to adapt and change as a response to rare events (learning).

8.2 Project Success/Failure versus Project Performance

A challenge of the extant literature, in particular the quantitative surveys of IT projects, is the confusion between performance and success/failure. The issue of what constitute success has been widely debated (e.g. Samset 2003, Cooke-Davies 2002, Shenhar et al. 2001, Müller and Jugdev 2012, Pinto and Slevin 1988, Freeman and Beale 1992, Jugdev and Müller 2005). De Wit (1988) suggested a differentiation between 'project success' in regards to the wider project objectives and 'project management success' in regards to the classic iron triangle of cost, schedule, and quality.

The first Standish Chaos report (Standish 1994) lost the distinction and reported a low success rate of 16% (53% were challenged, i.e., experienced cost overruns, schedule delays, and/or reductions in functionality, and 31% of the projects completely failed), which supposedly painted a software crisis. The Standish reports are noteworthy for their wide-spread impact despite questionable rigor. Dyba et al. (2005:64) went as far as calling the Standish reports "the most influential study of the early '90s". However, the findings have been widely questioned, for example by Jørgensen and Moløkken-Østvold (2006), Eveleens and Verhoef (2010), Glass (2006), Emam and Koru (2008), Boehm (2006). Most convincing was Eveleens and Verhoef's (2010) critique, where the authors were able to draw conclusions from a large dataset of 1,156 projects from three organizations, in which one organization used the Standish categories of successful/challenged/failed to manage the IT project portfolio. In this

organization forecasts were deliberately biased because the categorization pressured IT project managers to overestimate the project budget in order “to increase the safety margin of success” (Eveleens and Verhoef 2010, p.4) a strategy also labelled sandbagging (Blumentritt 2006). In the other two organizations in Eveleens and Verhoef’s (2007) analysis forecasts were biased differently. In one organization the estimates were biased in the opposite direction because forecasts were deliberately defined as the minimal amount of effort and the earliest possible date of completion. Eveleens and Verhoef concluded that because of organizational biases aggregated numbers of successful, challenged, and failed projects are useless and meaningless for benchmarks (2010, p. 7). In more general terms, the authors problematized the issue of comparing IT projects across organizations and abstracting from the organizational context when unknown biases are potentially present in the data.

This thesis deals with the two challenges of (1) success/failure and (2) biases through a change in perspective: cost overruns are cost underestimations. A conclusion first drawn by Kahneman and Tversky (1977/1979) in their review of empirical research of unaided, intuitive judgments of probabilities for uncertain events. Kahneman and Tversky coined the term planning fallacy, to describe the systematic errors of judgment (in the authors' words it is bias not confusion) that lead to underestimations of time, effort, and cost needed to complete a task. Kahneman and Tversky (1977) offered their own observations of academic publishing projects and a study of power plant repairs (Kidd 1970)

305

as empirical evidence. Other early empirical demonstrations were construction projects (Hall 1980), running errands (Hayes-Roth and Hayes-Roth 1979), and pioneer production plants (Morrow et al. 1981).

Changing the problem focus to forecast inaccuracies turns the problem of performance shortfalls into a problem of decision-making and inadequate planning. Morrow et al. (1981) found that "severe underestimation of capital costs is the norm for all advanced technologies" (page 2). They further found that the factors that account for underestimations were known early in the project before major expenditures were made towards detailed designs or construction had actually begun. Planning scholars have long established and demonstrated that the problem of delays, effort, and cost overruns is one of underestimation. Subsequent studies into the planning fallacy have shown that the fallacy was linked to planning activities, e.g., taking an inside view to forecasting problems (Kahneman and Tversky 1977, Kahneman and Lovallo 1993, Camerer and Lovallo 1999) and anchoring forecasts in plan-based scenarios, which diminish the relevance of past experiences (Buehler et al. 1994). This thinking is not new in the context of IT projects either, for example Ewusi-Mensah's (1989) study of decision support systems showed that a contributing factor of cost overruns in IT project portfolios is the underestimation of cost risks on the portfolio level⁸¹.

⁸¹ The study also rejected a fundamental assumption behind the idea of managing IT projects as a portfolio (first proposed by McFarlan 1981), which assumed that investments into IT projects are either risk free or the risks are small and balance each other out due to the law of large numbers. Ewusi-Mensah's (1989) empirical insight questions this assumption.

The change in perspective has two implications for this thesis in direct response to the objections raised in the discussions of success/failure and generalizing amidst unknown biases in forecasts.

First, the change in perspective from overruns to underestimations implies that cost overruns, schedule delays, and benefit shortfalls are not measuring success or failure of projects at large, but only success and failure of decision-making and planning. The documented tendency of projects to be over budget, to be late, and to fall short of their intended benefits raises the question whether the expenditures into the front-end process are contributing the expected value to the shareholders and taxpayers by better informing decision making and governance of investments. In short, this thesis adds to the body of work that argues for abandoning the over-simplified categorizations of project success and failure based on project performance (for example de Wit 1988, Fowler and Walsh 1999, Samset 2003, Thomas et al. 2008, Lech 2013). As far as success and failure is concerned, project performance can be seen as the success or failure of the political decision making and governance process.

Second, the change in perspective from overruns to underestimations shifts the question about the comparability of different project types to the question about the comparability of the decision-making process. The objection by Verhoef (2005) is valid that the performance result of a study into the cycle times of software development projects is not comparable to a study of the typical delay on a technology-lead business transformation project. However, this thesis

takes the position that if forecast are the results of similar planning activities, the measured performance can compare the accuracy of planning and decision-making processes.

In this vein, this thesis is going to operationalize project performance through the lens of cost benefit analyses. The theoretical roots of the cost benefit analysis reach back to Dupuit's discussion of the utility of public works published in 1844 (Prest and Turvey 1965). It became the standard for project appraisals first at the US Army Corps of Engineers as part of the 1936 New Deal (Prest and Turvey 1965). Later, cost benefit analysis was made mandatory throughout the US public sector through a 1981 executive order by the Reagan administration and two executive order issued in 1993 and 1994 by the Clinton administration (Adler and Posner 1999). The adoption of cost benefit analyses in the private sector is less well documented. However, early discussions stated that "Cost benefit analysis of computer-based information systems is a major concern of managers in public and private organizations using computers" (King and Schrems 1978:19). Innovations in financial analysis of capital investment decisions subsequently put cost benefit analysis into a pivotal role in the pace of adoption of information technology in the 1980s (Sassone 1988). Cost benefit analyses were not only used to justify investments into information technology, but cost benefit analyses were also used to schedule work, to staff projects, to monitor and control project progress, and to evaluate system integrators and project planners (Ginzberg 1979, Lederer et al. 1990).

However, quantifying the value of investments into information technology is difficult (Ahituv 1980). Reviewing 30 years of research into the subject Renkema and Berghout (1997) found 65 different suggested methods to quantify the value of IT investments with no consensus emerging. Similarly, Santos (1991) found that because of issues of applying conventional capital budgeting techniques to IT investments decision makers rely most often on intuition and gut feelings.

8.3 First Draft of Variables to Include in Study

For each project, we would like to obtain data for the items listed below. The more items that are covered for each project, the better.

Please include all IT Projects (regardless if they are named IT, IS, ICT), we define IT Projects as those whose solution is dependent on information and communication technologies.

Sampling criteria:

Scope	1,000	function points
Effort	2,000	man days
Duration	0.5	years
Budget	1	million USD
Date	1998-2008	
Sample Size	Min. 20	per participant

If these criteria result in too many projects to be manageable, I suggest shortening the list "from the bottom," i.e., excluding smaller projects until the number of projects is manageable, resulting in a list of the largest projects.

If the criteria result in too few projects to produce useful results for an individual participant, I suggest lowering the EUR 1 million threshold, for instance to EUR 0.25 million.

8.3.1 Technical data (for project as built)

- Project name and location.
- Project type

IT System deployment	Integration Project
	Standard Software

	Custom Development
IT Architecture	
IT Infrastructure	
Communication	
Outsourcing/Offshoring	
Other	

- If IT system replacement/new deployment please specify what kind of system typology used by Stahlknecht and Hasenkamp (1999)

User	Type of system	Example
Internal	Management Information/Support System	MIS, Decision making support system
	Planning and Control System	Reporting systems, ERP
	Office Information Systems	Workflow mgmt, document mgmt, knowledge mgmt, office automation
	Administration and Disposition	Core Banking System, Ticketing System, SCM, CAD
	Transaction Systems	Database Systems, EDI
External	B2B	eBusiness, B2B auctioning systems
	B2C	Web pages, eCommerce, eGovernment

- Project size/dimensions in man days, if measured additionally in function points, use case points [moved to Q5]
- Short description of the project, outlining objectives, name of software implemented and similar information
- Name of the department/government agency or subsidiary or country organisation that carried out the project
- Name of the programme the project was part of; if the project was part of a larger programme then include the name of the programme, thus projects that form strong dependencies can be grouped together
- Type of contracting – Fixed price or time and material to control for biases and explanation in the data; also include other contract types such as cpff, cpfc, cpif, ffix if the information is available
- Planning constraints – control variable to check if during the project cost or schedule escalation was constrained either as design-to-budget or design-to-schedule

- Context: Type of company/organization, division within company/organization.
- [Is type of organization not the same as type of ownership and finance?]

Type of organization	Public
	Private
	Public-Private Partnership
	Other (please specify)
Size of organization	No. of staff
	Annual revenue
Size of project	No. of users
	No. of staff involved full-time
	No. of staff involved part-time
	No. of vendors
Division within organization	IT
	Business
	Separate project entity

- Industry

Financial Services
Consumer Discretionary
Information Technology
Industrials
Telecommunication Services
Health Care
Consulting/Services
Consumer Staples
Materials
Manufacturing
Public Sector
Utilities
Energy
Other

— Type of ownership and finance (private, public, PPP, etc.).

— Other.

- Currency

8.3.2 Schedule

- — Year and month of formal decision to go ahead. A.k.a. "time of 'go' decision" or "full business case".

- Estimates of "go-live date" (defined as the date the system is made available to users), with indication of the time each estimate was made (year and month, stage in project cycle). The estimate of go-live date should be given for each of the following stages in the project cycle (estimates may be identical for several or all stages in the cycle):

Life cycle stage	Date (MM/YY)	Estimated go-live (MM/YY)	Comment
Pre-feasibility study (strategic outline case)			
Feasibility study (outline business case)			
Formal decision to build (full business case)			Date and estimate needed
Tender			
Development/ project execution start			Date is needed, estimate is highly desirable
Go-live		n/a	Date needed

— ~~Causes of schedule changes, if known.~~

8.3.3 Budget

- Costs are estimated excluding VAT, project costs are all costs needed to build the solution. This does not include operations costs, e.g. if parts of the solution went live during the project
- Costs should be given in constant prices, with indication of which year's level
- Estimates of costs from first to last estimate, with indication of the time each estimate was made (year and month, stage in project cycle). Costs estimates should be given for each of the following stages in the project cycle (estimates may be identical for several or all stages in the cycle):

Life cycle stage	Date (MM/YY)	Estimated cost	Comment
Pre-feasibility study (strategic outline case)			
Feasibility study (outline business case)			

Formal decision to build (full business case)			Date and estimate needed
Tender			
Development/ project execution start			Estimate is highly desirable
Go-live (actual)			Needed

- To quantify the optimism bias that goes into the initial cost estimate
- Range of estimate at go-decision (full business case)
- Probability of estimate falling into the range, e.g., 100% = range is max-min, 95% (2 std. deviations away from the mean), 80% that is 8 out of 10 projects deliver within the range (1.8 std. dev. from the mean)
- Size of risk buffer (in %) included in estimate at go-decision (full business case)
- Final actual project costs, in constant prices.
- If further costs may be expected at a later stage - e.g., costs from unsettled claims or unfinished features of the project - these costs should be listed. Broken down by

Hardware	
Software	
Internal personnel	
External personnel	
VAT rate (if included)	

- Breakdown of effort put into typical stages of the lifecycle

Feasibility/Pre-Study	
Requirements/Specification engineering	
Development/Customizing	
Testing	
Deployment/Integration	
Rollout	
Training	
Risk Buffer	
Project management	

8.3.4 *Causes of inaccurate cost estimates*

If cost estimates were inaccurate for a project (resulting in cost overrun or underrun), please have respondents indicate which of the following factors in their judgment caused the inaccuracy (more than one factor may be checked):

1. Scope changes.
2. Optimism bias (estimates that are too optimistic).
3. Deliberate (strategic) cost underestimations.
4. Inadequate feasibility study (for instance re. complexity).
5. Delays.
6. Unforeseen price fluctuations.
7. Unforeseen requirements changes.
8. Other unforeseen events.
9. Inadequate risk assessment and management.
10. Other (please specify)

8.3.5 *Relative project size*

Size of the total IT budget the project was part of – to analyze whether it was comparatively large or small

8.3.6 *Estimated project size*

Size of the project is measured in person days. A 220 work day year is assumed.

If size is measured in function points (according to ISO 20926, IFPUG etc.)⁸² this should be stated as well as which method was used, e.g., function points,

⁸² Function points are estimated from a pure user perspective, independently of implemented technology. The development follows 4-step process

- a) Count algorithms and logical files, as well as transactions
- b) Evaluate complexity of each algorithm, files and transaction
- c) Estimate the value adjustment factor to reflect complexity of the development environment (e.g. user requirements, technical requirements, quality standards)
- d) Calculate the Function Points as a product of the sum of all files and transactions, weighted for their complexity, multiplied by the value adjustment factor

Alternatively size can be measured in use case points. Use case point estimation is a simplified approach of function point estimation following a 5-step process. It is based on requirements definition in a use case format.

- a) Estimate the actor weight factor as weighted sum of all actors needed (simple interfaces, interactive, protocol driven interface, GUI)
- b) Count the use cases weighted by number of transactions inside each use case
- c) Estimate technical complexity factor for the whole project

normalized function points, use case points, lines of code, COCOMO. Published benchmarks will be used to convert other size estimates to function points.

Life cycle stage	Date (MM/YY)	Man days estimate	Size estimate	Comment
Pre-feasibility study (strategic outline case)				
Feasibility study (outline business case)				
Formal decision to build (full business case)				Date and estimate needed
Tender				
Development/ project execution start				Estimate is highly desirable
Go-live (actual)				Needed

8.3.7 Users

To compare different point estimates we would like to scale the different project sizes along their number of users along the category of IT users (admin, developers) and 4 grades of named users on the business side (complex, super, standard, lightweight)

Users	Number
Admin	
Developer	
Complex Users	
Super Users	
Standard End User	
Lightweight Users	

8.3.8 Actual Size Break-Down in Persondays

Feasibility/Pre-Study	
Requirements/Specification engineering	

-
- d) Estimate environmental factor for the whole project
e) $UCP = [a) + b)] * c) * d)$

Development/Customizing	
Testing	
Deployment/Integration	
Rollout	
Training	
Risk Buffer	
Project management	

8.3.9 Value and benefits estimates

Any available estimate should be given, e.g. NPV (including used discount factor), payback period (including how it was calculated, e.g., simplified payback by allocating costs to last project period, or cash flow based payback of project), internal rate of return, additional revenue, cost savings

Life cycle stage	Date (MM/YY)	Benefits estimate	Comment
Pre-feasibility study (strategic outline case)			
Feasibility study (outline business case)			
Formal decision to build (full business case)			Date and estimate needed
Tender			
Development/ project execution start			Estimate is highly desirable
Go-live (actual)			Needed

8.3.10 Complexity

~~Complexity of software requirements on a scale from 0 = not complex~~

~~(following the factors for the value adjustment in function points)~~

-	0	1	2	3	4	5
Data-communication with user	File interface	-	Remote data-entry	Online GUI	-	Multiple front-end and comm.-Protocols
Distributed data processing	No transfer	-	Data needs to be synchronized in one-direction (master-slave)	-	-	Dynamic-synchronisation (hot standby)

Performance requirements	None	-	High (performance is critical)	-	-	Very high (measurement required)
Heavily used configuration	No influence of hardware configuration on software	-	Security or timing influence	-	CPU balancing	Balancing distributed resources
Transaction rate	No	-	Weekly peaks	-	SLAs	Tools and performance measurement for SLA
Online data entry	0% of all entries	<20%	<40%	<60%	<80%	>80%
End-user efficiency	No add-ons	-	2-5 add-ons	-	-	All add-ons
	Navigation aids, menus, context help, automatic cursors, scrolling, remote printing, function keys, mouse interface, pop-up windows					
Online update of software	-	Few control files	-	Few logical files	-	Entire application
Complex processing	-	One critical area	-	A few critical areas	-	All areas
	Sensitive controls, extensive logical processing, extensive mathematical processing, exception processing, complex processing of multiple inputs/outputs (e.g., multimedia)					
Reusability	Not at all	-	Single elements	-	-	Whole application
Installation ease	Not critical	-	Some user requirements	-	-	Fully automated
Operational ease (startup, backup etc.)	No consideration	-	-	Considered in some aspects	-	Designed for unattended operations
Multiple sites	One site	-	-	Multiple sites with different hardware	-	Multiple sites, environments, documentations, support
Facility change	No change	1 additional change	2 additional changes	3 additional changes	-	-
	User adaptable queries and reports, processing logic can be configured, data structure can be changed online					

8.3.10.1 Technical complexity

Using the complexity measures used in the use case point methodology

-	0	1	2	3	4	5
Distributed system						
Performance objectives						
End-user efficiency						

Complex internal processing						
Reusability						
Easy to install						
Easy to use						
Portability						
Ease of change						
Concurrency						
Security objectives						
Third party access						
User training facilities						

8.3.10.2 Environmental complexity

	0	1	2	3	4	5
RUP familiarity						
Application experience						
Object orientation experience						
Lead analyst capability						
Motivation						
Stable requirements						
Part time staff						
Difficult programming language						

8.3.10.3 Project Management Complexity

—— Complexity as defined by Whitty & Maylor (2009):

- Structural complexity
- Dynamic complexity

(Complexity is usually characterized as open, dynamic, recursive, non-linear feedback, and emergent)

Complexity as measured by Orlitzky and Hirokawa (2001)

- Degree of clarity of goals (What?)
- Degree of clarity of goal paths (How?)
- Propensity of goals to change during project

- Propensity of goal path to change during project

Complexity after Maylor, Vidgen & Carver (2008): Original list, less frequent mentions are stricken out

— Mission	Objectives	Is there a clear vision? Are the goals, business case, success criteria clear? Are the goals aligned with the organisation's strategy? Are there measurement criteria for project success Are the requirements clear? Is the scope clearly defined?
	Scale	Long timescale Large number of resources Large number of deliverables High value/budget High profile, e.g., strategic/controversial Broad scope, e.g., product for a global market place
	Uncertainty	Are there interdependencies with other projects? Are there competing priorities between projects? Does it involve new technology? Has the project been done before? Is the timescale realistic? Are the implications and side effects understood? Is there a large number of unknowns? Have assumptions been tested? Is there a large solution space? Are there many ways of achieving the solution? Is there a high degree of urgency? Is the budget defined? Are there significant intangible benefits? Are there interdependencies within the project? Are there multiple technology interfaces? Is the product aimed at a new market? Is product highly customized?
Organis ation	Constraints	Are there legislative or compliance constraints? Confidentiality Security Health and Safety
	Time	Are there multiple time zones?
	Space	Are team members collocated? Is there face to face communication between team members?
	Geography	Are there multiple languages? Are there multiple currencies?
	Project/orga fit	Is there a mismatch between structures of project (matrix) and organisation (departments)? Is delivery to a non project based organisation?
	Orga change	Does the project involve organisational restructuring? Does the project involve cultural change?

		Is there ongoing organisational restructuring that impacts the project?
Delivery	Administration	Is the project data collection accurate?
		Is the project data collection true?
		Is the project data collection complete?
		Is the true progress/status of the project known?
		Is there information overload?
		Is status communicated effectively up, down, and sideways?
		Is the procurement process effective? Are inputs acquired in a timely manner?
	Decision making	Is there effective governance of project decision making?
		How many levels of management involved in decision-making?
	Change mgmt.	Is the change management process cost effective?
		Is it sufficiently flexible?
	Project processes	Are project processes defined, standardised but not over-bureaucratic?
		Is there a clear responsibility for tasks and deliverables?
		Are work packages well defined?
	PM methodology	Is there a common methodology used throughout the project?
		What is the maturity level of the organisation?
		How well does the project team understand the methodology?
		Is the methodology used for real?
		Is the methodology appropriately adapted to the project?
	Resources—HR	Are human resources shared across projects?
		Who controls human resources for the project?
		Does the project manager have control over resource-selection?
		Do human resources have sufficient %FTE allocated to the project?
		To what extent is there reliance on key experts?
		What is the availability of key experts?
	Resources—Tech	Does the project have tool support?
		Are the tools new?
		Are the tools home grown/unsupported?
	Resources—Finance	How flexible is the project budget?
Stakeholders	Stakeholder-identification	How many stakeholders are there?
		Are there any unidentified stakeholders?
	Support for-project	Have stakeholders buy in to the project?
		Have commitment to the project?
		Interfere with the project?
		Have sufficiently available time for the project?
		Resist the project?
		Take ownership of the project?
		Do they respond to project needs in a timely manner?
	Relationship-basis	Is the relationship between the project and stakeholders contractual?
	Experience	Do the stakeholders have realistic expectations of the project?
		Do they have domain experience?
		Do they have project management experience?

	Power	Do they understand the implications of the project?
		Do the stakeholders have power to make decisions.
		Have authority?
		Have accountability?
	Key stakeholders	Is there senior management support?
		Is there clear sponsor?
		Are unions involved?
	Socio-political	Have the stakeholder groups worked together before?
		Are there hidden agendas/unsurfaced assumptions?
		Is there conflict between stakeholders?
		Are there power struggles between stakeholders?
		Do the requirements of different stakeholders conflict?
		Are there competing priorities between stakeholders?
	Interdependencies	Is there a shared understanding of the aim of the project among stakeholders?
		Are there interdependencies between stakeholders? (e.g. between suppliers)
Team	Project staff	Are the team members motivated?
		Do the team members have a good attitude?
		Are the team members knowledgeable and competent in all aspects of the project (business, technical, and project management)?
		Are the team members committed to the project?
		What is the level of sickness/leaves?
		Do the team members have sufficient prior experience in all aspects of the project (business, technical, and project management)?
		Does the project involve multiple languages, and technical disciplines?
		Has sufficient training be conducted?
	Project manager	Is the project manager an effective communicator?
		Does the project manager display leadership?
		Does the project manager have authority?
	Group	Are there cultural differences between team members?
		Are there personality clashes/rivalry within the team?
		Have the team members worked together before?
		Do the team members trust each other?
		Does the team have a shared vision for the project?

8.3.11 Critical Success Factors for Senior Mgmt.⁸³

Quantitative Studies showing that top management support is a critical success factor.

⁸³ Based on the meta-analysis by Fortune and White (2002)

~~Best Practices (identified in qualitative research) for Senior Management Sponsorship are (Helm & Remington, 2005):~~

~~[Measurement should either be on a scale from 1-7 or 1-5 or 1-10, could be asked from the project manager or sponsor perspective]~~

- ~~•—Appropriate and power within the organization~~
- ~~•—Political knowledge of the organization and political savvy~~
- ~~•—Connect project and organization~~
- ~~•—Battle for the project~~
- ~~•—Motivate team~~
- ~~•—Partner with project team~~
- ~~•—Communication skills~~
- ~~•—Compatibility with project team~~
- ~~•—Provide objectivity and challenge project~~
- ~~•—General effectiveness in the role as project sponsor~~
- ~~•—Governance structure: Single Sponsor (if so — which department?), Joint Sponsorship, Steering committee, other (please specify)~~

~~Contingency approach (support focus vs. governance focus) [Crawford et al. (2008)], measured on a scale from 1-10/1-5/1-7:-~~

- ~~•—Does sponsor appropriately support the project if it has a need, e.g., when parent organisation fails to provide resources, resistance occurs, lack of decision making, or conflicting objectives are imposed on project?~~
- ~~•—Does sponsor represent appropriately the need of the organisation, —e.g., if project is high risk/high exposure, project performs poorly, market changes?~~

8.3.12 *Uncertainty/Dynamics of the project*

Measurement of uncertainty and changes in the project by getting

- ~~Total number of raised issues in the issue log at the end of the project~~
- ~~Total number of change requests put forward and change requests approved~~
- ~~Total monetary value of risk register at the business case stage~~
- ~~Value of reduction lists~~

8.3.13 *Qualitative characteristics*

From the previous deleted questions we selected the 15 most important qualitative characteristics (determining the social and technical complexity) and 14 managerial characteristics that assess underlying best practices explaining project escalation.

Qualitative Characteristics of the project (score 1-10*)	The goals, business case, and success criteria were clear.
	The project involved innovative technology.
	There were a large number of unknowns in the solution design.
	The project followed an incremental approach.
	The deliverables of the project were highly inter-dependent.
	The project faced resistance in the organization.
	The project had an effective governance structure.
	Senior management supported the project.
	The project had competing and shifting requirements.
	The project could influence its environment in order to achieve success.
	This project is unique and therefore it is difficult to compare with other projects.
	The solution requires complex internal processing in the system
	The solution needs to be reusable
	The solution needs to fulfill high performance requirements
	Security aspects are very important for the solution
Managerial Characteristics of the project	The initial plan and business case showed political intent instead of the most likely outcome
	If we had known the real costs of the project it would not have been built
	The organisational unit has a historic record of past projects that this business case was compared to
	Effectively monitored and controlled, e.g., by a PMO?
	Supported by senior management?
	Lead by competent project manager(s)?

(score 1-10*) "Was the project..."	Planned in detail and the plan updated regularly?
	Following a realistic schedule?
	Lacking suitable staff?
	Good in communication among team members?
	Supported by well performing suppliers?
	Managing change effectively?
	Addressing and managing risks effectively?
	Involving users?
	Using an effective management methodology?
	Had the project an effective project sponsor?
	Different scale! - How many years of experience did the most senior project leader have?

8.4 Technical definitions

8.4.1 Project Types

Project types were differentiated by the type of the IT solution and software types following Stahlknecht and Hasenkamp (1997). Table 31 shows Stahlknecht and Hasenkamp's typology, which is based on a business domain logic to differentiate projects and systems.

Table 31 Project and system types (Stahlknecht and Hasenkamp, 1997)

Project types	System types
System deployment - integration	Internal
System deployment - standard software	Management information/support systems
System deployment - bespoke solution	Enterprise requirements planning/ control systems
IT Architecture	Office information / automation, incl. document and content management systems
IT Infrastructure	Administration and scheduling, incl. core banking, supply chain management systems
Communication technology	Transaction, including database and electronic data interchange systems
Outsourcing and offshoring	External
IT post-merger management	Business-to-business
IT organizational change	Business-to-consumer
Other	Other

8.4.2 *Planning constraints*

Planning constraints defined whether the project was planned to cost, planned to schedule, planned to cost and schedule, or had no known constraints.

8.4.3 *Contract Types*

Contract types defined whether the project had a fixed price contract, time and materials, cost plus fixed fee, cost plus percentage of cost, cost plus incentive fee, or fixed price plus incentive fee.

8.4.4 *Project Lifecycle*

The actual project schedule as it had happened was captured through a simplified project lifecycle (cf. table 32), which followed Flyvbjerg et al. (2003, 2004, 2005). If the project followed an iterative methodology the latest date of the specific activity was collected. In other words, finish of the design stage means that no design activities happened after this date. The method was chosen because it allowed a more general comparison between different projects, at the expense of losing the richness of detail of different project methodologies.

Table 32 Simplified Project Phases

Project stage	Definition	Example documents
Pre-feasibility study	End of idea stage, when pre-feasibility study was finished (if any)	Strategic outline case
Feasibility study	End of feasibility study	Outline business case
Formal 'go decision' to build	End of planning, date when project received the green light	Full business case
Tender	Date of contract signature	Bidding documents, contract
Project execution start	End of a ramp up	Project initiation document
Specifications completed	End of requirements definition	Project progress reports, steering group reports
Design completed	End of design activities	Ditto

Development completed	End of coding, development or customization activities	Ditto
Testing completed	User acceptance test	Ditto
Go live	End of roll out, final go live of full operational capacity and geographic footprint	Ditto
Project close	End of the project	Post implementation review, audited project accounts

8.4.5 *Project Performance*

Following long-standing convention (Conte et al. 1986), cost, effort, schedule, and benefits performance are defined as the difference between actual and estimated cost/ effort/schedule/benefit expressed as percentage of estimated cost/effort/schedule/ benefits or as the ratio of actual over estimated cost/effort/schedule/benefits. The survey instrument collected estimates at the end of each project stage in table 32. Unless indicated otherwise, project cost/effort/schedule/benefits performance is calculated as the difference between actual cost/effort/schedule/benefits and estimated cost/effort/schedule/benefits at the formal go decision to build the project.

8.4.6 *Actual and Estimated Cost, Effort, Schedule, and Benefits*

Actual costs are defined as real, accounted cost determined at the time of completing the project (outturn costs). Estimated costs are defined as budgeted, or forecasted cost at the time of the estimate.

Actual effort is defined as accounted for effort in person days of the project at the time of completion. Estimated effort is defined as budgeted, or forecasted project effort, in person days, at the time of the estimate.

Actual schedule is defined as duration between the go-live of the project and the decision to build in calendar days. Estimated schedule is defined as the length of time in calendar days between the decision to build and the date of the planned go-live.

Actual benefits are defined as the quantified and accounted for benefits of the IT project at go-live or a later date, when the project was closed and/or a post-implementation review of the benefits achieved was completed. The survey instrument allowed to capture any metric used to quantify benefits. Estimated benefits were defined as the forecasted, or estimated, quantified benefits at the end of each project stage⁸⁴.

⁸⁴ This definition can lead to somewhat counter intuitive results because benefits shortfall can be smaller than -100%, for example if an IT project was started to reduce the operations costs of an IT solution but the actual costs of the replacement systems were higher than the operations costs of the legacy system.

8.5 Anchoring of Explanatory Variables and Anchoring of Variables

Table 33 Anchoring of Variables in Survey Instrument

Explanation	Variable	Item	(1)	(4)	(7)	(10)
Social complexity	H2.2.1 Business case	The goals, business case, and success criteria were clear.	No defined business case, goals nor success criteria	Project goals vaguely defined; only effort cost estimate no full business case	Project objectives stated w/o prioritization; rough business case exists	Clear, concise, prioritized goals and objectives; sound business case incl. benefits; clearly defined success measures
	H2.2.2 Environmental influence	The project could influence its environment in order to achieve success.	The project had not influence on what went on outside of it	The project could only influence the project environment through its sponsor/top-execs	The project leadership had some influence on the project environment, disturbances happened but were dealt with quickly	Effective influence over the environment, disturbances were anticipated ahead and dealt with before they happened
	H2.2.3 Resistance	The project faced resistance in the organization.	The project was uncontested in the organisation	A few users did not see the need for the project, needed more training and communication	Key users were on board, but some parts of the organisation did not want the project	Users did not understand the project, were highly critical of it, users did not adopt the solution
	H2.2.4 User involvement	Was the project involving users?	The project was done without user involvement; very limited training	Users were consulted, via business units; varying degree of usefulness; only basic training	Users were involved often and frequently but not continuously; user training happened just before go-live	Users actively involved as part of the project team, they are knowledgeable and motivated; users were trained early and sufficiently
	H2.2.5 Political intent	The initial plan and business case showed political intent instead of the most likely outcome	No political influence on the project plan	Some tried to influence the plan and business case; as a result it was somewhat aggressive but still realistic	The plan and business case were not allowed to be questioned; assumptions were very aggressive	The plan and business case was dictated from above/by the client and unfeasible
	H2.2.6 Schedule	Was the project following a realistic and reliable schedule?	Project plan unrealistic, mainly arbitrarily imposed and impossible deadlines, no buffer	Highly aggressive plan, with fixed imposed end date, minimal buffer	Challenging plan and deadlines, but includes adequate buffer	Plan and deadlines were realistic, no arbitrarily or imposed deadlines, plan includes adequate buffer
	H2.2.7 Uniqueness	This project is unique and therefore it is difficult to compare with other projects.	The project was not unique at all	Project had a few unique aspects to it; but not fundamentally different from other projects	Project was unique to parts of the team, users, e.g. experienced vendor or project leaders; difficult to compare to other projects	Truly one of a kind project, seen as unique by team, vendors, clients
Technical complexity	H2.3.1 Requirements volatility	The project had competing and shifting requirements.	Requirements were mutually exclusive and stable throughout the project	Very few trade-offs between requirements had to be made, changes were tightly controlled	Trade-offs between requirements had to be made, changes happened often	Requirements frequently conflicted and changed
	H2.3.2 Innovation	The project involved innovative technology.	Solution based on fully tame, off-the-shelf technology; no innovation needed	Technology is tried and tested but needs a few additions and customizations	Technology is somewhat new, project is early adopter; or technology delivered at unprecedented scale	Highly innovative and untested technology; project is new product development
	H2.3.3 Coupling	The deliverables of the project were highly inter-dependent.	The deliverables were fully decoupled	Few deliverables are coupled to others, very limited complexity of dependencies	Key deliverables coupled to others, some but still manageable complexity	The deliverables were tightly coupled, inter-dependencies very complex and difficult to foresee

Software complexity	H2.3.4	Scope volatility	The scope was minimal and stable.	Scope unknown or widest range possible, scope changes easy and frequent, "we don't know what we will be doing next week"	Scope w/o clear boundaries, some gold plating, changes happen often	Scope defined, scope control processes are in place, scope changes sometimes	Scope no larger than necessary, tight scope control, no changes to the work until completed
	H2.3.5	PM method	Was the project using an effective management methodology?	No methodology used, no tool support, no management processes in place	Management methodology was somewhat ineffective, e.g., lack of tools, capabilities, key management processes	Project management method was sound but could be more efficient and effective	Methodology project management were highly appropriate for the project and delivery method
	H2.3.6	Unknowns in design	There were a large number of unknowns in the solution design.	The design was fully specified; all conflicts were solved	Design specification done, but contains a few open questions and action points	Design principles and boundaries are defined	Design was not ready to be implemented
	H2.4.1	Processing	The solution requires complex internal processing in the system	No complex operations in the system for either data or functionality	Few, but limited, complex operations	Several complex operations, but optimization and design still manageable	Many highly complex operations
	H2.4.2	Reusability	The solution needs to be reusable	No reusability requirements	Basic functionality needs to be reusable	Few key/critical components need to be reusable later on	Solution needs to be highly flexible and widely reusable
	H2.4.3	Performance	The solution needs to fulfil high performance requirements	No particular requirements for response times, availability/up-times, or speed	Standard requirements for response times, availability, speed very few peaks	Above average speed, availability, response time requirements with frequent peaks	Very challenging requirements for response times, availability, speed
	H2.4.4	Security	Security aspects are very important for the solution	No particular security requirements	Few security considerations needed, e.g. basic role-based security concepts, no special data centre provisions	Solution is somewhat security sensitive, but easily manageable with standard data centre requirements	Security is highly important; data, access, functionality, data centre need tight control
	H2.5.1	Outside view	The organisational unit has a historical record of past projects that this business case was compared to	No comparison of the project to past projects was made/ could be made	Some team members had previous experience with similar projects	Project was compared to 1-2 previous projects	Project was systematically compared to previous projects
	H3.1	Risk management	Was the project addressing and managing risks effectively?	No risk management	Risk management, not useful to inform future work, box ticking exercise	Risk management somewhat effective to identify future changes and roadblocks, could be more efficient and effective	Effective risk management, able to foresee and deal with changes and roadblocks to the project
	H4.1.1	Monitor & control	Was the project effectively monitored and controlled, e.g., by a PMO?	Project had no control over spending, resource consumption, and progress	Key information was tracked, but not good enough to make decisions	Information sufficient to support decision-making	Tracking of true status of project, leading indicators enabled forward oriented decision-making
Project processes	H4.1.3	Plan	Was the project planned in detail and the plan updated regularly?	No reliable plan existed if one existed it was completely out-of-date	Project plan existed/was made in the beginning but rarely updated; not a useful tool to inform future work	Detailed project plan existed, was updated infrequently, helped in informing future work, but could have been more efficient or effective	Cost, schedule, benefits/functionality planned in detail (incl. dependencies) and tracked regularly to inform future work
	H4.1.4	Change management	Was the project managing change effectively?	No change management process was defined, change happened frequently and uncontrolled, users often surprised	Ineffective change management process, change was not controlled; change that the project brought to the organisation was not managed	Change was only managed internally in the project, but no effective change management for changes the project brought to the organisation	Change to the project was tightly controlled and well managed; as well as change that the project brought to the organisation

Staffing	H4.2.1	Sustainable res. mix	Was the project using a sustainable mix of internal & external resources?	Roles needed long-term staffed externally, no knowledge sharing	Some key roles are staffed externally, no clear plan how to internalize; collaboration and trust between int/ext. personnel slightly unsatisfactory	Some key roles staffed externally with a clear plan for knowledge transfer and succession; good collaboration between int/ext. resources	Roles needed long-term staffed internally; trust, collaboration, and knowledge exchange between int/ext
	H4.2.2	PM	Was the project led by competent project manager(s)?	Project leadership team lacked critical competencies and capabilities	Leadership team had some gaps in the competencies, e.g., lack of decision-making power over staffing; and/or capabilities, e.g., technical knowledge	Project leadership team had all the necessary competencies/capabilities, lacked a few but minor ones	Project leadership team was highly competent and motivated, good capabilities (incl. people skills), trusted and with good standing in the organisation
	H4.2.3	Team	Was the project carried out by a qualified and motivated team?	Team mostly inexperienced, lacking qualification, and/or not overly motivated	Some team members responsible for critical tasks are inexperienced, lack qualification and/or not fully motivated	Motivated and qualified and experienced team on the critical tasks, with the occasional gap in the wider team	Highly qualified, experienced, and motivated project team
	H4.2.4	Staffing	Was the project lacking suitable staff?	Sufficient number of experienced and qualified personnel are assigned full-time to key tasks	Personnel is experienced and qualified with a few gaps; most work full-time or 80% for the project	Personnel mostly 50% on the project with key roles in other projects, several competency gaps and lack of experience	Personnel responsible for key tasks is not fully qualified, lacks experience, and is assigned only part-time to the project
Governance	H4.3.1	Governance	The project had an effective governance structure.	There was virtually no or completely ineffective governance for the project	Governance was defined but slow and not suitably staffed	Governance was effective, but sometimes slow, or missed 1-2 key decision-makers from time to time	Highly effective governance, top-level decisions were made quickly and competently
	H4.3.2	Sponsor	Was the project having an effective project sponsor?	No project sponsor	Project sponsor not involved in the project other than distant meetings; ultimate decision-maker	Project sponsor somewhat available to the project, mainly involved in critical decisions, sponsor deals only with issues on user/client side if escalated	Highly involved project sponsor, available for problem-solving, sponsor highly active to work on the client/user side
Support of the project	H4.4.1	Senior management support	Senior management supported the project.	No support from/access to top execs; avoidance of blame, everything was projects fault	Senior management solely the ultimate decision-maker (e.g. in steering committee), issues were only escalated as last resort	Vocal exec support, but limited involvement in project and problem solving, mainly involved in final decisions	Vocal exec support, highly involved in project and problem-solving, team comfortable to escalate issues
	H4.4.3	Contracting	Was the project employing robust supplier contracts with clear responsibilities?	Contracts are a high risk, some contracts are not in line with project objectives anymore, no flexibility to react to changes; no incentives to align vendor's with project's objectives; no management of contracts	Parts of the contracted scope no longer useful (if ever), only end-products contracted, limited ability to deal with changes, incentives not aligned with goals	Some contracted scope is not needed anymore, rough definition of interim products and approval criteria; standard escalation process to react to changes; key incentives align with goals	Contracts executed professionally (e.g. by vendor management team), scope in line with project needs, interim products defined, with detailed approval and acceptance process, contracts are flexible to deal with foreseeable and unforeseeable changes; incentives fully aligned with project objectives

	H4.4.4	Stakeholders	Was the project able to align the major stakeholders?	No stakeholders identified OR Stakeholders are highly conflicting, no shared vision of goals and success criteria exist, many and complex groups with opposing views	Stakeholders fall into multiple camps with several conflicting visions of the goals and success criteria of the project	Stakeholders fall into basic camps with largely overlapping but somewhat conflicting vision of the goals and success criteria of the project	Major stakeholders share a joint vision of the goals and success criteria of the project
Communication	H4.5.1	Status communication	Was the project transparently communicating the project status?	Virtually no status reporting to speak of	Status reporting usually slow and out-of-date; reliance on gut feeling about progress, cost etc.	Transparency limited e.g. to FTE but not costs, status reporting only via queries to subprojects	Full transparency on budget, schedule, and benefits; insights into all sub-projects and their dependencies
	H4.5.2	Team com.	Was the project good in communication among team members?	No defined flow of information exists, members keep information to themselves, communication is rife with rumours	No defined flow of information, team members occasionally keep information to themselves	Defined flow of information, but irregular sharing of information; frequent informal sharing of information	Information is regularly shared in an organized way, project team members are communicate goals and act in the interest of the project
Viability	H5.1	Viability	If we had known the real costs of the project it would not have been built	No regrets; no or very limited cost overruns or delays	Less than anticipated cost/benefits ratio but no real challenge to the business case	With the true costs the business case is questionable but project would probably proceed	Definitely, if we had known the real costs of the project it would not have been built

8.6 Maximum Likelihood Estimation of a Power Law

Given are N real-valued data x_i with $i = 1, 2, \dots, N$, which are assumed to be independent and identically distributed (i.i.d.). Each observation is drawn from the probability density function (pdf) $p(x|\theta)$, in which θ is the unknown set of parameter values, which are to be estimated given the dataset $(x_i)_{i \in \mathbb{N}}$.

The Maximum Likelihood protocol⁸⁵ is to maximize the likelihood function⁸⁶

$$\mathcal{L}(\theta | (x_i)_{i=1}^N) \equiv \prod_{i=1}^N p(x_i | \theta) \quad (0.1)$$

Fitting a power law tail to the data assumes that $p(x|\theta) = A \cdot x^{-\alpha}$ for $x \geq x_0 > 0$.

In other words, it is assumed that the pdf has a power law tail above a minimum value x_0 . To fit the power law tail to the data satisfying $x_i \geq x_0$, the tail is normalized to unity with the normalization factor A so that the tail becomes the pdf restricted to $x \geq x_0 > 0$.

⁸⁵ Earliest reference is Fisher (1922)

⁸⁶ For i.i.d. data, the likelihood function factors into a product of individual likelihood functions, the log of this product is the sum of the individual logs following the logarithm rule that $\log(ab) = \log a + \log b$. If we have a parameterized family of pdfs for $x \mapsto f(x|\theta)$ with parameter values $\theta \mapsto f(x|\theta)$ then the likelihood function is $\mathcal{L}(\theta|x) = f(x|\theta)$. In other words, when $f(x|\theta)$ is a function of x with given parameter θ then it is a probability function; when viewed as a function of θ given a set of observations x it is a likelihood function. The maximum likelihood protocol is based on equivalence $\mathcal{L}(\theta|x) \equiv p(x|\theta)$, in other words, the likelihood of parameter values θ occurring given the observed outcomes x is equal to the probability of the observed outcomes x given parameter values θ . Fisher showed that if $f(x)$ is the pdf with unit area, then $p = f \mathbb{I}_x$ or the area under the pdf in this range is the chance of an observation falling within the range $\mathbb{I}_x$. Thus, $\int x^p f = \int x^p p$ with $\log P = \int \log p$ for the pdf, $\log P = \int \log f$ for the likelihood and therefore $\log P = \int \log f \mathbb{I}_x$.

$$1 = A \int_{x_0}^{\infty} x^{-a} dx = A \left[\frac{x^{-a+1}}{-a+1} \right]_{x_0}^{\infty} = A \frac{x_0^{1-a}}{a-1} \quad (0.2)$$

Solving for A gives

$$A = \frac{a-1}{x_0^{1-a}} \quad (0.3)$$

Putting the normalization factor A back into the pdf $Ax^{-a} = p(x|\theta)$ for $x \geq x_0 > 0$ gives

$$p(x|\alpha) = \frac{\alpha-1}{x_0^{1-\alpha}} \cdot x^{-\alpha} = \frac{(\alpha-1)x^{-\alpha}}{x_0^{1-\alpha}} = \frac{(\alpha-1)x^{-\alpha}}{x_0^{-\alpha} x_0} = \frac{\alpha-1}{x_0} \left(\frac{x}{x_0} \right)^{-\alpha} \quad (0.4)$$

This can be rewritten as $p(x|\alpha) = (\alpha-1) \cdot \frac{x_0^{\alpha-1}}{x^\alpha}$. The power law pdf differs from

the Pareto distribution pdf. The pdf of Pareto distributions is $p(x|\alpha) = \alpha \cdot \frac{x_0^\alpha}{x^{\alpha+1}}$.

Comparing both formulas shows that they differ by 1 in their exponents. The tail exponent of a power law distribution = 1 + of the tail exponent of the Pareto distribution.

Putting the power law pdf into the equivalence for the likelihood⁸⁷ gives for

$$x \geq x_0 > 0$$

⁸⁷ The pdf gives the probability of the data x given parameter α , i.e., $p(x|\alpha)$. The likelihood is needed to fit a power law tail to the data. The likelihood is the probability of the parameter α given the data x , i.e., $p(\alpha|x)$. Bayes law states that $P(\alpha|x) = P(x|\alpha) \frac{P(\alpha)}{P(x)}$. The prior

probability of x is fixed because the data is given. It is assumed that the prior probability of the parameter $P(\alpha)$ is uniform, i.e. independent of α , thus $P(\alpha|x) \propto P(x|\alpha)$.

$$\mathcal{L}(\alpha | (x_i)_{1..N}) \equiv \left(\frac{\alpha - 1}{x_0} \right)^N \cdot \prod_{i=1}^N \left(\frac{x_i}{x_0} \right)^{-\alpha} \quad (0.5)$$

It is easier to proceed with the log-likelihood function. Log-transformation is a monotone transformation because the logarithm is monotonically increasing, in other words, $\log \mathcal{L}$ achieves the maximum at the same point as $\mathcal{L}$:

$$\ell(\alpha | (x_i)_{1..N}) \equiv \ln \mathcal{L}(\alpha | (x_i)_{1..N}) \quad (0.6)$$

Taking the $pdf p(x | \alpha) = \frac{\alpha - 1}{x_0} \left(\frac{x}{x_0} \right)^{-\alpha}$ resulted in the likelihood

$$\mathcal{L}(\alpha | x) \equiv p(\alpha | x) = \prod_{i=1}^N \left(\frac{\alpha - 1}{x_0} \right) \left(\frac{x_i}{x_0} \right)^{-\alpha} \quad \text{the logarithm of the likelihood function}$$

gives $\ell = \ln \mathcal{L}(\alpha | x) \equiv \ln \prod_{i=1}^N \left(\frac{\alpha - 1}{x_0} \right) \left(\frac{x_i}{x_0} \right)^{-\alpha}$. Because of the factor rule of

logarithms, i.e., $\ln(ab) = \ln a + \ln b$, the log-likelihood simplifies to

$$\ell = \sum_{i=1}^N \left[\ln \left(\frac{a - 1}{x_0} \right) + \ln \left(\frac{x_i}{x_0} \right)^{-a} \right] \quad \text{and further to} \quad \ell = \sum_{i=1}^N \left[\ln(a - 1) - \ln x_0 - a \ln \left(\frac{x_i}{x_0} \right) \right].$$

The first two summands are constants, which leads to

$$\ell = N \cdot \ln(a - 1) - N \cdot \ln x_0 - a \cdot \sum_{i=1}^N \ln \left(\frac{x_i}{x_0} \right) \quad (0.7)$$

Finding the maximum of the log-likelihood function ℓ with regards to a

$$\frac{\partial \ell}{\partial a} = 0 = \frac{N}{(a - 1)} - 0 - a^0 \cdot \sum_{i=1}^N \ln \left(\frac{x_i}{x_0} \right) = \frac{N}{(a - 1)} - \sum_{i=1}^N \ln \left(\frac{x_i}{x_0} \right) \quad (0.8)$$

Isolating a on the left hand side of the equation

$$\begin{aligned}
0 &= \frac{N}{(a-1)} - \sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) \\
\sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) &= \frac{N}{(a-1)} \\
(a-1) \cdot \sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) &= N \\
a \cdot \sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) - \sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) &= N \\
a \cdot \sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) &= N + \sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) \\
a &= \frac{N}{\sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right)} + \frac{\sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right)}{\sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right)}
\end{aligned}$$

gives the estimator

$$\hat{a} = \frac{N}{\sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right)} + 1 \quad (0.9)$$

which can be rewritten to

$$\hat{a} = 1 + N \left[\sum_{i=1}^N \ln\left(\frac{x_i}{x_0}\right) \right]^{-1} \quad (0.10)$$

And which for $x_0 = 1$ simplifies to $\hat{a} = \frac{N}{\sum_{i=1}^N \ln x_i} + 1 = 1 + N \left[\sum_{i=1}^N \ln x_i \right]^{-1}$.

In some empirical datasets⁸⁸ the end of the tail deviates from power law. In

other words in some empirical settings the power law $p(x|\alpha) = A \cdot x^{-\alpha}$ is only

⁸⁸ For a discussion and empirical examples from Finance and fitting the magnitude of earthquakes see Sornette et al. (2000), Sornette and Ouillon (2012).

supported for data $(x_i)_{i=1}^N$ that fall within an interval $x_{sup} \geq x_i \geq x_{inf}$. To fit the supported part of the power law tail to the data requires a different normalization A so that the pdf between x_{sup} and x_{inf} is unitary. Empirical fits to the data should always be made using the whole tail unless a theoretical reason exists that defines x_{sup} and x_{inf} , as it is for example the case of earthquakes. For this case

$$1 = \int_{x_{inf}}^{x_{sup}} p(x | \alpha) dx = A \cdot \int_{x_{inf}}^{x_{sup}} x^{-\alpha} dx \quad (0.11)$$

The antiderivative of x^{-a} is $\frac{x^{1-a}}{1-a}$ and the area under the definite integral is calculated $\int_{x_{inf}}^{x_{sup}} f(x) dx = F(x_{sup}) - F(x_{inf})$, where $F(x)$ is the antiderivative of $f(x)$, which gives

$$1 = A \cdot \left(\frac{x_{sup}^{1-a}}{1-a} - \frac{x_{inf}^{1-a}}{1-a} \right)$$

$$1 = \frac{A}{1-a} \cdot (x_{sup}^{1-a} - x_{inf}^{1-a})$$

Reordering the minuend and subtrahend and solving for A

$$1 = \frac{-A}{1-a} (x_{inf}^{1-a} - x_{sup}^{1-a})$$

$$1-a = -A (x_{inf}^{1-a} - x_{sup}^{1-a})$$

$$A (x_{inf}^{1-a} - x_{sup}^{1-a}) = a - 1$$

$$A = \frac{a-1}{x_{inf}^{1-a} - x_{sup}^{1-a}} \quad (0.12)$$

Again, the likelihood is equivalent to $\mathcal{L}(\theta | (x_i)_{i=1}^N) \equiv p(\theta | (x_i)_{i=1}^N) = \prod_{i=1}^N p(x_i | \theta)$ and

$p(x_i | \theta) = A \cdot x_i^{-\alpha}$ for $x_{sup} \geq x_i \geq x_{inf}$, which gives

$$\mathcal{L}(\alpha | (x_i)_{i=1 \dots N}) \equiv \left(\frac{\alpha - 1}{x_{inf}^{1-\alpha} - x_{sup}^{1-\alpha}} \right)^N \cdot \prod_{i=1}^N x_i^{-\alpha} \quad (0.13)$$

Taking the log-likelihood simplifies the first factor according to the logarithm laws $\ln(a/b) = \ln a - \ln b$, $\ln a^b = b \ln a$, and transforming products into sums according to the rule $\ln a \cdot b = \ln a + \ln b$ results in

$$\ell = \ln(\mathcal{L}) = N \cdot \ln(\alpha - 1) - N \cdot \ln(x_{inf}^{1-\alpha} - x_{sup}^{1-\alpha}) - \alpha \cdot \sum_{i=1}^N \ln x_i \quad (0.14)$$

where N is the sample size of observations, which were drawn from the definite region between x_{sup} and x_{inf} . The next step is to maximize the log-likelihood

with regards to alpha $\frac{\partial \ell}{\partial \alpha} = 0$, which after some calculus results in

$$\hat{\alpha} = 1 + \left[\frac{1}{N} \sum_{i=1}^N \ln x_i - \frac{x_{inf}^{\alpha-1} - x_{sup}^{\alpha-1}}{x_{inf}^{\alpha-1} \ln x_{inf} - x_{sup}^{\alpha-1} \ln x_{sup}} \right]^{-1} \quad (0.15)$$

Expectancy Value and Confidence Interval of the Expectancy Value

If $\hat{q}$ is the maximum likelihood estimator of q then $\hat{q} \xrightarrow{D} N\left(q, \frac{1}{I(q)}\right)$, where $I(q)$

is the Fisher Information⁸⁹. Fisher Information, sometimes also simply called information, is defined as the expectation of the second derivative of the loglikelihood, i.e.,

⁸⁹ See Fisher (1922)

$$I(q) = \sum \left(- \frac{\partial^2 l}{\partial q^2} \right) \quad (0.16)$$

Earlier the likelihood was defined as $\ell = N \cdot \ln(a - 1) - N \cdot \ln x_0 - a \cdot \sum_{i=1}^N \ln \left(\frac{x_i}{x_0} \right)$ for

$x \geq x_0 > 0$. The first derivative of the likelihood was

$$\frac{\partial \ell}{\partial a} = \frac{N}{(a - 1)} - \sum_{i=1}^N \ln \left(\frac{x_i}{x_0} \right)$$

The second derivative simply is

$$\frac{\partial^2 \ell}{\partial a^2} = - \frac{N}{(a - 1)^2}$$

Thus the information of a is

$$I(a) = \frac{N}{(a - 1)^2} \quad (0.17)$$

And the empirically observed Fisher information is

$$\frac{N}{(\hat{a} - 1)^2} \quad (0.18)$$

Observed Fisher Information is a consistent estimator (because of the uniform

law of large numbers and the Slutsky theorem⁹⁰) in probability $I(\hat{a}) \xrightarrow{P} I(a)$.

Thus

⁹⁰ The theorem states that a condition converges in probability to a constant (Slutsky 1925).

$$\hat{\alpha} \sim N\left(\alpha, \frac{(\alpha-1)^2}{n}\right) \quad (0.19)$$

The invariance property of maximum likelihood estimator states that if $\hat{q}$ is the estimator of q then $f(q)$ is the maximum likelihood estimator of $f(\hat{q})$. Thus a maximum likelihood estimate is invariant under functional transformations.

The expectancy value is defined as $E = \int x \cdot p(x|\theta)$. Using the maximum likelihood estimator from above for x^3 $x_0 > 0$ gives

$$E = \int x \frac{a-1}{x_0} \left(\frac{x}{x_0}\right)^{-a} \partial x \quad (0.20)$$

which simply is

$$E(a) = \frac{a-1}{a-2} x_0 \quad (0.21)$$

Or, in more general terms, the m^{th} moment is

$$m = \frac{a-1}{a-1-m} x_0^m \quad (0.22)$$

The estimation of confidence intervals around the expectancy value $E(a)$ is also straight forward. The maximum likelihood estimator for the expectancy value is $\hat{m} = \frac{\hat{a}-1}{\hat{a}-2} x_0$ thus $f(\hat{a}) = \frac{\hat{a}-1}{\hat{a}-2} x_0$. Taylor expansion gives

$$f(\hat{a}) = f(a) + \frac{f'(a)}{1!}(\hat{a} - a) + \frac{f''(a)}{2!}(\hat{a} - a)^2 + \frac{f^{(3)}(a)}{3!}(\hat{a} - a)^3 + \dots$$

In the last formula $\frac{f^{\epsilon}(a)}{1!}(\hat{a} - a)^{+1/4}$ is nearly $1/n$. In a large n dataset the error is sufficiently small, so that it can be assumed

$$E(f(\hat{\alpha})) \approx f(\alpha) \quad (0.23)$$

The variance is defined as

$$V(f(\alpha)) \approx f'(\alpha)^2 V(\hat{\alpha}) \quad (0.24)$$

With

$$V(\hat{a}) = \frac{1}{I(\hat{a})} = \frac{(\hat{a} - 1)^2}{n} \quad (0.25)$$

The first derivative of $f(a) = \frac{a-1}{a-2}x_0$ is $f^{\epsilon}(a) = -\frac{x_0}{(a-2)^2}$. The square of the

first derivative is simply $f^{\epsilon}(a)^2 = \frac{x_0^2}{(a-2)^4}$. This gives the variance as

$$V(f(\hat{a})) = \frac{(\hat{a} - 1)^2}{n} \frac{x_0^2}{(\hat{a} - 2)^4} \quad (0.26)$$

The standard deviation is simply the square root of the variance

$$SD = \sqrt{V(f(\hat{a}))} = \sqrt{\frac{(\hat{a} - 1)^2}{n} \frac{x_0^2}{(\hat{a} - 2)^4}} \quad (0.27)$$

Which sets the 95% confidence interval for the expectancy value for large n and

$x^3 \ x_0 > 0$ as

$$CI_{95} = \frac{a - 1}{a - 2} \times x_0 \pm 1.96 \times \sqrt{\frac{(\hat{a} - 1)^2}{n} \frac{x_0^2}{(\hat{a} - 2)^4}} \quad (0.28)$$

8.7 Testing Power Laws in Disaggregating Observations

Table 34 Goodness of Fit Test for Different Subsamples of Cost Performance

Variable	Subsample	Alpha	x0	KS-distance	p-value
Data Source	Desk study	3.04	1.39	0.057	0.999
	Direct participants	2.47	1.26	0.061	0.999
	E300 2003-2010 manual entry of pdf forms	2.46	2.00	0.039	1.000
	E300 2003-2010 scan of pdf forms	1.78	1.11	0.080	0.720
	OMB reporting	2.15	2.00	0.054	1.000
Project type	Communication technology	2.97	1.54	0.064	0.988
	IT Architecture	2.11	1.57	0.103	0.826
	IT Infrastructure	2.19	1.23	0.054	0.994
	Other	2.64	1.02	0.072	0.875
	System implementation	2.27	2.00	0.040	1.000
Country	Denmark	3.00	1.59	0.124	0.085
	Germany	1.88	3.46	0.049	1.000
	Netherlands	7.41	1.65	0.145	0.296
	Other	3.40	2.01	0.091	0.945
	UK	5.91	1.26	0.132	0.348
	United Kingdom	3.50	1.01	0.059	0.742
Year of decision to build	United States	2.26	2.00	0.041	1.000
	2004	2.55	1.70	0.106	0.534
	2005	2.66	1.11	0.111	0.201
	2006	2.93	1.06	0.089	0.653
	2007	3.63	1.00	0.058	0.767
	2008	1.98	4.17	0.052	1.000
	2009	2.71	1.02	0.063	0.011
	2010	2.25	2.00	0.036	1.000
	2011	1.82	6.72	0.070	0.994
	2012	2.36	1.03	0.093	0.104

Table 35 Goodness of Fit Test for Different Subsamples of Schedule Performance

Variable	Subsample	Alpha	x0	KS-distance	p-value
Country	Denmark	3.63	1.19	0.057	0.978
	Germany	3.22	1.28	0.032	1.000
	Netherlands	6.92	2.15	0.125	0.508
	Other	3.42	1.20	0.105	0.065
	UK	2.13	1.00	0.132	0.240
	United Kingdom	4.91	1.61	0.092	0.849
	United States	2.64	1.05	0.119	0.314
Project type	Communication technology	NULL			
	IT Architecture	2.69	1.06	0.150	0.648
	IT Infrastructure	3.42	1.50	0.113	0.829
	Other	1.97	1.07	0.204	0.259
	System implementation	3.26	1.09	0.023	1.000
Year of decision to build	2004	5.64	1.29	0.069	0.982
	2005	2.31	1.04	0.169	0.182
	2006	8.35	2.06	0.123	0.717
	2007	2.94	1.23	0.119	0.518
	2008	3.38	1.03	0.111	0.232
	2009	2.81	1.35	0.047	1.000
	2010	3.15	1.08	0.061	0.077
	2011	3.76	1.08	0.049	0.751
	2012	2.74	1.31	0.117	0.659
Regions	Other	3.53	1.29	0.058	0.951
	Western Europe	3.24	1.07	0.031	0.997

345

[illegible]

8.9 Principal Component Analysis of the Explanatory Variables

Principal component analysis was done using regularized iterative PCA algorithm implemented in the package `missMDA` to impute the missing values in the questionnaire.

First, the numbers of dimensions for the principal component analysis were estimated by leave-one-out cross-validation. The approach removes each cell of the data matrix in turn and predicts the cell with a range of a number of dimensions. The number of dimensions that leads to the smallest mean square error of prediction is then chosen to impute the missing values. In this case the mean square error of prediction is smallest for using one component (cf. table 36).

Table 36 Determining Dimensions for Imputation of Missing Values

Number of dimensions	0	1	2	3	4	5
Means square error of prediction	4.090051	3.437570	4.724154	8.023370	19.975326	136.698594

Second, the missing values are imputed using regularized Principal Component Analysis. Then a Principal Component Analysis is conducted using the `PCA` function of the `FactoMineR` package.

Table 37 Eigenvalues and Percentage of Variance Extracted by Principal Component Analysis

Component	Eigenvalue	Percentage of variance	Cumulative percentage of variance
1	18.25856802	43.4727810	43.47278
2	2.73902770	6.5214945	49.99428
3	1.92119590	4.5742759	54.56855
4	1.42376880	3.3899257	57.95848
5	1.31725226	3.1363149	61.09479
6	1.12878187	2.6875759	63.78237

7	1.07717859	2.5647109	66.34708
8	1.00079366	2.3828420	68.72992
9	0.97973558	2.3327038	71.06262
10	0.87589497	2.0854642	73.14809
11	0.79794009	1.8998573	75.04795
12	0.75088891	1.7878307	76.83578
13	0.73084030	1.7400960	78.57587
14	0.65732908	1.5650692	80.14094
15	0.62559242	1.4895058	81.63045
16	0.60128854	1.4316394	83.06209
17	0.59162281	1.4086257	84.47071
18	0.54752981	1.3036424	85.77436
19	0.52620952	1.2528798	87.02724
20	0.50237744	1.1961368	88.22337
21	0.47345508	1.1272740	89.35065
22	0.41751333	0.9940794	90.34473
23	0.38958390	0.9275807	91.27231
24	0.35501574	0.8452756	92.11758
25	0.33769041	0.8040248	92.92161
26	0.30963906	0.7372359	93.65884
27	0.29845186	0.7105997	94.36944
28	0.29597241	0.7046962	95.07414
29	0.27503106	0.6548359	95.72897
30	0.24358875	0.5799732	96.30895
31	0.23195119	0.5522647	96.86121
32	0.21188925	0.5044982	97.36571
33	0.18240168	0.4342897	97.8
34	0.15670673	0.3731113	98.17311
35	0.13393412	0.3188908	98.492
36	0.12709535	0.3026080	98.79461
37	0.11182897	0.2662594	99.06087
38	0.09943078	0.2367400	99.29761
39	0.08540360	0.2033419	99.50095
40	0.07729451	0.1840345	99.68499
41	0.06899299	0.1642690	99.84925
42	0.06331294	0.1507451	100

The screeplot (cf. figure 48) suggests that only one component should be extracted from the data. The factor loadings for the first five factors (cf. figure 49) shows that only seven variables do not load on the first factor above the critical value of $|0.4|$. These seven variables are

- Environmental influence of the project (H2.2.2)
- Organizational resistance encountered (H2.2.3)
- Perceived uniqueness of the project (H2.2.7)

- Incremental approach of the project (H2.3.7)
- Dummy whether benefits were measured (H4.1.2)
- Years of experience of the most senior project manager (H4.2.5)
- Effectiveness of the team communication (H4.5.2)

The results show that no meaningful factor structure emerges. Moreover, the extracted components have no explanatory power with regards to cost overruns. If put in a linear model the first component holds some explanatory power (p-value of coefficient = 0.04, model: $F = 0.947$ on 5 and 3666 DF, model p-value = 0.449, adjusted $R^2 < 0.001$). When controlling for the organization as a first level this effect becomes non-significant for the first component ($t = 0.554$, $p = 0.58$).

Figure 48 Screeplot of the Principal Component Analysis

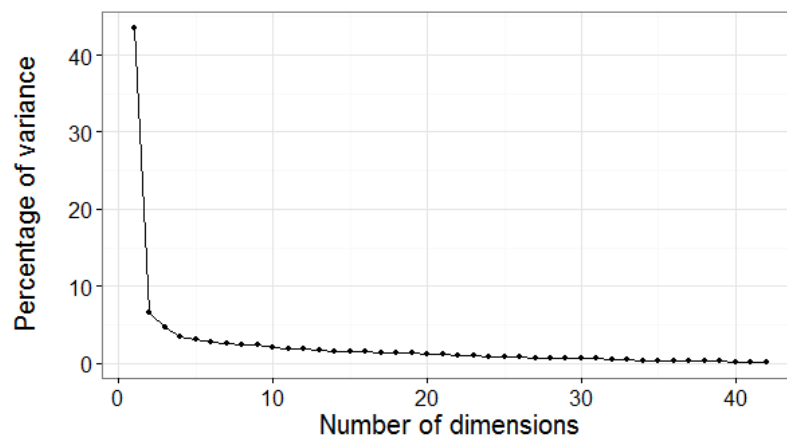
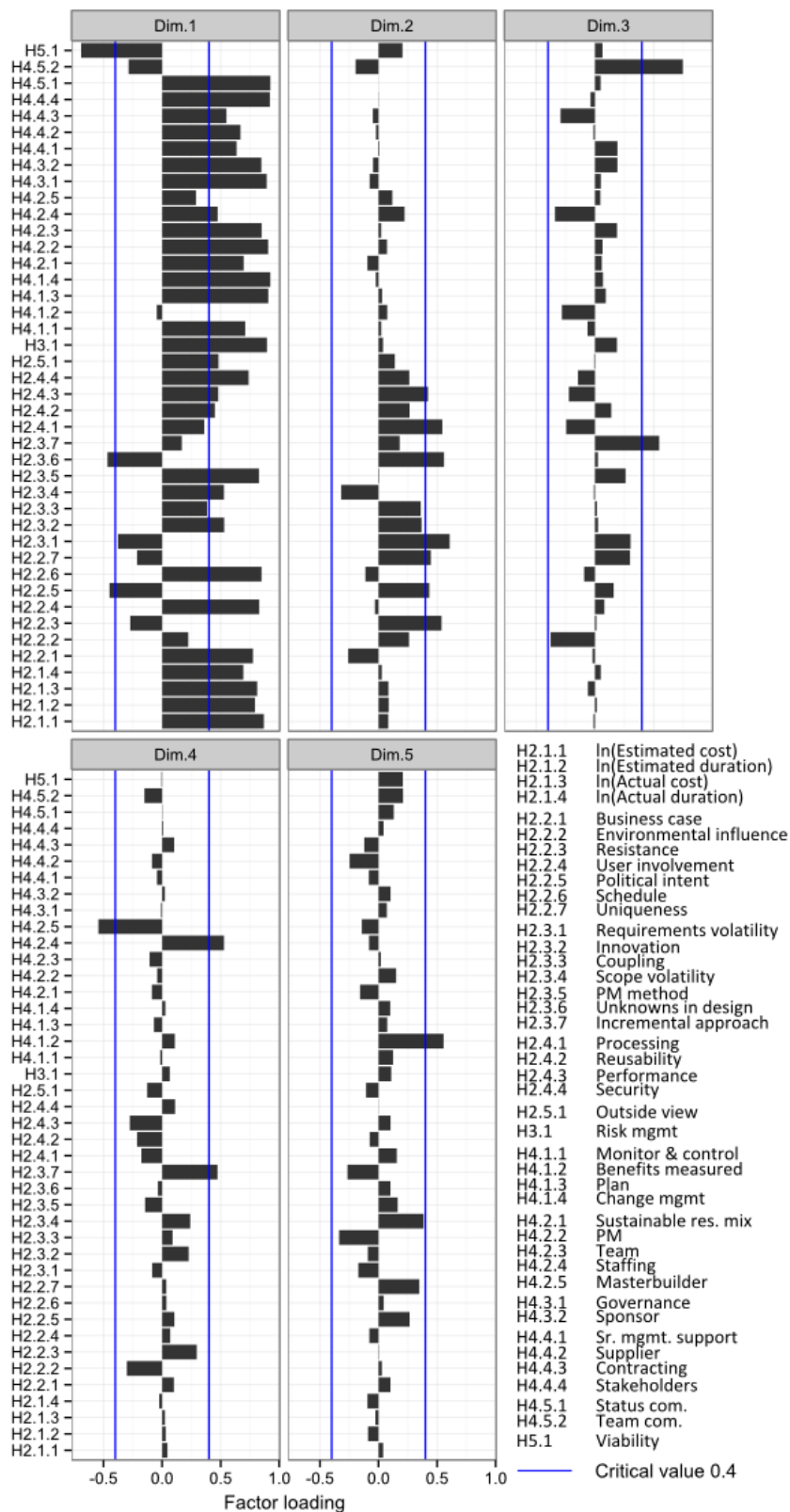


Figure 49 Factor Loadings for the First Five Dimensions Extracted by the Principal Component Analysis



8.10 Determinants of Effort Outliers IT Projects

This section follows the same structure of analyses as the previous two sections. The section will first test the determinants of effort overruns in IT projects. It will then test which determinants influence the thickness of the upper power law tail of the effort performance distribution. Finally, multilevel models will be fitted to explain both effort overruns and effort outliers while controlling for the differences in effort performance between organizations.

8.10.1 *Determinants of Effort Overruns in IT Projects*

Table 38 shows the results of the non-parametric test for a statistically significant shift in the location of the effort overrun distribution. As before the sample was split at the median of the potential determinant to test whether the effort performance significantly differs between the two subsamples. None of the hypotheses could be rejected with high statistical significance after controlling for the false discovery rate of multiple independent statistical tests. However, three hypotheses were initially rejected and might offer some exploratory value:

H4.1.2: Whether projects measured benefits or not does not vary with IT project effort overruns.

H4.2.2: Perceptions of the competence of the project manager(s) do not vary with IT project effort overruns.

H4.2.5: Numbers of years of experience of the most senior project manager do not vary with IT project effort overruns.

Other variables that might potentially have explanatory value are influence over the project environment, realistic and reliable schedule, and whether reusability was a requirement.

Table 38 Non-Parametric Wilcoxon Test for Location Shift in Effort Overrun Distribution

Abbreviated description	Median	IQR	N1	N2	M1	M2	IQR1	IQR2	W	p-value	adj. p*-value
H2.1.1 Estimated cost	1.2	8.4	89	69	1.208	1.337	0.994	1.206	2958	0.694	0.884
H2.1.2 Estimated duration	181	203	34	13	1.018	1.360	0.504	0.515	188	0.439	0.695
H2.1.3 Actual cost	0.8	6.7	83	32	1.494	1.431	1.342	0.959	1465	0.396	0.695
H2.1.4 Actual duration	1096	1461	5	107	2.002	1.064	1.552	0.590	352	0.236	0.550
H2.2.1 Clear goals, business case	8.0	2.3	18	44	1.348	1.056	0.880	0.408	496	0.125	0.529
H2.2.2 Environmental influence	8.0	3.0	34	28	1.034	1.254	0.437	0.820	355	0.087	0.529
H2.2.3 No resistance	3.0	4.0	34	28	1.096	1.106	1.006	0.343	519	0.548	0.746
H2.2.4 User involvement	8.0	2.0	16	46	1.000	1.148	0.644	0.510	300	0.274	0.554
H2.2.5 No political intent	3.0	3.0	17	18	1.096	1.295	0.562	0.729	139	0.655	0.860
H2.2.6 Realistic, reliable schedule	8.0	3.0	38	24	1.026	1.248	0.358	0.998	334	0.078	0.529
H2.2.7 Uniqueness	5.0	4.0	25	36	1.064	1.096	0.562	0.429	501	0.463	0.695
H2.3.1 Stable requirements	5.0	4.0	27	35	1.254	1.015	0.942	0.474	574	0.153	0.529
H2.3.2 Innovation	7.0	3.0	24	37	1.259	1.041	0.983	0.414	531	0.201	0.529
H2.3.3 Decoupled deliverables	7.0	3.0	29	33	1.064	1.132	0.597	0.494	481	0.983	0.994
H2.3.4 Minimal & stable scope	5.0	4.0	20	8	1.201	1.114	0.837	0.676	78	0.939	0.994
H2.3.5 Effective PM method	8.0	2.0	16	45	1.348	1.064	0.910	0.490	446	0.161	0.529
H2.3.6 Few unknowns in the design	5.0	4.0	29	33	1.132	1.048	0.579	0.461	493	0.849	0.964
H2.3.7 Incremental approach	6.0	6.0	36	25	1.056	1.137	0.378	0.895	372	0.256	0.554
H2.4.1 Low processing requirements	7.0	3.0	28	33	1.056	1.096	0.422	0.515	440	0.756	0.923
H2.4.2 Low reusability requirements	8.0	4.0	16	45	1.322	1.041	1.320	0.366	466	0.085	0.529
H2.4.3 Low performance requirements	8.0	3.0	22	39	1.148	1.041	0.724	0.433	472	0.523	0.746
H2.4.4 Low security requirements	8.0	4.0	38	23	1.044	1.254	0.355	0.739	364	0.277	0.554
H2.5.1 Outside view	5.0	4.8	7	27	1.360	1.137	1.021	0.538	101	0.798	0.931
H3.1 Effective risk management	8.0	2.0	14	47	1.301	1.064	0.930	0.501	401	0.220	0.543
H4.1.1 Monitor & control	8.0	2.0	35	26	1.041	1.296	0.339	0.888	361	0.173	0.529
H4.1.2 Benefits measured	0.0	0.0	83	75	1.066	1.461	0.736	1.556	2444	0.020*	0.295
H4.1.3 Detailed, up-to-date plan	8.0	2.0	10	17	1.199	1.137	0.652	0.917	83	0.940	0.994
H4.1.4 Effective change management	8.0	1.3	8	19	1.313	1.096	0.516	1.300	75	0.979	0.994
H4.2.1 Sustainable resource mix	8.0	2.0	3	22	1.801	1.114	3.653	0.550	53	0.103	0.529
H4.2.2 Competent project managers	9.0	2.0	12	48	1.358	1.026	2.631	0.505	415	0.019*	0.295
H4.2.3 Qualified & motivated team	9.0	2.0	34	28	1.072	1.116	0.399	0.645	418	0.416	0.695
H4.2.4 Suitably staffed	5.0	5.0	40	21	1.096	1.338	0.369	0.895	357	0.343	0.654
H4.2.5 Years of experience PM	10.0	7.0	13	17	1.562	1.041	2.315	0.266	166	0.021*	0.295
H4.3.1 Effective governance	8.0	2.0	13	49	1.041	1.132	0.336	0.579	284	0.551	0.746
H4.3.2 Effective sponsor	8.0	2.0	20	42	1.199	1.080	1.062	0.512	471	0.451	0.695
H4.4.1 Senior management support	8.5	3.0	31	31	1.048	1.137	0.399	0.701	386	0.186	0.529
H4.4.2 Well performing supplier	7.0	1.0	18	42	1.175	1.072	0.438	0.510	377	0.994	0.994
H4.4.3 Robust contracting	9.0	3.0	8	52	1.228	1.080	0.423	0.516	222	0.769	0.923
H4.4.4 Aligned stakeholders	9.0	3.0	8	52	1.348	1.056	0.252	0.547	268	0.195	0.529
H4.5.1 Transparent status comm.	8.0	2.0	15	12	1.360	1.048	1.410	0.569	119	0.164	0.529
H4.5.2 Good team communication	7.0	3.0	20	42	1.313	1.026	0.888	0.350	526	0.112	0.529
H5.1 If true costs were known...	1.0	2.0	19	15	1.336	1.031	0.629	0.844	166	0.435	0.695

Note: M1 is the median of the subsample where variable > median, M2 is the median of the subsample where the variable ≤ median, N1, N2 and IQR1, IQR2 are subsample sizes and IQRs of the subsamples respectively.

W is the non-parametric Wilcoxon test statistic, the adjusted p-value uses the Benjamini-

Hochberg procedure, significance levels are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1, . < 0.2

8.10.2 Determinants of the Thickness of the Upper Power Law Tail

As before, the sample was split along the median of each potential determinant. For each subsample the upper tail of the distribution for all effort overruns was fitted to a power law, i.e., the actual/estimated effort with $x_0 > 1$. A likelihood ratio test was then used to compare both fits and determine, whether the model with the two subsamples explains the distribution significantly better than a model with the pooled sample. Table 39 shows the results. Before penalizing the results for conducting multiple independent statistical tests, the likelihood ratio test is able to reject four null hypotheses, after adjusting the p*-values none of the hypotheses can be rejected. The four, initially rejected, hypotheses were:

H2.2.1: The power law exponent does not differ between projects with different perceptions of the clarity of the business case, goals, and success criteria.

H2.2.3: The power law exponent does not differ between projects with different perceptions of organizational resistance encountered by the project.

H4.1.1: The power law exponent does not differ between projects with different perceptions of the effectiveness of the monitoring and controlling.

H4.2.2: The power law exponent does not differ between projects with different perceptions of the competence of the project manager.

It is important to note that maximum likelihood estimators suffer from small sample bias. Splitting the sample into two subsamples reduced the power of the test. Penalizing the p-values for the false discovery rate further reduces the statistical power. Overall, the test will therefore err on the conservative side and might not reject null hypotheses that are not true but are simply not rejected because the data on effort performance in the upper tail is insufficient. Thus it is not surprising given the small sample size and the large number of hypotheses that none of the hypotheses rise beyond the reasonable doubt of being a false positive. However further explanatory analyses and additional data collection should not be discouraged.

Table 39 Test for Thickness of the Upper Tail of the Effort Performance Distribution

Abbreviated description	Median	IQR	N1	N2	Alpha1	Alpha2	LR	p-value	adj. p*-value
H2.1.1 Estimated cost	1.2	8.4	63	44	2.612	2.328	0.822	0.663	0.998
H2.1.2 Estimated duration	181	203	20	8	2.735	2.444	0.168	0.919	0.998
H2.1.3 Actual cost	0.8	6.7	64	21	2.438	2.379	0.023	0.989	0.998
H2.1.4 Actual duration	1096	1461	4	63	2.268	2.681	0.274	0.872	0.998
H2.2.1 Clear goals, business case	8.0	2.3	12	28	1.995	3.850	8.909	0.012 *	0.458
H2.2.2 Environmental influence	8.0	3.0	19	21	2.564	3.158	0.908	0.635	0.998
H2.2.3 No resistance	3.0	4.0	22	18	2.322	4.432	7.509	0.023 *	0.458
H2.2.4 User involvement	8.0	2.0	7	33	2.557	2.898	0.206	0.902	0.998
H2.2.5 No political intent	3.0	3.0	11	12	2.960	2.082	1.645	0.439	0.998
H2.2.6 Realistic, reliable schedule	8.0	3.0	23	17	2.982	2.654	0.281	0.869	0.998
H2.2.7 Uniqueness	5.0	4.0	15	24	2.230	3.654	4.950	0.084 ..	0.507
H2.3.1 Stable requirements	5.0	4.0	21	19	3.204	2.538	1.133	0.568	0.998
H2.3.2 Innovation	7.0	3.0	19	21	2.590	3.114	0.710	0.701	0.998
H2.3.3 Decoupled deliverables	7.0	3.0	18	22	2.405	3.425	2.598	0.273	0.955
H2.3.4 Minimal & stable scope	5.0	4.0	12	7	2.312	3.747	1.941	0.379	0.998
H2.3.5 Effective PM method	8.0	2.0	12	28	2.445	3.063	0.967	0.617	0.998
H2.3.6 Few unknowns in the design	5.0	4.0	20	20	3.366	2.490	1.861	0.394	0.998
H2.3.7 Incremental approach	6.0	6.0	23	17	3.669	2.282	4.675	0.097 ..	0.507
H2.4.1 Low processing requirements	7.0	3.0	18	21	2.716	2.953	0.143	0.931	0.998
H2.4.2 Low reusability requirements	8.0	4.0	12	27	2.372	3.161	1.568	0.457	0.998
H2.4.3 Low performance requirements	8.0	3.0	15	24	2.482	3.158	1.163	0.559	0.998

H2.4.4 Low security requirements	8.0	4.0	24	15	3.373	2.348	2.655	0.265	0.955
H2.5.1 Outside view	5.0	4.8	4	19	2.227	2.414	0.056	0.972	0.998
H3.1 Effective risk management	8.0	2.0	11	29	2.076	3.488	5.361	0.069..	0.507
H4.1.1 Monitor & control	8.0	2.0	23	17	3.867	2.227	6.266	0.044*	0.458
H4.1.2 Benefits measured	0.0	0.0	48	59	2.558	2.426	0.174	0.917	0.998
H4.1.3 Detailed, up-to-date plan	8.0	2.0	8	11	3.109	2.392	0.669	0.716	0.998
H4.1.4 Effective change management	8.0	1.3	6	13	3.422	2.410	0.968	0.616	0.998
H4.2.1 Sustainable resource mix	8.0	2.0	3	15	1.982	2.878	1.004	0.605	0.998
H4.2.2 Competent project managers	9.0	2.0	11	28	2.018	3.549	6.353	0.042*	0.458
H4.2.3 Qualified & motivated team	9.0	2.0	22	18	3.347	2.439	2.090	0.352	0.998
H4.2.4 Suitably staffed	5.0	5.0	27	13	3.504	2.171	4.702	0.095..	0.507
H4.2.5 Years of experience PM	10.0	7.0	11	9	1.921	3.662	4.318	0.115.	0.539
H4.3.1 Effective governance	8.0	2.0	7	33	2.875	2.818	0.005	0.998	0.998
H4.3.2 Effective sponsor	8.0	2.0	13	27	2.483	3.059	0.850	0.654	0.998
H4.4.1 Senior management support	8.5	3.0	18	22	2.643	3.013	0.359	0.836	0.998
H4.4.2 Well performing supplier	7.0	1.0	11	28	3.165	2.734	0.332	0.847	0.998
H4.4.3 Robust contracting	9.0	3.0	6	33	3.195	2.784	0.184	0.912	0.998
H4.4.4 Aligned stakeholders	9.0	3.0	7	32	2.739	2.794	0.005	0.998	0.998
H4.5.1 Transparent status comm.	8.0	2.0	13	6	2.528	2.881	0.148	0.929	0.998
H4.5.2 Good team communication	7.0	3.0	15	25	2.547	3.052	0.665	0.717	0.998
H5.1 If true costs were known...	1.0	2.0	15	8	3.062	1.849	3.570	0.168.	0.705

Note: Alpha1 is the estimated power law exponent alpha of the subsample where variable > median, Alpha2 is the estimated power law exponent of the subsample where the variable ≤ median, N1 and N2 are the subsample sizes respectively.

LR is the Likelihood Ratio test statistic, the adjusted p-value uses the Benjamini-Hochberg procedure, significance levels are shown as *** < 0.001, ** < 0.01, * < 0.05, .. < 0.1, . < 0.2

8.10.3 Exploratory Multilevel Regression Models

8.10.3.1 Multilevel Model of Effort Overruns

Following the same steps as before a multilevel model was fitted to the data explaining the effort overrun. The organization was entered as the random factor, the potential determinants of the effort overrun were entered as fixed factors. The dependent variable was log-transformed to correct for outliers and lack of symmetry in the data.

Table 40 shows the initial fit based on the three determinants that the Wilcoxon test identified. The random effect has a standard deviation of zero, thus might be excluded in steps to simplify the model. None of the three determinants, whether benefits were measured, perceived competence of the project

manager(s), nor the number of years of experience the most senior project manager had were statistically significant explanatory variables.

Table 40 Explanatory Fixed Effects Model for LN(Effort Overruns)

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>p-value</i>	<i>F-value</i>	<i>p-value</i>
<i>Intercept</i>	-1.606	1.431	24.0	-1.123	0.273	-	-
<i>Benefits measured</i>	-0.011	0.372	24.0	-0.029	0.977	0.001	0.977
<i>Competent project manager(s)</i>	0.196	0.195	24.0	1.003	0.326	1.006	0.326
<i>Years of experience project manager</i>	0.040	0.055	24.0	0.716	0.481	0.513	0.481

Note: t-values calculated using Wald test; F-values calculated using Satterthwaite approximation of degrees of freedom

However, further simplification of fixed and random effects did not lead to a model with statistically significant effect sizes.

8.10.3.2 Multilevel Model of Effort Outliers

Table 41 shows the final parsimonious logistic regression model, the initial model with the four identified factors was overfitted from the beginning. In the final model only two terms remain. The random effect with an intercept of 0.4378 (std. dev. = 0.6617) and organizational resistance as the fixed effect.

Table 41 Parsimonious Explanatory Fixed Effects Model for the Odds of Effort Outliers Occurring

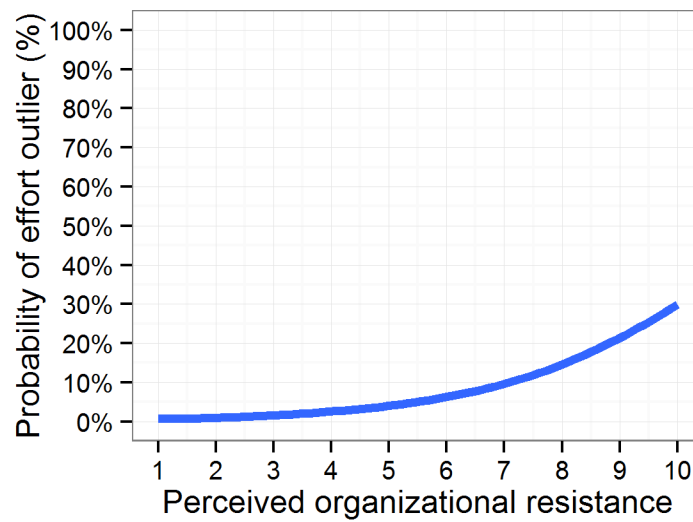
	<i>Estimate</i>	<i>Std. Error</i>	<i>z-value</i>	<i>p-value</i>	<i>F-value</i>
<i>Intercept</i>	-5.493	1.965	-2.796	0.005	-
<i>Organizational resistance</i>	0.465	0.269	1.727	0.084	2.9819

Note: t-values calculated using Wald test; F-values calculated using Satterthwaite approximation of degrees of freedom

Figure 50 shows the predicted effect strength of the organizational resistance. If a project was ranked 4 = “A few users did not see the need for the project, needed more training and communication” the probability of the project being an effort outlier is 12%. If a project was ranked 10 = “Users did not understand

the project, were highly critical of it, users did not adopt the solution” the probability of the project being an effort outlier was 30%.

Figure 50 Prediction of Average Probability of Effort Outliers Occurring



8.11 Participant Information

[On McKinsey letterhead]

[date]

[Name and address of Participant]

Dear _____:

McKinsey & Company, Inc. ("McKinsey") is undertaking a benchmarking study in IT Architecture in cooperation with the University of Oxford in order to collect historic performance data from ICT megaprojects with the purpose of developing tools for quantitative management of large-scale projects, which determine costs, schedule and benefits, identify key drivers and factors for project success and failure and assess project health and risk of new and ongoing projects ("the Study"). We are pleased that [*name of Participant*] is participating in the Study. Various other companies are interested or have already agreed to participate in the Study and are willing to provide McKinsey with certain information and data. For purposes of this agreement (the "Agreement"), all companies that have agreed in writing to participate in the Study are referred to individually as a "Participant" and collectively as the "Participants".

McKinsey and the undersigned Participant agree as follows:

1. Confidentiality

1.1 McKinsey agrees to keep confidential all information that Participant provides to McKinsey in connection with the Study ("Confidential Information"), but shall not include information which is:

- (i) at the time of disclosure in the public domain or generally available in the industry;
- (ii) or becomes publicly available other than as a result of a breach of this Agreement by McKinsey;

- (iii) already known to McKinsey, except when subject to the terms of an existing confidentiality agreement;
- (iv) independently acquired or developed by McKinsey without violating any of its obligations under this Agreement;
- (v) required to be disclosed by law or judicial process; or
- (vi) conspicuously marked by Participant as “Shareable Information”.

1.2 Except as provided in this Agreement or as required by law, McKinsey will not disclose Confidential Information to any persons other than those of its directors, officers, employees, shareholders, agents or advisors of McKinsey or the Participant and shall use it exclusively for the purpose of the Study.

1.3 Upon the request of Participant, McKinsey will return to Participant or destroy all documents and tangible Confidential Information (including all copies) provided to McKinsey by Participant pursuant to this Agreement, provided that McKinsey may retain a copy for its records subject to the obligation to maintain such copy confidential in accordance with this Agreement.

2. Development, use and disclosure of the Results and the Participant Report

2.1 McKinsey will develop and disclose to each of the Participants general results of the Study (the “Results”) in the form of a written and/or electronic report for each Participant (the “Participant Report”). Participant recognizes and confirms that McKinsey will use and rely primarily on the Confidential Information and on information available from public sources in preparing the Participant Report without having independently verified the same and does not assume responsibility for its accuracy or completeness.

2.2 In developing the Results, McKinsey will aggregate, blind and/or disguise all information to prevent that the Results attribute or identify Confidential Information to any of the Participants individually, except that each Participant will receive customized Results that may contain Confidential Information of such Participant. Where the Study requires McKinsey to collect price-related information from the Participants, McKinsey will always sufficiently aggregate such information as mentioned above and, where necessary, take additional measures to avoid dissemination among Participants of price-related information which (i) is less than 6 months old, (ii) is sliced in a way that results in a set of less than five Participants, and (iii) represents Participant data in excess of 25 percent of the aggregated information.

2.3 Participant agrees that McKinsey will provide the Results and the Participant Report solely for Participant’s information and lawful internal use and that it will not circulate, quote, disclose or otherwise reproduce for distribution outside of its organization the Results and the Participant Report without the prior written approval of McKinsey. The foregoing restriction does not prohibit Participant to paraphrase the Results and content of the Participant Report and use such paraphrases external to its organization provided it deletes all references to McKinsey there from and makes no attribution to McKinsey whatsoever in connection therewith.

3. Acknowledgement by the Participant

3.1 Participant acknowledges and agrees (i) that McKinsey, including its affiliates and the individual consultants assigned to the Study, may have served, may currently be serving, or may in the future serve any of the Participants or other companies on any matters, provided, however, that no consultant who is providing serviced on the Study and who receives Confidential Information will, for a period of one year following such consultant's services on the Study, be staffed by McKinsey to provide services on a competitively sensitive project if a Participant's Confidential Information acquired by such consultant in connection with the Study could be used to the Participant's competitive disadvantage in such other project, (ii) that the arrangements described herein do not constitute a conflict of interest on the part of McKinsey, its affiliates, or the employees or shareholders of either, and (iii) to forever waive any claims it might have at any time against McKinsey, its affiliates, and the employees or shareholders of either that such services constitute a conflict of interest.

3.2 Participant acknowledges that it remains solely responsible, and that McKinsey is not liable in any way for Participant's compliance with all applicable laws, including competition/antitrust laws, in relation to the exchange of information with other Participants in connection with the Study.

3.3 Participant further acknowledges that McKinsey (i) is providing the Participant Report "as is", without any express or implied warranty, including but not limited to warranties of merchantability or fitness for a particular purpose and (ii) shall in no event be liable to any of the Participants or any person claiming through any of the Participants, including without limitation its owners, parents, affiliates, security holders, or creditors, for any damages or losses relating to or arising out of the Participant Report or Participant's use thereof, including without limitation indirect, consequential, incidental or special damages (including without limitation lost profits), even if McKinsey knew or should have known of the possibility of the same.

4. McKinsey's rights

Participant acknowledges and agrees that McKinsey reserves the right to (i) prepare and publish reports outlining general conclusions from the Study, (ii) draw on disguised data from its internal benchmarking database and from case studies for its knowledge and client initiatives in this and other industry sectors, (iii) distribute and use the Results internally and in client work (except that no Confidential Information will be specifically attributed or identified to any of the Participants) and (iv) disclose the names of each Participant only to the other existing and prospective Participants.

5. General Provisions

5.1 The terms of this Agreement shall survive the completion or termination of the Study.

5.2 This Agreement shall be governed by and construed in accordance with the laws of Germany.

5.3 Any dispute, controversy or claim arising out of or in connection with, or relating to this Agreement, or the breach, termination or validity thereof, shall be finally settled

by arbitration. The arbitration shall be conducted in English in accordance with the rules of the Deutsche Institution für Schiedsgerichtsbarkeit e.V. (DIS) in effect at the time of the arbitration, which rules shall also apply to the appointment of arbitrators. The seat of the arbitration shall be Berlin.

If the above terms and conditions are acceptable to you, please indicate your acceptance by signing the enclosed copy of this letter and returning it to us. We look forward to your participation in this effort.

Sincerely,

McKinsey & Company, Inc.

By: _____

Title: _____

Date: _____

Accepted and Agreed:

[Insert name of Participant]

By: _____

Title: _____

Date: _____

8.12 Implications of Power Law Tails for (IT Project) Management Theory

Multiple theoretical models have been shown to produce power law tails⁹¹. This is going to briefly review those models in relation to the theoretical schools of thought found in organizational, project management, and IT project management theory.

8.12.1 *System-centric explanations*

System-centric explanations, as the literature review in chapter 2 identified, focus on the question of system design. Systems modelling has shown that power law behavior can result from systems design. Two models are commonly proposed: (self)-organized criticality and (over)-optimization.

First, *self-organized criticality* (cf. Bak et al. 1987) certain physical systems have been observed to be scale free at certain critical points in their parameter space, for example percolation models of cluster formation. Systems have been found to self-organize to a point close to the critical point. The classic example is the forest fire model. Trees grow as spatial clusters, which when lightning strikes burn down. A simple percolation model predicts that the whole system self-organizes to a structure where the size of clusters or the size of fires follows a power law. The system will oscillate to be above and below the critical point

⁹¹ For a review of different generating mechanisms see Gabaix (2009), Newman (2005), and Farmer and Geanakoplos (2004). Among the alternative explanations not discussed here are the mixture of distributions, sometimes also called a hierarchy or combination of exponentials, which chapter 4.4.5 tested as alternative hypotheses but which failed to reject the hypothesis that the tails fit power laws. Other explanations of power law tails are found in extreme value theory, which describes the behavior of extreme values over time, e.g. the economics of superstars and blockbusters.

where either trees grow back or burn down because of fire. Self-organized criticality has been suggested to explain other phenomenon such as file sizes, neuronal avalanches, and earthquakes (cf. Marković et al. 2013). In relation to project management a project could be conceptualized as a complex network of tasks, where certain nodes hold more links, e.g., dependencies between tasks.

Second explanation for power laws is the concept of *highly optimized tolerance*. The concept builds on Mandelbrot's interpretation of Zipf's law, which argued that word frequency follows power law tails because language is designed to be the most efficient system to transmit information (Mandelbrot 1953). Carlson and Doyle (1999) have argued that optimization of one design objective under uncertainty and with constraints present can lead to systems that are highly robust to known failures and, at the same time, highly fragile to unknown disruptions and design flaws (cf. Marković et al. 2013). In other words, their models tend to withstand small losses from known causes while being vulnerable to rare perturbations.

Both theoretical explanations account for outliers in the system-centric school of thought. Self-organized criticality explains outliers through amplification of random fluctuations. Highly optimized tolerance explains outliers as catastrophic system failure from rare perturbations.

8.12.2 *Event-centric explanations*

Theoretical models also offer event-centric explanations of why IT project performance has power law tails. Specifically two models exist that suggest

exogenous explanations for outliers: clustered volatility and random news processes.

Clustered volatility, for example the basis of ARCH⁹² models, is based on the observation by Mandelbrot (1963a) that the variation in a time series is not independent but highly correlated and thus tends to appear in clusters.

Mandelbrot suggested the model of random jumps as opposed to random walks. Other explanations have suggested that not only short memory processes, i.e., correlation of variation at points in time that decays quickly, but also long memory processes lead to power law behavior (Gabaix 2009). In relation to IT projects, the event-centric view could model the project based on clustered volatility as the collective reaction of agents to exogenous events.

A second explanation has been given by the random news process. As discussed before, true power law behavior is found when not only errors are multiplicative but also if a source of discontinuity is present. In stock market pricing the arrival of news has been modelled to be such a random noise process that creates power law behavior (Gabaix 2009). In the context of IT project management the random noise has been modeled through absenteeism of programmers that creates fat tails (Barseghyan 2009), other external events that create random noise are conceivable.

⁹² Autoregressive conditional heteroskedasticity

8.12.3 *Process-centric explanations*

Process-centric explanations of why outliers happen have also been suggested to generate power laws. Most prominent is the concept of proportional random growth, also known as Yule process, Gibrat's law, Zipf's law, the Matthew's effect or preferential attachment.

The concept can theoretically be explained by the behavioral roots found in the escalation of commitment, i.e., throwing good money after bad money. Which could be conceived as the opposite version of a rich-get-richer mechanism.

Rich-get-richer mechanisms have been found in several other areas of social science, e.g., book sales, scientific productivity, web links, city sizes. These models are based on two mechanisms that generate power law behavior:

growth and preferential attachment. Growth, in the simplest terms, means that the project increases continuously, e.g., continuously spends money, consumes effort, or time passes. Preferential attachment means that a new quantity, e.g., cost or effort, are allocated in accordance to the existing size. Yule (1925)

demonstrated this mechanism for the number of species per genus of flowering plants, Simon (1955) and Simon and Bonini (1958) for the size of cities and firms. Price (1976) called it the cumulative advantage process – in which larger size is a positive property, which has been argued for scientific citation counts, journal use or links to web pages. The theory of the escalation of commitment might explain a cumulative disadvantage process in the context of IT project performance.

The theory of man-made disasters explained the occurrence of outliers through a process of miscommunication and misunderstanding of failures over time.

The theory explained that errors multiplied over time. As discussed before (and by Gabaix 2009) such a system displays fat tails and under certain conditions, e.g., lower reflective barrier or random noise, a power law behavior.

Lastly, the theory of normalization of deviance explained how an organization slowly enlarges the circle of accepted risk over time. This is similar to the model of rational bubbles suggested by Lux and Sornette (2002), which produces fat tails. Blanchard and Watson (1983) describe the concept of the rational bubble as one where investors assume that asset prices will increase and thus collectively 'ride the bubble'. They argued that only in hindsight, i.e., after the (partial) collapse, can the concept of the bubble be established. The underlying principle of the rational bubble has also been termed copycat or herding behavior (Bouchaud 2008). The theory of normalization of deviance is the translation of these behaviors to project management.

9 Bibliography

- Abdelkader, Y. H., 2004. Evaluating project completion times when activity times are Weibull distributed. *European Journal of Operational Research*, 157(3), pp. 704–715.
- Addison, T. & Vallabh, S., 2002. Controlling software project risks: an empirical study of methods used by experienced project managers. In *Proceedings of the SAICSIT 2002*. pp. 128–140.
- Adler, M., & Posner, E., 1999. Rethinking cost-benefit analysis. *Yale Law Journal*, 109, pp. 165–247.
- Ahituv, N., 1980. A Systematic Approach toward Assessing the Value of an Information System. *MIS Quarterly*, 4(4), pp. 61–75.
- Ahmed, M. a., Ahmad, I., & AlGhamdi, J. S., 2013. Probabilistic size proxy for software effort prediction: A framework. *Information and Software Technology*, 55(2), pp. 241–251.
- Alderman, N., & Ivory, C., 2007. Partnering in major contracts: Paradox and metaphor. *International Journal of Project Management*, 25(4), pp. 386–393.
- Andersen, O., 1990. The use of software engineering data in support of project management. *Software Engineering Journal*, 5(6), pp. 350–356.
- Anderson, J. R., 2000. *Learning and memory*. (2nd Edition), New York: John Wiley.
- Anderson, P., Meyer, A., Eisenhardt, K. M., Carley, K., & Pettigrew, A., 1999. Introduction to the special issue: Applications of complexity theory to organization science. *Organization Science*, 10(3), pp. 233–236.
- Ansar, A., Flyvbjerg, B., Budzier, A., & Lunn, D., 2014. Should we build more large dams ? The actual costs of hydropower megaproject development. *Energy Policy*, in press, pp. 1–14.
- Anscombe, F. J., 1960. Rejection of outliers. *Technometrics*, 2(2), pp. 123–147.
- Ansoff, H., 1975. Managing strategic surprise by response to weak signals. *California management review*, 18(2), pp. 21–34.
- Atkinson, R., Klakegg, O. J., & Westerveld, E., 1999. Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, 17(6), pp. 337–342.

- Auerbach, F., 1913. Das Gesetz der Bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, 59, pp. 74–76.
- Augustine, N. R., 1982. *Augustine's Laws and Major System Development Programs*. Washington, DC: American Institute of Aeronautics & Astronautics.
- Armstrong, J.S. & Collopy, F., 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1), pp. 69–80.
- Artto, K. A., 2001. Management of project-oriented organization–conceptual analysis. *Project portfolio management: strategic management through projects*, Helsinki: Project Management Institute Finland, pp. 5–22.
- Azaron, A., Katagiri, H., Sakawa, M., Kato, K., & Memariani, A., 2006. A multi-objective resource allocation problem in PERT networks. *European Journal of Operational Research*, 172(3), pp. 838–854.
- Baccarini, D., 1996. The concept of project complexity a review. *International Journal of Project Management*, 14(4), pp. 201–204.
- Baccarini, D., 2004. Modelling Complex Projects. *International Journal of Project Management*, 22(6), pp. 519–520.
- Bak, P., Tang, C., & Wiesenfeld, K., 1987. Self-organized criticality: an explanation of 1/f noise. *Physical review letters*, 59(4), pp. 381–384.
- Banker, R. D., & Kauffman, R. J., 1991. Reuse and Productivity in Integrated Computer-Aided Software Engineering : An Empirical Study. *MIS Quarterly*, 15(3), pp. 375–401.
- Barabasi, A. L., 2005. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039), pp. 207–211.
- Barabási, A. L., 2010. *Bursts: the hidden patterns behind everything we do, from your e-mail to bloody crusades*. London: Penguin.
- Barnett, V., 1978. The study of outliers: purpose and model. *Applied Statistics*, 27(3), pp. 242–250.
- Barros, M. D. O., Werner, C. M. L., & Travassos, G. H., 2004. Supporting risks in software project management. *Journal of Systems and Software*, 70(1–2), pp. 21–35.

- Barsalou, L. L. W., 2008. Grounded cognition. *Annual Review of Psychology*, 59(1), pp. 617–45.
- Barsalou, L. W., Huttenlocher, J., & Lamberts, K., 1998. Basing categorization on individuals and events. *Cognitive Psychology*, 272(36), pp. 203–272.
- Barseghyan, P., 2009. Human Effort Dynamics and Schedule Risk Analysis. *PM World Today*, 11(3).
- Barton, M. A., & Sutcliffe, K. M., 2009. Overcoming dysfunctional momentum: Organizational safety as a social achievement. *Human Relations*, 62(9), pp. 1327–1356.
- Bartunek, J. M., 1987. First-Order, Second-Order, and Third-Order Change and Organization Development Interventions: A Cognitive Approach. *The Journal of Applied Behavioral Science*, 23(4), pp. 483–500.
- Bates D., Maechler M., Bolker B., & Walker S., 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-6.
- Baumard, P. & Starbuck, W.H., 2005. Learning from failures: Why it May Not Happen. *Long Range Planning*, 38(3), pp. 281–298.
- Beckman, R., & Cook, R., 1983. Outlier.....s. *Technometrics*, 25(2), pp. 119–149.
- Bessel, F. W., & Baeyer, J. J., 1838. *Gradmessung in Ostpreußen und ihre Verbindung mit Preußischen und Russischen Dreiecksketten* (p. 452). Berlin: Königliche Akademie der Wissenschaften.
- Belassi, W. & Tukel, O., 1996. A new framework for determining critical success/failure factors in projects. *International Journal of Project Management*, 14(3), pp. 141–151.
- Belout, A., & Gauvreau, C., 2004. Factors influencing project success: the impact of human resource management. *International Journal of Project Management*, 22(1), pp. 1–11.
- Benaroch, M., & Kauffman, R. J., 1999. A Case for Using Real Options Pricing Analysis to Evaluate Information Technology Project Investments. *Information Systems*, 10(1), pp. 70–86.
- Bendell, A., Solomon, D., & Carter, J., 1995. Evaluating project completion times when activity times are Erlang distributed. *Journal of the Operational Research Society*, 46(7), pp. 867–882.

- Benjamini, Y., & Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), pp. 289–300.
- Benjamini, Y., & Hochberg, Y., 2000. On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. *Journal of Educational and Behavioral Statistics*, 25(1), pp. 60–83.
- Bergeron, F. & St. Arnaud, J.-Y., 1992. Estimation of information systems development efforts. *Information & Management*, 22(4), pp. 239–254.
- Blanchard, O., & Watson, M. W., 1983. *Bubbles, Rational Expectations and Financial Markets* (NBER Working Paper No. 945). Lexington, MA: National Bureau of Economic Research, Inc.
- Bliese, P. D., 2000. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In Klein, K. J., Kozlowski, S. W. J. (Eds). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco, CA: Jossey-Bass, pp. 349–381.
- Blumentritt, T., 2006. Integrating strategic management and budgeting. *Journal of Business Strategy*, 27(6), pp. 73–79.
- Boehm, B., 1991. Software risk management: principles and practices. *Software, IEEE*, (January), pp. 32–41.
- Boehm, B., 2006. A view of 20th and 21st century software engineering. In *Proceedings of the 28th international conference on Software engineering*, ACM, pp. 12–29.
- Boin, A., & 't Hart, P., 2003. Public Leadership in Times of Crisis: Mission Impossible? *Public Administration Review*, 63(5), pp. 544–553.
- Boisot, M., & McKelvey, B., 2010. Integrating modernist and postmodernist perspectives on organizations: A complexity science bridge. *Academy of Management Review*, 35(3), pp. 415–433.
- Boland, R. J., Singh, J., Salipante, P., Aram, J. D., Fay, S. Y., & Kanawattanachai, P., 2001. Knowledge representations and knowledge transfer. *The Academy of Management Journal*, 44(2), pp. 393–417.
- Booth, P., & Schulz, A. K.-D., 2004. The impact of an ethical environment on managers' project evaluation judgments under agency problem conditions. *Accounting, Organizations and Society*, 29(5-6), pp. 473–488.

- Bouchaud, J. P., 2008. Economics needs a scientific revolution. *Nature*, 455(7217), pp. 1181-1181.
- Brändström, A., Bynander, F., & Hart, P., 2004. Governing by looking back: historical analogies and crisis management. *Public Administration*, 82(1), pp. 191-210.
- Brockner, J., & James, E. H., 2008. Toward an Understanding of When Executives See Crisis as Opportunity. *The Journal of Applied Behavioral Science*, 44(1), pp. 94-115.
- Brooks, F.P., 1975. *The mythical man-month: Essays on software engineering*, Reading, MA: Addison-Wesley.
- Brown, T., & Long, D., 1992. *A Study of Organizational Downsizing and Information Management Strategies*. Wright-Patterson AFB, OH: Air Force Institute of Technology.
- Bruijn, H. de & Leijten, M., 2008. Management Characteristics of Mega-Projects. In H. Priemus, B. Flyvbjerg, & B. Van Wee, eds. *Decision-Making on Mega-Projects: Cost-Benefit Analysis, Planning, and Innovation*. Cheltenham, UK: Elgar, pp. 23-29.
- Buehler, R., Griffin, D., & Ross, M., 1994. Exploring the “planning fallacy”: Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67(3), pp. 366-381.
- Byers, Alex, 2012. Budget battle hobbles W.H. tech tool, *Politico*, 18.12.2012.
- Camerer, C., & Lovallo, D., 1999. Overconfidence and excess entry: An experimental approach. *The American Economic Review*, 89(1), pp. 306-318.
- Campbell, D. T., & Fiske, D. W., 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), pp. 81-105.
- Cantarelli, C. C., Flyvbjerg, B., & Buhl, S. L., 2012. Geographical variation in project cost performance: the Netherlands versus worldwide. *Journal of Transport Geography*, 24, pp. 324-331.
- Carlson, J. M., & Doyle, J., 1999. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E*, 60(2), pp. 1412-1427.
- Caron, F., Ruggeri, F., & Merli, A., 2013. A Bayesian Approach to Improve Estimate at Completion in Earned Value Management. *Project Management Journal*, 44(1), pp. 3-16.

- Carson, S. J., Madhok, A., Varman, R., & John, G., 2003. Information Processing Moderators of the Effectiveness of Trust-Based Governance in Interfirm R&D Collaboration. *Organization Science*, 14(1), pp. 45-56.
- Chang, S.-J., van Witteloostuijn, A., & Eden, L., 2010. From the Editors: Common method variance in international business research. *Journal of International Business Studies*, 41(2), pp. 178-184.
- Chapman, R. J., 1998. The role of system dynamics in understanding the impact of changes to key project personnel on design production within construction projects. *International Journal of Project Management*, 16(4), pp. 235-247.
- Chapman, C., & Ward, S., 2003. Constructively simple estimating: a project management example. *Journal of the Operational Research Society*, 54(10), pp. 1050-1058.
- Charette, R.N., 2005. Why Software Fails. *IEEE Spectrum*, (September 2005).
- Christianson, M.K., Farkas, M. T., Sutcliffe, K. M., & Weick, K. E., 2008. Learning through Rare Events: Significant Interruptions at the Baltimore & Ohio Railroad Museum. *Organization Science*, 20(5), pp. 846-860.
- Cicmil, S., & Marshall, D., 2005. Insights into collaboration at the project level: complexity, social interaction and procurement mechanisms. *Building Research & Information*, 33(6), pp. 523-535.
- Clauset, A., Shalizi, C. R., & Newman, M. E., 2009. Power-law distributions in empirical data. *SIAM review*, 51(4), pp. 661-703.
- Collopy, F., 2007. Difficulty and complexity as factors in software effort estimation. *International Journal of Forecasting*, 23(3), pp. 469-471.
- Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Psychology Press.
- Comfort, L. K., 2005. Risk, Security, and Disaster Management. *Annual Review of Political Science*, 8(1), pp. 335-356.
- Conlon, D. E., 1999. Escalation in Decision-Making: The Tragedy of Taurus. *Administrative Science Quarterly*, 44(1), pp. 180-183.
- Conroy, G., & Soltan, H., 1998. ConSERV, a project specific risk management concept. *International Journal of Project Management*, 16(6), pp. 353-366.

- Conte, S. D., Dunsmore, H. E., Shen, Y. E., & Shen, V. Y., 1986. *Software engineering metrics and models*. Redwood City: Benjamin-Cummings Publishing.
- Cooke-Davies, T., 2002. The “real” success factors on projects. *International Journal of Project Management*, 20(3), pp. 185–190.
- Cooke-Davies, T., & Cicmil, S., 2008. We’re not in Kansas anymore, Toto: Mapping the Strange Landscape of Complexity Theory, and its Relationship to Project Management. *Journal of Project Management*, 38(2), pp. 50–61.
- Coombs, W., & Holladay, S., 1996. Communication and attributions in a crisis: An experimental study in crisis communication. *Journal of Public Relations Research*, 8(4), pp. 279–295.
- Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C., 1988. Entrepreneurs’ perceived chances for success. *Journal of Business Venturing*, 3(2), pp. 97–108.
- Crawford, L., Cooke-Davies, T., Hobbs, B., Labuschagne, L., Remington, K., & Chen, P., 2008. Situational Sponsorship of Projects and Programs: An Empirical View, *Project Management Journal*, 39(S1), S43-S55
- Creed, W. E. D., Langstraat, J. A., & Scully, M. A., 2002. A Picture of the Frame: Frame Analysis as Technique and as Politics. *Organizational Research Methods*, 5(1), pp. 34–55.
- Czuchry, A. J., & Yasin, M. M., 2003. Managing the project management process. *Industrial Management & Data Systems*, 103(1), pp. 39–46.
- Dalcher, D., 1999. Disaster in London the LAS Case study. In *Engineering of Computer-Based Systems, 1999. Proceedings. ECBS '99. IEEE Conference and Workshop on*. pp. 41–52.
- Daniel, D. R., 1961. Management information crisis. *Harvard Business Review*, 39(5), pp. 111–121.
- Denrell, J., 2004. Random Walks and Sustained Competitive Advantage. *Management Science*, 50(7), pp. 922–934.
- Denzin, N., & Keller, C., 1981. Frame analysis reconsidered. *Contemporary Sociology*, 10(1), pp. 52–60.
- De Wit, A., 1988. Measurement of project success. *International Journal of Project Management*, 6(3), pp. 164–170.

- Dixon, W. J., 1950. Analysis of Extreme Values. *The Annals of Mathematical Statistics*, 21(4), pp. 488–506.
- Dosi, G., & Lovallo, D., 1997. Rational entrepreneurs or optimistic martyrs? Some considerations on technological regimes, corporate entries, and the evolutionary role of decision biases, in Garud, R., Nayyar, P. R., Shapira, Z. B. (eds.) *Technological Innovation: Oversights and Foresights*, Cambridge, UK: Cambridge University Press, pp. 41–70.
- Douglas, M., 1986. *How institutions think*, Syracuse, NY: Syracuse University Press.
- Drummond, H., 1997. *Escalation in Decision-Making: The Tragedy of Taurus*, Oxford: Oxford University Press.
- Dumke, R., & Abran, A., 2011. *Cosmic Function Points: Theory and Advanced Practices*. Boca Raton: Taylor & Francis Ltd.
- Dutton, J., & Jackson, S., 1987. Categorizing strategic issues: Links to organizational action. *Academy of Management Review*, 12(1), pp. 76–90.
- Dyba, T., Kitchenham, B. A., & Jørgensen, M., 2005. Evidence-Based Software Engineering for Practitioners. *IEEE Software*, 22(1), pp. 58–66.
- Ekstedt, E., Lundin, R. A., Soderholm, A., & Wirdenius, H., 1993. Project organization in the squeeze between short-run flexibility and long-run inertia. *Proceedings of the Project Management Institute Annual Seminar Symposium*, Newton Square, PA: Project Management Institute, pp. 586–590.
- Elsbach, K. D., Barr, P. S., & Hargadon, A. B., 2005. Identifying Situated Cognition in Organizations. *Organization Science*, 16(4), pp. 422–433.
- Emam, K. El, & Koru, A., 2008. A replicated survey of IT software project failures. *IEEE Software*, 25(5), pp. 84–90.
- Embrechts, P., Furrer, H., & Kaufmann, R., 2003. Quantifying regulatory capital for operational risk. *Derivatives Use, Trading & Regulation*, 9(3), pp. 217–233.
- Engwall, M., 2003. No project is an island: linking projects to history and context. *Research Policy*, 32(5), pp. 789–808.
- Eppel, E., 2012. What Does it Take to Make Surprises Less Surprising? *Public Management Review*, 14(7), pp. 881–902.

- Eveleens, J.L., van der Pas, M. & Verhoef, C., 2012. Quantifying forecast quality of IT business value. *Science of Computer Programming*, 77(3), pp. 314–354.
- Eveleens, J.L. & Verhoef, C., 2009. Quantifying IT forecast quality. *Science of Computer Programming*, 74(11-12), pp. 934–988.
- Eveleens, J.L. & Verhoef, C., 2010. The rise and fall of the Chaos report figures. *IEEE Software*, 27(1), pp. 30–36.
- Ewusi-Mensah, K. & Przasnyski, Z.H., 1991. On information systems project abandonment: an exploratory study of organizational practices. *MIS Quarterly*, 15(1), pp. 67–86.
- Fama, E., 1965. The behavior of stock-market prices. *Journal of Business*, 38(1), pp. 34–105.
- Farmer, J. D., & Geanakoplos, J., 2004. *Power laws in finance and their implications for economic theory*. Santa Fe, NM: Santa Fe Institute.
- Feeny, D. F., Edwards, B. R., & Simpson, K. M., 1992. Understanding the CEO/CIO Relationship. *MIS Quarterly*, 16(4), pp. 435–448.
- Festinger, L., 1962. *A theory of cognitive dissonance (Vol. 2)*. Stanford, CA: Stanford University Press.
- Fink, S., 1986. *Crisis management: Planning for the inevitable*. New York, NY: American Management Association.
- Finnie, G. R., Wittig, G. E., & Desharnais, J. M., 1997. A comparison of software effort estimation techniques: using function points with neural networks, case-based reasoning and regression models. *Journal of Systems and Software*, 39(3), pp. 281–289.
- Fisher, R. A., 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604), pp. 309–368.
- Fleming, Q. W., & Koppelman, J. M., 2000. *Earned Value Project Management, Second Edition*. Newton Square, PA: Project Management Institute.
- Fleming, P., & Zyglidopoulos, S., 2008. The escalation of deception in organizations. *Journal of Business Ethics*, 4(81), pp. 837–850.
- Flyvbjerg, B., 2007. Cost Overruns and Demand Shortfalls in Urban Rail and Other Infrastructure. *Transportation Planning & Technology*, 30(1), pp. 9–30.

- Flyvbjerg, B., Bruzelius, N., & Rothengatter, W., 2003. *Megaprojects and Risk*. Cambridge, UK: Cambridge University Press.
- Flyvbjerg, B., & Budzier, A., 2011. Why your IT project may be riskier than you think. *Harvard Business Review*, 89(9), pp. 23–25.
- Flyvbjerg, B., Holm, M. S., & Buhl, S. S., 2002. Underestimating Costs in Public Works Projects: Error or Lie? *Journal of the American Planning Association*, 68(3), pp. 279–295.
- Fortune, J. & White, D., 2006. Framing of project critical success factors by a systems model. *International Journal of Project Management*, 24(1), pp. 53–65.
- Fowler, A., & Walsh, M., 1999. Conflicting perceptions of success in an information systems project. *International Journal of Project Management*, 17(1), pp. 1–10.
- Freeman, M., & Beale, P., 1992. Measuring project success. *Project Management Journal*, 23(1), pp. 8–17.
- Gabaix, X., 2009. Power Laws in Economics and Finance. *Annual Review of Economics*, 1(1), pp. 255–294.
- Gaddis, P. O., 1959. The Project Manager. *Harvard Business Review*, 37(3), pp. 89–97.
- GAO, 2005. *OMB Can Make More Effective Use of its Investment Reviews*. Washington, DC: Government Accountability Office (GAO).
- GAO, 2010. *Accuracy Has Improved, and Additional Efforts Are Under Way to Better Inform Decision Making*. Washington, DC: Government Accountability Office (GAO).
- Garland, H., & Conlon, D., 1998. Too Close to Quit: The Role of Project Completion in Maintaining Commitment. *Journal of Applied Social Psychology*, 28(22), pp. 2025–2048.
- Gareis, R., 1991. Management by projects: the management strategy of the 'new' project-oriented company. *International Journal of Project Management*, 9(2), pp. 71–76.
- Gelman, A., & Hill, J., 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

- Genuchten, M. van, 1991. Why is software late? An empirical study of reasons for delay in software development. *IEEE Transactions on Software Engineering*, 17(6), pp. 582–590.
- Ghaziani, A., & Ventresca, M. M. J., 2005. Keywords and Cultural Change: Frame Analysis of Business Model Public Talk, 1975–2000. *Sociological Forum*, 20(4), pp. 523–559.
- Ginzberg, M. J., 1979. Improving MIS project selection. *Omega*, 7(6), pp. 527–537.
- Glass, R.L., 1997. *Software Runaways: Monumental Software Disasters*, London: Prentice Hall.
- Gioia, D., & Chittipeddi, K., 1991. Sensemaking and sensegiving in strategic change initiation. *Strategic Management Journal*, 12(1), pp. 433–448.
- Goffman, E., 1974. *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Goldstein, M. L., Morris, S. A., & Yen, G. G., 2004. Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 41(2), pp. 255–258.
- Golenko-Ginzburg, D., 1988. On the Distribution of Activity Time in PERT. *The Journal of the Operational Research Society*, 39(8), pp. 767–771.
- Grant, K. P., Cashman, W. M., & Christensen, D. S., 2006. Delivering projects on time. *Research Technology Management*, 49(6), pp. 52–58.
- Grote, G., Weichbrodt, J. C., Günter, H., Zala-Mezö, E., & Künzle, B., 2008. Coordination in high-risk organizations: the need for flexible routines. *Cognition, Technology & Work*, 11(1), pp. 17–27.
- Grubbs, F., 1969. Procedures for detecting outlying observations in samples. *Technometrics*, 11(1), pp. 1–21.
- Gumbel, E. J., 1958/2004. *Statistics of Extremes*. Mineola, NY: Dover Publications Inc.
- Hahn, E., 2008. Mixture densities for project management activity times: A robust approach to PERT. *European Journal of Operational Research*, 188(2), pp. 450–459.
- Hall, P., 1980. *Great planning disasters*. Harmondsworth, U.K.: Penguin Books

- Hallgren, M., & Wilson, T. L., 2008. The nature and management of crises in construction projects: Projects-as-practice observations. *International Journal of Project Management*, 26(8), pp. 830–838.
- Handcock, M. S., & Morris, M., 1999. *Relative Distribution Methods in the Social Sciences*. New York: Springer.
- Hardin, G., 1968. The tragedy of the commons. *Science*, 162(3859), pp. 1243–1248.
- Hassan, A. E., Hindle, A., Runeson, P., Shepperd, M., Devanbu, P., & Kim, S., 2013. Roundtable: What's Next in Software Analytics. *IEEE Software*, 30(4), pp. 53–56.
- Hayes-Roth, B., & Hayes-Roth, F., 1979. A Cognitive Model of Planning. *Cognitive Science*, 3(4), pp. 275–310.
- Heemstra, F.J., 1992. Software cost estimation. *Information and Software Technology*, 34(10), pp. 627–639.
- Heltewig, S., 2008. *Improving the Use Case Point Method*. PhD thesis. Universität Kaiserslautern.
- Hellström, M., & Wikström, K., 2005. Project business concepts based on modularity–improved manoeuvrability through unstable structures. *International Journal of Project Management*, 23(5), pp. 392–397.
- Helm, J., & Remington, K., 2005. *Effective Sponsorship: An Evaluation of the Role of the Executive Sponsor in Complex Infrastructure Projects by Senior Project Managers*. Newton Square, PA: Project Management Institute.
- Heracleous, L., & Barrett, M., 2001. Organizational Change As Discourse: Communicative Actions and Deep Structures in the Context of Information Technology Implementation. *Academy of Management Journal*, 44(4), pp. 755–778.
- Herrerias-Velasco, J. M., Herrerias-Pleguezuelo, R., & van Dorp, J. R., 2011. Revisiting the PERT mean and variance. *European Journal of Operational Research*, 210(2), pp. 448–451.
- Hodge, V. J., & Austin, J. I. M., 2004. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22(2), pp. 85–126.
- Hollnagel, E., & Woods, D., 1999. Cognitive systems engineering: New wine in new bottles. *International Journal of Human-Computer Studies*, 51(6), pp. 339–356.

- Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), pp. 65–70.
- Hopkins, A., 2007. *The problem of defining high reliability organisations* (Working Paper No. 51). National Research Center for Occupational Safety and Safety Regulation (p. 15). Canberra: The Australian National University.
- Huchzermeier, A., & Loch, C. H., 2001. Project Management Under Risk: Using the Real Options Approach to Evaluate Flexibility in R&D. *Management Science*, 47(1), pp. 85–101.
- Hubert, M., & van der Veeken, S., 2008. Outlier detection for skewed data. *Journal of Chemometrics*, 22(1), pp. 235–246.
- ICMR, 2003. Kmart: Fall of a Retailing Giant. *Business*, p. 16.
- Isidore, C., 2004. Comeback for Kmart, Lampert. *CNN Money*.
- Jaafari, A., 2007. Project and program diagnostics: A systemic approach. *International journal of project management*, 25(8), pp. 781–790.
- James, E. H., Wooten, L. P., & Dushek, K., 2011. Crisis Management: Informing a New Leadership Research Agenda. *The Academy of Management Annals*, 5(1), pp. 455–493.
- Janis, I. L., 1973. Groupthink and Group Dynamics: A Social Psychological Analysis of Defective Policy Decisions. *Policy Studies Journal*, 2(1), pp. 19–25.
- Jenkins, A.M., Naumann, J.D. & Wetherbe, J.C., 1984. Empirical investigation of systems development practices and results. *Information & Management*, 7(2), pp. 73–82.
- Jensen, M., & Meckling, W., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 4(3), pp. 305–360.
- Jiang, J., & Klein, G., 2000. Software development risks to project effectiveness. *Journal of Systems and Software*, 52(1), pp. 3–10.
- Jick, T. D., 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), pp. 602–611.
- Jolivet, F., & Navarre, C. (1996). Large-scale projects, self-organizing and meta-rules : towards new forms of management. *International Journal of Project Management*, 14(5), 265–271.

- Jones, C., 1996. *Patterns of software system failure and success*, London: International Thomson Computer Press.
- Jones, C., 2000. *Software assessments, benchmarks, and best practices*, London: Addison-Wesley.
- Jones, C., 2007. *Estimating Software Costs*, McGraw-Hill Prof Med/Tech.
- Jones, C. 2008. *Applied Software Measurement: Global Analysis of Productivity and Quality* (p. 662). McGraw-Hill Prof Med/Tech.
- Jørgensen, M., 2004. A review of studies on expert estimation of software development effort. *Journal of Systems and Software*, 70, pp. 37–60.
- Jørgensen, M., Halkjelsvik, T., & Kitchenham, B., 2012. How does project size affect cost estimation error? Statistical artifacts and methodological challenges. *International Journal of Project Management*, 30(7), pp. 839-849.
- Jørgensen, M. & Moløkken-Østvold, K., 2006. How large are software cost overruns? A review of the 1994 CHAOS report. *Information and Software Technology*, 48(4), pp. 297–301.
- Jugdev, K., & Müller, R., 2005. A retrospective look at our evolving understanding of project success. *Project Management Journal*, 36(4), pp. 19-31.
- Kahneman, D., & Lovallo, D., 1993. Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, 39(1), pp. 17–31.
- Kahneman, D., & Tversky, A., 1977/1979. Intuitive prediction: Biases and corrective procedures. *TIMS Studies in the Management Sciences*, 12, 313–327. Reprint of a 1977 research report for the Defence Advanced Research Projects Agency (DARPA)
- Kahneman, D., & Tversky, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 47(2), pp. 263–291.
- Kaplan, S., 2008. Framing Contests : Strategy Making Under Uncertainty. *Organization Science*, 19(5), pp. 729–752.
- Kaplan, S., 2011. Research in Cognition and Strategy: Reflections on Two Decades of Progress and a Look to the Future. *Journal of Management Studies*, 48(3), pp. 665–695.

- Kazanjian, R. K., Drazin, R., & Glynn, M. A., 2000. Creativity and technological learning: the roles of organization architecture and crisis in large-scale projects. *Journal of Engineering and Technology Management*, 17(3-4), pp. 273-298.
- Keefer, D., & Bodily, S., 1983. Three-point approximations for continuous random variables. *Management Science*, 29(5), pp. 595-609.
- Keil, M., Mixon, R., Saarinen, T., & Tuunainen, V., 1994. Understanding runaway information technology projects: results from an international research program based on escalation theory. *Journal of Management Information Systems*, 11(3), 65-85.
- Keil, M., 1995. Pulling the Plug: Software Project Management and the Problem of Project Escalation. *MIS Quarterly*, 19(4), 421-447.
- Keil, M. & Mann, J., 1997. Understanding the Nature and Extent of IS Project Escalation: Results from a Survey of IS Audit and Control Professionals. In *The Thirtieth Annual Hawaii International Conference on System Sciences*. p. 10.
- Keil, M., Mann, J. & Rai, A., 2000. Why Software Projects Escalate: An Empirical Analysis and Test of Four Theoretical Models. *MIS Quarterly*, 24(4), pp. 631-664.
- Keil, M., & Mähring, M., 2010. Is Your Project Turning into a Black Hole? *California Management Review*, 53(1), 6-31.
- Keil, M., & Truex, D. P., 1995. The effects of sunk cost and project completion on information technology project escalation. *IEEE Transactions on Engineering Management*, 42(4), pp. 372-381.
- Kemerer, C. F., 1987. An empirical validation of software cost estimation models. *Communications of the ACM*, 30(5), pp. 416-429.
- Kenny, D. A., & Judd, C. M., 1986. Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, 99(3), pp. 422-431.
- Kidd, J., 1970. The utilization of subjective probabilities in production planning. *Acta Psychologica*, 34, pp. 338-347.
- Kieser, A., & Nicolai, A. T., 2005. Success Factor Research Overcoming the Trade-Off Between Rigor and Relevance?. *Journal of Management Inquiry*, 14(3), pp. 275-279.

- King, J. L., & Schrems, E. L., 1978. Cost-benefit analysis in information systems development and operation. *ACM Computing Surveys (CSUR)*, 10(1), pp. 19-34.
- King, G. & Zeng, L., 2001. Logistic regression in rare events data. *Political Analysis*, 9(2), pp. 137-163.
- Kitchenham, B., 1998. A procedure for analyzing unbalanced datasets. *Software Engineering, IEEE Transactions on*, 24(4), pp. 278-301.
- Kitchenham, B. A., Pickard, L. M., MacDonell, S. G., & Shepperd, M. J., 2001. What accuracy statistics really measure [software estimation]. *IEE Proceedings In Software*, 148(3), pp. 81-85.
- Knight, F., 1921. *Risk, uncertainty and profit*. New York: AM Kelley.
- Konicki, S., 2002. Now In Bankruptcy, Kmart Struggled With Supply Chain. *InformationWeek*.
- Kotiah, T., & Wallace, N., 1973. Another look at the PERT assumptions. *Management Science*, 20(1), pp. 44-49.
- Kreiner, K., 2013. Projects, Promises, and Performances. In *11th Conference of International Research Network on Organizing by Projects (IRNOP)*, pp. 1-11. Oslo, Norway.
- Kulk, G.P., 2009. *IT Risks in Measure and Number*. PhD thesis. Vrije Universiteit Amsterdam.
- Kulk, G.P., Peters, R.J. & Verhoef, C., 2009. Quantifying IT estimation risks. *Science of Computer Programming*, 74(11-12), pp. 900-933.
- Kulk, G.P. & Verhoef, C., 2008. Quantifying requirements volatility effects. *Science of Computer Programming*, 72(3), pp. 136-175.
- Kulkarni, V., & Adlakhia, V. 1986. Markov and Markov-regenerative PERT networks. *Operations Research*, 34(5), pp. 769-781.
- Kumar, V., Persaud, a. N. S., & Kumar, U., 1996. To terminate or not an ongoing R&D project: a managerial dilemma. *IEEE Transactions on Engineering Management*, 43(3), pp. 273-284.
- Kusters, R.J., van Genuchten, M.J.I.M. & Heemstra, F.J., 1990. Are software cost-estimation models accurate? *Information and Software Technology*, 32(3), pp. 187-190.

- Lagadec, P., 1993. *Preventing chaos in a crisis*. London: McGraw-Hill.
- Lampel, J., Shamsie, J., & Shapira, Z., 2009. Experiencing the Improbable: Rare Events and Organizational Learning. *Organization Science*, 20(5), pp. 835–845.
- Lampel, J., & Shapira, Z., 1995. Progress and its discontents: Data scarcity and the limits of falsification in strategic management. In Shrivastava P., Huff A., Dutton J. E. (eds). *Advances in Strategic Management Vol. 12A*, JAI Press: Greenwich, CT, pp. 113-150.
- LaPorte, T., & Consolini, P., 1991. Working in practice but not in theory: theoretical challenges of “high-reliability organizations.” *Journal of Public Administration Research and Theory: J-PART*, 1(1), pp. 19–48.
- Lech, P., 2013. Time, Budget, and Functionality? IT Project Success Criteria Revised. *Information Systems Management*, 30(3), pp. 263-275.
- Lederer, A. L., 1995. Causes of Inaccurate Cost Estimates Software Development. *Journal of Systems and Software*, 31(2), 125–134.
- Lederer, A. L., Mirani, R., Neo, B. S., Pollard, C., Prasad, J., & Ramamurthy, K., 1990. Cost Estimating : A Management Perspective. *MIS Quarterly*, 14(2), pp. 159–176.
- Lederer, A. L., & Prasad, J., 1991. The validation of a political model of information systems development cost estimating. *ACM SIGCPR Computer Personnel*, 13(2), 47–57.
- Lederer, A. L., & Prasad, J., 2000. Software management and cost estimating error. *Journal of Systems and Software*, 50(1), 33–42.
- Legendre, A. M., 1805. *Nouvelles Méthodes Pour La Détermination Des Orbites Des Comètes*. Paris: Didot.
- Leijten, M., 2013. Chapter 4: Real-world decision-making on mega-projects: politics, bias and strategic behaviour. Priemus, H., & van Wee, B. (Eds.). *International Handbook on Mega-Projects*, pp. 57–82. Cheltenham, UK: Edward Elgar Publishing.
- Leslie, K. J., & Michaels, M. P., 1997. The real power of real options. *McKinsey Quarterly*, (3), pp. 97–108.
- Leveson, N., 2004. A new accident model for engineering safer systems. *Safety Science*, 42(4), pp. 237–270.

- Leveson, N., Dulac, N., Marais, K., & Carroll, J., 2009. Moving Beyond Normal Accidents and High Reliability Organizations: A Systems Approach to Safety in Complex Systems. *Organization Studies*, 30(2-3), pp. 227-249.
- Levinthal, D., & Rerup, C., 2006. Crossing an Apparent Chasm: Bridging Mindful and Less-Mindful Perspectives on Organizational Learning. *Organization Science*, 17(4), pp. 502-513.
- Lindstrom, M.J. and Bates, D.M., 1990. Nonlinear Mixed Effects Models for Repeated Measures Data, *Biometrics*, 46, pp. 673-687.
- Little, T., & Landmark Graphics, 2006. Project Estimation the Cone of Uncertainty. *IEEE Software*, 23(3), pp. 48-54.
- Louis, M. R., 1991. Switching Cognitive Gears: From Habits of Mind to Active Thinking. *Human Relations*, 44(1), pp. 55-76.
- Lundin, R. A., & Söderholm, A., 1995. A Theory of the Temporary Organization. *Scandinavian Journal of Management*, 1(4), pp. 437-455.
- Lundin, R. A., & Söderholm, A., 1998. Conceptualizing a projectified society discussion of an eco-institutional approach to a theory on temporary organisations. *Projects as arenas for renewal and learning processes*, Heidelberg/Berlin: Springer, pp. 13-23.
- Lux, T., & Sornette, D., 2002. On rational bubbles and fat tails. *Journal of Money, Credit, and Banking*, 34(3), pp. 589-610.
- MacCrimmon, K., & Ryavec, C., 1964. An analytical study of the PERT assumptions. *Operations Research*, 12(1), pp. 16-37.
- Madsen, P. M., 2009. These Lives Will Not Be Lost in Vain: Organizational Learning from Disaster in U.S. Coal Mining. *Organization Science*, 20(5), pp. 861-875.
- Mahaney, R., & Lederer, A., 1999. Runaway information systems projects and escalating commitment. *Proceedings of the 1999 ACM SIGCPR*, 291-296.
- Malcom, D. G., Roseboom, J. H., Clark, C. E., & Fazar, W., 1959. Application of a Technique for Research and Development Program Evaluation. *Operations Research*, 7(5), pp. 646-669.
- Mallak, L. A., Kurstedt, H. A., & Patzak, G. A., 1997. Planning for crises in project management. *Project Management Journal*, 28(2), pp. 14-24.

- Mallak, L. A., & Watts, P. S., 1997. Delayering in a US Army R&D organization. *TQM Journal*, 9(4), p. 260.
- Marković, D., Gros, C., & Schuelein, A., 2013. Criticality in conserved dynamical systems: Experimental observation vs. exact properties. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(1), pp. 013106.
- Mandelbrot, B., 1953. An information theory of the structure of language based upon statistical matching of messages and coding. In Jackson, W. (Ed.), *Proceedings of symposium on applications of communication theory*. London: Butterworth Scientific Publications.
- Mandelbrot, B., 1960. The Pareto-Levy law and the distribution of income. *International Economic Review*, 1(2), pp. 79–106.
- Mandelbrot, B., 1961. Stable Paretian random functions and the multiplicative variation of income. *Econometrica: Journal of the Econometric Society*, 29(4), pp. 517–543.
- Mandelbrot, B., 1963a. The variation of certain speculative prices. *Journal of Business*, 36(4), pp. 394–419.
- Mandelbrot, B. 1963b. New methods in statistical economics. *The Journal of Political Economy*, 71(5), pp. 421–440.
- Mandelbrot, B. B., 1997. *Fractals and Scaling in Finance*. New York: Springer.
- Marais, K., Dulac, N., & Leveson, N., 2004. Beyond normal accidents and high reliability organizations: The need for an alternative approach to safety in complex systems. In *Engineering Systems Division Symposium*. Cambridge, MA: Massachusetts Institute of Technology.
- March, J. G., 1991. Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1), pp. 71–87.
- March, J., & Olsen, J., 1975. The Uncertainty of the Past: Organizational Learning under Ambiguity*. *European Journal of Political Research*, 3(1), pp. 147–171.
- March, J. G., Sproull, L. S., & Tamuz, M., 1991. Learning from samples of one or fewer. *Organization Science*, 2(1), pp. 1–13.
- March, J. G., & Sutton, R. I., 1997. Crossroads — Organizational Performance as a Dependent Variable. *Organization Science*, 8(6), pp. 698–706.

- Marchewka, J. T., & Keil, M., 1995. Portfolio theory approach for selecting and managing IT projects. *Information Resources Management Journal (IRMJ)*, 8(4), pp. 5-16.
- Maylor, H., Vidgen, R., & Carver, S., 2008. Managerial complexity in project-based operations: A grounded model and its implications for practice. *Project Management Journal*, 39(S1), pp. S15-S26.
- MacDonell, S. G., & Shepperd, M. J., 2003. Combining techniques to optimize effort predictions in software project management. *Journal of Systems and Software*, 66(2), pp. 91-98.
- McConnell, S. C., 1996. *Rapid Development: Taming Wild Software Schedules*. Redmond, WA: Microsoft Press.
- McFarlan, F.W., 1971. Problems in Planning the Information System. *Harvard Business Review*, 49(3), pp. 75-89.
- McFarlan, F.W., 1981. Portfolio approach to information systems. *Harvard Business Review*, 59(5), pp. 142-151.
- McGuire, W. J., 1960. Cognitive consistency and attitude change. *The Journal of Abnormal and Social Psychology*, 60(3), pp. 345-353.
- McKeen, J., 1983. Successful development strategies for business application systems. *MIS Quarterly*, 7(3), 47-65.
- McKelvey, B., 2004. Complexity science as order-creation science: New theory, new method. *Emergence: Complexity & Organization*, 6(4), pp. 2-27.
- McKelvey, B., & Andriani, P., 2005. Why Gaussian statistics are mostly wrong for strategic organization. *Strategic Organization*, 3(2), pp.219-228.
- McNamara, G., & Bromiley, P., 1997. Decision making in an organizational setting: cognitive and organizational influences on risk assessment in commercial lending. *Academy of Management Journal*, 40(5), pp. 1063-1088.
- Merrow, E., Phillips, K., & Myers, C., 1981. *Understanding cost growth and performance shortfalls in pioneer process plants*. Santa Monica, CA: Rand Corporation.
- Merrow, E. W., 2011. *Industrial Megaprojects: Concepts, Strategies, and Practices for Success*. London: John Wiley & Sons.

- Miller, R. & Lessard, D., 2000. *The strategic management of large engineering projects: Shaping institutions, risks, and governance*, Boston, MA: The MIT Press.
- Miller, R. & Lessard, D., 2001. Understanding and managing risks in large engineering projects. *International Journal of Project Management*, 19(8), pp. 437–443.
- Miller, R. & Lessard, D.R., 2008. Evolving strategy: risk management and the shaping of mega-projects. In H. Priemus, B. Flyvbjerg, & B. Van Wee, eds. *Decision-Making on Mega-Projects: Cost-Benefit Analysis, Planning, and Innovation*. Cheltenham, UK: Edward Elgar Publishing, pp. 145–172.
- Mitchell, V. L., & Zmud, R. W., 1999. The Effects of Coupling IT and Work Process Strategies in Redesign Projects. *Organization Science*, 10(4), 424–438.
- Moløkken, K., & Jorgensen, M., 2003. A review of software surveys on software effort estimation. *Proceedings of the International Symposium on Empirical Software Engineering*, 2003, pp. 223–230.
- Moløkken-Østfold, K., Jørgensen, M., Tanilkan, S. S., Gallis, H., Lien, A. C., & Hove, S. E. (2004). A survey on software estimation in the Norwegian industry. In *Proceedings of the 10th International Symposium on Software Metrics (METRICS'04)*, pp. 208–219.
- Montealegre, R. & Keil, M., 1998. Denver International Airport's Automated Baggage Handling System: A Case Study of De-escalation of Commitment. In *Academy of Management Proceedings*. Academy of Management, pp. D1–D9.
- Montealegre, R., & Keil, M., 2000. De-Escalating Information Technology Projects: Lessons from the Denver International Airport. *MIS Quarterly*, 24(3), pp. 417–447.
- Morris, P. W. G., & Hough, G. H., 1987. *The Anatomy of Major Projects: Study of the Reality of Project Management* (pp. 338). London: John Wiley & Sons.
- Müller, R., & Jugdev, K., 2012. Critical success factors in projects: Pinto, Slevin, and Prescott – the elucidation of project success. *International Journal of Managing Projects in Business*, 5(4), pp. 757–775.
- Murre, J. M., & Chessa, A. G., 2011. Power laws from individual differences in learning and forgetting: mathematical analyses. *Psychonomic Bulletin & Review*, 18(3), pp. 592–597.

- Narbaev, T., & De Marco, A., 2014. An earned schedule-based regression model to improve cost estimate at completion. *International Journal of Project Management*, in press.
- Natovich, J., 2003. Vendor Related Risks in IT Development: A Chronology of an Outsourced Project Failure. *Technology Analysis & Strategic Management*, 15(4), pp. 409–419.
- Neumann, J. von, & Morgenstern, O., 1947. *Theory of games and economic behavior*, 2nd ed., p. 625. Princeton: Princeton University Press.
- Newman, M. E. J., 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), pp. 323–351.
- Newell, S., Bresnen, M., Edelman, L., Scarbrough, H., & Swan, J., 2006. Sharing Knowledge Across Projects Limits to ICT-led Project Review Practices. *Management Learning*, 37(2), pp. 167–185.
- Nutt, P. C., 1990. Preventing decision debacles. *Technological Forecasting and Social Change*, 38(2), pp. 159–174.
- Nutt, P. C., & Wilson, D. C., 2010. *Handbook of decision making*. Chichester, UK: John Wiley & Sons, Ltd.
- Ocasio, W., 2012. Situated Attention, Loose and Tight Coupling, and the Garbage Can Model. In Alessandro Lomi, J. Richard Harrison (ed.) *The Garbage Can Model of Organizational Choice: Looking Forward at Forty (Research in the Sociology of Organizations, Volume 36)*, Emerald Group Publishing Limited, pp. 293–317
- Orlitzky, M., & Hirokawa, R. Y., 2001. To Err is Human, to Correct for it Divine A Meta-Analysis of Research Testing the Functional Theory of Group Decision-Making Effectiveness. *Small Group Research*, 32(3), pp. 313–341.
- Orton, J., & Weick, K., 1990. Loosely coupled systems: A reconceptualization. *Academy of Management Review*, 15(2), pp. 203–223.
- Oz, E. & Sosik, J.J., 2000. Why information systems projects are abandoned: A leadership and communication theory and exploratory study. *The Journal of Computer Information Systems*, 41(1), pp. 66–79.
- Ozkaya, E. H., Dabas, C., Kolev, K., Hult, G. T. M., Dahlquist, S. H., & Manjeshwar, S. A., 2013. An assessment of hierarchical linear modeling in international business, management, and marketing. *International Business Review*, 22(4), pp. 663–677.

- Packendorff, J., 1995. Inquiring into the temporary organization: new directions for project management research. *Scandinavian Journal of Management*, 11(4), pp. 319–333.
- Pan, G. S. C., & Flynn, D., 2003. Information Systems Project Abandonment: A Case of Political Influence by the Stakeholders. *Technology Analysis & Strategic Management*, 15(4), pp. 457–466.
- Pan, G., Hackney, R., & Pan, S., 2008. Information Systems implementation failure: Insights from prism. *International Journal of Information Management*, 28(4), pp. 259–269.
- Pan, G., & Pan, S. L., 2006. Examining the coalition dynamics affecting IS project abandonment decision-making. *Decision Support Systems*, 42(2), pp. 639–655.
- Pareto, V., 1964. *Cours d'Economie politique* (Nouvelle e.). Geneve: Librairie Droz.
- Patanakul, P., 2014. Managing large-scale IS/IT projects in the public sector: Problems and causes leading to poor performance. *The Journal of High Technology Management Research*, 25(1), pp. 21–35.
- Pich, M. T., Loch, C. H., & Meyer, A. de, 2002. On uncertainty, ambiguity, and complexity in project management. *Management Science*, 48(8), pp. 1008–1023.
- Pearson, C., & Clair, J., 1998. Reframing crisis management. *Academy of Management Review*, 23(1), pp. 59–76.
- Pedersen, M., 2014. *A quantitative examination of critical success factors comparing agile and waterfall project management methodologies*, Doctoral dissertation, Minneapolis, MN: Capella University.
- Perminova, O., Gustafsson, M., & Wikström, K., 2008. Defining uncertainty in projects—a new perspective. *International Journal of Project Management*, 26(1), pp. 73–79.
- Perrow, C., 1981. Normal Accident at Three Mile Island. *Society*, (July/August), pp. 8–26.
- Perrow, C., 1984. *Normal Accidents: Living with High Risk Technologies*, New York, NY: Basic Books.
- Perrow, C., 2004. A personal note on normal accidents. *Organization & environment*, 17(1), pp. 9–14.

- Perry, C., & Greig, I., 1975. Estimating the mean and variance of subjective distributions in PERT and decision analysis. *Management Science*, 21(12), pp. 1477-1480.
- Phalp, K.T. & Cox, K., Supporting Communicability with Use Case Guidelines : An Empirical Study. *Empirical Software Engineering*, pp. 1-16.
- Phan, D., 1990. *Information systems project management: an integrated resource planning perspective model*, Ph.D. Thesis, department of Management Information Systems, University of Arizona, Tucson, AZ.
- Pich, M.M.T., Loch, C.H. & Meyer, A. De, 2002. On uncertainty, ambiguity, and complexity in project management. *Management Science*, 48(8), pp. 1008-1023.
- Pidgeon, N., 1997. The Limits to Safety? Culture, Politics, Learning and Man-Made Disasters. *Journal of Contingencies and Crisis Management*, 5(1), pp. 1-14.
- Pidgeon, N., & O'Leary, M., 2000. Man-made disasters: why technology and organizations (sometimes) fail. *Safety Science*, 34(1-3), pp. 15-30.
- Pinto, J. K., & Mantel, S. J., 1990. The causes of project failure. *IEEE Transactions on Engineering Management*, 37(4), pp. 269-277.
- Pinto, J. & Prescott, J., 1988. Variations in critical success factors over the stages in the project life cycle. *Journal of management*, 14(1), pp. 5-18.
- Pinto, J.K. & Slevin, D.P., 1987. Critical factors in successful project implementation. *IEEE Transactions on Engineering Management*, EM-34(1), pp. 22-27.
- Pinto, J. & Slevin, D.P., 1988. Critical success factors in effective project implementation. Cleland, D. I., & King, W. R. (Eds.). *Project management handbook (Vol. 2)*. New York: Van Nostrand Reinhold. pp. 167-190.
- Pisarenko, V., & Sornette, D., 2003. Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution. *Pure and Applied Geophysics*, 160(12), pp. 2343-2364.
- PMI, 2004. *A guide to the project management body of knowledge*, (3rd ed.). Newton Square, PA: Project Management Institute.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P., 2003. Common method biases in behavioral research: a critical review of the literature and

- recommended remedies. *The Journal of Applied Psychology*, 88(5), pp. 879–903.
- Prest, A., & Turvey, R., 1965. Cost-Benefit Analysis: A Survey. *The Economic Journal*, 75(300), pp. 683–735.
- Price, D. D. S., 1976. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5), pp. 292–306.
- Priemus, H., & Wee, B. van, 2013. *International Handbook On Mega-Projects* (p. 480). Cheltenham, UK: Edward Elgar Publishing.
- Raffo, D., & Kellner, M., 2000. Empirical analysis in software process simulation modeling. *Journal of Systems and Software*, 53, pp. 31–41.
- Rai, A., Mann, J. E. C., Zhang, G. P., Keil, M., Ellen, J., & Mann, C., 2003. Why Software Projects Escalate: The Importance of Project Management Constructs. *IEEE Transactions on Engineering Management*, 50(3), pp. 251–261.
- Ram, J., & Corkindale, D., 2014. How “critical” are the critical success factors (CSFs)? Examining the role of CSFs for ERP. *Business Process Management Journal*, 20(1), pp. 151–174.
- Ramana, M. V., 2009. Nuclear Power: Economic, Safety, Health, and Environmental Issues of Near-Term Technologies. *Annual Review of Environment and Resources*, 34(1), pp. 127–152.
- Renkema, T. J., & Berghout, E. W., 1997. Methodologies for information systems investment evaluation at the proposal stage: a comparative review. *Information and Software Technology*, 39(1), pp. 1–13.
- Rerup, C., 2009. Attentional Triangulation: Learning from Unexpected Rare Crises. *Organization Science*, 20(5), pp. 876–893.
- Roberts, K.H., 1989. New challenges in organizational research: high reliability organizations. *Organization & Environment*, 3(2), pp. 111–125.
- Roberts, K.H., 1990. Some characteristics of one type of high reliability organization. *Organization Science*, 1(2), pp. 160–176.
- Roberts, K. H., & Rousseau, D. M., 1989. Research in Nearly Failure-Free, High Reliability Organizations: Having the Bubble. *IEEE Transactions on Engineering Management*, 36(2), pp. 132–139.

- Roberts, K. H., Stout, S. K., & Halpurn, J. J., 1994. Decision dynamics in two high reliability military organizations. *Management Science*, 40(5), pp. 614-624.
- Rockart, J. F., 1978. Chief executives define their own data needs. *Harvard Business Review*, 57(2), pp. 81-93.
- Rogers, E. M., & Shoemaker, F. F., 1971. *Communication of innovations: A cross-cultural approach*. New York: Free Press.
- Roux-Dufort, C., 2007. Is crisis management (only) a management of exceptions? *Journal of Contingencies and Crisis Management*, 15(2), pp. 105-114.
- Ross, J. & Staw, B.M., 1986. Expo 86 : An Escalation Prototype. *Administrative Science Quarterly*, 31(2), pp. 274-297.
- Ross, J. & Staw, B., 1993. Organizational escalation and exit: Lessons from the Shoreham nuclear power plant. *Academy of Management Journal*, 36(4), pp. 701-732.
- Rubin, J. Z., & Brockner, J., 1975. Factors affecting entrapment in waiting situations: The Rosencrantz and Guildenstern effect. *Journal of Personality and Social Psychology*, 31(6), pp. 1054-1063.
- Rubin, I. M., & Seelig, W., 1967. Experience as a factor in the selection and performance of project managers. *IEEE Transactions on Engineering Management*, 14(3), pp. 131-135.
- Ruhe, M., Jeffery, R., & Wiecek, I., 2003. Cost estimation for web applications. In *Proceedings of the 25th International Conference on Software Engineering, ICSE'03*, pp. 285-294.
- Sagan, S. D., 2004. Learning from Normal Accidents. *Organization & Environment*, 17(1), pp. 15-19.
- Samset, K., 2003. *Project Evaluation: Making Investments Succeed*. Trondheim: Tapir Academic Press.
- Santos, B. L. Dos., 1991. Justifying Investments in New Information Technologies. *Journal of Management Information Systems*, 7(4), pp. 71-90.
- Sassone, P. G., 1988. Cost benefit analysis of information systems: A survey of methodologies. *ACM SIGOIS Bulletin*, 9(2-3), pp. 126-133.

- Sauder, M., & Espeland, W. N., 2009. The Discipline of Rankings: Tight Coupling and Organizational Change. *American Sociological Review*, 74(1), pp. 63–82.
- Sauer, C., 1993. *Why information systems fail: a case study approach*, Henley-on-Thames: Alfred Waller.
- Sauer, C. & Cuthbertson, C., 2003. *The State of IT Project Management in the UK 2002-2003*, Templeton College, University of Oxford.
- Sauer, C., Gemino, A. & Reich, B.H., 2007. The impact of size and volatility on IT project performance. *Communications of the ACM*, 50(11), pp. 79–84.
- Schmidt, R., Lyytinen, K., Keil, M., & Cule, P., 2001. Identifying Software Project Risks: An International Delphi Study. *Journal of Management Information Systems*, 17(4), pp. 5–36.
- Schoemaker, P., 1990. Strategy, complexity and economic rent. *Management Science*, 36(10), pp. 1178–1192.
- Schwenk, C., 1988. The cognitive perspective on strategic decision making. *Journal of Management Studies*, 25(1), pp. 41–55.
- Shenhar, A. J., & Dvir, D., 1996. Toward a typological theory of project management. *Research Policy*, 25(4), pp. 607–632.
- Shenhar, A. J., Dvir, D., Levy, O., & Maltz, A. C., 2001. Project success: a multidimensional strategic concept. *Long Range Planning*, 34(6), pp. 699–725.
- Shenhar, A. J., Tishler, A., Dvir, D., Lipovetsky, S., & Lechler, T., 2002. Refining the search for project success factors: a multivariate, typological approach. *R&D Management*, 32(2), pp. 111–126.
- Simon, H. A., 1955. On a class of skew distribution functions. *Biometrika*, 42(314), pp. 425–440.
- Simon, H. A., 1962. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6), pp. 467–482.
- Simon, H. A., 1991. Bounded rationality and organizational learning. *Decision and Organization*, 2(1), pp. 161–176.
- Simon, H. A., & Bonini, C., 1958. The size distribution of business firms. *The American Economic Review*, 48(4), pp. 607–617.

- Sleesman, D.J., Conlon, D. E., McNamara, G. & Miles, J. E., 2012. Cleaning Up the Big Muddy: A Meta-Analytic Review of the Determinants of Escalation of Commitment. *Academy of Management Journal*, 55(3), pp. 541–562.
- Slutsky, E., 1925. Über stochastische Asymptoten und Grenzwerte. *Metron*, 5(3), pp. 3–89.
- Snook, S. A., 2002. *Friendly Fire: The Accidental Shootdown of U.S. Black Hawks over Northern Iraq* (p. 280). Princeton N.J.: Princeton University Press.
- Sornette, D., Andersen, J., & Simonetti, P., 2000. Portfolio theory for “fat tails.” *International Journal of Theoretical and Applied Finance*, 3(3), pp. 523–535.
- Sornette, D., & Ouillon, G., 2012. Dragon-kings: Mechanisms, statistical methods and empirical evidence. *The European Physical Journal Special Topics*, 205(1), pp. 1–26.
- Standish Group, 1994. *Chaos*. Boston, MA.
- Standish Group, 1999. *Chaos: A Recipe for Success*. Boston, MA.
- Standish Group, 2001. *Extreme Chaos*. Boston, MA.
- Standish Group, 2010. *Chaos Manifest*. Boston, MA.
- Stahlknecht, P., & Hasenkamp, U., 1997. *Einführung in Die Wirtschaftsinformatik (Springer-Lehrbuch)*. Berlin/Heidelberg: Springer.
- Starbuck, W. H., 2009. Perspective--Cognitive Reactions to Rare Events: Perceptions, Uncertainty, and Learning. *Organization Science*, 20(5), pp. 925–937.
- Starbuck, W. H., & Farjoun, M., 2005. *Organization at the Limit: Lessons from the Columbia Disaster*. Malden, MA: Blackwell Publishing.
- Starbuck, W. H., & Milliken, F. J. , 1988. Challenger: Fine-Tuning the Odds Until Something Breaks. *Journal of Management Studies*, 25(4), pp. 319–340.
- Staw, B. M. & Ross, J., 1978. Commitment to a Policy Decision: A Multi-Theoretical Perspective. *Administrative Science Quarterly*, 23(1), pp. 40–64.
- Staw, B. M., 1981. The Escalation of Commitment to a Course of Action. *The Academy of Management Review*, 6(4), pp. 577–587.
- Staw, B. M. & Ross, J., 1989. Understanding Behavior in Escalation Situations. *Science*, 246(4927), pp. 216–220.

- Staw, B. M., Sandelands, L. E., & Dutton, J. E., 1981. Threat-Rigidity Effects in Organizational Behaviour: A Multilevel Analysis. *Administrative Science Quarterly*, 26(4), pp. 501-524.
- Steele, M., & Huber, W., 2004. Exploring data to detect project problems. In *AACE International Transactions*, 32(12), pp. PM.21.1-PM.21.7.
- Stouffer, D. B., Malmgren, R. D., & Amaral, L. A. N., 2005. Comments on "The origin of bursts and heavy tails in human dynamics". *arXiv preprint physics/0510216*.
- Strum, D. P., May, J. H., & Vargas, L. G., 2000. Modeling the uncertainty of surgical procedure times: comparison of lognormal and normal models. *Anesthesiology*, 92(4), pp. 1160-1167.
- Tadikamalla, P. R., 1979. The lognormal approximation to the lead time demand in inventory control. *Omega*, 7(6), pp. 553-556.
- Taudes, A., Feurstein, M., & Mild, A., 2000. Options analysis of software platform decisions: a case study. *MIS Quarterly*, 24(2), pp. 227-243.
- Thomas, G., Fernandez, W., & Ferna, W., 2008. Success in IT projects: A matter of definition? *International Journal of Project Management*, 26(7), pp. 733-742.
- Tiwana, A., & Keil, M., 2004. The one-minute risk assessment tool. *Communications of the ACM*, 47(11), pp. 73-77.
- Tiwana, A., Wang, J., Keil, M., & Ahluwalia, P., 2007. The Bounded Rationality Bias in Managerial Valuation of Real Options: Theory and Evidence from IT Projects. *Decision Sciences*, 38(1), pp. 157-181.
- Topping, J., 1985. *Consultant Involvement in Information Systems Design*, Working Paper 85-01. Université Laval, Quebec.
- Trendowicz, A., Heidrich, J., Münch, J., Ishigai, Y., Yokoyama, K., & Kikuchi, N., 2006. Development of a hybrid cost estimation model in an iterative manner. *Proceeding of the 28th International Conference on Software Engineering - ICSE '06*, pp. 331-340.
- Tukey, J.W., 1977. *Exploratory Data Analysis*, London: Addison-Wesley.
- Turner, B., 1976. The organizational and interorganizational development of disasters. *Administrative Science Quarterly*, 21(3), pp. 378-397.
- Turner, B., 2005. Causes of disaster: sloppy management. *British Journal of Management*, 5(September 1993), pp. 215-219.

- Turner, B. & Pidgeon, N., 1997. *Man-Made Disasters, Second Edition*, Oxford: Butterworth-Heinemann.
- Tversky, A., & Kahneman, D., 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk & Uncertainty*, 5(4), pp. 297–323.
- United States Presidential Commission on the Space Shuttle Challenger Accident, 1986. *Report to the President: Actions to Implement the Recommendations of the Presidential Commission on the Space Shuttle Challenger Accident*. Washington, DC: National Aeronautics and Space Administration.
- Vaughan, D., 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*, Chicago, IL: University of Chicago Press.
- Venables, W. N., & Ripley, B. D., 2002. *Modern Applied Statistics with S* (4th ed.). Heidelberg: Springer.
- Verhoef, C., 2002. Quantitative IT portfolio management. *Science of Computer Programming*, 45(1), pp. 1–96.
- Verhoef, C., 2004. *Quantifying Software Process Improvement* (pp. 1–148). Amsterdam.
- Verhoef, C., 2005. Quantifying the value of IT-investments. *Science of Computer Programming*, 56(3), pp. 315–342.
- Verner, J. M., Evancho, W. M., & Cerpa, N., 2007. State of the practice: An exploratory analysis of schedule estimation and software project success prediction. *Information and Software Technology*, 49(2), pp. 181–193.
- Vogus, T. J., Rothman, N. B., Sutcliffe, K. M., & Weick, K. E., 2014. The affective foundations of high-reliability organizing. *Journal of Organizational Behavior*, (in press), pp. 5.
- Vuong, Q. H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 57(2), pp. 307–333.
- Wallace, L., Keil, M. & Rai, A., 2004a. How Software Project Risk Affects Project Performance: An Investigation of the Dimensions of Risk and an Exploratory Model. *Decision Sciences*, 35(2), pp. 289–321.
- Wallace, L., Keil, M. & Rai, A., 2004b. Understanding software project risk: a cluster analysis. *Information & Management*, 42(1), pp. 115–125.

- Walsh, J. P., 1995. Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane. *Organization Science*, 6(3), pp. 280–321.
- Watkins, M. D., & Bazerman, M. H., 2003. Predictable surprises: the disasters you should have seen coming. *Harvard Business Review*, 81(3), pp. 72–80.
- Weick, K.E., 1988. Enacted Sensemaking in Crisis Situations. *Journal of Management Studies*, 25(4), pp. 305–317.
- Weick, K.E., 1990. The Vulnerable System: An Analysis of the Tenerife Air Disaster. *Journal of Management*, 16(3), pp. 571–593.
- Weick, K. E., 1993. The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative Science Quarterly*, 38(4), pp. 628–652.
- Weick, K.E., 2004. Normal Accident Theory As Frame, Link, and Provocation. *Organization & Environment*, 17(1), pp. 27–31.
- Weick, K.E., 2006. Organizing for Mindfulness: Eastern Wisdom and Western Knowledge. *Journal of Management Inquiry*, 15(3), pp. 275–287.
- Weick, K. & Roberts, K., 1993. Collective mind in organizations: Heedful interrelating on flight decks. *Administrative science quarterly*, 38(3), pp. 357–381.
- Weick, K. E., & Sutcliffe, K. M., 2006. Mindfulness and the Quality of Organizational Attention. *Organization Science*, 17(4), pp. 514–524.
- Weick, K.E. & Sutcliffe, K.M., 2007. *Managing the Unexpected: Resilient Performance in an Age of Uncertainty* 2nd ed., San Fransico, CA:Jossey-Bass.
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D., 2005. Organizing and the Process of Sensemaking. *Organization Science*, 16(4), pp. 409–421.
- Whyte, G., 1986. Escalating Commitment to a Course of Action: A Reinterpretation. *Academy of Management Review*, 11(2), pp. 311–321.
- Whyte, G., 1989. Groupthink reconsidered. *Academy of Management Review*, 14(1), pp. 40–56.
- Whitty, S.J. & Maylor, H., 2009. And then came Complex Project Management (revised). *International Journal of Project Management*, 27(3), pp. 304–310.
- Wilcox, R., 1998. How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53(3), pp. 300–314.

- Wilks, S. S., 1937. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1), pp. 60-63.
- Williams, T., 1999. The need for new paradigms for complex projects. *International Journal of Project Management*, 17(5), pp. 269-273.
- Williams, T., & Samset, K., 2010. Issues in front-end decision making on projects. *Project Management Journal*, 41(2), pp. 38-49.
- Williams, V. S., Jones, L. V., & Tukey, J. W., 1999. Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics*, 24(1), pp. 42-69.
- Woodhouse, E. J., 1988. Sophisticated trial and error in decision making about risk. In M.E. Kraft, N.J. Vig (Eds.), *Technology and Politics*. Durham, NC: Duke University Press.
- Xia, W., & Lee, G., 2004. Grasping the complexity of IS development projects. *Communications of the ACM*, 47(5), pp. 69-74.
- Yoon, K.-A., Kwon, O.-S., & Bae, D.-H., 2007. An Approach to Outlier Detection of Software Measurement Data using the K-means Clustering Method. *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, pp. 443-445.
- Yourdon, E., 1997. *Death march: managing "mission impossible" projects*, Upper Saddle River, NJ: Prentice Hall.
- Yule, G. U., 1925. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 213(402-410), pp. 21-87.
- Zapata, A. H., & Chaudron, M. R. V., 2012. An analysis of accuracy and learning in software project estimating. In *2012 38th Euromicro Conference on Software Engineering and Advanced Applications*, pp. 414-421.
- Zaretski, R. L., Briggs, W. M., Sterling, M., & Shankar, M., 2009. Distributions of large-scale power outages: Extreme values and the effect of truncation. *International Journal of Power and Energy Systems*, 29(1).
- Zhang, G. P., Keil, M., Rai, A., & Mann, J., 2003. Predicting information technology project escalation: A neural network approach. *European Journal of Operational Research*, 146(1), pp. 115-129.

- Zimmerman, D. W., 1994. A note on the influence of outliers on parametric and nonparametric tests. *The Journal of General Psychology*, 121(4), pp. 391–401.
- Zipf, G., 1949. *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge Mass.: Addison-Wesley Press.
- Zmud, R., 1980. Management of large software development efforts. *MIS Quarterly*, 4(2), pp. 45–55.