Making-Sense of the Impact and Importance of Outliers in Project Management through the Use of Power Laws

Alexander Budzier[a,*], Bent Flyvbjerg[a]

[a] *BT Centre for Major Programme Management, Saïd Business School, University of Oxford*

Abstract

The academic literature and popular press has chronicled large IT project failures for the last 40 years. Two points of contention surround this debate. First, quantitative studies found mixed support of a wide-spread crisis, questioning the representativeness of failure cases. Second, organizational theories disagreed on underlying assumptions about the nature of uncertainty, in particular about stability, locus of control, and controllability of the causes of IT project disasters. To advance the understanding of these two gaps four hypotheses were tested with a sample of 4,227 IT projects. The findings showed that outliers are stable phenomena following power laws, occurrence and impact of outliers differs between public and private sector, benefits management is associated with thinner tails and lower risk, and agile delivery methods do not statistically significantly influence the thickness of the tails. In sum, outliers are stable and non-random phenomena. They matter more than medians or means when it comes to IT project risk. Second, the notion of outliers bridges the gap between qualitative and quantitative studies. The findings also show that causes of outliers are, at least to some extent, internal and controllable by organizations. Lastly, the paper draws implications for organizational decision-making, learning, and risk management.

Keywords: Outlier, project failure, cost overruns, schedule delay, benefits shortfall, project risk, risk management

1. Outliers and Projects

> We cannot truly plan, because we do not understand the future – but this is not necessarily bad news. We could plan while bearing in mind such limitations. (Taleb, 2007b, p. 157)

The academic literature is rife with accounts of large failed IT projects. Among the most notorious cases are for example the Denver Airport Baggage Handling System (Montealegre and Keil, 1998), London Stock Exchange's Taurus (Drummond, 1997), London Ambulance Service

*Corresponding author

Email addresses: `alexander.budzier@sbs.ox.ac.uk` (Alexander Budzier), `bent.flyvbjerg@sbs.ox.ac.uk` (Bent Flyvbjerg)

(Beynon-Davies, 1999), and Fox Meyer's ERP project (Scott, 1998). Case of extreme organizational failure fascinate and attract substantial attention from journalists, politicians, and academics alike (Lampel et al., 2009). The problem of large IT project failures has been labelled "runaway projects" (Mahaney and Lederer, 1999; Glass, 1997), "White Elephants" (Shapira, 1997), "Black Swans" (Flyvbjerg and Budzier, 2011), or "Black Holes" (Keil and Mähring, 2010). But how representative are these cases for the field of IT projects?

Quantitative studies of project performance paint a slightly different picture (cf. table 1). They showed that the cry of a wide-spread crisis in IT project delivery is unwarranted (Glass, 2006). Nevertheless, the surveys report a wide range of project performance. Cost overruns span on average from -15% (Little and Graphics, 2006) to +286% (Eveleens and Verhoef, 2009). Similarly, the reported schedule slippage ranges from -8% (Jones, 2008) to +105% (Little and Graphics, 2006). The benefits shortfall ranges on average from -7%, a benefits over-delivery (Sauer et al., 2007), to +10% (Eveleens and Verhoef, 2009). The large variability, particularly of cost overruns, warrants attention.

The reported cost overruns (cf. table 1) show on the one hand, that extreme failures are rare. On the other hand, the large variability might be due to the impact of extreme failure cases on averages. Because of the low averages extreme failures are expected to be low frequency, high impact occurrences – in other words *outliers*. An outlier, in statisticians' language, is an observation that: "appears to deviate markedly from other members of the sample in which it appears" (Grubbs, 1969, p. 1). It is a subjective post-data manifestation (Barnett, 1978), which only reveals itself after the sample has been drawn and some analysis has taken place. Therefore, outliers are an outcome of sense-making and sense-giving to observations through statistical analysis. Outliers have been given many labels: unrepresentative, wild, straggler, rogue, spurious, maverick, alien contaminants (Anscombe, 1960; Barnett, 1978). They are usually thought of as aberrant, discordant, or anomalous observations (Anscombe, 1960; Hoaglin et al., 1986) that cause troubling distortions in datasets (Hoaglin et al., 1986; Hoaglin and Iglewicz, 1987). It is this troubling distortion, which might explain the variability in prior findings.

The large variability of reported means is a first indicator for the presence of outliers. A second indicator is the discrepancy between reported means and medians. Statistical rule-of-thumb suggests that the discrepancy between the mean and the median indicates skewness and the length of the tails of distributions (Agresti and Finlay, 2008). Only eight of the previous studies reported both median and mean cost overruns (cf. table 1). The discrepancies, nevertheless, indicate that the skew in the data might be considerable. The most extreme finding reported median cost performance of −38%, whereas the average cost performance was +286% (Eveleens and Verhoef, 2009). A very strong indication that outliers were present. Schedule outcomes show similar patterns. The largest reported difference is between +105% average and +79% median schedule overrun (Little and Graphics, 2006).

The high variability in means and the discrepancy of median and means in prior studies show that outliers are indeed a troublesome problem in studies of quantitative IT project performance. This observation is in line with Kitchenham and Pickard (1987), who found that the problem of outliers is often overlooked in software productivity studies. For instance, data used in estimating effort and cycle times in software development are not only positively skewed but also highly problematic because of outliers (Kitchenham and Pickard, 1987; Raffo and Kellner, 2000; MacDonell

Table 1: Quantitative Studies of Project Performance

Author	Year	Sample size	Project size	Success rate	Cost overrun	Schedule slippage	Benefits shortfall
Addison & Vallabh	2002	36 organizations	No performance data reported				
Augustine	1979	100 projects				33%	
Bergeron & St. Arnaud	1992	67 organizations 89 projects	1251 PD		33%		
Conte et al.	1986	No performance data given					
Emam Koru	2008	288 projects		84% (2005) 88% (2007)			
Eveleens & Verhoef	2009	Data not reported by organization surveyed					
Org X		867 projects			286% (-38%)		
Org Y		168 projects			16% (0%)		10% (0%)
Org Z		307 projects			0%		
Little	2002	1 organization 121 projects			-15%	105% (79%)	
Ewusi-Mensah & Przasnyski	1994	49 organizations 49 projects		28%			
Glass	2005	No performance data reported					
Glass	2006	No performance data reported					
Heemstra	1989	598 organizations 2659 projects		30%			
Huber, Sauer & Cuthbertson	2003	421 projects		16%	18%	23%	7%
Jenkins et al.	1984	23 organizations 72 projects	$103k	61%	67% (34%)	22%	
Jones	2007			62%			
Jones	2008	8290 projects		71%		-8%	
Jones	2000	9155 projects 560>10k FP		69%			
Kampstra & Verhoef	2009	No original data collection, based on Jones (2000) data					
Keil et al.	2010	579 projects		62%	119%, 18% in non-escalated projects	103% 22% in non-escalated projects	
Kulk et al.	2009	1 organization 165 projects	$2.24m		-4% (-10%)		
Lederer	1991		Total ~202 PD	40%			
Little	2006	1 organization 106 projects	329 days		80% (100%)		
McAulay	1987	120 organizations 280 projects	No performance data reported				
McKeen	1983	5 organizations 32 projects	$35k		28% (22%)	18% (22%)	
Moløkken-Østvold et al.	2004	11 organizations 36 projects	~400PD	11%	41% (21%)	25% (9%)	
Moores	2000	54 organizations 115 projects	No performance data reported				
Oz & Sosnik	2000	78 organizations 27 projects		2.66 abandoned projects per 5 years per organization			
Phan	1988	191 organizations 827 projects			33%		
Sauer et al.	2007	412 projects			13%	20%	-7%
Standish	1994	365 organizations			189%	222%	
Standish	1996				142%	131%	
Standish	1998				69%	79%	
Standish	2000				45%	63%	
Standish	2002	80,000 projects in total (ca. 5000 per annual survey)	Average $3-5m Median $1-3m	34%	43%	82%	67%
Standish	2004			29%	56%	84%	64%
Standish	2006			35%	47%	72%	68%
Standish	2008			32%	54%	79%	67%
Standish	2010			37%	46%	71%	74%
Topping	1985	22 projects			40% (26%)		
Verhoef	2002	No original data collected using Jones (2000)					
Wydenbach	1995	213 organizations 515 projects	No performance data reported				

and Shepperd, 2003).

Only few prior studies report medians and averages, even fewer quantify the occurrence of outliers. In these five studies outlier occurrences ranged from 33% (Kulk, 2009), 10% in Yoon et al. (2007), 2-5% (Mitchell and Zmud, 1999; Banker and Kauffman, 1991), to as low as 0.2% (Grant et al., 2006). Under the assumption of thin tails, for instance in a normal distribution, outliers would be expected to occur in less than 0.07% of all cases in both ends of the tail (Tukey, 1977; McGill et al., 1978). These studies challenge the thin-tail assumption. At the same time the 33% outlier rate reported by Kulk (2009) is extremely high.

Four conventional strategies exist to deal with outliers: rejection, accommodation, incorporation, and identification (Barnett, 1978).

Rejection is the most common method to deal with outliers. This holds true for project management studies, most academic studies simply exclude outliers, e.g., 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); Banker and Kauffman (1991); Andersen (1990); Mitchell and Zmud (1999).

Only few studies argue that "an outlier is not to be thrown out (due to its unusualness) but rather might be the clue to data behaviors that are not revealed by the rest of the information." (Steele and Huber, 2004, p. PM21.3). *Accommodation* means that the inferences are made by adjusting the method with which the data is analyzed and interpreted. One example of accommodating outliers is to use median instead of averages to describe centrality and inter-quartile ranges instead of standard deviation to describe variability or to log-transform variables prior to regression analyses (an example is Eveleens and Verhoef, 2009).

Incorporation replaces the underlying assumption of thin tails with a new model that brings the outliers back into the expected range of observations. This strategy is found predominantly in operations management. Scheduling studies, particularly studies advancing the PERT method, abandon the model of thin-tailed distributions and replace it with fat tail models, most often log-normal, beta or Weibull distributions (Hahn, 2008; Chapman and Ward, 2003; Keefer and Bodily, 1983; Abdelkader, 2004). Less common are other distributions such as triangular (Williams, 1992; Johnson, 1997; MacCrimmon and Ryavec, 1964), exponential (Kulkarni and Adlakha, 1986), or polynomial (Schmidt and Grossmann, 2000).

Identification, the final strategy, makes inferences while treating outliers as their own self-contained group. For example, MacDonell and Shepperd (2003) first classify outliers through n-means clustering and then make inferences about outliers as a special subgroup of the sample. An alternative approach to clustering is to forecast outliers through neural networks (Finnie et al., 1997).

In sum, prior studies of large IT project failure cases established the existence of low frequency high impact events. Quantitative studies of IT project performance have debunked the alarmist claims of a crisis in IT project management. They show that, if any, high impact events occurred with a low frequency. Nevertheless, the reported cost, schedule, and benefit performance show a large amount of variability. Moreover, the studies reported discrepancies between median and means. Both, variability and discrepancy, are indicators for outliers. Outliers are defined as observations that markedly deviate from the sample. Outliers are problematic not only in the field of IT project management, but in management in general. However, only few prior studies report the rate of outliers. Four basic strategies exist to deal with outliers: rejection, accommodation, incorporation, and identification. Prior studies show that outlier rejection is the most common choice.

The gap between qualitative studies, with thick descriptions and root cause analyses of notorious IT project failures, and between quantitative studies of IT project performance is not purely academic. In practice, the extreme impact of outliers also isolates those individual instances (Roux-Dufort, 2007). Rare events in organization are typically portrayed by managers as unique, unprecedented, or even uncategorizable (Christianson et al., 2008). Outliers are often treated as "accidental manifestations of underlying organizational processes" (Lampel et al., 2009, p. 835). In hindsight, they are often declared to have been experiments all along, which prevents organizational learning and improvements (Baumard and Starbuck, 2005). Additionally, labeling an event a 'perfect storm' or a 'Black Swan' has been criticized for implying that the event is uncontrollable and that, because nothing can be done about it, nothing is done to prevent or manage the risks of rare events (Paté-Cornell, 2012).

Lastly, the theoretical and practical importance of the gap between qualitative and quantitative studies is not purely phenomenological. Organisational theory points towards fundamentally different understandings of the nature and causes of uncertainty.

2. Organizational Theories of Outliers

Three schools of thought of organizational theory have theorized outliers: the system-, event-, and process-centric view. Each school differs in their understanding and assumption about the nature and causes of uncertainty.

The *system-centric* school of thought depicts outliers as normal accidents (Perrow, 1981). Normal Accident Theory postulates that high system complexity, in terms of interactive complexity and tight coupling, inevitably leads to system failures. These failures are often wrongly attributed to operator error but their true root causes lie in organizational and technical design choices that pre-date the operational phase of these systems (Perrow, 1981, 1984, 2004; Weick, 2004).

The second school of thought is the *event-centric* view. These theories view organizational failure as the result of the occurrence of an external event followed by an insufficient response (Lampel et al., 2009; Starbuck, 2009; Rerup, 2009; Quarantelli, 1988). Dominant among this view is the rich literature on crisis management (c.f. reviews by Quarantelli, 1988; Roux-Dufort, 2007; Weick, 1988; Boin, 2004), which also has found its way into the discussion of IT project management (for example Angell and Smithson, 1990; Brown et al., 1998; Chartier et al., 2010).

The third school of thought is the *process-centric* view, where organizational failure is the result of a slow build-up of smaller errors over time. Three prominent process models exist. First, the man-made disaster model argues that technical errors are miscommunicated or misunderstood and thus allowed to build-up over time. At the end of this incubation period small errors have amplified into large disaster (Turner and Pidgeon, 1997). The second model describes how the organizational and cultural process of risk managements leads to a normalization of deviance (Vaughan, 1996, 1997). Over time, the circle of accepted risks slowly enlarges. The third process model is the escalation of commitment. In this model locally rational decision cause an organization to persevere on a dysfunctional course of action despite negative feedback (Staw and Ross, 1986; Staw, 1981;

Staw and Ross, 1989; Drummond, 1997; Keil et al., 1994, 2000). In all three process models the true causes of the organizational failure remain hidden to the organization until after fact.

Each school of thought developed different core concepts to describe outliers, in particular accidents, crises, and disasters. The communality between these conceptualizations is that the future does not go according to plan. The future either holds unintended consequences of social action (Merton, 1936) or "intendedly rational action fails" outright (Turner and Pidgeon, 1997, p. 6). In the system-centric view, the failures is with the intention of designing a safe, highly reliable system (Perrow, 1981; Weick and Sutcliffe, 2007). In the event-centric view, the intention to be resilient and prepared, if an unforeseen event occurs, fails (Roux-Dufort, 2007; Starbuck, 2009). In the process-centric view, the organizational decision making and risk mitigation actions have unintended negative consequences (Vaughan, 1997; Staw and Ross, 1986).

The three different schools of thought point to a more profound difference. Each makes fundamentally different assumptions about the future and the nature of uncertainty. The differences can be unpacked using the multi-dimensional model of spontaneous causal thinking (Weiner, 1985). The model groups assumptions about causes in three dimensions. The first dimension is *stability*: Are cause-effect relationships stable? The second dimension is *locus of control*: Are causes of outliers internal or external to the organization? The third dimension is *control*, also referred to as controllability: Are causes of outliers influenced by the organization? (Weiner et al., 1976; Weiner, 1985, 2008).

Table 2 unpacks the assumptions made by the three different schools of thought in organizational theory. First, the system-centric and process-centric view see the cause-effect relationship as stable. Normal Accident Theory, for example, postulates that high complexity inevitably causes outliers (Perrow, 1984). The escalation of commitment shows how difficult it is to pull the plug on a failing outlier (Keil, 1995). Conversely, the event-centric view describes the causes of outliers as unique, idiosyncratic, and not equivalent (Christianson et al., 2008).

Second, the system-centric view and the process-centric view locate the cause of outliers within the organization, in complex designs or dysfunctional organizational processes (Vaughan, 1997; Staw and Ross, 1986). The event-centric view locates it external to the organization (Christianson et al., 2008).

Lastly, the system-centric view sees the causes of outliers as controllable, they are unintended consequences of design decisions (Perrow, 1984; Turner and Pidgeon, 1997) and can be prevented by mindful attention to early warning signs (Levinthal and Rerup, 2006; Weick and Sutcliffe, 2006). Whereas the event-centric view argues they are uncontrollable external occurrences (Pearson and Clair, 1998) and the process-view portrays the causes as dysfunctional but rational outcome of local action (Vaughan, 1997; Staw and Ross, 1986).

3. Research questions

This paper sets out two goals: (1) bridge the gap between qualitative and quantitative research in large failures of (IT) projects and (2) bridge the gap between the three theoretical schools of thought. Following a multi-dimensional model of spontaneous causal thinking questions arise that inquire into the different assumptions about the nature and the causes of uncertainty. The first question is: Are outliers a stable phenomenon?

Table 2: Differences between system-, event-, process-centric views

	System-centric	**Event-centric**	**Process-centric**
Intention	Safe design	Resilience	Robustness
Assumptions			
- Stability	Stable	Unstable	Stable
- Locus	Internal	External	Internal
- Control	Yes	No	No
Implications	(Design) decisions	Planning & Response	Practice & behavior

If outliers are not a stable phenomenon they manifest because of random chance events. A conceptualization that is supported by the event-centric view. The opposing conceptualization, held by the process- and system-centric view, is that outliers are not idiosyncratic, but that outliers are a stable phenomenon. In statistical terms, the problem translates into question whether outliers are from a different population of uncertainty than the rest of the sample. If this were the case, the appropriate strategy is to reject and exclude the outliers. If this is not the case, then outliers are stable extrapolations of smaller overruns: they are from the same population of risk albeit from a different position in the spectrum.

> *Hypothesis 1:* Outliers are from a different population of uncertainty than smaller failures.

The second question is whether the causes of outliers are internal or external to the organization. The third question asks whether the causes are controllable or not. Theories from the event-centric view argue that the causes and trigger events are external and uncontrollable. They are chance events that occur and require a response. The process-centric view locates the causes internal but not controllable, because the organization is blind to slow processes that build-up risk or is unable to stop a failing course of action despite receiving negative feedback. The system-centric view also locates the causes internally but views them as controllable through choices of organizational and technological design.

To test these questions three organizational parameters have been chosen that are internal to the organization but have a varying degree of controllability: public versus private sector, whether benefits were quantified, and whether an agile project method was used.

First, empirical studies that compare the between the public and the private sector identify three characteristics how both sectors differ. Public sector organizations are more bureaucratic, public manager are less materialistic, and public sector managers have a weaker organizational commitment (Boyne, 2002). Others have noted different incentive structures and that the private sector tends to have more complex goals which subsequently are more difficult to quantify (Rainey et al., 1976). Others argued that both sectors are culturally different, for instance that the public sector makes more risk averse decisions (Bozeman and Kingsley, 1998). Lastly, some scholars argued that public projects are more political (Pitsis et al., 2003). In sum, while differences between the public and the private sector exist, they tend to be structural and cultural, no discussion could be found that shows that public and private sector projects differ in how they are managed and executed or differ in management function and processes. In terms of controllability, it can be argued that the sector setting is invariant to project management decisions. In other words, the organization has little control over the sector it operates in. Even if there would be none but a cultural difference, changing the public culture it is a difficult long-term process (Fernandez and Rainey, 2006) and attempts such as public-private partnerships have had mixed success (Goldberg, 2006).

Hypothesis 2: Outliers are differ between the public versus private sector.

Second, benefits management is an under-researched topic (Cooke-Davies, 2002, p. 186), yet an important practice. Benefits management is the umbrella of a wide variety of activities, including measurement and modeling of benefits, as well as planning and management activities to harvest benefits (Morris, 2004). Sauer et al. (2009, p. 5) describe benefits management as «...understanding exactly what the project is trying to achieve in terms of business benefits, what it's trying to deliver.» They point out that it is a planning activity that is typically carried out in the front-end planning process and is the extension of business strategy into project management (Sauer et al., 2009). Cooke-Davies' analysis shows that if a benefits management process is present in organizations the chances of success improve (Cooke-Davies, 2000, 2002). Although important, benefits quantification is difficult in practice. Not all projects that theoretically could quantify and manage benefits actually do so (Ward et al., 2007). Nevertheless, it can be argued that the decision to quantify and manage benefits is controllable by the organization and to some extent under the influence of a project's management team.

Hypothesis 3: Outliers are associated with benefits management.

Thirdly, agile project delivery is an organizational practice related to how the project delivers its functionality. IT projects have been at the forefront of agile practices (Geraldi et al., 2008). The key objective of agile methods is to react to change by focussing on creativity and problem-solving rather than focussing management attention on deterministic engineering processes (Dyba and Dingsoyr, 2008). It has been argued that agile methods improve project performance and thus reduce the risks of IT projects (Opelt et al., 2013). Dyba and Dingsoyr (2008) review the academic literature and find that of 1,996 published articles only 33 present empirical evidence. Not only does little empirical evidence exist it is also inconclusive regarding commonly made claims by the agile practitioner community (Dyba and Dingsoyr, 2008, pp. 851-853). Nevertheless, the project is in control over the decision to adopt an agile development method.

Hypothesis 4 Outliers are associated with agile methodology.

4. Data and Methodology

This study collected data on estimated and actual cost and schedule of 4,227 projects. The project estimate is defined as the cost, schedule, and benefits baseline at the date of the decision to build. The date of the decision to build was probed to be the full business case, typically the basis of a top management decision to begin tendering for implementation partners. Estimated project size and duration is defined as the total forecasted project budget, from the owner's not the supplier's perspective. Actual data is defined as the final reported project cost, schedule, and benefits achieved of the project.

These data were collected from documentary evidence through 4 key channels: (1) project archaeology, (2) desk study, (3) freedom of information requests, and (4) budgeting data. First, the data of 1,299 projects were collected from 27 organizations; those organizations volunteered to take part in this study and the data collection asked organizations to provide 20-30 of their recently finished projects. Second, a desk study of audit reports and reliable academic studies collected data on 275 projects from 37 organizations. Third, freedom-of-information requests were used to collect data on 1,029 projects from 58 organizations in the American public sector. Fourth, the data from recent budget requests was obtained from the U.S. Office of Management and Budget; this yielded another 1,624 projects from 120 organizations.

3,605 projects had a reliable cost baseline and reliable actual outturn cost. For only 944 projects we established a reliable schedule baseline and actual schedule.

The projects are comparatively large. The average planned project budget is USD 24.6 million (in 2010 terms), the median planned size is USD 1.2 million. The actual project costs were on average USD 21.3 million, with a median actual cost of USD 1.2 million. The total actual value of the sample is USD 85.0 billion. The average estimated duration of the projects in the sample is 280 days; median duration is 182 days. The average actual duration of the projects is 309 days with a median of 207 days.

The projects span a wide field of activities. Most are system implementation (89.9%), followed by IT infrastructure (5.4%), communication technology (2.7%), IT architecture (0.5%), and others (1.4%), which includes among others IT post merger-management, IT outsourcing. The most common system types among the implementation projects are supply chain management systems (26.9%), enterprise resource planning system (19.7%), human resource management systems (9.5%), and management information systems (4.9%).

5. Analysis and Discussion

Table 3 shows the project performance. Median cost overrun, schedule overrun, and benefit shortfall are ±0.00%. The averages differ quite widely from the median. The average cost overrun is +107.2%, the average schedule overrun is +37.3%, the average benefits shortfall is 29.3%. This finding is noteworthy for three reasons. First, the average cost overrun is in line with previously reported findings (e.g., Eveleens and Verhoef, 2009; Jenkins et al., 1984; Keil and Mähring, 2010; Little and Graphics, 2006). Nevertheless the median of ±0% is surprising. On the one hand, it is in line with findings by Little and Graphics (2006); Eveleens and Verhoef (2009). On the other hand, the difference between the mean and the median is considerable. A magnitude previously found only in one of the three organizations surveyed by Eveleens and Verhoef (2009). Secondly, the schedule overrun is in line with findings in Sauer and Cuthbertson (2003); Jenkins et al. (1984); Augustine (1982); Keil and Mähring (2010). Thirdly, the benefits shortfalls are slightly greater than the reported figure in Eveleens and Verhoef (2009). Again, the difference between mean and median schedule and benefits performance is a clear indication that considerable skew might be present in the data.

The conventional definition is that outliers are 1.5 inter-quartile ranges away from the edge of the box in a box plot (Tukey, 1977). Thus outliers are observations with a cost overrun greater than +105.9% and a schedule delay of more than +90.0% (cf. table 4). Data on benefits is very sparse, but the few projects indicate that a shortfall of -149.1% or more signifies an outlier. Benefit shortfalls of more than 100% are possible, if, for instance, a project is planned to reduce operations expenditures through replacing a legacy system but the actual running cost of the replacement are higher than the running cost of the legacy system. In total, 12.8% of the projects are classified as cost overrun outliers and 12.6% as schedule overrun outliers (cf. table 4).

Table 3: Project Outcomes

	Mean	**Median**	**IQR**	**n**
Cost overruns	+107.2%	±0.0%	0.458	3,650
Schedule overruns	+37.3%	±0.0%	0.360	946
Benefits shortfalls	-29.3%	±0.0%	0.994	64
Effort overruns	+119.8%	+24.8%	1.143	142

Table 4: Thresholds and Outlier Rates

	Cost	**Schedule**	**Benefits**
Upper whisker	+105.9%	+90.0%	-149.1%
Absolute frequency of outliers in the right tail	466	119	1
n	3650	946	64
Relative frequency of outliers in the right tail	12.8%	12.6%	1.6%

Hypothesis 1: Outliers are from a different population of uncertainty than smaller failures.

The first null-hypothesis states that large overruns are not from a different population of uncertainty than smaller overruns. To test the first hypothesis the tail of the distribution needs to be analyzed (Mandelbrot, 1960, 1963; Fama, 1965). Outliers are not outliers if the observed, deviant values can be fitted by a fat tailed distribution (Pisarenko and Sornette, 2003). A common family of extreme value distributions are power laws or pareto distributions. Power laws were initially discovered by geographers describing city sizes (Auerbach, 1913) and later first used by Economists to study the concentration of wealth (Pareto, 1964). Power laws are commonly found in physics, biology, and social sciences (reviews for example in Clauset et al., 2000; Virkar and Clauset, 2012; Mitzenmacher, 2001), for instance in word frequency, citations of academic papers, visits to a website, or the magnitude of earthquakes (Newman, 2005). Power laws were introduced into organizational studies to explain the size distribution of business firms (Simon and Bonini, 1958).

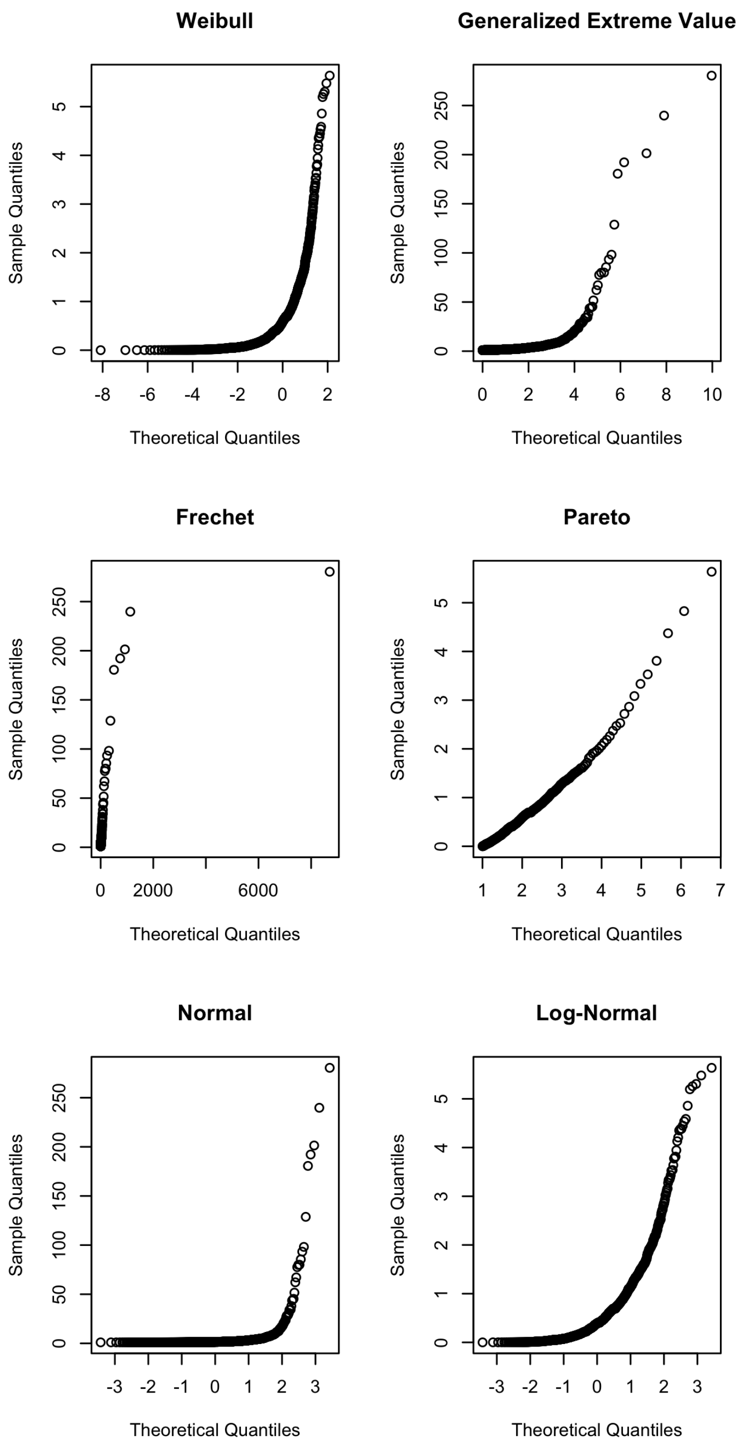
A random variable is said to have a power law distribution if $Pr[X \geq x] \sim cx^{-\alpha}$. The exponent α represent the fatness of the tails of the distribution. The smaller the power law exponent the higher the impact of outliers on the center of the distribution (Taleb, 2007b,a).

The most unbiased estimator of power laws is the Maximum Likelihood estimator (Clauset et al., 2000; Goldstein et al., 2004). The cumulative density function is defined as $F(X) = 1 - \left(\frac{x_m}{x}\right)^{\alpha}$ for $x \in [x_m, \infty)$. If $x_m = 1$ it follows that for $x > 1$ the probability density function simplifies to $f(x) = \frac{\alpha}{x^{\alpha+1}}$. The log-likelihood function of the distribution is $l(\alpha) = \log(L(\alpha)) = n \cdot \log(\alpha) - (\alpha + 1) \sum \log(x_i)$. Maximizing the log-likelihood $\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} - \sum \log(x_i) \stackrel{!}{=} 0$ gives the parameter estimator $\hat{\alpha} = \frac{n}{\sum \log(x_i)}$.

First, taking all projects with a positive cost overrun and excluding missing values results in 1,631 observations in the right tail of the distribution. Second, the Q-Q plots of the normal distribution and common extreme value distributions (cf. figure 1) confirm that a Pareto distribution is the most suitable fit. If the empirical distribution corresponds to a theoretical distribution the points of the Q-Q plot resemble a straight line, whose slope is influenced by the fitted parameters (Karian and Dudewicz, 2010). Even if not a perfect 45 degree angle, only the Q-Q plot for a pareto distribution resembles a straight line. This indicates the suitability of fitting a power law. Third, computing the estimator $\hat{\alpha} = \frac{n}{\sum \log(x_i)}$ results in $\hat{\alpha} = 1.609759$.

The result shows, that the cost overruns of IT projects have fat tails. Because the estimated exponent is between 1 and 2, the distribution has an expectancy value but does not converge for higher order moments such as variance, skewness, and kurtosis. Thus, the distribution is fat tailed but not heavy tailed.

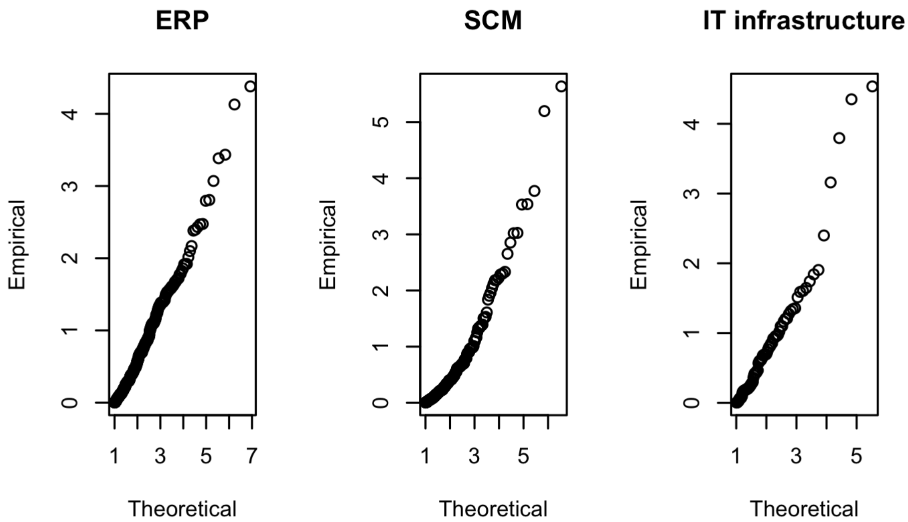
Figure 1: Q-Q Plots for normal distribution and common extreme value distributions



The observation of fat tails is commonly challenged by two alternative explanations (Fama, 1965). First, fat tails can be the spurious result of mixing normal distributions with the same mean and a different variance structure. Secondly, fat tails can be the spurious result of mixing observations over a non-stationary parameter.

To test the first alternative explanation the subsample is split by project type. Figure 2 shows the Q-Q plot for the power law distribution for ERP projects, SCM projects, and IT infrastructure projects. All three subsamples show the presence of fat tails.

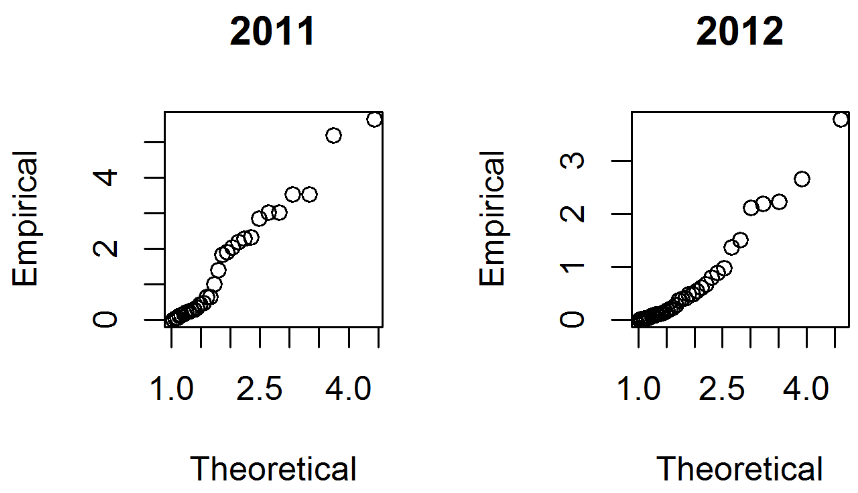
Figure 2: Q-Q Power Law Plot for Different Project Types



The second alternative explanations is that fat tails are an outcome of non-stationary parameters. For instance, in the analysis of stock market prices fat tails appear if the risk of the company changes over time. In the context of project management, fat tails could be caused by changes in project performance that come from changes in the organization. In other words changes of the underlying generating processes. To test the second alternative explanation we analyzed the project portfolio of one participating private sector organization. This organization conducted a transformation of their IT department during the data collection. 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 difference is nearly significant at $p = 0.061$ (using a non-parametric Wilcoxon test). Thus the underlying performance has changed. So did pooling the data prior and post this organizational change create spuriously fat tails? The 2011 and 2012 Q-Q plots both indicate the presence of fat tails (cf. figure 3).

Both alternative explanations failed to make the potentially spurious fat tails disappear. Thus the finding supports the conclusion that the observed fat tails are neither an artifact of aggregating normal distributions with same means but different variances nor of aggregating outcomes across non-stationary parameters.

Figure 3: Q-Q Plot before and after Reorganization



In sum, the analysis of the first hypothesis validated that the data is far from being normally distributed. The tails of the distribution are fatter than log-normal or Weibull tails, which are both distributions commonly used to incorporate outliers. The tail can be fitted with a Pareto distribution, with a shape parameter of 1.6. Two common alternative explanations fail to account for the fat tails. First, fat tails can be an artifact of mixing normal, thin-tailed distributions with the same mean and different variance structures. Q-Q plots show that subsamples of different project types all show similar fat tails. Second, fat tails can be an artifact of mixing thin-tailed distributions over time, when an underlying parameter changes. The portfolio of one participating organization shows, that despite improvements in project delivery after a reorganization of the IT department, fat tails are continuously present. Again, the alternative explanation demonstrates that fat tails are a stable phenomenon. The fat tails are not an artifact from aggregating performance over non-stationary parameters. Thus the analysis can reject hypothesis one that outliers are from a different population of uncertainty.

Hypothesis 2: Outliers are differ between the public versus private sector

Hypotheses 2-4 question whether the organization can influence the risk of outliers. The test strategy for these hypotheses is three-fold.

First, a Q-Q plot is used to visually confirm the suitability of power laws to describe the subsamples indicated by the hypothesis. Second, a likelihood ratio test assesses whether a model with two subgroups is significantly better than a model for the pooled distribution. A model with more parameters, in the case of two subgroups with two shape parameters instead of one, will always fit better. The likelihood ratio test answers the question whether a model then is significantly better. The likelihood ratio test statistic D is defined as $D = 2(log(likelihood\, alternative\, model) - 2(log(likelihood\, null\, model)$. The null hypothesis states that the pooled model offers the better fit of the distribution. If H_0 can not be rejected the sample should not be split into subgroups.

In the case of the alternative model being a two subgroup model and the null model being the pooled sample D is $D = 2\left(n_1 log(\alpha_1) - (\alpha_1 + 1)\sum \log(x_1) + n_2 log(\alpha_2) - (\alpha_2 + 1)\sum \log(x_2)\right) - 2(n_1 + n_2)log(\alpha) + (\alpha + 1)\sum \log(x)$, with n_1 and n_2 the sizes of the two subsamples, and the total sample size of $n = n_1 + n_2$. D is approximately χ^2 distributed with degrees of freedom equal to $df\,(alternative\,model) - df\,(null\,model)$. In the case where H_0 compares a two subsamples model with a model of one sample $df = 2$. Third, the difference of shape parameters is tested. The null hypothesis states that the shape parameters in both groups are equal: $H_0 : \alpha_1 = \alpha_2$. The maximum likelihood estimator used to estimate α from the data is asymptotically normal distributed. Thus under H_0 the difference between the estimators is normally distributed around 0 with $\hat{\alpha}_1 - \hat{\alpha}_2 \dot{\sim} N\left(0, \frac{\alpha_1}{n_1} + \frac{\alpha_2}{n_2}\right)$. The difference can be standardized so that the resulting test statistic z is $N(0,1)$ distributed: $z = \frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\sqrt{\frac{\alpha_1^2}{n_1} + \frac{\alpha_2^2}{n_2}}} \dot{\sim} N(0,1)$.

Hypothesis 2 asks whether outliers are associated with the organization's sector. An important cultural context of projects, which is difficult to control. First, a Q-Q plot confirms that a Pareto distribution is appropriate for the private and the public sector subsample (cf. figure 4). Splitting the tails of the distribution results in two subsamples with $n_{Private} = 368$ and $n_{Public} = 1,263$. The empirical average cost overrun of the subsamples is +93.3% in the private sector and +111.6% in the public sector. The medians are ±0% in both cases. The respective IQRs are 0.323 and 0.554. While both sectors have the same median the variation in the public sector is greater than in the private sector. Moreover, the higher average cost overrun in the public sector indicates a stronger influence of outliers. The estimated distributional parameters are $\hat{\alpha}_{Private} = 2.003$ and $\hat{\alpha}_{Public} = 1.523$. The greater alpha estimate in the private sector shows that the tails in the private sector are thinner than in the public sector. The expectancy value $\frac{\alpha}{\alpha - 1}$, for the private sector is +99.7% for the public sector it is +191.4%. Thus if a project escalates, i.e., overruns its budget, the cost are expect to nearly double in the private sector to nearly triple in the public sector. A finding that is in line with the escalated projects reported by Keil and Mann (1997).

Secondly, the likelihood ratio tests the the null-model of the pooled sample against the alternative model of splitting the sample between the public and the private sector. The likelihood ratio test is statistically significant ($D = 20.380$, one-tailed χ^2 test, $p < 0.001$).

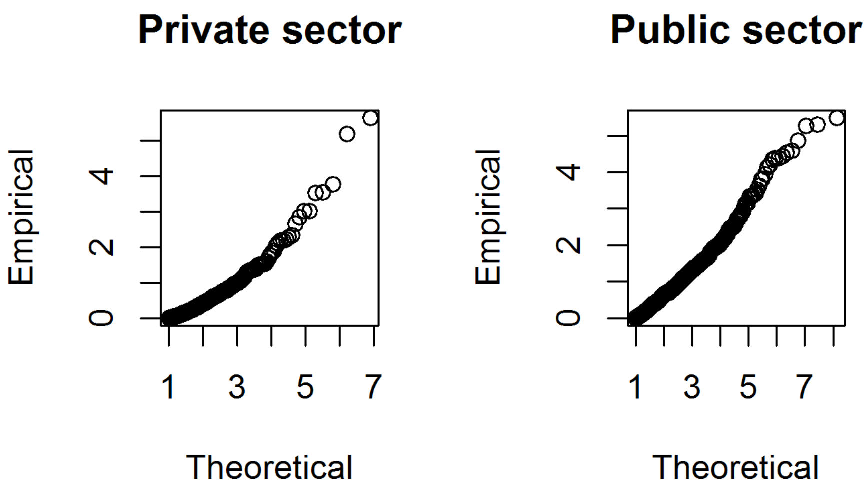
Thirdly, the standardized difference between the maximum likelihood estimators is $z = \frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\sqrt{\frac{\alpha_1^2}{n_1} + \frac{\alpha_2^2}{n_2}}} = 4.259$. The test value z is asymptotically normally distributed with $p < 0.001$.

Replicating the analysis for schedule overruns, results in an estimated $\hat{\alpha}_{Private} = 2.107$ and $\hat{\alpha}_{Public} = 2.361$. The much smaller difference in the thickness of the tails results in a likelihood ratio between the split and the pooled model of $D = 0.250$, which corresponds to a $p = 0.882$. Similarly the standardized difference between the estimators has a test value of $z = 0.483$, which corresponds to $p = 0.629$.

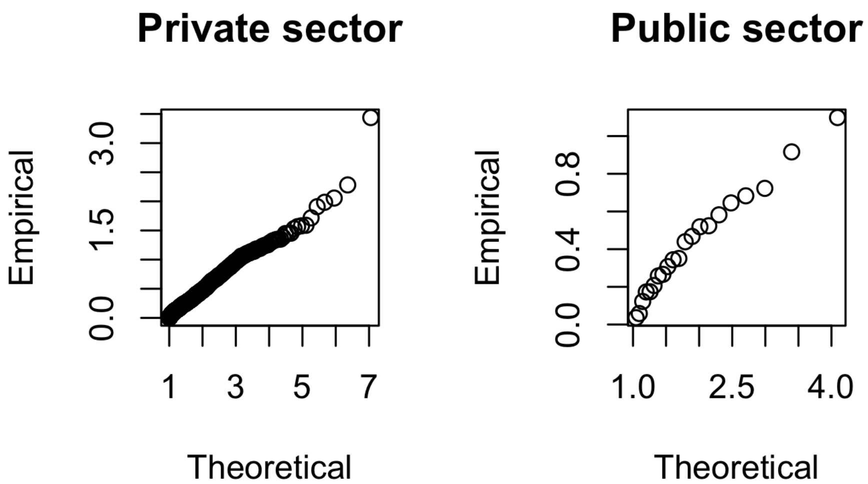
Thus the public and the private sector have statistically significantly different tails when it comes to cost performance of their projects, but not in terms of schedule performance.

Figure 4: Q-Q plot sector comparison

(a) Cost overruns



(b) Schedule overruns



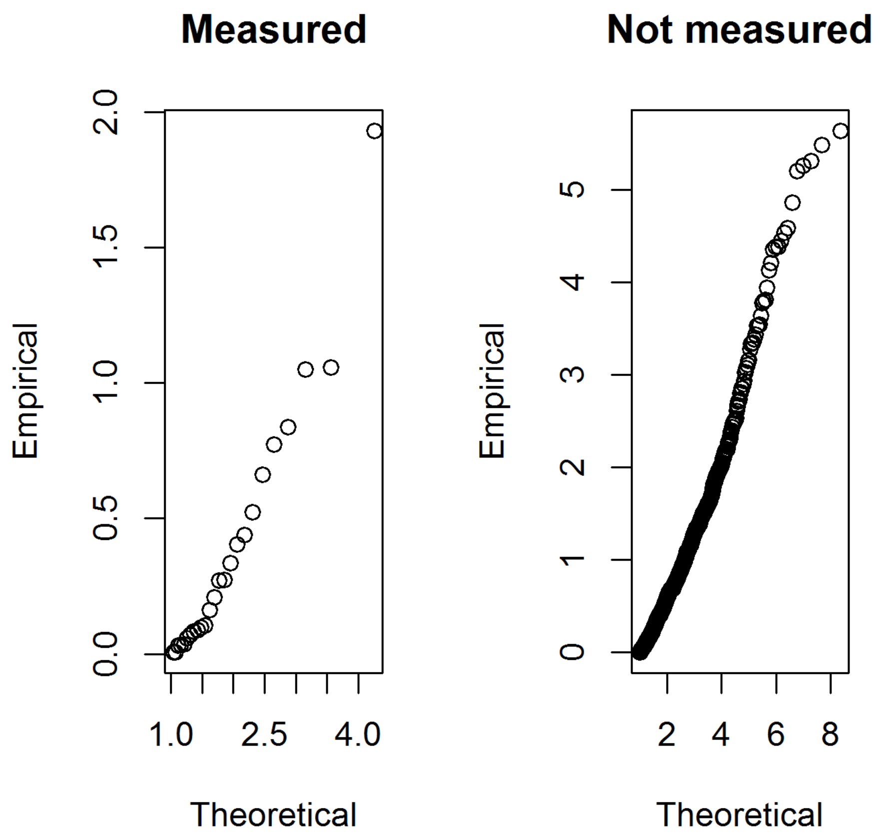
Hypothesis 3: Outliers are associated with benefits management

The third and fourth hypotheses test whether, organizational practices are associated with the fatness of tails. Hypothesis three inquires into benefits management, a project practice often problematized in the literature. To test the effect of this practice, the sample is split between projects that measured benefits and projects that did not measure benefits. This results in two sub samples with 25 and 1,606 projects respectively. The number is noteworthy, as only 25 projects that measured benefits had a cost overrun. Both subsamples show fat tails that might be fitted with a pareto distribution (cf. figure 5).

The maximum likelihood estimator for the shape parameters are $\hat{\alpha}_{Benefits\,measured} = 2.619729$ and $\hat{\alpha}_{Benefits\,not\,measured} = 1.600156$. The likelihood ratio test is nearly significant, thus providing some statistical evidence that the alternative model of splitting the sample into two subsamples is statistically significantly better than the null model of a pooled sample ($D = 5.130699$, one-tailed χ^2-test, $p = 0.077$). The standardized difference between the two shape parameters is $z = 1.940325$, which corresponds to $p = 0.052$. Because only 12 reliable data points of schedule overruns are available for projects that measured benefits, a reliable replication of this analysis for schedule performance is not feasible.

However, despite the very small sample size, the analysis offers some statistical evidence to the hypothesis that benefits management, an organizational practice, is associated with the thickness of the tails.

Figure 5: Q-Q plot benefits measurement



Hypothesis 4: Outliers are associated with agile methodology

The fourth hypothesis analyzes whether the delivery method has an influence on the thickness of the tails. Delivery methods are often the choice of the project management team. Agile development methods have been suggested to improve project performance and thus reduce the risks of projects (for example Dyba and Dingsoyr, 2008; Blunden, 2010; Tengshe and Noble, 2007; Opelt et al., 2013). The subgroups tested by this hypothesis are formed based on information provided by the project manager about the project. One question in the post mortem questionnaire, used in the data collection, asked whether the project was delivered incrementally or as a big bang. The scale from 1-10 was anchored as follows:

- Q4 The project followed an incremental approach.
  - (1) The project was a classic Big-Bang
  - (4) The project is done in a few big releases
  - (7) The project is done in multiple, short (<6 months) release cycles
  - (10) The project was highly agile and iterative

199 projects answered the question, of those cost performance data was reliably available for 147 projects. 9 of these 147 projects turned into outliers based on the IQR definition. The median answer to the question was 7. The median was used to split the sample into two groups of projects: those with less agile methodologies, projects who answered 1-7, and those with more agile methodologies, projects that were scored 8-10 by their project managers in the post-mortem questionnaire.

Table 5 shows the project performance of both subgroups. Non-parametric Wilcox tests showed that both subgroups do not significantly differ in their median cost, schedule, or benefits performance.

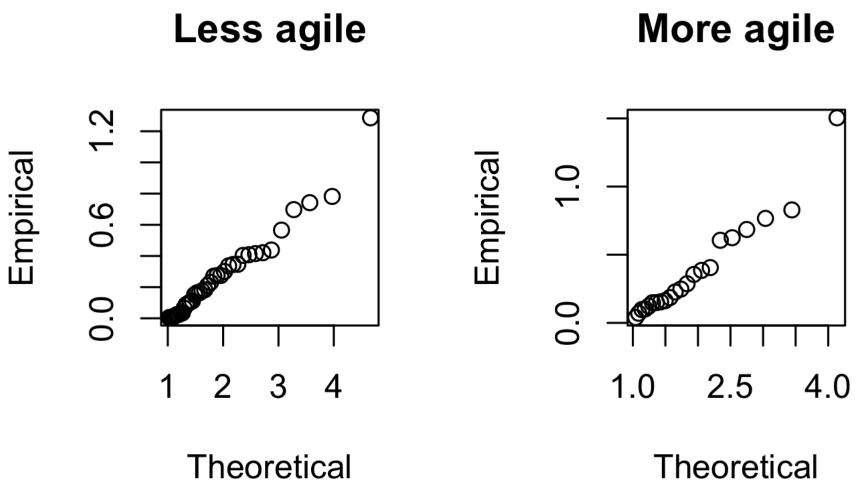
Table 5: More agile versus less agile projects

Dimension		Less agile	More agile	p-value
Cost	Median	±0.0%	±0.0%	0.8315
	Mean	+13.2%	+18.1%	
Schedule	Median	+14.3%	+3.7%	0.2643
	Mean	+45.0%	+27.2%	
Benefits	Median	±0.0%	-98.8%	0.1146
	Mean	-13.6%	-77.5%	
	n	131	61	

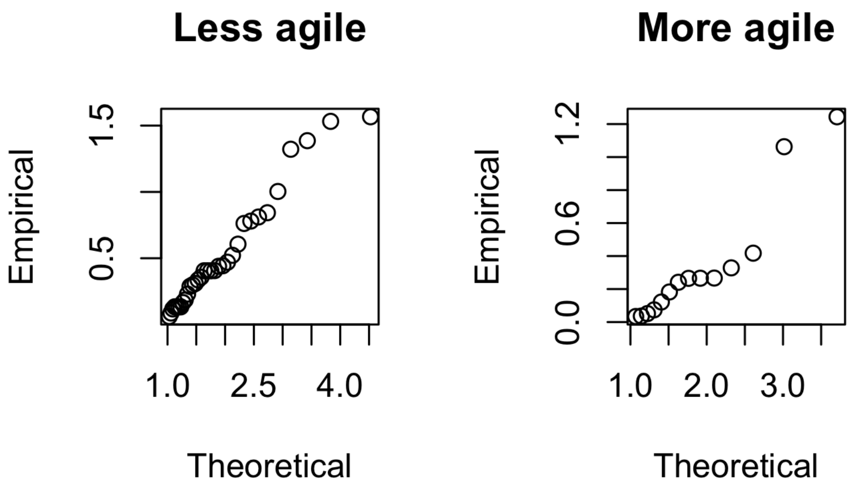
The Q-Q plots for both subgroups (cf. figure 6) confirmed the suitability of fitting a pareto distribution. Fitting the tail results in the estimate of $\hat{\alpha}_{Less\,agile} = 3.72$ and $\hat{\alpha}_{More\,agile} = 2.70$. The comparison between the two nested models has a likelihood ratio of $D = 1.474$, which is approximately χ^2 distributed with 2 degrees of freedom, which corresponds to $p = 0.478$. The result does not support splitting the model into two models for each subgroups. The final step tests whether the shape parameters of the two groups are significantly different. The test statistic $z = 1.2263$ corresponds to a $p = 0.220$, again the test can not reject the null hypothesis and offers no support that the shape parameters are significantly different.

Figure 6: Pareto Q-Q plot distribution of agile projects

(a) Cost overruns



(b) Schedule overruns



Replicating the same analysis for the schedule performance results in ML estimates of $\hat{\alpha}_{Less\,agile} = 1.93$ and $\hat{\alpha}_{More\,agile} = 3.05$. The comparison between the two nested models has a likelihood ratio of $D = 1.915$, which is approximately χ^2 distributed with 2 degrees of freedom, which corresponds to $p = 0.384$. The result does not support splitting the model into two models for each subgroups. Testing the standardized difference of the shape parameters yields the test statistic $z = 1.268$, which corresponds to a $p = 0.205$. Again the test can not reject the null hypothesis and offers no support that the shape parameters of the schedule performance are significantly different.

6. Conclusions

The analyses rejected the first hypothesis. It showed that projects with very large cost overruns or schedule delays are not statistical outliers from a different population of uncertainty than smaller overruns. The 'outliers' are smooth extrapolations of smaller overruns. The analysis found support for the second hypothesis that the thickness of tails differs between the private and the public sector. We argued based on prior research that the divide between the public and the private sector is predominantly cultural more than an actual difference how projects are managed and planned. The analysis also found some statistical evidence to support the third hypothesis that benefits management thins out the tails. Yet, it failed to support the fourth hypothesis that agile delivery methods have a similar impact. These findings have theoretical and practical implications.

Theoretical implications

This research project set out to bridge the methodological gap in the existing literature. Where, on one side, case studies documented and analyzed large project failure and, on the other side, quantitative studies survey and explain project performance by mostly excluding those outliers. The focus on qualitative methods to understand outliers corresponds to the theoretical stance that large-scale failures are unique and an idiosyncratic phenomena. This notion is most pronounced in the literature on crisis management but the notion is also part of process-oriented theories such as the man-made disasters theory and the theory of escalation of commitment.

Statisticians have argued that what is perceived to be a statistical outlier under one assumed regime of uncertainty is a 'normal' or 'non-outlier' observation in other regimes of uncertainty. Building on the notion of fat tail distributions, the analysis showed that cost and schedule overruns of IT projects follow a power law. This result showed that large cost and schedule overruns are not fundamentally different from smaller overruns.

The first theoretical implication of this finding is that projects that suffered from large overruns are not unique. Overruns themselves are not idiosyncratic chance events. They show regular and stable patterns.

Second, the analysis implies that rejecting outliers or grouping outliers into a cluster of special observations is statistically unsound. Outliers need to and can be incorporated into analyses. The findings showed stable patterns in the occurrence of outliers. These patterns can be used to bridge the divide between qualitative and quantitative studies of outlier projects. Further research is needed to better understand the generators of uncertainty in projects.

Third, the analysis explored the locus of control and the controllability of large overruns. It showed that the thickness of tails differs in the public and the private sector. Prior literature has argued that both sectors are set apart by a cultural divide not a difference in how public and private organizations manage and plan their IT projects. The findings also underline the need for more research into better understanding not only the differences in project management in both sectors, but how they generate and manage uncertainty.

Fourth, the study explored the impact of two practices of project management: benefits management and agile delivery methodology. Only benefits management showed significantly different tails. The findings imply that at least some of the explanatory factors of outliers of cost and schedule overruns are internal to the organization and controllable. This supports the argument that large-scale failures of IT projects are not random chance events, a notion which is dominant in the crisis and disaster management literature.

Additionally, the finding, that benefits management thins out the tail, has implications for theories of the escalation of commitment. The analysis showed that projects that quantified benefits had thinner tails, in other words fewer and smaller outliers. Escalation of commitment explains large-scale IT project failures through a string of rational decisions that lead to a dysfunctional course of action despite negative feedback. The findings showed that framing effects (Kaplan, 2011), particularly framing decisions as not only a trade-off between cost and schedule but also benefits, reduces the likelihood and impact of escalation. Prior discussion focussed on framing effects in terms of gains or losses (Bazerman and Samuelson, 1983; Brockner et al., 1986; Brockner, 1992), this analysis implies that structural elements also play a role.

Lastly, the research has also shown that the thickness of the tail offers a dependent variable that is suitable for explanatory analyses. The thickness of the tail of power laws or pareto distributions is expressed by the exponent alpha. However, one can argue that the value of such a model is limited for two reasons. Firstly, fitting a tail parameter has limited value in predictive models. A small change in the parameter value leads to large change in forecasted risk (Taleb, 2013). For example, the P-80 level of overruns has been recommended as a conservative estimate of risk (Flyvbjerg, 2006). The P-80 level corresponds to an overrun that in 8 of 10 cases was not exceed. On the flip side, in 2 out of 10 cases the cost overruns exceeded the P-80 level. The P-80 level cost contingency required for a distribution with an α of 1.5 is +197%, the same contingency for a distribution with α of 1.6 is +175%. The small change from 1.5 to 1.6 adds 22 percentage points of contingency, which is large for purposes of risk predictions.

Secondly, the tail exponent has a statistical error of $\sigma = \frac{\hat{\alpha}-1}{\sqrt{n}}$. Directly, following from the previous point, the width of the confidence interval of an estimate of $\alpha = 1.6$ is only smaller than 0.1 for samples of 600 observations and above. Thus the ML estimator, even though being the best available, retains a large error. Again limiting the predictive power of any such model. For the case in point, the large required sample sizes might explain the failure to reject hypothesis four. The analysis shows that unless more large-n research into the aspects of project execution is done research into outliers might be limited.

Practical implications

The most important implication is that cost and schedule risk of IT projects have fat tails. The fatter the tails the greater the impact of outliers on the center of the distribution (Taleb, 2013). Thus organizations that steer the performance of their projects on averages alone are blind to outliers and their impact. The challenger disaster (Vaughan, 1996) has shown that hidden processes slowly enlarge the risk accepted by organizations. Measuring performance and managing risks without problematizing outliers is again one way how organizations hide and accept large risks without them knowing.

Second, large failures attract journalistic, academic, and popular attention. Despite this attention and fascination failed projects, particularly if large, do not lead to organizational learning (Baumard and Starbuck, 2005). One reason lies in the framing of failure. Framing a failed project and its managers as victims to external, random, and idiosyncratic circumstances inhibits learning. The analysis showed that outliers are neither unique nor freak events. Furthermore, the analysis demonstrated that to understand, learn, and prevent failure managers need not to look externally or at large failures alone. Small overruns foreshadow large overruns.

Moreover, the comparison of the private and the public sector made the case for transfer learning from the private to the public sector. However, more research is needed to understand the specific differences between public and private sector project management and how that generates such different populations of uncertainty.

Lastly, the findings showed that benefits management is an important practice to manage uncertainty. Projects that manage benefits have thinner tails. The finding also underlined the importance of the front-end process. Benefits management might well be the single biggest deficiency in project management.

While the results for agile projects were mixed. The average schedule overrun significantly decreased; yet neither cost nor schedule fat tails were reduced. Agile methods seem to reduce schedule risk, but neither benefits nor cost risk. On the one hand it shows, quite expectedly, that agile methods are neither panacea nor silver bullet. However, more research is needed particularly because of the limited statistical power of smaller sample sizes.

In sum, three important findings emerge: Outliers are more prevalent than previously thought. They are part of the phenomenon, not a unique random event that should be rejected from quantitative studies and can only be analyzed qualitatively. The causes of outliers are at least partially found to have internal and controllable contributors, particularly the sector, i.e., cultural differences, and organizational practices, such as benefits management, matter.

References

Abdelkader, Y. H., Sep. 2004. Evaluating project completion times when activity times are Weibull distributed. European Journal of Operational Research 157 (3), 704–715.

Agresti, A., Finlay, B., 2008. Statistical Methods for the Social Sciences (4th Edition). Pearson.

Ahmed, M. a., Ahmad, I., AlGhamdi, J. S., Feb. 2013. Probabilistic size proxy for software effort prediction: A framework. Information and Software Technology 55 (2), 241–251.

Andersen, O., 1990. The use of software engineering data in support of project management. Software Engineering Journal (November), 350–357.
URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=61749`

Angell, I. O., Smithson, S., 1990. Managing information technology: A crisis of confidence? European management journal 8 (1), 27–36.
URL `http://www.sciencedirect.com/science/article/pii/0263237390900469`

Anscombe, F., 1960. Rejection of outliers. Technometrics 2 (2), 123–147.
URL `http://amstat.tandfonline.com/doi/abs/10.1080/00401706.1960.10489888`

Auerbach, F., 1913. Das Gesetz der Bevölkerungskonzentration. Petermanns Geographische Mitteilungen 59, 74–76.

Augustine, N. R., 1982. Augustine's Laws and Major System Development Programs. American Institute of Aeronautics & Astronautics, Reston, VA.
URL http://www.amazon.co.uk/Augustines-Major-System-Development-Programs/dp/0915928620

Banker, R. D., Kauffman, R. J., 1991. Reuse and Productivity in Integrated Computer-Aided Software Engineering : An Empirical Study. MIS Quarterly 15 (3), 375–401.

Barnett, V., 1978. The study of outliers: purpose and model. Applied Statistics 27 (3), 242–250.
URL http://www.jstor.org/stable/10.2307/2347159

Barros, M. D. O., Werner, C. M. L., Travassos, G. H., Feb. 2004. Supporting risks in software project management. Journal of Systems and Software 70 (1-2), 21–35.
URL http://linkinghub.elsevier.com/retrieve/pii/S0164121202001553

Baumard, P., Starbuck, W. H., Jun. 2005. Learning from failures: Why it May Not Happen. Long Range Planning 38 (3), 281–298.
URL http://linkinghub.elsevier.com/retrieve/pii/S0024630105000348

Bazerman, M., Samuelson, W., 1983. I won the auction but don't want the prize. Journal of Conflict Resolution 27 (4), 618–634.
URL http://jcr.sagepub.com/content/27/4/618.short

Beynon-Davies, P., 1999. Human error and information systems failure: the case of the London ambulance service computer-aided despatch system project. Interacting with Computers 11 (October 1992), 699–720.
URL http://www.sciencedirect.com/science/article/pii/S0953543898000502

Blunden, C., 2010. Clean delivery: An experience report of collaborative lean software delivery.

Boin, A., 2004. Crisis Management Vol. II. SAGE, Los Angeles, CA.
URL http://onlinelibrary.wiley.com/doi/10.1111/1477-7053.00085/abstract

Boyne, G., 2002. Public and private management: what's the difference? Journal of management studies 39 (1), 97–105.

Bozeman, B., Kingsley, G., 1998. Risk culture in public and private organizations. Public Administration Review 58 (2), 109–118.
URL http://www.jstor.org/stable/10.2307/976358

Brockner, J., Jan. 1992. The Escalation of Commitment to a Failing Course of Action: Toward Theoretical Progress. Academy of Management Review 17 (1), 39–61.
URL http://www.jstor.org/stable/10.2307/258647 http://www.jstor.org/stable/258647

Brockner, J., Houser, R., Birnbaum, G., Lloyd, K., Deitcher, J., Nathanson, S., Rubin, J. Z., 1986. Escalation of Commitment to an Ineffective Course of Action: The Effect of Feedback Having Negative Implications for Self-Identity. Administrative Science Quarterly 31 (1), 109–126.

Brown, W. J., Malveau, R. C., McCormick, H. W. I., Mowbray, T. J., 1998. Anti-Patterns - Refactoring Software , Architectures , and Projects in Crisis Executive Summary Overview. John Wiley & Sons, Inc., New York, New York, USA.

Chapman, C., Ward, S., Oct. 2003. Constructively simple estimating: a project management example. Journal of the Operational Research Society 54 (10), 1050–1058.
URL http://www.palgrave-journals.com/doifinder/10.1057/palgrave.jors.2601610

Chartier, A., Banville, C., Landry, M., 2010. Crisis management in information system projects: lessons of a case study. Canadian Journal of Administrative Sciences 27 (2), 148–160.
URL http://onlinelibrary.wiley.com/doi/10.1002/cjas.151/abstract

Christianson, M. K., Farkas, M. T., Sutcliffe, K. M., Weick, K. E., Oct. 2008. Learning Through Rare Events: Significant Interruptions at the Baltimore & Ohio Railroad Museum. Organization Science 20 (5), 846–860.
URL http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1080.0389

Clauset, A., Shalizi, C. R., Newman, M. E. J., 2000. Power-law distributions in empirical data. Arxiv preprint (1), 1–26.
URL http://arxiv.org/abs/0706.1062

Cooke-Davies, T., Apr. 2002. The ‚Äúreal‚Äù success factors on projects. International Journal of Project Management 20 (3), 185–190.
URL http://linkinghub.elsevier.com/retrieve/pii/S0263786301000679

Cooke-Davies, T. J., 2000. Towards Improved Project Management Practice. Uncovering the evidence for effective practices through empirical research. Ph.D. thesis, Leeds Metropolitan University.

Drummond, H., 1997. Escalation in Decision-Making: The Tragedy of Taurus. Oxford University Press, Oxford, UK.

Dyba, T., Dingsoyr, T., Aug. 2008. Empirical studies of agile software development: A systematic review. Information and Software Technology 50 (9-10), 833–859.
URL http://linkinghub.elsevier.com/retrieve/pii/S0950584908000256

Eveleens, J., Verhoef, C., Nov. 2009. Quantifying IT forecast quality. Science of Computer Programming 74 (11-12), 934–988.

Fama, E., 1965. The behavior of stock-market prices. Journal of business 38 (1), 34–105.
URL http://www.jstor.org/stable/10.2307/2350752

Fernandez, S., Rainey, H., 2006. Managing successful organizational change in the public sector. Public Administration Review (April), 168–175.

Finnie, G., Wittig, G., Desharnais, J.-M., Dec. 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), 281–289.
URL http://linkinghub.elsevier.com/retrieve/pii/S0164121297000551

Flyvbjerg, B., Aug. 2006. FROM NOBEL PRIZE TO PROJECT MANAGEMENT: GETTING RISKS RIGHT. Project Management Journal 37 (3), 5–15.

Flyvbjerg, B., Budzier, A., 2011. Why your IT project may be riskier than you think. Harvard Business Review 89 (9), 23–25.

Geraldi, J. G., Rodneyturner, J., Maylor, H., Soderholm, a., Hobday, M., Brady, T., Turner, J. R., So, A., Jul. 2008. Innovation in project management: Voices of researchers. International Journal of Project Management 26 (5), 586–589.
URL http://linkinghub.elsevier.com/retrieve/pii/S0263786308000756

Glass, R. L., 1997. Software Runaways: Monumental Software Disasters. Prentice Hall.

Glass, R. L., Aug. 2006. The Standish Report: Does It Really Describe a Software Crisis? Communications of the ACM 49 (8), 15–16.

Goldberg, J. L., 2006. The Fatal Flaw in the Financing of Private Road Infrastructure in Australia. In: Australasian Transport Research Forum. Brisbane, p. 20.

Goldstein, M. L., Morris, S. a., Yen, G. G., Sep. 2004. Problems with fitting to the power-law distribution. The European Physical Journal B 41 (2), 255–258.
URL http://www.springerlink.com/index/10.1140/epjb/e2004-00316-5

Grant, K. P., Cashman, W. M., Christensen, D. S., 2006. Delivering projects on time. Research-Technology … (Nov-Dec), 52–60.
URL http://www.ingentaconnect.com/content/iri/rtm/2006/00000049/00000006/art00008

Grubbs, F., 1969. Procedures for detecting outlying observations in samples. Technometrics 11 (1), 1–21.

Hahn, E., 2008. Mixture densities for project management activity times: A robust approach to PERT. European Journal of Operational Research, 1–18.

Hoaglin, D., Iglewicz, B., 1987. Fine-tuning some resistant rules for outlier labeling. Journal of the American Statistical … 82 (400), 1147–1149.
URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1987.10478551

Hoaglin, D. C., Iglewicz, B., Tukey, J. W., 1986. Performance of Some Resistant Rules for Labeling Outlier. Journal of the American Statistical Association 81 (396), 991–999.

Jenkins, A., Naumann, J. D., Wetherbe, J. C., Apr. 1984. Empirical investigation of systems development practices and results. Information & Management 7 (2), 73–82.
URL http://linkinghub.elsevier.com/retrieve/pii/0378720684900120

Johnson, D., 1997. The triangular distribution as a proxy for the beta distribution in risk analysis. Journal of the Royal Statistical Society: Series D (The Statistician) 46 (3), 387–398.
URL http://onlinelibrary.wiley.com/doi/10.1111/1467-9884.00091/abstract

Jones, C., 2008. Applied Software Measurement: Global Analysis of Productivity and Quality. McGraw-Hill Prof Med/Tech.

URL http://books.google.com/books?id=mj7yQiQEusUC&pgis=1
Kaplan, S., May 2011. Research in Cognition and Strategy: Reflections on Two Decades of Progress and a Look to the Future. Journal of Management Studies 48 (3), 665–695.
URL http://doi.wiley.com/10.1111/j.1467-6486.2010.00983.x
Karian, Z. A., Dudewicz, E. J., 2010. Handbook of Fitting Statistical Distributions with R. CRC Press, Boca Raton.
Keefer, D., Bodily, S., 1983. Three-point approximations for continuous random variables. Management Science 29 (5), 595–609.
URL http://mansci.journal.informs.org/content/29/5/595.short
Keil, M., Dec. 1995. Pulling the Plug: Software Project Management and the Problem of Project Escalation. MIS Quarterly 19 (4), 421–447.
Keil, M., Mähring, M., 2010. Is Your Project Turning into a Black Hole? California Management Review 53 (1), 6–31.
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 Hawwaii International Conference on System Sciences. Vol. 00. p. 10.
Keil, M., Mann, J., Rai, A., Dec. 2000. Why Software Projects Escalate: An Empirical Analysis and Test of Four Theoretical Models. MIS Quarterly 24 (4), 631–664.
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.
Kitchenham, B., Pickard, L., 1987. Towards a constructive quality model. Part 2: Statistical techniques for modelling software quality in the ESPRIT REQUEST project. Software Engineering Journal (July).
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4807895
Kulk, G. P., 2009. IT Risks in Measure and Number. Ph.D. thesis, Vrije Universiteit Amsterdam.
Kulkarni, V., Adlakha, V., 1986. Markov and Markov-regenerative PERT networks. Operations Research 34 (5), 769–781.
URL http://or.journal.informs.org/content/34/5/769.short
Lampel, J., Shamsie, J., Shapira, Z., Sep. 2009. Experiencing the Improbable: Rare Events and Organizational Learning. Organization Science 20 (5), 835–845.
URL http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1090.0479
Levinthal, D., Rerup, C., Jul. 2006. Crossing an Apparent Chasm: Bridging Mindful and Less-Mindful Perspectives on Organizational Learning. Organization Science 17 (4), 502–513.
URL http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1060.0197
Little, T., Graphics, L., 2006. project estimation the Cone of Uncertainty, 48–54.
MacCrimmon, K., Ryavec, C., 1964. An analytical study of the PERT assumptions. Operations Research 12 (1), 16–37.
URL http://or.journal.informs.org/content/12/1/16.short
MacDonell, S. G., Shepperd, M. J., May 2003. Combining techniques to optimize effort predictions in software project management. Journal of Systems and Software 66 (2), 91–98.
URL http://linkinghub.elsevier.com/retrieve/pii/S0164121202000675
Mahaney, R., Lederer, A., 1999. Runaway information systems projects and escalating commitment. Proceedings of the 1999 ACM SIGCPR . . . , 291–296.
URL http://dl.acm.org/citation.cfm?id=299719
Mandelbrot, B., 1960. The Pareto-Levy law and the distribution of income. International Economic Review 1 (2), 79–106.
URL http://www.jstor.org/stable/10.2307/2525289
Mandelbrot, B., 1963. The variation of certain speculative prices. Journal of business 36 (4), 394–419.
URL http://www.jstor.org/stable/10.2307/2350970
McGill, R., Tukey, J., Larsen, W., 1978. Variations of box plots. The American Statistician 32 (1), 12–16.
URL http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1978.10479236
Merton, R., 1936. The unanticipated consequences of purposive social action. American sociological review 1 (6),

894–904.
URL http://www.jstor.org/stable/10.2307/2084615
Mitchell, V. L., Zmud, R. W., Jul. 1999. The Effects of Coupling IT and Work Process Strategies in Redesign Projects. Organization Science 10 (4), 424–438.
URL http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.10.4.424
Mitzenmacher, M., 2001. A brief history of generative models for power law and lognormal distributions. INTERNET MATHEMATICS 1, 226 – 251.
Montealegre, R., Keil, M., Aug. 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.
Morris, P. W. G., 2004. The irrelevance of project management as a professional discipline Peter W G Morris.
Newman, M., Sep. 2005. Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46 (5), 323–351.
URL http://www.tandfonline.com/doi/abs/10.1080/00107510500052444
Opelt, A., Gloger, B., Pfarl, W., Mittermayr, R., 2013. Agile Contracts: Creating and Managing Successful Projects with Scrum. John Wiley & Sons, Hoboken, NJ.
Pareto, V., 1964. Cours d'Economie politique, nouvelle e Edition. Librairie Droz, Geneve.
URL http://books.google.com/books?hl=en&lr=&id=s249oYSJbVMC&pgis=1 http://www.worldcat.org/title/cours-deconomie-politique/oclc/10080936
Paté-Cornell, E., Nov. 2012. On "black swans" and "perfect storms": risk analysis and management when statistics are not enough. Risk analysis : an official publication of the Society for Risk Analysis 32 (11), 1823–33.
Pearson, C., Clair, J., 1998. Reframing crisis management. Academy of Management Review 23 (1), 59–76.
URL http://www.jstor.org/stable/10.2307/259099
Perrow, C., 1981. Normal accident at three mile island. Society (July/August), 8–26.
URL http://www.springerlink.com/index/B5861783763H2123.pdf
Perrow, C., 1984. Normal Accidents: Living with High Risk Technologies. Basic Books, New York.
Perrow, C., 2004. A personal note on normal accidents. Organization & environment 17 (1), 9–14.
URL http://oae.sagepub.com/content/17/1/9.short
Pisarenko, V., Sornette, D., 2003. Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution. Pure and Applied Geophysics.
URL http://www.springerlink.com/index/5m256kecpyalju5q.pdf
Pitsis, T. S., Clegg, S. R., Marosszeky, M., Rura-Polley, T., 2003. Constructing the Olympic dream: a future perfect strategy of project management. Organization Science 14 (5), 574–590.
Quarantelli, E., 1988. Disaster crisis management: A summary of research findings. Journal of Management Studies (July).
URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-6486.1988.tb00043.x/abstract
Raffo, D., Kellner, M., 2000. Empirical analysis in software process simulation modeling. Journal of Systems and Software 53, 31–41.
URL http://www.sciencedirect.com/science/article/pii/S0164121200000066
Rainey, H. G., Backoff, R. W., Levine, C. H., 1976. Comparing Public and Private Organizations. Public Administration Review 36 (2), 233–244.
Rerup, C., Aug. 2009. Attentional Triangulation: Learning from Unexpected Rare Crises. Organization Science 20 (5), 876–893.
URL http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1090.0467
Roux-Dufort, C., 2007. Is crisis management (only) a management of exceptions? Journal of Contingencies and Crisis Management 15 (2), 105–114.
URL http://onlinelibrary.wiley.com/doi/10.1111/j.1468-5973.2007.00507.x/full
Ruhe, M., Jeffery, R., Wieczorek, I., 2003. Cost estimation for web applications. In: ICSE'03 Software Engineering, 2003. Vol. 6. IEEE Computer Society, p. 10.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1201208
Sauer, C., Cuthbertson, C., 2003. The State of IT Project Management in the UK 2002 - 2003.
Sauer, C., Gemino, A., Reich, B. H., Nov. 2007. The impact of size and volatility on IT project performance. Com-

munications of the ACM 50 (11), 79–84.
Sauer, C., Reich, B. H., Horner, B., Feb. 2009. Rethinking IT project management: Evidence of a new mindset and its implications. International Journal of Project Management 27 (2), 182–193.
Schmidt, C. W., Grossmann, I. E., Nov. 2000. The exact overall time distribution of a project with uncertain task durations. European Journal of Operational Research 126 (3), 614–636.
URL http://linkinghub.elsevier.com/retrieve/pii/S0377221799003161
Scott, J. E., 1998. The FoxMeyer Drugs ' Bankruptcy : Was it a Failure of ERP ? Organizational factors. Delta, 223–225.
Shapira, Z., 1997. Risk Taking: A Managerial Perspective. Russell Sage Foundation.
URL http://books.google.com/books?id=HuIaCKWhyd4C&pgis=1
Simon, H., Bonini, C., 1958. The size distribution of business firms. The American Economic Review 48 (4), 607–617.
URL http://www.jstor.org/stable/10.2307/1808270
Starbuck, W. H., Jun. 2009. Perspective–Cognitive Reactions to Rare Events: Perceptions, Uncertainty, and Learning. Organization Science 20 (5), 925–937.
URL http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1090.0440
Staw, B. M., Ross, J., 1989. Understanding Behavior in Escalation Situations. Science 246 (4927), 216–220.
Staw, B. M., Ross, S. A., 1986. Expo 86: An Escalation Prototype. Administrative Science Quarterly 31 (2), 274–97.
Staw, B. M. B., Oct. 1981. The Escalation of Commitment to a Course of Action. The Academy of Management Review 6 (4), 577–587.
URL http://www.jstor.org/stable/257636 http://www.jstor.org/stable/10.2307/257636
Steele, M., Huber, W., 2004. Exploring data to detect project problems. In: AACE International Transactions. pp. PM.21.1–PM.21.7.
Taleb, N. N., 2007a. Black Swans and the domains of statistics. American Statistician 61 (3), 198–200.
Taleb, N. N., 2007b. The Black Swan. The Impact of the Highly Improbable. Penguin, New York, NY.
Taleb, N. N., 2013. Lectures on Risk and(Anti)fragility.
Tengshe, A., Noble, S., 2007. Establishing the agile PMO: Managing variability across projects and portfolios. In: Proceedings - AGILE 2007. AGILE 2007. Capital One Auto Finance, pp. 188–193.
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, 331.
URL http://portal.acm.org/citation.cfm?doid=1134285.1134332
Tukey, J. W., 1977. Exploratory Data Analysis. Addison-Wesley, Reading, MA.
Turner, B., Pidgeon, N., 1997. Man-Made Disasters, Second Edition. Butterworth-Heinemann.
URL http://www.amazon.com/Man-Made-Disasters-Second-Edition-Turner/dp/0750620870
Vaughan, D., 1996. The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA. University of Chicago Press, Chicago, IL.
Vaughan, D., 1997. The trickle-down effect: policy decisions, risky work, and the Challenger accident. California Management Review 39 (2), 80–102.
Virkar, Y., Clauset, A., Aug. 2012. Power-law distributions in binned empirical data. Arxiv preprint arXiv:1208.3524, 33.
URL http://arxiv.org/abs/1208.3524
Ward, J., Hertogh, S. D., Viaene, S., 2007. Managing benefits from IS/IT investments: An empirical investigation into current practice. In: System Sciences, 2007. pp. 1–10.
Weick, K. E., 1988. Enacted Sensemaking in Crisis Situations. Journal of Management Studies 25 (4), 305–317.
Weick, K. E., Mar. 2004. Normal Accident Theory As Frame, Link, and Provocation. Organization & Environment 17 (1), 27–31.
URL http://oae.sagepub.com/cgi/doi/10.1177/1086026603262031
Weick, K. E., Sutcliffe, K. M., Jul. 2006. Mindfulness and the Quality of Organizational Attention. Organization Science 17 (4), 514–524.
URL http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1060.0196

Weick, K. E., Sutcliffe, K. M., 2007. Managing the Unexpected: Resilient Performance in an Age of Uncertainty, 2nd Edition. Jossey-Bass, San Francisco, CA.

Weiner, B., Jan. 1985. "Spontaneous" causal thinking. Psychological bulletin 97 (1), 74–84.

Weiner, B., Jan. 2008. Reflections on the History of Attribution Theory and Research. Social Psychology 39 (3), 151–156.

Weiner, B., Nierenberg, R., Goldstein, M., 1976. Social learning (locus of control) versus attributional (causal stability) interpretations of expectancy of success. Journal of Personality 44 (1), 52–68.

Williams, T., 1992. Practical use of distributions in network analysis. Journal of the Operational Research Society 43 (3), 265–270.
URL `http://www.jstor.org/stable/10.2307/2583716`

Yoon, K.-A., Kwon, O.-S., Bae, D.-H., Sep. 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) (1), 443–445.
URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4343773`