Relative performance of judgmental methods for forecasting the success of megaprojects

Litsiou, Konstantia; Polychronakis, Yiannis ; Karami, Azhdar; Nikolopoulos, Kostas

International Journal of Forecasting

DOI:
10.1016/j.ijforecast.2019.05.018

Published: 01/07/2022

Peer reviewed version

Relative performance of judgemental methods

for forecasting the success of megaprojects[a]

Mrs Konstantia Litsiou, BA (Athens), PGCE (Athens) , MA (Lancaster) is a Senior Lecturer in Retail Management in Manchester Metropolitan University Business School. Konstantia is also pursuing a PhD in "Forecasting with Judgmental Methods the success of Megaprojects" in Salford Business School. Konstantia has a background in arts and has worked in big scale art projects and events in Greece for many years. Her work has appeared in international journals and conferences.

Dr. Yiannis Polychronakis is a Senior Lecturer in Operations, Logistics and Project Management at the Salford Business School. Yiannis holds a BSc (Hons) in Manufacturing Systems, an MSc in Quality Management and a PhD by Published Works. He has published several articles in international journals (International Journal of Production Economics, International Journal of Quality and Reliability Management, Journal of Manufacturing and Technology Management) in the area of project management application for supply chain development, and has a number of papers included in refereed International Conference Proceedings. He has also assisted a number of organisations in developing and implementing best practices in the areas of process management, supply chain development and project management. Organisations include Silent Night, Cleveland Guest, Trafford Council, Manchester City Council and more.

Dr Azhdar Karami is a Senior Lecturer in Strategy and Management. He started his academic career in 1991 when he obtained his MBA and has worked in different teaching posts in overseas universities, University of Bradford, and ultimately as a Senior Lecturer in the Bangor Business School. Azhdar has been and is currently responsible for organizing two modules in undergraduate and two more modules in postgraduate levels at the Bangor Business School. Dr Karami's research interests include strategic management; entrepreneurship and innovation in SMEs. He has published several books, journal articles and conference papers. His latest book titled "strategy development in entrepreneurial firms" appeared in the book market in late 2007. He has published in a number of leading international peer refereed journals including Journal of Management Development; International Journal of Management and Corporate Governance: International Journal of Business in Society, International Journal of Technology Management, and International Journal of Business Information Systems.

Professor Konstantinos (Kostas) Nikolopoulos is the Director of forLAB www.forlab.eu. He is an expert in Time Series Analysis & Forecasting, Decision Support Systems and Management Judgment. Professor Nikolopoulos received his Engineering Doctorate in 2002 from National Technical University of Athens. He has worked in the past for the University of Manchester and other prestigious institutions in U.K. and Greece, and now holds the Chair in Business Analytics in Bangor Business School. His work has appeared in JOM, EJOR, IJF, JoF, JORS, Omega, IJPE, IJPR, IMA JMM, JBR, JCIS, IMDS, AE among other journals, and he is an Associate Editor of Oxford IMA Journal of Management Mathematics and Supply Chain Forum: An International Journal. He is co-originator of the Theta forecasting method and the ADIDA temporal aggregation method-improving framework.

[a] Accepted for publication in *International Journal of Forecasting* on 18th May 2019

Relative performance of judgemental methods for forecasting the success of megaprojects

Abstract

Forecasting the success of megaprojects, such as the Olympic Games or space exploration missions, is a very difficult and important task because of the complexity of such projects and the large capital investment they require. Megaproject stakeholders do not typically employ formal forecasting methods, relying instead on Impact Assessments and/or Cost Benefit Analysis; these tools do not necessarily include forecasts, and thus there is no accountability. This study evaluates the effectiveness of judgemental methods towards successfully forecasting the accomplishment of specific megaproject objectives – when the measure of success is the collective accomplishment of such objectives. We compare the performance of three judgemental methods used by a group of 55 semi-experts: Unaided Judgement (UJ), semi-Structured Analogies (s-SA), and Interaction Groups (IG). The empirical evidence reveals that the use of s-SA leads to accuracy improvement compared with UJ. This improvement is amplified further when introducing pooling of analogies through teamwork in IG.

Key words: Judgemental Forecasting; Megaprojects; Semi-Experts; Structured Analogies; Interaction Groups

1. Introduction

Forecasting the success of major projects like the Olympic Games or a space exploration mission is a very difficult but also extremely important task. A significant amount of resources is allocated to such projects, and there are great expectations of them. Success would typically entail finishing on time (duration), within budget, as well as benefits realisation. Unfortunately, success lies in the eyes of the stakeholder(s), and that view can be very different for the funder, the project manager, or the public (Talbot, 2009). All these constitute a very challenging forecasting task and are the main motivation for our engaging in this research and finding the appropriate horses for that course (Petropoulos et al. 2014), that is, the forecasting methods most appropriate for the task.

These major projects are often called a) megaprojects (Merrow 2011, 1988; Flyvbjerg, Bruzelius & Rothengatter, 2003), b) complex service-led projects (Alderman, Ivory, McLoughlin and Vaughan, 2005), c) large capital projects (Bekker & Steyn, 2007), d) large complex projects (Miller & Hobbs, 2005), and e) large engineering projects (Miller & Lessard, 2000, 2007). For the sake of consistency, from this point on, we will refer to them as *megaprojects*.

Megaprojects are temporary in nature and characterised by large investment commitment, enormous organisational complexity, a long-lasting impact on the environment, the economy, and society (Sanderson, 2012). The US Department of Transportation defines megaprojects as projects with at least a budget of USD 1 billion (Capka, 2006). In EU countries, the International Project Management Association (IPMA) (2011) describe EUR 1 billion as the threshold defining megaprojects across all industries.

Flyvbjerg, Garbuio & Lovallo (2014) argue vividly how difficult it is to forecast the success of megaprojects. They claim that:

> *'Large capital investments that are completed on schedule and within their budgets are probably the exception rather than the rule—and even when completed many fail to meet expected revenues. Executives often blame project underperformance on foreseeable complexities and uncertainties having to do with the scope of and demand for the project, the technology or project location, or even stakeholder opposition. No doubt, all of these factors at one time or another contribute to cost overruns, benefit shortfalls, and delays.'*

Turner and Zolin (2012) claim that we cannot even properly define what success is – or what it will be when the megaprojects' targets are materialised to some extent. They argue that we need to employ reliable scales in order to predict multiple perspectives by multiple stakeholders over multiple time frames. This could be done via a set of leading performance indicators that will enable managers of megaprojects to forecast during project execution how various stakeholders will perceive success months or even years into the operation. Megaprojects have many stakeholders with different objectives for the project, its output, and the business objectives they will deliver. The output of a megaproject may have a lifetime that lasts for years, or even decades, and ultimate impacts that go beyond its immediate operation. How different stakeholders perceive success can change with time.

Megaprojects are mostly implemented for the first time on such a scale, and there is no previous experience (or data per se) for the respective duration, budget and potential socio-economic impact. Thus, quantitative methods are not the first choice of weapons in our forecasting arsenal, and we must rely on experts and judgemental forecasting methods for the aforementioned challenging task. These forecasts will consequently drive decisions before and during the megaproject for all the stakeholders: funders, clients and project managers.

The rest of the paper is structured as follows: section two revisits and critically synthesises the relevant literature on forecasting for megaprojects; section three exposes our methodological approach. Section four presents our results and discussion while the last section offers concluding remarks and roadmaps for future research.

2. Literature review

2.1 Forecasting in Project Management

Forecasting is vital in project management for predicting accurately the actual duration and cost of a project in progress. This is even more essential in the case of megaprojects where the duration and costs are of substantial magnitude.

As stated by Batselier and Vanhoucke (2015), earned value management (EVM) is the most frequently used and best performing methodology for obtaining a project's actual duration and cost forecasts. EVM assists project managers to systematically measure variances in projects based on the comparison of work completed versus work planned. EVM is used on the cost and schedule control and can be very useful in project forecasting via projecting the measured cost and time variances. The EVM technique is generally deemed a feasible and valued basis for forecasting a project's duration and cost based on ongoing data collected from the same project. Various novel EVM-based time forecasting approaches have been developed in recent years, and these techniques can be categorised into deterministic and probabilistic approaches (Barraza et al., 2004). Deterministic approaches yield a point estimate of the eventual project duration, whereas probabilistic techniques provide confidence intervals and/or distributions of possible durations (Batselier & Vanhoucke 2015).

Lipke (2003) introduced a new concept in EVM known as earned schedule (ES) where instead of using cost for measuring schedule performance, the unit is time: thus EVM could also be used for measuring and forecasting the duration of a project: this is commonly referred as ESM (Earned Schedule Method). Various forecasting approaches have also emerged over time but largely as an extension to EVM (Anbari 2003; Kim & Reinschmidt, 2010; Lipke, 2011; Elshaer, 2013; Khamooshi & Golafshani, 2014; Mortaji et al., 2014; Baqerin et al., 2015; Chen et al., 2016).

Jacob and Kane (2004) developed earned duration method (EDM). Traditionally cost, schedule and duration are considered correlated in projects, but this can be sometimes misleading. EDM proposed to decouple schedule and cost performance measures and developed a number of indices to measure progress respectively. Elshaer (2013) developed an approach which integrates activity sensitivity information in ESM time forecasting to calculate project duration forecasts comparable to Lipke (2011), which are both extensions of the traditional ESM. Khamooshi and Golafshani (2014) developed an approach which though sprung from ESM had a different definition of the key metrics. They proposed EDM, which uses time-based rather than cost-based metrics. They opined that using ESM for time forecasting could still produce ambiguous results as the method continuously uses costs as a proxy to measure schedule performance. Thus, ES is calculated based on EV and PV values, which are both expressed in cost units. They therefore developed the technique replacing the ES metric by earned duration ED(t), which is calculated as the projection of the total earned duration on the total planned duration based on metrics, expressed in time units instead of the projection of EV on PV, yielding ES. Studies on EVM forecasting accuracy by Batselier and Vanhoucke (2015b) and Vanhoucke and Vandevoorde (2007) have however found ESM to be dominant over EVM and EDM.

Batselier and Vanhoucke (2015) integrated EVM methodology with the exponential smoothing forecasting approach to propose XSM (eXponential Smoothing-based Method). Rationalising the use of the exponential smoothing technique, they implied that since the data collected during a project represent a time series, exponential smoothing which is applied to any time series can therefore be utilised to forecast project duration and cost. The forecasting approach developed by Batselier and Vanhoucke (2015) for both project duration and cost is also an extension of the established EVM and earned schedule (ES) cost and time forecasting methods. The approach requires only one smoothing parameter to calculate the enhanced EVM performance factor, which can be adjusted during the project's growth based on information of past performance and/or anticipated management actions. The XSM is built by integrating the known EVM metrics into the exponential smoothing formulas. Batselier and Vanhoucke (2015) additionally emphasised that XSM demonstrates a significant improvement in overall performance compared with the most accurate project forecasting methods proposed by previous research.

They further argue that the XSM can be applied in both a static and a dynamic way. Static approach chooses the value for β before the project begins and remains constant throughout the project duration. Additionally, Batselier and Vanhoucke (2015) compared the forecast accuracies of the static and the dynamic approach of the XSM with the accuracies of the most known and best performing EVM forecasting methods for both time and cost. They found XSM could potentially produce forecasts that are on average 14.8% more accurate than the best EVM time forecasting methods and 25.1% more accurate compared to the best EVM cost forecasting method.

2.2 Forecasting megaprojects

It has to be noted that all the methods mentioned in section 2.1, do use data collected during a project, predominantly in order to spot differences in between planned and achieved intermediate targets, and thus forecast if a project will be finishing on time and/or on budget. There is however no use of any a priori information from that specific project, or any other project(s) for that matter, similar or not. This is a major pitfall of these models, as they are reactive in nature and cannot provide forecasts in advance.

This latter deficit renders these models (EVM, EDM, ESM, XSM, etc) inapplicable for forecasting the success (especially) of megaprojects, where any deviations in budgets and duration are of significant importance and thus need to be known in advance. To address this issue, we discuss hereafter models that can forecast a priori variances in a project's cost and/or duration, as well as forecast adequately the overall success in terms of benefits realisation.

Quantitative forecasting models and, in general, models that best fit past data may be the best methods to predict such future outcomes (Nikolopoulos & Thomakos, 2019; Makridakis et al. 1998). However megaprojects are usually one-off projects, and as such usually, no past data are available. Thus, if we want to forecast the success of megaprojects we can neither rely on models requiring past data such as quantitative forecasting methods, as advanced as these may be: multiple regression, data mining, neural nets, and big data analytics.

In the aforementioned context, a promising alternative is judgemental forecasting methods, that are increasingly being recognised as on par or even advantageous to quantitative forecasting approaches (Armstrong & Green, 2018; Lawrence et al., 2006, Makridakis & Gaba, 1998), especially for very complex problems where the views of experts might be the only way to estimate future outcomes (Tetlock & Gardner, 2015).

Furthermore, judgemental forecasts are prescribed in situations that are suitable for the characteristics of megaprojects. For example, Makridakis, Gaba and Hogarth (2009) suggested that judgemental forecasting is suitable where there are scarce quantitative data and where the level of uncertainty is very high. Similarly, O'Connor and Lawrence (1998) suggested judgemental forecasting where expert knowledge is believed to be needed to improve forecasting accuracy. Judgemental methods are quick to use and typically inexpensive (Makridakis et al. 1998). However, selecting the best judgemental method may be contingent on the requirements of the forecasting situation (Meyer & Booker, 2001).

In Unaided Judgement, individuals are not provided with any form of guidance about forecasting; this is the standard benchmark of judgemental forecasting (Green & Armstrong, 2007). However, it is not without flaws, which has prompted academic researchers to suggest using structured judgemental forecasting methods or tools to predict the outcome of projects over less structured methods (Armstrong, 1986; 2001; Nikolopoulos et al. 2015). In principle, although unaided judgement can provide useful information, this method of forecasting produces inaccurate forecasts as the forecasters may not always be able to recall analogous cases correctly (Green, 2002; Lee, Goodwin, Fildes, Nikolopoulos & Lawrence, 2007).

Therefore, the adoption of structured approaches to judgemental forecasting is considered a way to overcome the limitations of unaided judgement and fully utilise expert judgement (Green & Armstrong, 2007). Taking this up a notch, Armstrong and Green (2018) demonstrated that incorporating evidence-based methods is more useful to processing complex information reliably. Similarly, Green and Armstrong (2004) suggested that an expert's understanding of their own analogies may help them to provide accurate forecasts.

Specifically, there is evidence to show that structured analogies and interaction groups provide more accurate forecasts than unaided judgement up to about 54% when the necessary conditions of forecasting are met (Nikolopoulos et al., 2015). These requirements and conditions, include but not limited to employment of experts from diverse backgrounds, using more related analogies, engaging experts with a high level of experience, and encouraging the interaction of experts, are contained in the checklist for forecasting methods and principles proposed by Armstrong and Green (2018). In this study an analysis across ten comparative tests from three studies shows an average 40% reduction in the error of forecasts made using structured analogies (Armstrong & Green, 2018). Nevertheless, the success of judgemental forecasting methods also rests upon careful examination and management of the strengths and weaknesses of the methods chosen (Lawrence et al., 2006; Parackal et al., 2007).

One effective way to combine the wisdom of the crowds is to form Interaction Groups; this method suggests active interaction with a group of experts until a consensus forecast is reached through deliberation and discussion. The ability to pool information from these deliberations is a crucial factor that could make or mar the process. This method is not without its flaws. Potential problems could arise from group biases introduced by the face-to-face contact of the experts, such as the ‘central tendency’ and the ‘dominant personalities’ effects (Van de Ven & Delbecq, 1971). Besides, group-based approaches tend to attract extra costs from multiple rounds in the Delphi set-up or the need for meetings in the formulation of Interaction Groups. This fact renders these methods relatively more costly than other methods that group-based approaches are competing against and could be a potential disadvantage. Finally, there is mixed evidence about the forecasting potential of Interaction Groups (Armstrong, 2006; Boje & Murnighan, 1982; Graefe & Armstrong, 2011).

Literature Gap

From the synthesis of the literature review in sections 2.1 and 2.2 the following gap can be identified: there is a need to provide accurate forecasts for the duration, the budget and the realised benefits of megaprojects. These usually are one-off projects without sufficient history to build quantitative models, and as such qualitative methods need to be employed. To that end, structuring has proven to help so we need formal and structured judgemental methods. There is also evidence that combining the wisdom of the crowds works in that context too, so employing interaction groups could improve forecasting performance.

Research Aim

This is exactly the stream of research and the body of theory that this research corroborates: employing the class of judgemental methods for the very complex task of forecasting the specific objectives and overall success of megaprojects.

3. Methodology

We are going to use semi-experts in order to evaluate various judgmental forecasting methods for forecasting megaprojects. In order to be able to evaluate the alternative methods, we need a megaproject that has already happened so we know the outcomes. However, we cannot name it, as some of the semi-experts might remember the exact outcome of it. Thus we need to sufficiently disguise it, and design a control experiment where different samples of semi-experts are asked to forecast with different methods the outcomes of the megaproject. The methods employed vary from very simple and unstructured to structured, and from individual forecasting to interaction groups.

The real megaproject examined in this research is about space exploration. The project is sufficiently disguised so the semi-experts cannot – and should not – identify it. It appears as if the project has not commenced yet. The detail of the project description and the experimental set-ups for UJ is provided in detail in Table 1; the actual required forecasts as well the actual outcomes are in Table 2.

Megaproject: Space Exploration

Description

A number of space probes left Earth for planets in the past few years. One of the missions is estimated to cost £250m to £300m and it will become a European-built probe on a spacecraft touching down on another planet. The aim is always simple – to find evidence of life, past or present, on another planet. The mission carries scientific instruments that will study the geology of planets and search for water under the surface. Research institutes throughout Europe have provided the instruments. A consortium of more than 20 companies from more than a dozen European countries and the USA built the spacecraft. The spacecraft will fly around the target planet for an entire planet year. Scientists are confident that if water is present on the planet, the spacecraft with the probe will find it.

European scientists want the mission to:

a) map the composition of the surface at 100-m resolution
b) map the composition of the atmosphere and determine its global circulation
c) determine the structure of the sub-surface to a depth of a few kilometres

d) determine the effect of the atmosphere on the surface, and,

e) determine the interaction of the atmosphere with the solar wind

On landing, cameras on the probe's robotic arm will take close-up images of soil and rocks to look for interesting specimens. The samples will be analysed for chemical signs of life using a package of instruments on the probe.

The Launch

The spacecraft carrying the probe will be launched from earth and placed on the right trajectory for the interplanetary voyage. If all goes well, the journey will take a few months.

Table 1. Disguised description of the Megaproject

1). To what extent do you think objectives a-e will be achieved?

a. 0% - 20% [ ]

b. 21% - 40% [ ]

c. 41% - 60% [ ]

d. 61% - 80% [ ]

e. 81% - 100% [ ]

2). Do you think water will be found?

Yes [ ]   No [ ]

3). Do you think close-up images will be captured?

Yes [ ]   No [ ]

Table 2. Actual questions and 'Shaded' the correct forecast for the Megaproject

In order to finalise the exact phrasing of the narrative of the disguised case and the respective questions a pilot experiment was run in an executive MBA class at Salford University with six participants in November 2017. The detail of the project description and the experimental setup for forecasting with the modified version of structured analogies s-SA is provided in detail in appendix A.

3.1 The semi-experts

Following the classification of the Savio and Nikolopoulos studies (2010, 2009b, 2013), we consider MSc and MBA students to be semi-experts. The MBA students[b] were attending a top-30 MBA programme (Global MBA FT 2017 rankings) and had at least three years of industrial experience and full training in quantitative forecasting methods. Many of them also were industrial engineers by training and had moderate experience of managing industrial projects. In total – from a class of 69 experienced and excellently educated students – 53 responded positively to the call and participated in the research. These semi-experts were sourced from a wide variety of sectors, including academia, industry, financial services and consultancy firms; all however were of south-east Asian origin and almost all were raised in India.

No monetary incentive was provided. However, an in-kind incentive was provided to the participants for taking part in the experiment: a bonus grade of 0.5 in case students fell below 2.5 (with a maximum of 4.0) in their grade for the Forecasting analytics course – more of a ‘safety net’ than a bonus per se.

[b] Honouring the consent given from the participating students, we cannot disclose the academic institution in which the behavioural experiment took place in early 2018.

3.2 Judgemental forecasting methods

Three methods have been evaluated in this study; the first – Unaided Judgement – is the benchmark[c]. The methods that were deployed included the following:

Group A - (53 semi-experts from a pool of 69 students), Unaided Judgement (UJ): This method is a simple and popular Judgemental Forecasting approach. Semi-experts are given no guidance except for a general description of the megaproject. The task lasted for 5 minutes

Group B - (45[d] semi-experts from the same pool of 69 students), semi-Structured Analogies (s-SA): The Structured Analogies approach was proposed by Green and Armstrong (2007) and is based on forecasting by analogy by exploiting the similarities of past events or experiences. These past events/situations have the same or similar characteristics as the problem to be forecasted and can be used as templates. These types of mental templates are the analogies. The semi-experts are first asked to recall as many analogies as possible. Subsequently, they produce a quantitative similarity rating between each analogy and the problem to be forecasted and state the outcome of that analogy. The administrator uses the semi-experts' data to produce a final forecast. In this study, a slightly simpler version of the method, called semi-Structured

[c] This is the standard benchmark in Judgmental forecasting (Nikolopoulos et al. 2015), as it is Naïve for time series forecasting (Nikolopoulos & Thomakos, 2019). In empirical forecasting investigations the simplest method is usually used as benchmark, despite being not very accurate. It is also computationally cheap, almost effortless.

[d] Some of the 53 students that were in group A, did not took part in groups B (45mstudentS) and C as were not present in the second and third phase of the experiment, that took part two and seven days later respectively.

Analogies (s-SA, Savio & Nikolopoulos, 2013; 2010; 2009a; 2009b) was implemented. In this approach, similarity ratings and outcomes are not used by the administrator to generate forecasts because the final forecasts are produced by the semi-experts. The task lasted for 15 minutes

Group C - (6-7 semi-experts per group - from the same pool of 69 students), 8 (eight) Interaction Group (IG): These groups met in a restaurant/cafeteria for an hour with their laptops, and internet connection was available. The entire process was supervised by a relatively inexperienced facilitator – the team captain. The meeting lasted three hours and was recorded. The first hour was spent with introductions and a light dinner. In the next two hours, the group forecasting exercise occurred, in which the semi-experts were first given the questionnaires, then encouraged to recall analogies and their corresponding outcomes, and then to rate those analogies in terms of similarity. Finally, the semi-experts were asked to select the most appropriate analogies to produce point forecasts as well as 90% prediction intervals. This process was first performed individually and was then followed by the group interaction in which the semi-experts repeated the process aloud and exchanged their information until a consensus group forecast was reached.

4. Results

Measuring Performance

Forecasting accuracy was measured through a [% success] metric of how often the correct answer was achieved from every group. Given the nature of the question and respective answer/forecast as % in steps of 20%, we considered the calculation of any other metrics such as MAPE (Mean Absolute Percentage Error) unnecessary. For the three questions presented in

Table 2 (with the realised outcomes listed in the same Table 2), all errors for the semi-experts' forecasts were calculated. For each of the methods and questions, the % success was estimated.

All groups forecasted perfectly questions 2 and 3, which were the yes/no ones, so everybody agreed that pictures would be taken while water would not be found on the unexplored planet. So these were perceived and proved to be the easy ones given both the Boolean nature and the recent memories of most space projects.

Therefore, our focus was on question one, where the extent of the success of the mission could be judged across five objectives on a scale of 1-100% with steps of 20%. The results are as follows where the IG group method clearly outperformed the alternatives:

- **Unaided Judgement**

  The accuracy for the UJ (Group A) for Q1 – forecasting accurately that 80% of the objectives was achieved – was **22.64%**.

- **Semi-Structured Analogies**

  The results for s-SA for Q1 was **27.27%** so better by almost 5% in absolute terms and as a performance improvement in the range of 20%.

  Many semi-experts recalled one to two analogies per policy, whereas others provided no analogies at all.

- **Interaction Group (IG)**

  The results for IG was a success rate of **57.14%** so better by almost 30% in absolute terms and as a performance improvement by over 100%

Participants' Expertise

The participants' expertise was rated based on the self-administered questionnaire provided with the SA method – see appendix; however, this was a very homogeneous group given the admission nature of the MBA programme and thus most of the candidates had 3-5 years of experience with very little deviation.

5. Discussion

The proposed judgemental algorithms are very simple so very easy and cheap/cost-effective to use in practice. So for forecasting megaproject semi-structured analogies used in interaction groups of a small number of semi-experts, proved to be more than enough in order to provide in advance sufficient forecasts for the success of megaprojects.

Given that *simple* is often passed for with *simplistic* in scientific research, we think we need to elaborate further on the importance of simplicity in derived scientific findings; for being a desirable property, rather the other way round. In fact, we live in an era where complex AI solutions are perceived to be superior before even tested against simpler - and computationally cheaper - benchmarks. However, this is exactly the raison d'être of the forecasting discipline: statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones (Makridakis & Hibon, 2000).

Of these simple approaches, the more structured one seemed to prove more accurate and the teaming of semi-experts really paid off – a result consistent with the overall body of literature and especially the results of the recent and widely popularised superforecasting project on the aspects of training (here the SA training and respective use of methods) and teaming up (Tetlock & Gardner, 2015).

5.1 Simplicity in scientific research

Simplicity should not be a negative factor in the evolution and promotion of science; to the contrary, the application of the simplicity principle to theories is sometimes defended as an application of Occam's Razor, that is, ‘accept the simplest theory that works’ (Simon, 1979). Zellner (2007), a leading economist, believed that complicated problems could be solved by the application of a few powerful, simplifying concepts, which he called ‘sophisticated simplicity’. These powerful and simplifying concepts have been implemented in a myriad of industries and services. Simplicity also plays an integral role in shaping decision-making heuristics. Gigerenzer (1996) argues that biases that stem from heuristics can be eliminated by utilising particular methods in a suitable context. In our case, this aforementioned methodological approach translates into using structured judgemental forecasting methods in a very complex and long-term forecasting problem

5.2 Simplicity in forecasting

Green and Armstrong (2015) have elaborated further on this topic by compiling a volume of articles in the subject field. They concluded that simplicity in forecasting requires that (1) method, (2) representation of cumulative knowledge, (3) relationships in models, and (4) relationships among models, forecasts, and decisions are all sufficiently uncomplicated as to be easily understood by decision-makers. Their review found 97 comparisons in 32 papers where none provided a balance of evidence that complexity improves forecast accuracy. To the contrary, they argue that complexity increases forecast error by 27 percent on average based on evidence from 25 academic studies. Nevertheless, complexity remains, and ‘incomprehensibility’ might be the element of the acceptance of forecasts from complex models.

5.3 Generalisation

Although the empirical evidence in this study was derived within a megaproject context, the results may be generalised and applied to a variety of other project situations in which the proposed forecasting methods might be used to forecast the critical success factors of projects. Also, these judgemental methods could well be used for any forecasting problem where limited cross-sectional data may be available, and there is an absence of any historic trends (Nikolopoulos et al. 2007).

6. Conclusion

Forecasting megaprojects is very important from a 'social good' perspective. Megaprojects are designed, run and completed in order to serve a bigger purpose; not just achieve some short-term financial goals. The bigger picture includes the benefits realisation, and these are usually aiming for the broader and long–term social good. For example when you run the Olympic games in a city, one thing is finishing the stadia on time, another - and much bigger - is what do you aspire to do (and transfort the city to) through this new state-of-the-art infrastructure . So, being able to forecast in advance the benefits realised is of fundament importance and thus we consider this research adds also to the literature in 'forecasting for social good'.

Forecasting megaprojects is also very challenging. This study utilises a space exploration mission, one of the most challenging, complex and longitudinal type of megaprojects, in which available historical information is limited and the forecasting horizon is extremely long. The results presented here could well be generalised and applied to many other megaprojects; however, more research should be carried out in the subject field.

The empirical evidence reveals that the use of s-SA Analogy leads to accuracy improvement compared with UJ. This improvement in accuracy is greater when introducing pooling of analogies through interaction in IG. The results also corroborate the stream of forecasting research in the presence of information cues (Nikolopoulos et al. 2007). The preliminary empirical findings suggest that overall actual forecasting improvement might exceed 100%. These results are consistent with the previous body of literature; however, the exact effect size varies depending on the context of each study.

Forecasting Principles

With the aforementioned results, it can be claimed that this study corroborates the existing body of evidence that supports the forecasting principles as maintained by J.S. Armstrong (2001a) at www.forprin.com. In further detail, empirical evidence is provided in favour of the following forecasting principles (Armstrong, J. S., 2001b).

Principle 3.5: *Obtain information from similar (analogous) series or cases.*

Principle 6.3: *Use structured forecasting methods rather than unstructured.*

Principle 7.1: *Keep methods simple.*

Principle 8.3: *Ask experts to justify their forecasts.*

Principle 12.2: *Use many approaches (or forecasters), preferably at least five.*

Principle 13.26: *Use out-of-sample (ex ante) error measures.*

Generalisation

The results presented herein are based on small-sized samples of semi-experts, a fact that might be an impediment for generalising the findings, or not (Armstrong 2007a, 2007b). However, if the context of this megaproject were to be taken into account, and how megaprojects are managed and more importantly a priori forecasted in real-life conditions, these results might provide valid insights into the performance and usability – real-life usability – of each forecasting method. Repetition in other case studies might help to prove the validity of the findings and provide a generalised output for the superiority of some these methods, especially the simpler ones, such as Structured Analogies.

The Future

The proposed approaches could also be tested for smaller and bigger megaprojects in order to gather further evidence that would allow for the full generalization of the results.

Moreover, an evaluation of other judgemental approaches, such as the Delphi methods (Rowe & Wright, 2001; 1999) and Nominal Group Technique (Van de Ven & Delbecq, 1971), could be explored (Graefe & Armstrong, 2011).

In addition, sampling more experts would offer the opportunity to test more treatments, such as IGs with UJ versus IGs with s-SA, direct comparisons of IGs and Delphi as well as versus SA as it was originally designed by Green and Armstrong (2007b).

Finally it is worth investigating the extent of the moderation from different levels of expertise.

References

Alderman, N., Ivory, C., McLoughlin, I., Vaughan, R. (2005). Sense-making as a process within complex service-led projects. International Journal of Project Management 23 (5), 380–385.

Anbari, F.T. (2003). Earned value project management method and extensions. Project management journal, 34(4), 12-23.

Armstrong, J. S. (1986). The Ombudsman: Research on Forecasting: A Quarter Century Review, 1960-1984. Interfaces 16 (1), 89-109.

Armstrong, J. S. (2001). Combining Forecasts. In J. S. Armstrong, Principles of Forecasting: A Handbook for Researchers and Practitioners. Norwell, Massachusetts: Kluwer Academic Publishers, pp. 417-439.

Armstrong, J. S. (2001b). Standards and practices for forecasting. Available from http://www.forecastingprinciples.com/

Armstrong, J. S. (2006). How to make better forecasts and decisions: Avoid face-to-face meetings. Foresight-The International Journal of Applied Forecasting, 5, 3–8.

Armstrong, J. S. (2007a). Significance tests harm progress in forecasting. International Journal of Forecasting, 23, 321–327.
Armstrong, J. S. (2007b). Statistical significance tests are unnecessary even when properly done and properly interpreted: Reply to commentaries. International Journal of Forecasting, 23, 335–336.
Armstrong, J. S. & Green, K. C. (2018). Forecasting methods and principles: Evidence-based checklists. Journal of Global Scholars of Marketing Science, 28(2), 103-159.
Baqerin, M.H., Shafahi, Y. & Kashani, H. (2015). Application of Weibull analysis to evaluate and forecast schedule performance in repetitive projects. Journal of Construction Engineering and Management, 142(2), 04015058.
Barraza, G.A., Back, W.E. & Mata, F. (2004). Probabilistic forecasting of project performance using stochastic S curves. Journal of Construction Engineering and Management, 130(1), 25-32.
Batselier, J. & Vanhoucke, M., 2015. Evaluation of deterministic state-of-the-art forecasting approaches for project duration based on earned value management. International Journal of Project Management, 33(7), 1588-1596.
Bekker, M. C. & Steyn, H. (2007, September). Defining 'project governance' for large capital projects. In AFRICON 2007 (pp. 1-13). IEEE.
Boje, D. M. & Murnighan, J. K. (1982). Group confidence pressures in iterative decisions. Management Science, 28, 1187–1196.
Capka, J.R. (2006). Issuance of Interim Major Project Guidance, Department of Transportation Federal Highway Administration, U.S.
Elshaer, R. (2013). Impact of sensitivity information on the prediction of project's duration using earned schedule method. International Journal of Project Management, 31(4), 579-588.
Flyvbjerg, B. (2005). Policy and Planning for Large Infrastructure Projects: Problems, Causes, Cures. World Bank Policy Research Working. Paper 3781
Flyvbjerg, B., Bruzelius, N., & Rothengatter, W. (2003). Megaprojects and risk: An anatomy of ambition. Cambridge: Cambridge University Press.
Flyvbjerg, B., Garbuio, Massimo, Lovallo, D. (2014). Better forecasting for large capital projects. McKinsey Reports. Available on https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/better-forecasting-for-large-capital-projects
Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. Psychological Review, 103, 592–596.
Graefe, A., & Armstrong J. S. (2011). Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. International Journal of Forecasting, 27, 183–195.
Green, K. C. (2002). Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgement. International Journal of Forecasting, 18(3), 321-344.
Green, K. C. & Armstrong, J. S. (2004). Structured Analogies for Forecasting. Monash University Econometrics and Business Statistics Working Paper No. 17/04. Available at http://dx.doi.org/10.2139/ssrn.675363
Green, K. C. & Armstrong, J. S. (2007). Structured Analogies for forecasting. International Journal of Forecasting, 23, 365–376.

Green, K. C. & Armstrong, J. S. (2007b). Value of expertise for forecasting decisions in conflicts. Interfaces, 37, 287–299.

Green, K. C. & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. Journal of Business Research, 68, 1678-1685.

International Project Management Association (IPMA), IPMA Project Excellence (PE) Awards: PE Award Categories. Accessed on 26 September 2018, from: http://www.ipma.ch/awards/projexcellence/Pages/PEAwardCategories.aspx.

Jacob, D.S. & Kane, M. (2004). Forecasting schedule completion using earned value metrics revisited. The Measurable News, 1(11), 7.

Khamooshi, H. & Golafshani, H. (2014). EDM: Earned Duration Management, a new approach to schedule performance management and measurement. International Journal of Project Management, 32(6), 1019-1041.

Kim, B.C. & Reinschmidt, K.F. (2010). Probabilistic forecasting of project duration using Kalman filter and the earned value method. Journal of Construction Engineering and Management, 136(8), 834-843.

Lawrence, M., Goodwin, P., O'Connor, M. & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. International Journal of Forecasting, 22(3), 493-518.

Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K. & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks. International Journal of Forecasting, 23, 377–390.

Lipke, W. (2003). Schedule is different. The Measurable News, 31(4), 31-34.

Lipke, W. (2011). Earned schedule application to small projects. PM World Today, 13, pp.1-12.

Makridakis, S. & Gaba, A. (1998). Judgment: Its Role and Value for Strategy. In: G. Wright & P. Goodwin, Forecasting with Judgment. London: John Wiley & Sons, pp. 1- 38.

Makridakis, S., Gaba, A., & Hogarth, R. M., (2009). Dance with Chance: Making Luck Work for You, Oxford: Oneworld Publications.

Makridakis, S. & Hibon, M. (2000). The M-3 Competition: results, conclusions, and implications. International Journal of Forecasting. 16 (4), 451–476.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting: Methods and Applications, London: John Wiley & Sons.

Meyer, M. A. and Booker, J. M. (2001). Eliciting and Analyzing Expert Judgment: A Practical Guide, American Statistical Society and the Society for Industrial and Applied Mathematics.

Merrow, E. (1988). Understanding the Outcomes of Megaprojects. RAND Corporation.

Merrow, E. W. (2011). Industrial megaprojects: concepts, strategies, and practices for success (Vol. 8), Hoboken, NJ: Wiley.

Miller, R. & Hobbs, B. (2005). Governance Regimes for Large Complex Projects. Project Management Journal, 36, 42-50.

Miller, R. & Lessard, D. R. (2000). The Strategic Management of Large Engineering Projects: Shaping Institutions, Risks and Governance, Massachusetts, Massachusetts Institute of Technology

Miller, R. and D. Lessard. (2007). Evolving strategy: Risk management and the shaping of large engineering projects. MIT Sloan School of Management. Working Paper: 4607-4639.

Mortaji, S.T.H., Noorossana, R. and Bagherpour, M. (2014). Project completion time and cost prediction using change point analysis. Journal of Management in Engineering, 31(5), p.04014086.

Nikolopoulos, K. & Thomakos, D. D. (2019). Forecasting with the Theta Method: Theory & Applications, Wiley: New Jersey: Wiley.

Nikolopoulos, K., Litsa, A., Petropoulos, F., Bougioukos, V., & Khammash, M. (2015). Relative performance of methods for forecasting special events. Journal of Business Research, 68(8), 1785-1791.

Nikolopoulos, K., Goodwin, P., Patelis, A., & Assimakopoulos, V. (2007). Forecasting with cue information: a comparison of multiple regression with alternative forecasting approaches. European Journal of Operational Research, 180, 354–368.

Parackal, M., Goodwin, P. & O'Connor, M. (2007). Judgment in Forecasting. International Journal of Forecasting, 23 (3) 343-345.

Petropoulos, F., Makridakis, S., Assimakopoulos, V. & Nikolopoulos, K. (2014) 'Horses for Courses' in demand forecasting. European Journal of Operational Research 237 (1): 152-163.

Rowe, G. & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. International Journal of Forecasting, 15, 353–375.

Rowe, G. & Wright, G. (2001). Expert opinions in forecasting. Role of the Delphi technique. In J. S. Armstrong (Ed.), Principles of forecasting: A handbook for researchers and practitioners (pp. 125–144). Norwell, MA: Kluwer Academic Publishers.

Sanderson, J. (2012). Risk, uncertainty and governance in megaprojects: A critical discussion of alternative explanations. International journal of project management, 30(4), 432-443.

Savio, N. D. & Nikolopoulos, K. (2009a). Forecasting the economic impact of new policies. Foresight, 11(2), 7–18.

Savio, N.D. and Nikolopoulos, K. (2009b) "Forecasting Effectiveness of Policy Implementation Strategies: working with semi-experts", *Foresight* 11(6), 86-93.

Savio, N. D. & Nikolopoulos, K. (2010). Forecasting the effectiveness of policy implementation strategies. International Journal of Public Administration, 33, 88–97.

Savio, N. D. & Nikolopoulos, K. (2013). A strategic forecasting framework for governmental decision-making and planning. International Journal of Forecasting, 29, 311–321.

Simon, H. A. (1979). Rational decision making in business organizations. The American Economic Review, 69, 493–513.

Talbot, C. (2009). Public value – The next "big thing" in public management? International Journal of Public Administration, 32, 167–170.

Tetlock, P. & Gardner, D. (2015). Superforecasting: The Art and Science of Prediction Paperback, Crown;

Turner, T. and Zolin, R. (2012). Forecasting Success on Large Projects: Developing Reliable Scales to Predict Multiple Perspectives by Multiple Stakeholders Over Multiple Time Frames, Project Management Journal – October issue

Van de Ven, A. & Delbecq, A. L. (1971). Nominal versus interacting group processes for committee decision-making effectiveness. The Academy of Management Journal, 14, 203–212.

Zellner, A. (2007). Philosophy and objectives of econometrics. Journal of Econometrics, 136, 331–339

Appendix A. Structured Analogies Judgement (Group B).

Megaproject: Space Exploration

Description

A number of space probes left Earth for planets in the past few years. One of the missions is estimated to cost £250m to £300m and it will become a European-built probe on a spacecraft touching down on another planet. The aim is always simple – to find evidence of life, past or present, on another planet. The mission carries scientific instruments that will study the geology of planets and search for water under the surface. Research institutes throughout Europe have provided the instruments. A consortium of more than 20 companies from more than a dozen European countries and the USA built the spacecraft. The spacecraft will fly around the target planet for an entire planet year. Scientists are confident that if water is present on the planet, the spacecraft with the probe will find it.
European scientists want the mission to:

f) map the composition of the surface at 100-m resolution
g) map the composition of the atmosphere and determine its global circulation
h) determine the structure of the sub-surface to a depth of a few kilometres
i) determine the effect of the atmosphere on the surface, and
j) determine the interaction of the atmosphere with the solar wind

On landing, cameras on the probe's robotic arm will take close-up images of soil and rocks to look for interesting specimens. The samples will be analysed for chemical signs of life using a package of instruments on the probe.

The Launch

The spacecraft carrying the probe will be launched from earth and placed on the right trajectory for the interplanetary voyage. If all goes well, the journey will take a few months.

Judgemental Forecasting

We are interested in the following **Forecasts**:
1). To what extent do you think objectives a-e will be achieved?
2). Do you think water will be found?
3). Do you think close-up images will be captured?

You are going to follow the process for **Structures Analogies** for producing your forecasts as in the **following pages**

Judgemental Forecasting with Structured Analogies

In the tables provided below, please describe any analogous project to the one described. Please include details on:

- the similarities and differences between your analogous project and the target projects.
- their source (e.g. your own experience, media reports, history, literature, etc.)
- a similarity rating between your analogous project and the target projects (0 = no similarity… 5 = similar… 10 = high similarity)

- the outcome of your analogous project (which of the outcomes a-e found at the bottom, is most similar, in terms of effectiveness, to the outcome of your analogy?).

Example analogy

Description	Landing on the Moon – Apollo mission
Similarities and differences	Similarities: same objective Differences: different budget available

Source __ Media __ Similarity rating __8__ OUTCOME:
Q1.To what extent do you think objectives have been achieved?
a. 0% - 20% [ ] b. 21% - 40% [ ] c. 41% - 60% [V]
d. 61% - 80% [ ] e. 81% - 100% [ ]
Q2. Was water found?
Yes [ ] No [V]
Q3.Have close-up images been captured?
Yes [V] No [ ]

1. Your Analogies

Analogy 1

Description	
Similarities and differences	

Source ______________ Similarity rating ______ OUTCOME:
Q1.To what extent do you think objectives have been achieved?
a. 0% - 20% [ ] b. 21% - 40% [ ] c. 41% - 60% [ ]
d. 61% - 80% [ ] e. 81% - 100% [ ]
Q2. Was water found?
Yes [ ] No [ ]
Q3.Have close-up images been captured?
Yes [ ] No [ ]

Analogy 2

Description	
Similarities and differences	

Source ______________ Similarity rating ______ OUTCOME:
Q1.To what extent do you think objectives have been achieved?
a. 0% - 20% [ ] b. 21% - 40% [ ] c. 41% - 60% [ ]
d. 61% - 80% [ ] e. 81% - 100% [ ]
Q2. Was water found?
Yes [ ] No [ ]
Q3.Have close-up images been captured?

Yes [ ]   No [ ]

Analogy 3

Description	
Similarities and differences	

Source _______________ Similarity rating _______ OUTCOME:

Q1.To what extent do you think objectives have been achieved?

a. 0% - 20% [ ] b. 21% - 40% [ ] c. 41% - 60% [ ]

d. 61% - 80% [ ] e. 81% - 100% [ ]

Q2. Was water found?

Yes [ ]   No [ ]

Q3.Have close-up images been captured?

Yes [ ]   No [ ]

Analogy 4

Description	
Similarities and differences	

Source _______________ Similarity rating _______ OUTCOME:

Q1.To what extent do you think objectives have been achieved?

a. 0% - 20% [ ] b. 21% - 40% [ ] c. 41% - 60% [ ]

d. 61% - 80% [ ] e. 81% - 100% [ ]

Q2. Was water found?

Yes [ ]   No [ ]

Q3.Have close-up images been captured?

Yes [ ]   No [ ]

if you need MORE analogies reprint this page

2. Your OWN Forecast

Q1.To what extent do you think objectives a-e will be achieved?

a. 0% - 20% [ ] b. 21% - 40% [ ] c. 41% - 60% [ ]
d. 61% - 80% [ ] e. 81% - 100% [ ]

Q2. Do you think water will be found?

Yes [ ] No [ ]

Q3.Do you think close-up images will be captured?

Yes [ ] No [ ]

How **confident** you are about your Forecast in
Q1[ ]%,
Q2 [ ]% and,
Q3[ ]%?

3. Questionnaire

(1) Roughly, how long did you spend on this task?
{include the time spent reading the description and instructions} [__] mins.
(2) How likely is it that taking more time would change your forecast?
{0 = almost no chance (1/100) … 10 = practically certain (99/100)} [__] 0-10.
(3) If you knew that this case was from the UK, how likely would you be to change your forecast?
{0 = almost no chance (1/100) … 10 = practically certain (99/100)} [__] 0-10.
(4) How many people did you discuss this forecasting problem with? [__] people.
(5) Roughly, how many years' experience do you have working in a project management (PM) issues setting?
[__] years.
(6) Roughly, please rate (out of 10)
- your experience with project management (PM). [___] 0-10
- your experience with projects similar to this one. [___] 0-10
- your suitability for predicting the success of major projects. [___] 0-10

(7) If you were contracted to produce such a forecast, what process/processes would you adopt?
[____] In what sort of time-scale? [___]