The distribution of project performance: initial investigations of its nature and what we can learn from it

Tristan Butterfield[1,2], Dr Mike Yearworth[1]

[1]Systems Centre, Faculty of Engineering, University of Bristol, Bristol, UK

[2]Thales Research and Technology, Thales UK, Reading, UK

tristan.butterfield@bristol.ac.uk, mike.yearworth@bristol.ac.uk

Abstract— *All too often projects perform very differently to how they are expected to, and indeed, how organisations would wish them to. There have been many attempts to further our understanding of the reasons for this behaviour by looking at the dynamics that exist within a single project environment. This paper looks to investigate the nature of the distribution of the performance of a number of projects, across a portfolio, and poses the question of whether unexpected project performance should in fact be expected. The aim of this work is to provide an initial framework to explore this question by investigating the distribution of simulated results from a System Dynamics project model based on the rework loop. Furthermore this research gives an opportunity to further investigate the hypothesis that this distribution may follow a power law, as has been discussed in literature and also within our own research team.*

Keywords—*Project Performance; Portfolio; Rework; Power Law; System Dynamics.*

1 INTRODUCTION

This paper details the authors' investigations into the nature of how project performance might be distributed across a portfolio of projects within an organisation. The purpose of conducting this work was to try and illuminate the issue of whether it is reasonable to expect that it is possible to reduce the variance of actual project performance against what is estimated and also how one might go about achieving this.

This has been carried out by running simulations using a System Dynamics project model and contrasting against basic estimates for the performance of these simulated projects using parameters devised from an analogy to past experience. From this the 'magnitude' of project performance (i.e. $\frac{actual\ performance}{estimated\ performance}$) can be calculated and plotted to show how this is distributed across many projects. The hypothesis we propose, in line with the work of Budzier & Flyvbjerg (2011), is that this distribution should, at least partially, follow a power law. This is a distribution of the form: $P(X \geq x) = Cx^{-\alpha}$.

The testing of this hypothesis, through the failure to falsify, could be seen to indicate that the poorly performing projects often regarded as 'outliers' or 'one-offs' are indeed of the same nature as other projects and should hence not be considered as different.

2 BACKGROUND

This section explores the background and reasons for this research, the validity of the System Dynamics model used and also the concept of power laws in the context of project performance.

2.1 Research Agenda

This piece of research has evolved from the wider research agenda of trying to address the issue of improving the estimation of complex project performance within Thales UK by the use of parametric models. This wider research agenda has already yielded a System Dynamics project model based on the rework loop (Walworth et al. 2013).

In the ongoing effort to calibrate this model for use with actual projects in Thales UK many questions have emerged regarding the issues surrounding the estimation and improvement of project performance. One of these strands of question has been the notion of what the distribution of project performance *should* look like across a portfolio of projects within an organisation. Or in other words, how much variance from estimations of future performance should one really expect?

Hence, with these thoughts and recent developments in literature (Budzier & Flyvbjerg 2011; Budzier & Flyvbjerg 2013) the research reported here was devised in order to investigate what the existing System Dynamics project model, which has been shown to produce valid behaviour (Walworth et al. 2013), predicts for this distribution.

2.2 Theoretical Background of the Rework Loop Model

The System Dynamics project model used, shown in Figure 1, is an extension of the model under development at Thales UK and already presented by Walworth et al (2013). The rework loop models work flowing through a project and explicitly acknowledges that a portion of this work is done incorrectly and will need correcting (i.e. reworking). The desired use of this model is the creation of planned lines (i.e. estimates) for projects against which collected metrics can be reviewed to give an indication of performance. Of course in this application the model is being used to generate the 'real' performance of randomly generated project simulations.

The model used here was, first shown by Cooper & Mullen (1993), has been well documented. There is an established acceptance of the rework loop as the base of a System Dynamics project model (Lyneis & Ford 2007) and there are numerous examples of this structure in use (Cooper & Lee 2009; Ford et al. 2007; Ford & Sterman 1998; Lyneis et al. 2001; Reichelt & Lyneis 1999). Indeed there have been claims that the rework loop is the "canonical structure" (Lyneis & Ford 2007, p.159) for System Dynamics project models.

The outputs from the model, as illustrated in Figure 2, are not calibrated but have already been shown (Walworth et al. 2013) to be of the nature of what has been shown to be expected in literature (e.g. Mawby & Stupples 2002; Putnam 1978) and also what has been observed from Thales UK data.

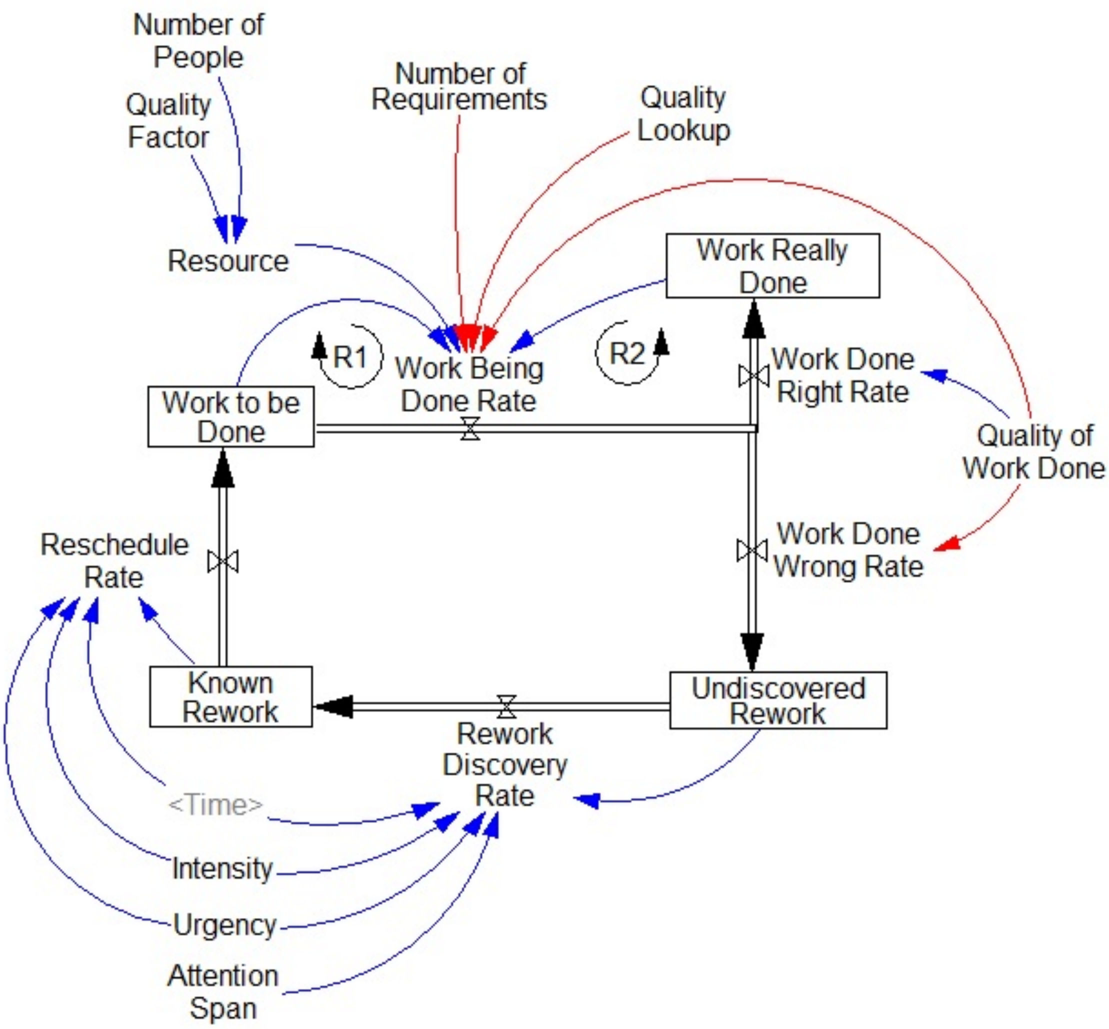

Figure 1 – System Dynamics Project Model

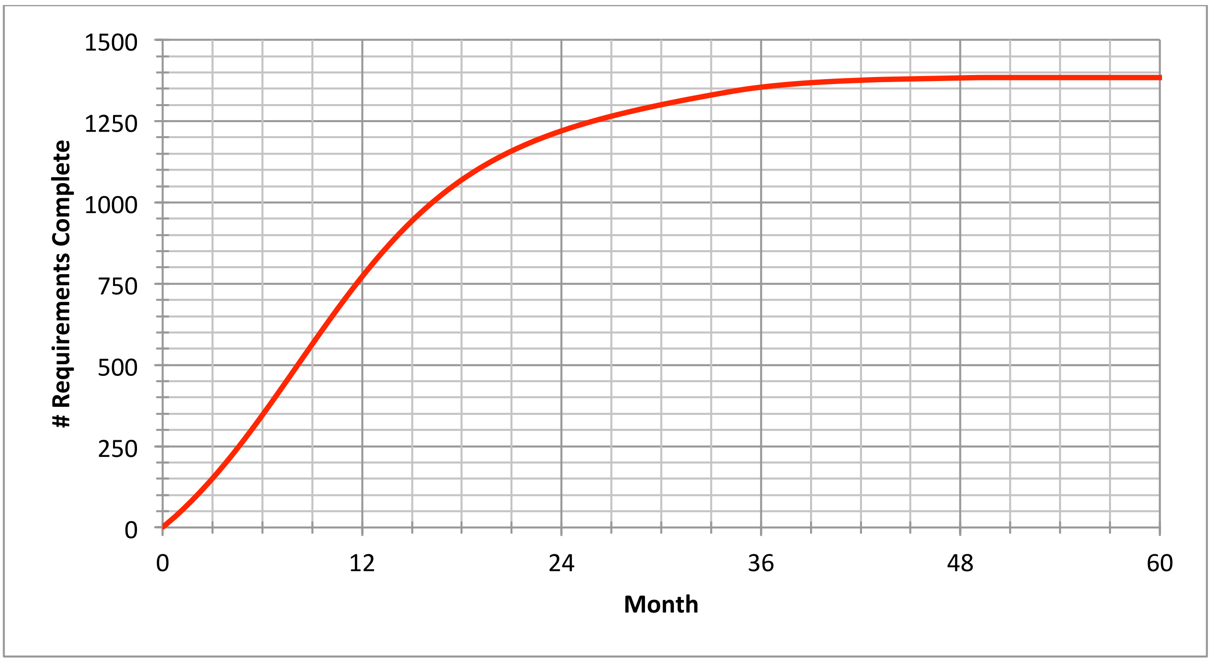

Figure 2 – Example Output from Model (Requirements Completed against Time – assuming constant level of effort applied)

2.3 The Distribution of Project Performance & Power Laws

Budzier & Flyvbjerg have shown that the distribution of ICT project performance is "far from normally distributed" (2013, p.14) and can be fitted to a power law distribution. They reject the notion that poorly performing 'outlier' projects come from a different probability regime to other projects and hence conclude that they must be considered in the same frame of reference. They go on to conclude that these outlier projects are not fundamentally any different to other 'normal' projects and hence should not be dismissed as one-off events. This is an important point for improving industrial practice, where too often these poorly performing project outliers are dismissed as such (March & Shapira 1987). Budzier & Flyvbjerg also argue that this can be attributed to "Black Swan Blindness" (2011, p.13) in addition to the bias towards optimism for project performance previously reported (Flyvbjerg 2008; Jørgensen & Grimstad 2005). A "Black Swan" is a rare, high impact event (Taleb 2005; Taleb 2007) and so in terms of a project would be a very poorly performing project, probably considered outside of the normal bounds of performance, that would have potentially ruinous impacts for the project organisation.

Whilst this background work is concerned with the ICT sector the authors of this paper believe that this is an analogous scenario to the Systems Engineering sector where they are conducting their research. Furthermore they believe that this is novel research in a relatively sparse topic that has the potential to have significant insight into the problem of addressing poor problem performance.

A power law is a probability distribution that follows the identity (Mitzenmacher 2004):

$$P(X \geq x) = Cx^{-\alpha}$$

Where $P(X \geq x)$ is the probability that a random variable X is greater than the given variable x, α is the exponent of the equation and C is a constant. In the case of investigating the distribution of project performance $P(X \geq x)$ would refer to the probability that any random project would have a greater magnitude of performance ($\frac{actual\ performance}{estimated\ performance}$) than the specific magnitude in question. Hence, for example, $P(X \geq 1.0)$ could be calculated giving the probability of any project performing worse than the estimate.

Power law behaviour has been shown to be present in many different environments (Newman 2005; Clauset et al. 2009; Mitzenmacher 2004) and their presence can be identified via the creation of histograms of given data using logarithmic binning (Newman 2005). The presence of a governing power law can then be easily identified by a straight trend line when this histogram is plotted on logarithmic scales, as can be seen in the examples of the population of cities or the magnitude of earthquakes (Newman 2005).

3 METHOD

In order to investigate how project performance might be distributed the existing System Dynamics model, as detailed in Section 2.1, was used to generate data for simulated project performance across a randomised portfolio of projects. These simulations were taken as the 'actual' performance of the project and then compared against performance estimates for the project. Hence the magnitude ($\frac{actual\ performance}{estimated\ performance}$) of the each project's performance was calculated and the distribution of this was plotted.

3.1 Model Setup and Assumptions

The Vensim® software package was used for modelling and simulation. The model was set up to run from time 0→240 (months) in time steps of 0.0625 using integration type RK4

Auto.

As part of this process the following assumptions were made:

- The System Dynamics model used produces valid behaviour based on the inputs despite not producing calibrated numerical outputs.
- The estimation method described in Section 3.4 is analogous to general practice, namely using past experience to predict future performance.
- The distribution of project performance produced with this method would be of the same nature, although skewed, to that of one produced using a calibrated model and estimation method.

3.2 Randomised Parameterisation of Simulations

There are 7 input parameters for the System Dynamics model, for each simulation 4 of these input parameters were randomly generated (within Microsoft Excel) according to the following rules:

- Number of Requirements: Uniformly distributed between 100 and 5,000 in intervals of 50. This parameter is simply the size of the project.
- Number of People: Uniformly distributed between 2 and 50. This parameter equates to the number of full-time project staff available to work on the project.
- Quality Factor: Normally distributed with a mean of 5.5 and standard deviation of 1 within a permitted range of 1→10. This parameter takes into account the relative suitability of the project team for the project.
- Quality of Work Done: Normally distributed with a mean of 0.75 and standard deviation of 0.05 within a permitted range of 0.5→1.0. This parameter is the fraction of work done correctly (i.e. 1 – rework fraction). For example if the Quality of Work Done = 0.8 then 80% work being done is done correctly and 20% will require rework. It should be noted that this fraction applies to all work, including requirements that have already been reworked.

For both the Quality Factor and the Quality of Work Done parameters any randomly generated values which fell outside the permitted range were manually rationalised to the upper or lower bound of the input range as appropriate.

The input parameters for Intensity, Urgency and Attention Span (which combined determine the rate at which rework is discovered) were all set to their default values for each of the simulation runs. These parameters are a direct analogy to the Jelinski-Moranda equation for defect discovery in software (Jelinski & Moranda 1972).

3.3 Simulation & Processing

Vensim® was used to simulate System Dynamics model using the randomised parameters and was set up to export the simulation data which were then processed and collated manually. The following results were captured for each simulation run: the project time to complete; the number of requirements actually completed; and the number of requirements ‘considered’ to have been completed (i.e. requirements actually completed + the level of undiscovered rework).

Each project was subject to the following stopping rule: the project was deemed to have finished when the change in the number of requirements ‘considered’ to be done fell below 1 for a single month. This was except in the cases where the Δ requirements ‘considered’ done falls below 1 early on in the lifecycle due to a large amount of rework being discovered. This was considered normal and not taken as the stopping point for the project.

For some simulation runs the total number of requirements actually done or 'considered' done exceeded the total number of requirements for the project due to integration and rounding errors. In these cases these numbers were rationalised to the total number of requirements.

3.4 'Estimation'

In addition to the generation of the simulated 'real' project performance an estimated performance was also needed for each of the simulation runs in order to calculate the project performance magnitude.

Estimates were created for each of the simulation runs based on the results of a smaller sub-set of 10 simulations. From these a figure for the average number of requirements / staff member / month was calculated. This figure was based on the number of requirements 'considered' done and subject to the same stopping rule as described in section 3.4.

This figure was then applied to each of the simulation runs in order to calculate an estimate for the project duration. The simulated 'real' results were then compared against the estimate and hence the magnitude of the project performance ($\frac{actual\ performance}{estimated\ performance}$) was calculated.

4 RESULTS

Figure 3 gives the estimated values for $P(X \geq x)$, meaning the probability that the magnitude of a random project X will be larger than the magnitude of a given project x. For example when x = 1.0 (performed exactly as expected) $P(X \geq x) \approx 60\%$ (i.e. there is a 60% chance that any project will perform worse than expectations).

Figure 4 shows the log-log graph of the data presented in Figure 3. If the distribution of project performance was governed by a power-law (or at least a section of the distribution) we would expect to see a straight line in this graph. Although far from conclusive one could make the argument for a straight trend line for the data between the log magnitude of 0.00 and 0.60. This corresponds to the data for runs where the magnitude was between 1 and 4 (approximately 75% of the runs).

We can also extract the following raw statistics for the magnitude of performance from the simulations:

- Median = 1.19
- Mean = 1.26
- Standard deviation = 0.79
- Minimum = 0.15
- Maximum = 6.42

Whilst it is hard to draw any real conclusions from these statistics it is very noticeable that the range for simulated project performance is very large. The projects performing much better than estimates particularly stand-out due to the scarcity of analogous examples in the real world.

A further point to draw from this is in comparison to a previous iteration of these simulations, which applied a uniform distribution to both the Quality of Work Done & Quality of Staff rather than a normal distribution. This change has pulled the statistics into a much more believable territory – previously the average magnitude was considerably higher and there was also little evidence to suggest the presence of a power-law governing the distribution.

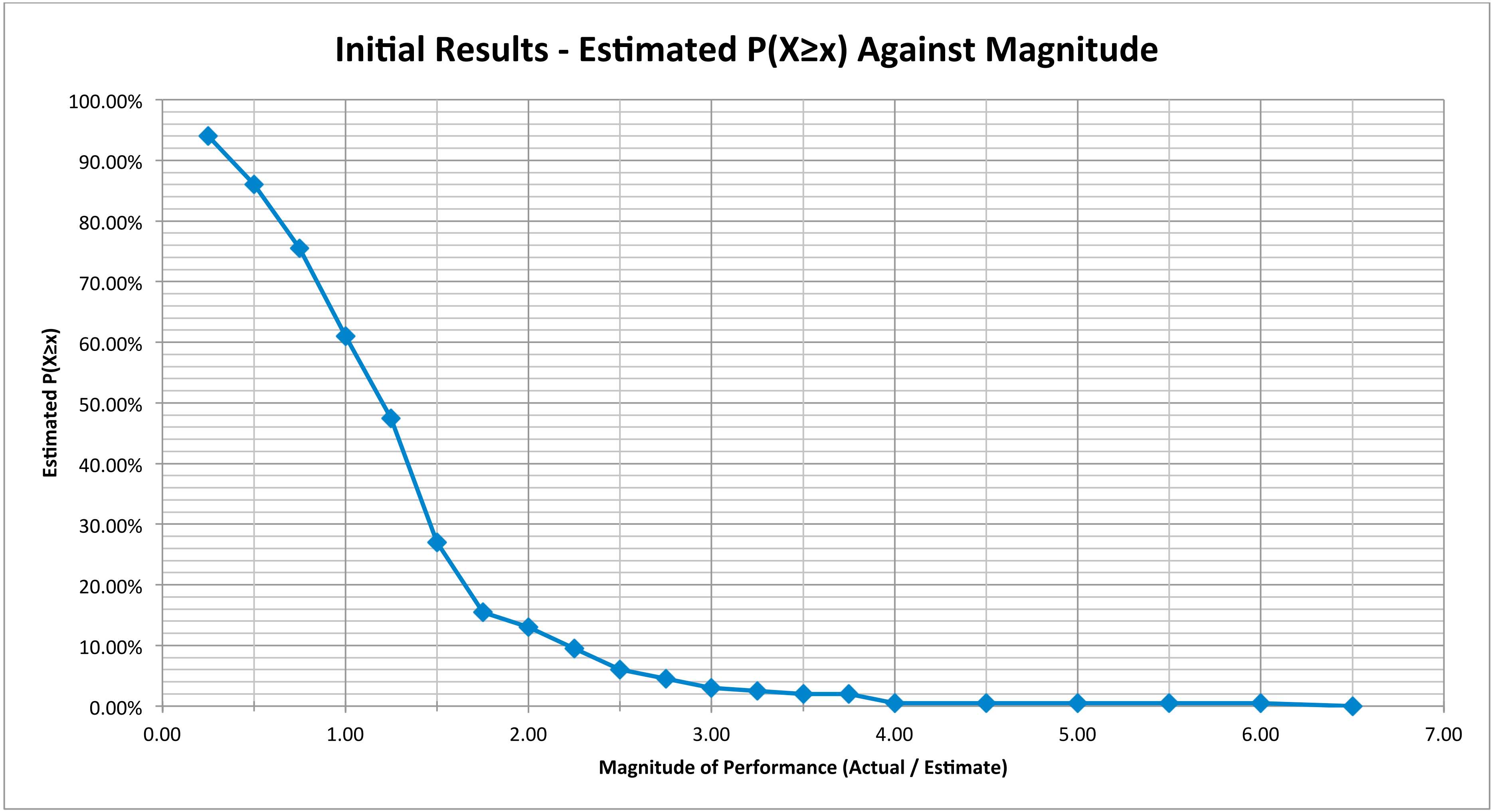
Initial Results - Estimated P(X≥x) Against Magnitude
Estimated P(X≥x)
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
Magnitude of Performance (Actual / Estimate)

Figure 3

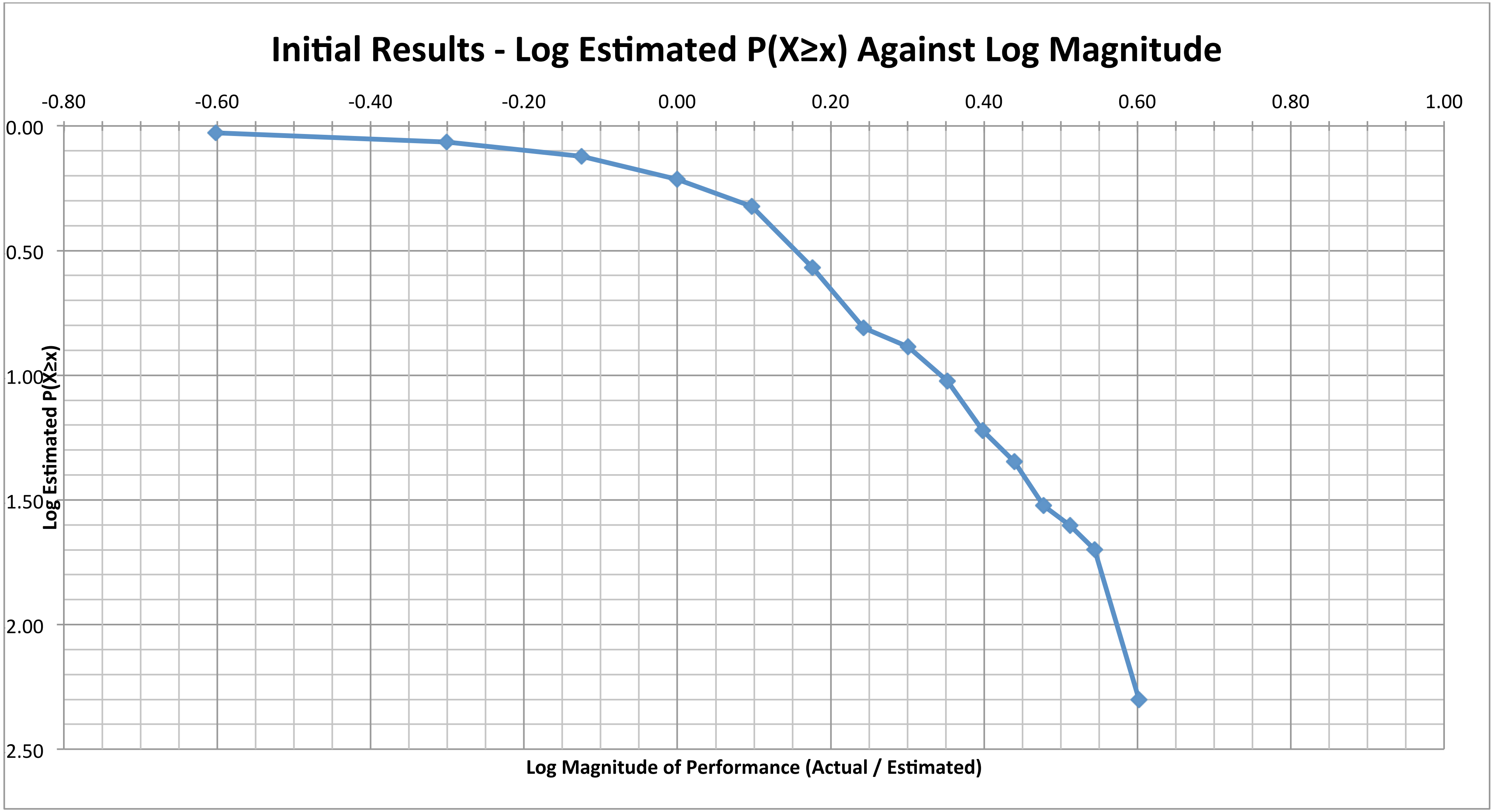
Initial Results - Log Estimated P(X≥x) Against Log Magnitude
-0.80
-0.60
-0.40
-0.20
0.00
0.20
0.40
0.60
0.80
1.00
0.00
0.50
1.00
1.50
2.00
2.50
Log Estimated P(X≥x)
Log Magnitude of Performance (Actual / Estimated)

Figure 4

5 Discussion

From the interim results presented in Section 4, a result of 200 simulations, it appears that the assertion that this distribution is at least partially governed by a power law cannot be dismissed as this stage and it appears that it may be correct. Further work in analysing a larger sample of the order of 10,000 simulations will lead to a clearer conclusions for this point.

If this were to be proved the case then it could be said that the System Dynamics model is a power law generating model and, as we believe that the behaviour generated by the model is valid, that we would expect to see analogous behaviour in the real world data. Hence we believe that this work provides a suitable framework to further investigate the claims of Budzier & Flyvbjerg (Budzier & Flyvbjerg 2011; Budzier & Flyvbjerg 2013) in a Systems Engineering environment and also as a comparison against real project data from organisations.

The impact for industry of these conclusions is not, perhaps, immediately evident. However the sheer awareness that the poorly performing projects, which are so often treated as outliers or one-offs, are in fact of the same nature as 'normal' projects is an important step in itself. With this awareness it might be possible to reduce the "Black Swan Blindness" (Budzier & Flyvbjerg 2011, p.13) of an organisation and perhaps reduce the impacts of any optimism bias that may be present. This could lead to an organisation with a much more insightful view of future performance.

One explanation for this behaviour could be that, although all parameters of the model can be seen explicitly in this example, some of the inputs that drive this behaviour are intangible and unmeasurable. Furthermore, estimates, in both this work and also in reality, are generated based on known quantities and past experience without taking these hidden values into account.

It is proposed that it is a combination of these hidden parameters (Quality Factor, Quality of Work Done & Rework Discovery Rate), particularly with a large project, can lead to a *tipping point* that causes the project to perform particularly badly against expectations and hence classification as an outlier of 'black swan'. One practical proposal from this finding is the recommendation for making these hidden parameters more visible, or at least explicitly acknowledging their hidden nature when making estimations. It is also proposed that an extension of this work could be an investigation into these tipping points within the model with the approach discussed by Gross & Feudel (2006).

If we consider addressing the issue of poor project performance as a complex problem, or indeed as a *wicked* problem (Rittel & Webber 1973), and that we are most likely operating in an environment where there are divergent views of what the problems are and how they should be tackled (i.e. we are operating in a Complex-Pluralist environment (Jackson 2003)) then a Problem Structuring Method (PSM) approach may be appropriate (Rosenhead 1996; Mingers & Rosenhead 2004; Yearworth & White 2014).

It is proposed that the value of the framework that this work sets out, for investigating the distribution of project performance, might best be realised by incorporating it into a PSM approach which would enable greater understanding of the system where there is not necessarily a consensus about the problem or the solution (Rosenhead 1996) or indeed where the problem may not be solvable (Yearworth et al. 2013) if it is *wicked*.

Further work is proposed as to how the framework proposed and inherent modelling could be included in a PSM approach that would be suitable and palatable for a project based organisation, such as Thales UK.

Of course there are a number of limitations to this work and these include the fact that the current model used is based on a single-stage rework loop. This is in contrast to how most projects would be made up of multiple rework loops corresponding to different stages of project development. This abstraction may change the dynamics of the proposed tipping points and stability within the system.

Also, the research does not consider or take into account size of the project – i.e. a small project vastly over budget may not be of great risk to a company but a very large project twice over budget could be potentially ruinous.

Furthermore, if comparing against real project data then values for actual performance and estimated performance will most likely be in the form of financial data. Whilst this is comparable it is not identical to the technical performance discussed in this paper. In order to compare against data from an organisation the System Dynamics model would have to be calibrated and parameterised sufficiently if there was a desire to extract numerical outputs from the distribution (i.e. probabilities) rather than simply a comparison of the nature of the distribution. The task of acquiring suitable data to calibrate the model is non-trivial.

6 CONCLUSIONS

This section provides some brief conclusions based on the interim results from the System Dynamics project model and will be refined following further work.

It appears that this distribution may be governed by a power law, and at the least this hypothesis has not been falsified.

Very poorly performing projects can be explained in the same probability regime as ‘normal’ projects and should not be dismissed as one-offs and not considered when planning for the future.

ACKNOWLEDGMENTS

The authors would like to thank their fellow members of the ‘metrics’ research team at the University of Bristol and at Thales UK. We would especially like to thank Thomas Walworth for his work in developing the System Dynamics model and also Hillary Sillitto for his contribution to the discussion about the relevance of power laws in this environment. Tristan Butterfield is registered on the Engineering Doctorate (EngD) in Systems programme at the University of Bristol. This research is supported by the EPSRC funded Industrial Doctorate Centre in Systems (Grant EP/G037353/1) at the University of Bristol and by Thales UK.

REFERENCES

Budzier, A. & Flyvbjerg, B., 2011. Double Whammy – How ICT Projects are Fooled by Randomness and Screwed by Political Intent. Available at: http://www.ssrn.com/abstract=2238057 [Accessed January 14, 2014].

Budzier, A. & Flyvbjerg, B., 2013. Making-Sense of the Impact and Importance of Outliers in Project Management through the Use of Power Laws. In *Proceedings of IRNOP (International Research Network on Organizing by Projects), At Oslo*. pp. 1–28.

Clauset, A., Shalizi, C.R. & Newman, M.E.J., 2009. Power-Law Distributions in Empirical Data*. *SIAM Review*, 51(4), pp.661–703.

Cooper, K. & Lee, G., 2009. Managing the Dynamics of Projects and Changes at Fluor. In *Proceedings of the International System Dynamics Conference*. Albuquerque, pp. 1–27.

Cooper, K.G. & Mullen, T.W., 1993. Swords & Plowshares - The Rework Cycles of Defense and Commercial Software. *American Programmer*, 6.

Flyvbjerg, B., 2008. Curbing Optimism Bias and Strategic Misrepresentation in Planning : Reference Class Forecasting in Practice. *European Planning Studies*, 16(1), pp.3–21.

Ford, D.N., Lyneis, J.M. & Taylor, T.R.B., 2007. Project controls to minimize cost and schedule overruns: a model, research agenda, and initial results. In *2007 International System Dynamics Conference*. pp. 1–27.

Ford, D.N. & Sterman, J.D., 1998. Dynamic modeling of product development processes. *System Dynamics Review*, 14(1), pp.31–68.

Gross, T. & Feudel, U., 2006. Generalized models as a universal approach to the analysis of nonlinear dynamical systems. *Physical Review E*, 73(1), p.016205. Available at: http://link.aps.org/doi/10.1103/PhysRevE.73.016205 [Accessed March 10, 2014].

Jackson, M.C., 2003. *Systems Thinking: Creative Holism for Managers*, Chichester: John Wiley & Sons.

Jelinski, Z. & Moranda, P.B., 1972. Software reliability reseach. In W. Freiburger, ed. *Statistical Computer Performance Evaluation*. Academic Press, pp. 465–484.

Jørgensen, M. & Grimstad, S., 2005. Over-optimism in Software Development Projects: "The winner's curse." In *Proceedings of the 15th International Conference on Electronics, Communications and Computers (CONIELECOMP 2005)*. pp. 280–285.

Lyneis, J.M., Cooper, K.G. & Els, S.A., 2001. Strategic management of complex projects: a case study using system dynamics. *System Dynamics Review*, 17(3), pp.237–260. Available at: http://doi.wiley.com/10.1002/sdr.213 [Accessed March 15, 2013].

Lyneis, J.M. & Ford, D.N., 2007. System dynamics applied to project management : a survey , assessment , and directions for future research. *System Dynamics Review*, 23(2), pp.157–189.

March, J.G. & Shapira, Z., 1987. Managerial Perspectives on Risk and Risk Taking. *Management Science*, 33(11), pp.1404–1418.

Mawby, D. & Stupples, D., 2002. Systems Thinking for Managing Projects. In *Engineering Management Conference, 2002. IEMC'02. 2002 IEEE International*. IEEE, pp. 344–349.

Mingers, J. & Rosenhead, J., 2004. Problem structuring methods in action. *European Journal of Operational Research*, 152, pp.530–554. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221703000560 [Accessed October 18, 2013].

Mitzenmacher, M., 2004. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics*, 1(2), pp.226–251. Available at: http://www.tandfonline.com/doi/abs/10.1080/15427951.2004.10129088 [Accessed January 14, 2014].

Newman, M.E.J., 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), pp.323–351. Available at: http://www.tandfonline.com/doi/abs/10.1080/00107510500052444 [Accessed January 27, 2014].

Putnam, L.H., 1978. A General Empirical Solution to the Macro Software Sizing and Estimating Problem. *IEEE Transactions on Software Engineering*, SE-4(4), pp.345–361. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1702544.

Reichelt, K. & Lyneis, J.M., 1999. The dynamics of project performance: benchmarking the drivers of cost and schedule overrun. *European Management Journal*, 17(2), pp.135–150.

Rittel, H.W.J. & Webber, M.M., 1973. Dilemmas in a General Theory of Planning*. *Policy Sciences*, 4, pp.155–169.

Rosenhead, J., 1996. What's the Problem? An Introduction to Problem Structuring Methods. *Interfaces*, 26(6), pp.117–131.

Taleb, N.N., 2007. Black Swans and the Domains of Statistics. *The American Statistician*, 61(3), pp.198–200. Available at: http://www.tandfonline.com/doi/abs/10.1198/000313007X219996 [Accessed March 3, 2014].

Taleb, N.N., 2005. The Roots of Unfairness: the Black Swan in Arts and Literature. *Literary Research/Recherche Litteraire*, 21, pp.241–254.

Walworth, T. et al., 2013. Early estimation of project performance : the application of a system dynamics rework model. In *IEEE International Systems Conference*.

Yearworth, M. et al., 2013. Integrating Problem Solving and Research Methods Teaching for Systems Practice in Engineering. *Procedia Computer Science*, 16, pp.1072–1081. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1877050913001142 [Accessed October 13, 2013].

Yearworth, M. & White, L., 2014. The Non-Codified Use of Problem Structuring Methods and the Need for a Generic Constitutive Definition. *European Journal of Operational Research*. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221714001301 [Accessed February 21, 2014].