PRODUCTION PLANNING & CONTROL
https://doi.org/10.1080/09537287.2025.2578708

Reference class forecasting: promises, problems, and a research agenda moving forward

Chantal C. Cantarelli[a], Kate Davis[a], Jeffrey K. Pinto[b], and Neil Turner[a]

[a]Faculty of Business and Management, Cranfield University, Cranfield, Bedfordshire, UK; [b]Black School of Business, Penn State – Erie, Erie, PA, USA

ABSTRACT
Reference Class Forecasting (RCF) has emerged as a prominent tool to counter chronic optimism in project planning. By applying an 'outside view' – using distributions of comparable past projects – RCF seeks to correct overly optimistic cost and schedule estimates that undermine megaproject delivery. While its adoption has expanded in policy and practice, particularly following Flyvbjerg's work, questions remain about its validity, applicability and limitations. This article critically reviews the promises and problems of RCF and sets out a research agenda for moving the debate forward.

ARTICLE HISTORY
Received 31 July 2025
Accepted 17 October 2025

KEYWORDS
Projects; overruns; reference class forecasting; optimism bias; forecasting

1. Reference class forecasting – origins

Project organizations and funding agencies are currently experiencing one of the largest booms in human history. Investment in megaprojects (defined as being of at least $1 billion USD in value (Flyvbjerg 2014, 6)) worldwide are common, as projects are being earmarked to address a variety of grand societal challenges (Ika and Munro 2022), including climate change, green transitions, infrastructure development and repair. One of the more common developments worldwide with most of these projects is (often) significant cost and schedule overruns (underperformance) (Cantarelli, Flyvbjerg, and Buhl 2012). In fact, within some classes – or type – of project, past research has suggested that these cost overruns average: 157% for Olympic Games (Flyvbjerg, Stewart, and Budzier 2016b), 73% for Information Technology (IT) (Flyvbjerg et al. 2022), 40% for rail (Flyvbjerg and Bester 2021), 96% for dams (Ansar et al. 2014) and time overruns average: 23% for transportation (Park 2021a), 74% for hydroelectric power (Callegari, Szklo, and Schaeffer 2018) and 45% on dams (Ansar et al. 2014). Flyvbjerg and Gardner (2023) provide an overview of cost overruns across 25 project types, with mean overruns ranging from 1 to 238%. Because of the persistent nature of errors of estimation, coupled with the pattern that these errors generally fall within the excessively optimistic range, efforts to address and, where possible, correct such estimates remain of high priority to scholars and public policy officials alike.

In trying to understand the roots of our inability to generate accurate estimates for project costs, schedules, and benefits, scholars cite, among other potential causes, the Prospect Theory arguments originally advanced by Kahneman and Tversky (1979a, 1979b) in a series of papers that addressed the distinctions between "normative" behaviour (how a system acts based on what is objectively correct or how people should want to behave), "descriptive" behaviour (how people actually behave in reality), and "prescriptive" behaviour (developing policies to bridge the two previous states; that is, creating methods or steps that help people behave in accordance with the normative "should"). Their arguments, applied to behavioural economics, offered the first real steps at addressing the perceptual and behavioural gaps between these three psychological states and the economic implications of repeated failure to bridge these gaps. In practical terms, when describing these three states of behaviour, Kahneman and Tversky (1979b) argued that human decision-making often occupies a descriptive state characterized by undue optimism that ignores distributional evidence (for example, statistical data) of likely outcomes. In other words, descriptive behaviour is typically predicated on objective evidence of past examples or states that should help us re-evaluate the logic or our preferred normative behaviour or assumptions. The fact that we often remain immune to descriptive data is the result of a set of built-in psychological biases.

One of the key schools of thought on the causes of these overruns is the role played by bias in initial estimates, often referred to as the planning fallacy (Kahneman and Tversky 1979a). Kahneman and Tversky (1979a, 1979b) contend that human judgement is generally optimistic due to overconfidence and insufficient consideration of distributional information about outcomes (i.e. the record of past experiences). Therefore, decision makers tend to underestimate the costs, completion times and risks of planned actions, whereas they

CONTACT Jeffrey K. Pinto jkp4@psu.edu Black School of Business, Penn State – Erie, Erie, PA 16563, USA

tend to overestimate the benefits of those same actions. Such an error is caused by actors taking a biased, “inside view”, where the focus is on the constituents of the specific planned action instead of on the actual outcomes of similar ventures that have already been completed (Kahneman and Tversky 1979a).

Kahneman and Tversky’s work on optimism bias is profound, but it is based on a limited set of laboratory experiments. Flyvbjerg gathered large data sets and employed Kahneman and Tversky’s theories to public planning and development and found that the net effect of their ideas, when applied to project estimation and planning, is significant. Drawing upon a history of poor project performance (across multiple classes of project) relative to original cost and benefit estimates, Flyvbjerg, Holm, and Buhl (2002) posited their cause as due to inherent flaws in human decision-making that include: (1) optimism bias, due to overconfidence and an unwillingness to consider historical data (distributional information) of the sort that Kahneman and Tversky (1979b) referred to as descriptive (Bracha and Brown 2012). Flyvbjerg defines optimism bias as “a cognitive predisposition found with most people to judge future events in a more positive light than is warranted by actual experience” (Flyvbjerg 2006, 6). Flyvbjerg and Sunstein (2016) extended the Planning Fallacy concept, by referring to the pattern of underestimating the costs, time and risks to complete tasks coupled with an overestimation of a project’s likely benefits as the “planning fallacy writ large.” (2) uniqueness bias, which prevents decision-makers from considering divergent information because “our project is different,” regardless of the degree of similarity with previous examples (Flyvbjerg et al. 2025; Flyvbjerg and Gardner 2023) and (3) strategic misrepresentation, or the deliberate distortion of information to steer a decision, such as the commitment of time and money to a new project, in the direction the advocate supports (Flyvbjerg 2011).

Kahneman and Tversky (1979a) argue that optimism bias can be mitigated by adopting an outside view, which relies on historical distributions rather than future scenarios. This principle underpins Reference Class Forecasting (RCF), first outlined by Kahneman and Tversky (1979a) and later tested in the context of large-scale transportation infrastructure projects by Flyvbjerg and colleagues. A “reference class” of similar projects offers a real, corrective benchmark against which a new project’s cost estimation can be compared. Thus, “reference class forecasting” provides an independent means of offering more accurate cost estimates for projects under consideration for funding. For a detailed description of the steps in RCF, see Lovallo, Cristofaro, and Flyvbjerg (2023, 142).

In practical terms, Flyvbjerg and colleagues contend that RCF has the effect of moderating the embedded and unreasonable optimism at the heart of most project forecasts, by refusing to take estimates at face value, but first submitting them to a statistical correction procedure that compares them against historical data. This corrective reanalysis of original project cost estimates is then likely to give a much more reasonable estimate of project costs at completion. RCF is primarily applied in project appraisal to counter optimism bias by situating forecasts within distributions of comparable past projects, and its results are subsequently used to inform realistic budget setting through appropriate cost and time uplifts.

Moreover, RCF has gained steady support within the UK and other European countries (Flyvbjerg, Hon, and Fok 2016a; HM Treasury, 2022, 2013). The UK was the first European government to make RCF mandatory for new government-sponsored transport project proposal cost estimation, quickly followed by Denmark. Governments of Australia, Germany, Norway, South Africa, Sweden, Switzerland, the Netherlands have also used RCF on individual projects (Flyvbjerg, Hon, and Fok 2016; Park 2021b).

Nevertheless, arguments supporting optimism bias as a main cause of project cost overruns and benefit underperformance and the principles of RCF as a method to address optimism bias are not without their critics. Several scholars have argued against a number of these underlying contentions, including the degree to which the planning fallacy is present as a primary cause of underperformance and the role that behavioural bias actually plays in decision making for major projects (Love, Ika, Sing, et al. 2019; Love, Ika and Ahiaga-Dagbui, 2019; Love and Ahiaga-Dagbui 2018; Mak and Raftery 1992). Love and colleagues argue, based on extensive data, that cost and schedule overruns in megaprojects are more readily explained by project complexity, pervasive uncertainties, and human limitations in forecasting rather than by systemic bias (Ika, Love, and Pinto 2022). They challenge the “bias school” by arguing that cost overruns may often result from error rather than systematic bias, noting that estimators can be subject to pessimism bias as well as optimism (Batselier and Vanhoucke 2016; Bayram and Al-Jibouri 2018; Love, Ika, Sing, et al. 2019), that collective estimation in megaprojects may mitigate individual bias (Caffieri et al. 2018), and that the assumption of bias itself risks being biased. Another frequently mentioned criticism of RCF has to do with the nature of simply accepting post hoc “smoothing” of cost projects based on previous class histories; that is, the problem of focusing on effect data at the expense of causal reasons for project overruns (Pinto 2023).

It is not our purpose to argue the efficacy of bias as a cause of cost estimation error, but to note instead that the “science” here is far from settled. To illustrate, a recent article by Ika, Love, and Pinto (2022) refutes the “either/or” dichotomy proposed by previous scholars as to which cause is more compelling of the underperformance of estimating, arguing that both bias and error, considered jointly, offer a more thorough and reasonable approximation of the causes of project underperformance.

Given the mainstream use of RCF, it seems particularly relevant to offer a view that acknowledges both its potential and ongoing problems, as policy makers and project planners alike continue to embrace the practice. Therefore, the purpose of this article is to offer a current assessment of the state of RCF: its empirical support, advantages and disadvantages and unanswered questions. To achieve this, we conduct a systematic literature review (SLR) structured around two research questions: (1) How does RCF compare with conventional and hybrid forecasting methods, and how has it been applied in practice? (2) What strengths and problems with RCF are identified in the literature? This paper extends previous reviews by moving from providing a broad overview of how RCF has been applied and discussed to a more critical comparison and evaluation. Whereas other studies have identified the potential of RCF and the methodological challenges it faces, we provide a structured assessment of its relative effectiveness, its documented strengths and problems, and its implications for practice. In doing so, we not only consolidate existing knowledge but also translate it into five propositions that set a clear agenda for future research.

Despite its growing prominence in both academic research and public policy, particularly in the UK and Europe, where RCF has been institutionalized, there remains little systematic and critical comparison of how the method performs relative to alternatives. Most prior reviews have catalogued applications of RCF but have not interrogated its strengths, limitations, and assumptions in a structured way. Given the increasing reliance on RCF for major investment decisions, coupled with mounting critiques of its theoretical basis and practical implementation, a comprehensive and critical synthesis is needed.

We have structured this article as follows. Section 2 describes the methodology for conducting an SLR. Then, Sections 3–5 will discuss the key themes of the literature, including a comparison of RCF to other, conventional forecasting techniques and hybrid forecasting methods, practical applications of RCF and the strengths and challenges of using RCF. Section 6 presents directions for future research on this methodology. As a part of evidence for our assessment, we offer a compendium of recent papers that employed RCF, with a summary of their findings.

2. Methodology

This article employs an SLR conducted in the form of a qualitative content analysis, which enables an objective and systematic description of the explicit content of publications (Gold, Seuring, and Beske 2010). Adopting this approach helps address common criticisms of literature reviews, particularly those concerning insufficient critical assessment, rigour, relevance and comprehensiveness (Tranfield, Denyer, and Smart 2003). The review follows the four-step process model of qualitative content analysis proposed by Mayring (2000) and applied by Seuring and Gold (2012): (1) Material collection, (2) Descriptive analysis, (3) Category selection and (4) Material evaluation.

2.1. Material collection

Our literature sample comprises English-speaking peer-reviewed articles on RCF covering the period 2001 to 2025. We did not restrict our search by time period, as RCF is a relatively recent concept with a manageable body of literature, and limiting by year could exclude early foundational studies.

The primary source of data in content analysis of SLRs consists of peer-reviewed journal articles (Gold, Seuring, and Beske 2010; Seuring and Gold 2012). However, given the emergent nature of the field, conference papers were included to capture the most recent relevant contributions that may not yet have appeared in journal form. The unit of analysis is therefore research articles, including both peer-reviewed journal publications and conference papers. Hence, book chapters, conference reviews and a paper erratum were excluded.

We adopted a hybrid search strategy, combining database searches and snowball searches. According to Wohlin et al. (2022), such a hybrid search approach performs better than database-driven searches. For the database search, we conducted a structured keyword search in SCOPUS and Web of Science (WoS), as these are among the most comprehensive and widely used databases for peer-reviewed research in the social sciences and management. Using both databases increases coverage, since they collectively index a broad range of journals and vendor databases in a single location (Thomé, Scavarda, and Scavarda 2016).

The central concept in this review is the method of RCF; therefore, "reference class forecast*" was used as the primary keyword, with the asterisk applied to capture relevant derivatives. This focused approach was justified, as the term precisely captures the relevant body of literature. Alternative searches using abbreviations such as "RCF" or broader related terms were tested, but either produced large numbers of irrelevant results or risked omitting relevant publications. Given the novelty of RCF as a research domain, the body of literature remains limited and clearly identifiable, making a single, well-defined keyword search both appropriate and sufficient for the purpose of this review. The literature search process is shown in Figure 1.

This initial search resulted in 82 articles from Scopus and 73 articles from WoS (Figures 2 and 3). We removed any duplications (a total of 70 articles), after which 85 articles remained in the initial sample. After reading the title and abstract, we excluded 26 articles, and a further 7 articles were excluded after reading the full text. Table 1 shows the exclusion criteria. Where both a conference and journal version of a study were identified, the journal version was retained and the conference version excluded, unless it contained unique information.

The snowballing procedure (backward search and forward search of retrieved papers) (Wohlin 2014) resulted in an additional 10 articles, leading to a final sample of 61 articles.

2.2. Descriptive analysis

SLRs often begin with a descriptive analysis of publication patterns, such as time of publication, journal, geographical focus and methodological approach. In this study, only the distribution of articles by year and journal is reported to provide context, while the main emphasis is placed on content analysis synthesize the conceptual and methodological contributions of the literature.

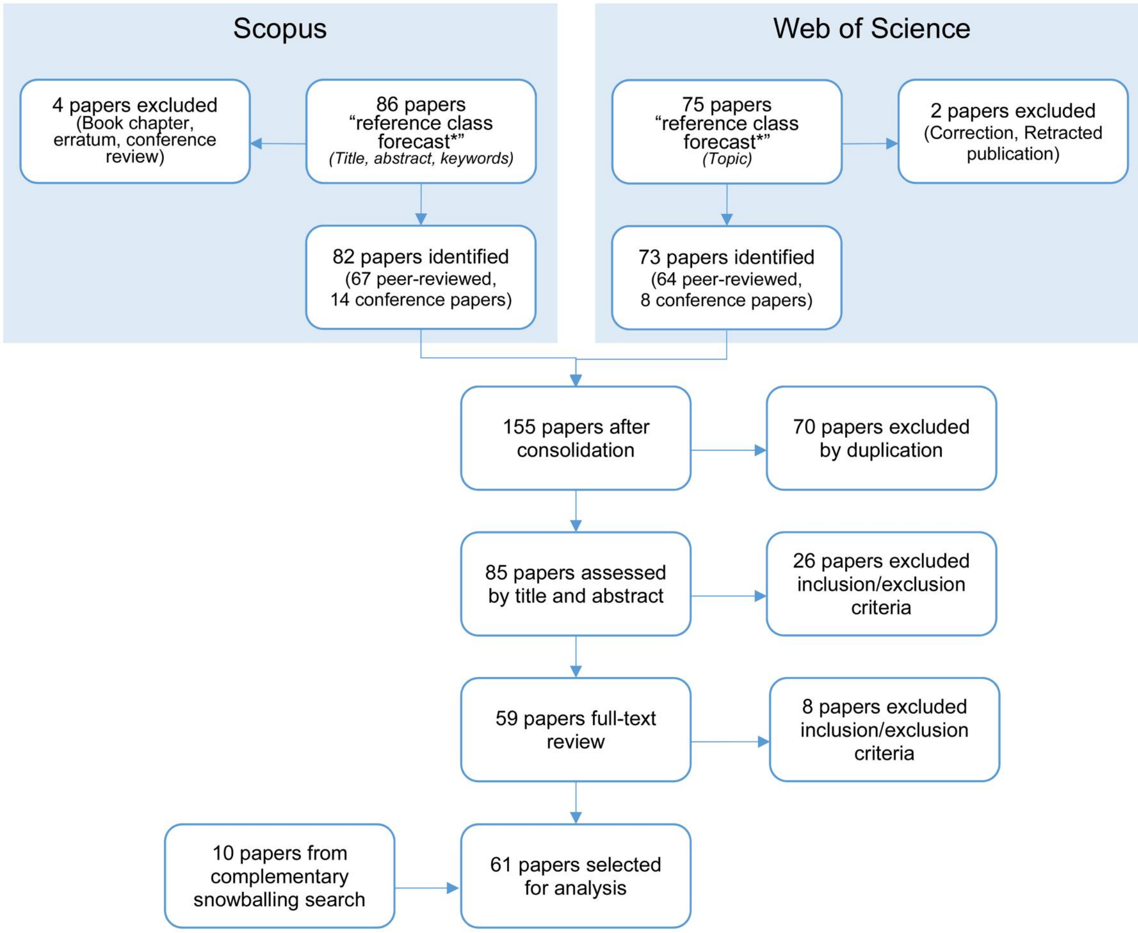

Figure 1. Literature search process.

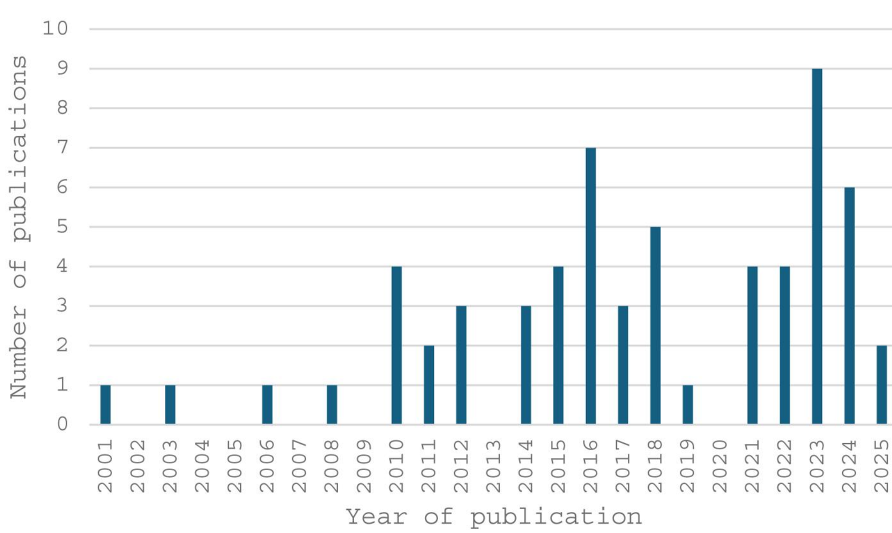

Figure 2. Distribution of publications by year.

Our SLR covers articles between 2001 and 2025, with the majority of the articles (74%) published since 2015. In total, 23 of the 61 articles appeared in journals with more than one contribution, mainly across project management (International Journal of Project Management, Project Management Journal), construction and engineering management (Construction Management and Economics, Journal of Construction Engineering and Management, European Journal of Transport and Infrastructure Research, IEEE Transactions on Engineering Management) and energy and sustainability (Renewable and Sustainable Energy Reviews, Energy Policy) (see Appendix).

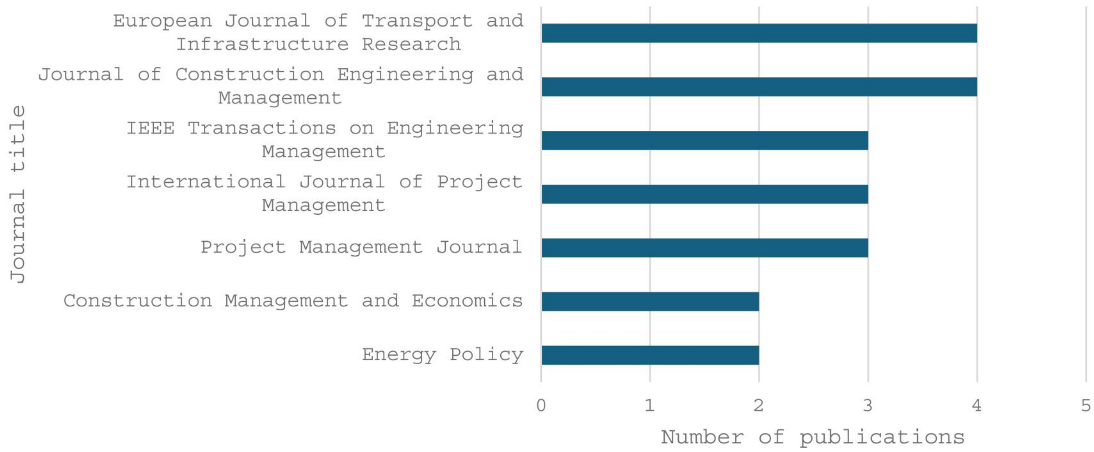

Figure 3. Distribution of articles by journal.

Table 1. Exclusion criteria.

Criteria	Description
Focus on RCF (23 excluded)	Insufficient engagement with RCF: Studies that discuss forecasting approaches with little or no substantive reference to RCF or that mention RCF only in passing without developing, applying or critically assessing it (e.g. only mentions RCF as a recommendation).
Scope (3 excluded)	RCF but not relevant to forecasting
Document type (8 excluded)	No access or duplicate of a journal paper

2.3. Category selection

The content analysis was structured around five analytic categories derived from the review objectives: (1) RCF and conventional forecasting methods, (2) hybrid approaches, (3) practical applications, (4) strengths and (5) problems. These categories provided a systematic framework for synthesizing the studies.

2.4. Material evaluation

The articles were analysed according to the selected categories, whereby articles were coded against one or multiple categories depending on the focus of the article. The findings of the content analysis are described in subsequent sections.

3. Methodological evaluation of RCF as a forecasting technique

Past studies have often focused on offering RCF as a superior estimation method to alternative options, or to test the accuracy of RCF in controlled settings or using archival data. In this section, we evaluate the literature and discuss how RCF performs compared to conventional forecasting methods and hybrid forms of RCF. We also discuss some of the practical applications of RCF covered by the papers.

3.1. RCF versus conventional forecasting methods

Several studies have compared RCF with conventional forecasting methods to test whether RCF performs better. For example, Chadee et al. (2023) show that RCF outperforms traditional bottom-up approaches by producing estimates closer to actual costs. Their back-testing indicates that RCF provides a more reliable representation of project costs and improves budget adherence in public sector social housing projects.

Based on a database of 420 completed building projects in Turkey, Bayram and Al-Jibouri (2016a) compare cost estimates (based on contract sums) produced using the conventional method against cost forecasts using RCF. The study confirms that forecast accuracy can be improved by using the RCF method using various uplifts representing different levels of risk (Bayram and Al-Jibouri 2016a). Bayram and Al-Jibouri (2018) find similar results. In a separate study, Bayram and Al-Jibouri (2016b) find that RCF provided the most accurate forecasts in the early stages of the project, whereas simple linear regression and radial basis function forecasting methods provided more accurate forecasts if estimates were based on reference point later in the project life cycle (in the post-design stage).

Batselier and Vanhoucke (2016) compared the accuracy of RCF as a project cost estimation technique to earned value management (EVM) and Monte Carlo simulation, using data from one construction project. They created four reference classes for comparison, ranging from broadest (construction projects) to most specific (office finishing works). They found that RCF outperforms EVM and Monte Carlo Simulation on all three criteria of accuracy, stability, and timeliness (Batselier and Vanhoucke 2016). Similarly, for transport infrastructure projects in the UK and US, Park (2021b) found that RCF provided more accurate forecasts compared to Monte Carlo simulation.

While the previous studies have found supporting evidence of improved performance of RCF, Fridgeirsson (2016) did not find a clear reason to adopt RCF as the current method seemed to work well enough for transport infrastructure projects in Iceland. This could be explained by the low forecasting inaccuracy in the first place.

For project cost estimation, regression analysis and Monte-Carlo simulation are some of the primary statistical modelling techniques, along with case-based reasoning, artificial neural network, fuzzy logic and support vector machine (Atapattu, Domingo, and Sutrisna 2023). However, given our stated intent, this paper will not consider these in depth but focuses on RCF in comparison to other techniques.

3.2. Hybrid forecasting methods

Building on the foundational strengths of RCF, several studies have explored hybrid approaches that integrate RCF with other forecasting methods, aiming to address some of its limitations. For example, Liu, Wehbe, and Sisovic (2010) considered a hybrid of RCF and the conventional fixed contingency approach and compared this to the conventional approach and the risk-based estimating (RBE) approach. They found the hybrid approach more accurate compared to the conventional approach, but the RBE can be more accurate than the hybrid approach. However, it should be noted that the RBE forecast was based on a small sample.

Several studies demonstrate the effectiveness of integrating RCF with Bayesian methods to improve forecast accuracy. Bordley (2014) uses a Bayesian approach to combine RCF with model-based forecasting, where the reference class informs the prior distribution. This hybrid approach resulted in a lower forecast error (and lower variance) than either the model-based forecast or the reference class forecast alone. Zangeneh and McCabe (2022) employ a Bayesian Network framework to incorporate expert-informed socio-technical risk factors alongside reference class data. By treating the reference class as a prior and updating it with project-specific evidence, their approach supports more differentiated and context-sensitive forecasts, particularly well-suited to the complexities and uncertainties inherent in industrial megaprojects. Similarly, Kim and Reinschmidt (2011) propose a Bayesian adaptive forecasting method that combines inside-view estimates, outside-view reference class data, and real-time performance information from EVM. Their model uses Bayesian model averaging to revise cost forecasts dynamically during project execution, thereby enhancing the accuracy of predictions as new information becomes available.

In line with recent efforts to formalize RCF through statistical techniques, Lordan-Perret et al. (2023) propose a hybrid method that, like Bayesian approaches, models uncertainty using empirical data. Rather than relying on prior distributions, they apply adaptive kernel density estimation to construct a smoothed empirical distribution of cost overruns and use Wilks' formula to quantify confidence in threshold estimates. El Yamami et al. (2018) propose a hybrid model for IT project financial risk prediction that combines analytical cost estimation with an analogous approach, using k-nearest neighbour (k-NN) to identify comparable past projects. These projects form the reference class, from which probability distributions of financial risk are derived to adjust estimates and improve predictive accuracy. Natarajan (2022) integrates RCF with Gradient Boosted Regression Trees (GBRT) to improve cost and schedule forecasting for offshore oil and gas megaprojects. By applying GBRT, Natarajan's hybrid approach captures both the distributional insight of RCF and the adaptive learning capacity of machine learning. This results in more accurate and context-sensitive forecasts, particularly in complex, high-uncertainty environments.

Recent studies have also developed new forecasting models. For example, Salling and colleagues embed RCF within increasingly complex decision-support models. In the CBA-DK framework (Salling and Banister 2010), RCF is introduced through optimism bias uplifts to supplement cost–benefit analysis. This then evolved into a new forecasting technique Reference Scenario Forecasting (RSF), combining RCF and quantitative risk analysis with scenario forecasting (Salling and Leleur 2012, 2017). The outcome of adopting this technique entails graphs based on scenarios, illustrating the likelihood of attaining specific Benefit-Cost (BC) ratios. The benefit of this approach lies in its capacity to consider the probability of implementing a non-feasible project or overlooking a viable one. The UNITE-DSS model (Salling and Leleur 2015) further integrates this approach by linking RCF to a large empirical database, while the SUSTAIN-DSS model (Salling and Pryn 2015) embeds sustainability criteria through multi-criteria decision analysis. Nevertheless, none of these studies demonstrate whether this integrated method outperforms RCF on its own.

Some hybrid models have been proposed that combine risk analysis with RCF. For example, Allahaim, Liu, and Kong (2016) develop a risk-based cost contingency estimation model (RBCCEM) based on a survey of project managers in Saudi Arabia. They found this method to provide better forecasts compared to RCF or the conventional forecasting model. Similarly, concerned with enhancing risk analysis, Zarikas and Kitsos (2015) extend the RCF methodology by incorporating statistical tolerance regions, offering a more rigorous framework for quantifying uncertainty within the reference class.

Another new forecasting method, similarity-based forecasting (SBF), was proposed by Lovallo, Clarke, and Camerer (2012). SBF is a combination of two known methods for implementing an outside view, that is, RCF and case-based decision theory (CBDT). SBF uses the reference classes from RCF, and the method of assigning more weight to the most similar cases in the reference class from CBDT. Thus, SBF is similar but expanding RCF by using similarity weighting instead of equal weighting. Like SBF, weighted reference class forecasting (WRCF), as proposed by Zani and Adey (2025) and Zani, Adey, and Carroll (2024) uses weighting mechanisms to account for the relevance or similarity of past projects. While SBF derives the weights from expert judgement, WRCF uses empirical distribution as weight basis. This method uses work package-level cost deviations when reference data are unavailable. A first empirical study into the effectiveness of WRCF found that it results in higher certainty and more conservative estimates. However, the results should be treated with caution as they have not yet been widely tested.

Additional research that advocates the use of expert opinion in combination with RCF is by Leleur et al. (2015) who used this forecasting method in cost estimation efforts aimed at modelling three scenarios for an airport construction project. They also found it necessary to apply overconfidence theory to the expert judgments to modify the statistical distribution of the projected possible outcomes. Their findings suggested that under certain circumstances a combination of RCF and expert opinion will offer more accurate cost estimations but caution against over-aggressive use of a reference class without first vetting the data for similarity. Garbuio and Gheno (2023) also aim to improve the construction of the initial reference class by combining RCF with a clustering-based method. Recognizing that the accuracy of RCF depends heavily on the relevance and granularity of the reference class, they develop a novel biclustering algorithm that allows project groupings based on multiple, interrelated variables, rather than one single dimension. This method enables a more nuanced segmentation of historical project data, thereby improving the precision of probabilistic forecasts. The resulting hybrid method, termed ECrfBimax, combines the statistical rigour of RCF with pattern recognition capabilities of biclustering to produce more context-sensitive estimates. Similarly, Zarghami (2023) aims to improve the reliability of RCF by proposing a quantitative similarity measure that determines how closely past projects align with a new project in terms of cost, duration, and scope. This measure uses the radius of gyration, a statistical tool that shows how far the reference projects differ from the project at hand. By using this method to exclude outliers and select more comparable reference classes, the risk of biased selection is reduced, and the accuracy of RCF-based forecasts is enhanced.

Lastly, based on an empirical study of 52 projects, Servranckx, Vanhoucke, and Aouam (2021) argued that the accuracy of RCF was successfully demonstrated by considering both the intra- and inter-accuracy of reference classes. The accuracy of RCF increases when more project properties are considered; however, not surprisingly, the greater the number of project properties (i.e. the narrowing of the reference class), the less reliable are the results of the analysis, simply due to the fewer number of projects with similar properties. The conundrum is that a broad reference class – for example, "road" projects – will yield a large data sample, but at the same time, a highly variable one. To accurately identify the most appropriate reference class in their research, Servranckx, Vanhoucke, and Aouam (2021) noted, is to lose data points, with the clear trade-off being a sacrifice in reliability.

3.3. Practical applications of RCF in specific project classes

Rather than testing the utility of RCF simply as an estimation method, several other studies have applied RCF as a means to address the challenges with building various classes of megaproject (Flyvbjerg, Hon, and Fok 2016; Steininger, Groth, and Weber 2020). For example, several studies have applied RCF to projects in the energy industry. Sovacool, Nugent, and Gilbert (2014) studied construction cost overruns for electricity infrastructure projects and used the RCF technique as a diagnostic tool to illustrate that construction cost overruns in electricity infrastructures are unavoidable, given the plethora of risks in planning. Awojobi and Jenkins's (2016) study of the construction of the Bujagali hydroelectric dam in Uganda involved massive uncertainty and risk at the front-end, for which they applied RCF. The application of RCF resulted in a more reliable cost estimate, and it increased the accuracy of the cost-benefit analysis. Their findings suggested that RCF can offer a reliable estimator of the final cost of dam construction, provided a reasonable fit into a suitable reference class can be created. Where the alternatives (the reference class employed) are not sufficiently similar, RCF will not materially aid in more accurate cost estimation. Ansar et al. (2014) reaches a similar conclusion for hydropower mega-dams, recommending RCF-based de-biasing in planning. Brown, Lux, and Cowan (2024) use RCF as part of a best practice approach to costing fusion power plants. They emphasize that RCF is a powerful tool to help project teams make more informed decisions about budget, timeline, and risk management by leveraging historical data from similar projects.

Recently, a study of 57 World Bank-financed hydropower projects by Jenkins, Olasehinde-Williams, and Baurzhan (2022) also found that when RCF was used to develop cost estimates at the outset of these projects, it helped to reduce net losses by preventing some hydropower projects with negative economic net present values from being undertaken. However, because of the increased ex-ante rejection of projects, the loss of potentially economically positive projects from the portfolio of hydro dam projects is greatly increased. Jenkins, Olasehinde-Williams, and Baurzhan (2022) concluded that the more narrowly defined a project's cost-benefit analysis, the greater the likelihood of missing the opportunity to create net long-run societal benefits. Also concerned with the economic impact, Callegari, Szklo, and Schaeffer (2018) conclude that probabilistic approaches such as RCF can provide more realistic estimates and improve risk management in project planning. Importantly, the authors highlight that megaprojects often fail to deliver anticipated economies of scale, as their disproportionate exposure to risk undermines financial efficiencies. Separately, and independently, RCF has been proven to be an effective method for improving the accuracy of cost forecasts for both Nuclear Power decommissioning projects (Lordan-Perret et al. 2023) and wind farms (Kaiser and Snyder 2012; Koch and Søndergaard 2010).

The transport infrastructure industry has also seen notable applications of RCF, particularly in the work of Flyvbjerg, which demonstrates how the use of uplift curves and risk thresholds can improve the reliability of cost forecasts (Flyvbjerg 2008; Flyvbjerg, Hon, and Fok 2016). Other studies confirmed the effectiveness of RCF in transport infrastructure projects. For instance, Steininger, Groth, and Weber (2020) investigated the railway project Stuttgart 21, and they found significantly improved forecast accuracy when using RCF.

Similarly, Park (2021b) concluded that RCF had delivered substantial benefits in curbing cost overruns in UK transport infrastructure, particularly in highways. Liu et al. (2018) applied RCF in the Chinese context. Using data from 30 large-scale Chinese long bridges, they construct a reference class and generate uplift values at different probability levels. When these are applied to the Hangzhou Bay Bridge, the adjusted estimates align more closely with actual final costs than the original estimates. Some studies suggested that RCF is more successful with engineering projects than other types (Prater, Kirytopoulos, and Ma 2017), presumably because the databases for engineering and construction are larger and the data have been easier to collect.

In the context of chemical projects, RCF was effective for large homogenous projects but not for various smaller projects (Walczak and Majchrzak 2018). In housing projects, RCF was effective, and in particular, Chadee et al. (2023) and Chadee, Hernandez, and Martin (2021) showed that the P50 estimate provides a closer estimate to the actual costs of the project compared to the contracted costs. RCF also resulted in lower errors compared to conventional forecasting techniques in the oil and gas industry (Natarajan 2022; Nesvold and Bratvold 2022).

Table 2 offers a summary of a sample of studies investigating various aspects of RCF, several of which were summarized above. While not intended to be exhaustive, the summary offers a sense of the variety of settings and project classes that have been investigated, with a brief description of key findings.

So, based on the review of the literature on RCF, what conclusions can reasonably be drawn? As the studies identified demonstrate, the efficacy of RCF seems to reflect a generally positive response, though with some important caveats and provisos. Thus, for our analysis, it seems reasonable to recognize both strengths and weaknesses of this technique, given the mixed results from empirical studies. In the following sections, we will offer both advantages and cautions that policy makers and scholars alike should consider when applying RCF for project cost estimation.

4. Strengths of RCF

4.1. RCF can improve estimates for large projects that face high levels of risk

The cost estimates produced at different stages of the project carry different levels of uncertainties/risks and thus different estimation accuracies. As a project progresses through its life cycle, more information about the project becomes available, reducing the levels of uncertainty, and increasing the accuracy of the estimate. As the roots of underestimation lie in the front-end stages of project development, RCF is recommended to be applied in the early stages of projects where risk is highest (Ansar et al. 2014; Awojobi and Jenkins 2016; Bayram and Al-Jibouri 2016a). Moreover, outside-view approaches are more commonly used during the early phases because the lack of detailed information makes it difficult to implement the bottom-up approach (Oberlender and Trost 2001).

Batselier and Vanhoucke (2016) found that RCF clearly outperforms EVM on timeliness, that is, the ability of a forecasting method to produce accurate forecasts in different stages of the project life cycle. This is important because it allows adequate corrective actions to be taken in a timely manner. Similarly, Bayram and Al-Jibouri (2016a) showed, with varied levels of efficacy, that RCF produced more accurate and realistic forecasts than other methods, however, RCF particularly produced more accurate forecasts in the pre-design stage, while simple linear regression analysis (SLRA) and radial basis function (RBF) were favoured in the post-design stage. Moreover, Flyvbjerg (2006) emphasizes that "the comparative advantage of the outside view is most pronounced for nonroutine projects, understood as projects that managers and decision makers in a certain locale or organization have never attempted before" (2006, 9). These projects are the ones in which biases towards optimism and strategic misrepresentation are likely to be largest (Themsen 2019).

Although RCF is a robust approach to addressing optimism bias, its effectiveness in practice can be limited by implementation challenges. For example, its application is less feasible for small companies dealing with a limited number of project types (Walczak and Majchrzak 2018), routine projects using well-known technologies (Jenkins, Olasehinde-Williams, and Baurzhan 2022) or situations where cost overruns are the result of late-stage further scope changes, technical and managerial difficulties, or external material price changes (Love, Ika, Sing, et al. 2019b). These cases do not represent shortcomings of the method itself but rather illustrate its sensitivity to reference class selection and data availability. Because project environments are dynamic, unforeseen developments may compromise the comparability of the original reference class, thereby reducing the relevance of earlier forecasts and uplifts. These practical constraints highlight that RCF's reliability depends on the availability of sufficiently stable and representative data throughout the project lifecycle.

4.2. RCF can provide a better-informed basis for the contingency level compared to conventional inside-view techniques

Due to the higher uncertainty in the front-end stages, a higher risk contingency percentage (Koch 2012; Oberlender and Trost 2001) is often applied at the earlier stages than in the later stages (Liu et al. 2018). There are various techniques that can be used to determine the risk contingency level, including Monte Carlo simulation, and sensitivity analysis, but the deterministic or conventional contingency approach is the most used method (Allahaim, Liu, and Kong 2016; Liu, Wehbe, and Sisovic 2010). The conventional contingency approach adds a fixed percentage of the overall project budget, based on the type of project and the life cycle stage, but it does not consider component risks. Moreover, the contingency percentage is often based on estimator's experience and hence it is argued to be particularly prone to optimism bias and strategic misrepresentation (Allahaim, Liu, and Kong

Table 2. Summary of key articles.

Source	Methods employed	Key findings
Allahaim, Liu, and Kong (2016)	A survey of infrastructure PM in Saudi Arabia, Development of an RBC contingency estimation model (RBCCEM).	The mean adjusted cost overrun percentage (as well as the variance) using the RBCCEM approach is lower compared to that of the RCF P50, RCF P90, and the hybrid.
Awojobi and Jenkins (2016)	Single-case study, the Bujagali dam in Uganda.	RCF resulted in a more reliable cost estimate and increased the accuracy of the cost-benefit analysis. While the use of RCF would lower overrun likelihood, there are two practical challenges: (1) finding a suitable reference class to generate meaningful projections, and (2) the a priori requirement for this extra contingency could also limit funding availability for the project; in effect, scare off potential funding agencies.
Batselier and Vanhoucke (2016)	Single-case archival data.	Compared to Monte Carlo simulation and Earned Value forecasting, RCF provides more accurate estimates to completion for sample construction projects. RCF only outperforms the other techniques when the degree of similarity between the considered project and the projects in the reference class is sufficiently high.
Batselier and Vanhoucke (2017)	Archival analysis and modelling of 23 completed construction projects.	RCF outperforms traditional methods of forecasting EVM and Monte Carlo Simulation on all three accuracy criteria of accuracy, stability, and timeliness.
Bayram and Al-Jibouri (2016a)	Single-case quantitative study using archival data from 420 Turkish public building projects to apply and validate RCF.	The study revealed improved cost forecast accuracy when using RCF.
Bayram and Al-Jibouri (2016b)	Single-case quantitative study using archival data from 420 Turkish public building projects to apply and validate RCF.	Generally, RCF produces more accurate forecasts of the project's final cost based on the unit area method (carried out in the predesign stage), while simple linear regression analysis (SLRA) and radial basis function (RBF) produce better forecasts based on detailed estimates (carried out in the post-design stage).
Bayram and Al-Jibouri (2018)	Archival analysis of 369 public building projects in Turkey with comparative application of RCF.	Estimates for both traditional methods Unit Area Cost (UAC) and Unit Price Analysis (UPA) are inaccurate. Estimates based on UPA are not better than those produced using UAC, despite more detailed project information. Applying the various optimism bias uplift values obtained from the RCF method to original total project cost estimates based on traditional methods of unit area costs (UAC), the accuracy of the forecast of the final project is improved considerably.
Bordley (2014)	Bayesian modelling study integrating RCF and statistical simulation. Modelling is based on simulated decision scenarios rather than observational data from real-world projects.	The resulting Bayesian posterior forecast (combining RCF and the model-based forecast) had lower variance (and lower forecast error) than either the model-based forecast or the reference-class forecast.
Flyvbjerg , Hon, and Fok (2016)	Analysis of 25 roadwork projects in Hong Kong.	The analysis established and verified the statistical distribution of the forecast accuracy against a benchmark of 863 similar projects, supporting RCF.
Fridgeirsson (2016)	Archival analysis of 110 Icelandic transport infrastructure projects using RCF with risk-based uplift modelling.	There is no urgent need for the Icelandic Road Administration to adopt RCF as its current methodology, which is based on time series data, seems to work well enough.
Garbuio and Gheno (2023)	Conceptual and experimental simulation study using synthetic project data to evaluate reference class selection strategies.	Combines RCF with a new biclustering algorithm, ECrfBimax. This approach helps design value propositions for IoT devices through a data-driven selection of initial classes.
Jenkins, Olasehinde-Williams, and Baurzhan (2022)	Studied 57 World Bank-financed hydropower projects constructed between 1975 and 2015.	The use of RCF reduced net losses by preventing some bad projects from being executed. However, RCF also caused net benefits to be forfeited by rejecting some projects with positive net present value. Finally, RCF did not improve the economic performance of hydropower projects financed by the World Bank.
Kaiser and Snyder (2012)	Archival analysis using secondary data on US Offshore wind capital costs.	Compares the conventional (bottom-up engineering) approach to the RCF (top-down) approach. The lower SD in capital costs within the reference class compared to the full sample suggests RCF is a more accurate estimator for costs.
Koch (2012)	Desk research, Case study of London Array.	Forecasting can be further improved by combining RCF with inside approaches, appreciating the socio-technical content, "taking into account, for instance, similarity in terms of technology, geography, nation state and operational, economic and regulatory conditions" (Koch 2012, 618). Using this socio-technical perspective may result in smaller, narrower reference classes.
Koch and Søndergaard (2010)	Desk research, Case study of London Array.	There is a much higher probability of delivering the project on time and budget if the budget is increased and the schedule is extended. RCF and a life cycle scope are needed to improve future investments in offshore wind energy
Leleur et al. (2015)	Single-case modelling of airport construction.	Applied RCF in tandem with expert judgments to focus on formulating the best possible reference pool of projects. Also applied overconfidence theory to interpret expert judgments about costs and demand as relating to a specific project up for examination. The greatest challenge lies in finding a suitable, representative database for the reference class.

(continued)

Table 2. Continued.

Source	Methods employed	Key findings
Liu, Wehbe, and Sisovic (2010)	Single case study (of the Australian Transport Authority) with multiple road projects as embedded units of analysis. The study uses secondary archival data.	The hybrid approach performs better compared to the conventional contingency approach. The hybrid approach has lower accuracy (mean estimation error and variance) compared to the RBE approach. The large variation in the project type that is used in the different samples and the small sample size for the RBE sample could have influenced the results.
Liu et al. (2018)	Analysis of a planned bridge project (Hangzhou Bay Bridge); 30 interviews.	RCF was used to suggest that the cost overrun probability distribution fell within the guidelines of previous similar projects. Traditional cost/benefit metrics are inadequate for assessing true benefits from megaprojects and can skew the data accordingly.
Lordan-Perret et al. (2023)	Empirical extension of RCF applied to decommissioning cost data from nine US nuclear power plants.	The study showed mixed results. Traditional methods can outperform RCF under certain conditions, especially at moderate confidence levels, and RCF tends to perform better at higher levels of risk exposure.
Lovallo, Clarke, and Camerer (2012)	Experimental study.	The study finds that SBF, which combines elements of RCF and case-based decision-making, produces better forecasts than regression models.
Natarajan (2022)	Mixed-methods study using a questionnaire-based survey with 26 Oil & Gas Project Professionals, and archival data.	RCF is better than conventional forecasting. Machine learning models trained on RCF data can significantly outperform RCF alone in forecasting accuracy.
Park (2021b)	Archival analysis of 107 road, rail, and building projects in UK and US.	The study of post-project records suggests that RCF offers more accurate estimates (for UK projects) than Monte Carlo simulation (for US projects).
Prater, Kirytopoulos, and Ma (2017)	Systematic quantitative literature review.	Apart from engineering projects, there has been no experimental and statistically validated research into the effectiveness of RCF.
Servranckx, Vanhoucke, and Aouam (2021)	Interviews with 76 project managers and a study of 52 projects.	The accuracy of RCF increases when more project properties are considered; however, a higher number of project properties decreases the size of the reference classes and reduces the reliability of the results.
Sovacool, Nugent, and Gilbert (2014)	Archival study of 401 power plants and electrical transmission projects in 57 countries.	Hydro and nuclear power plants had the highest cost and schedule escalation of all projects studied. Causes of overruns are multi-causal: both accidental (error) and strategic (misrepresentation)
Themsen (2019)	Longitudinal single-case study.	RCF does not improve project cost estimation. Estimates are always based on the relational network of human and non-human actors who are themselves biased in their efforts to establish them.
Walczak and Majchrzak (2018)	Single-case application of RCF in a chemical industry company using archival project data.	RCF is suitable primarily for large, homogeneous project environments where distinct, uniform groups can be established; however, its applicability is limited in real business contexts that usually involve a small number of heterogeneous project types.
Zangeneh and McCabe (2022)	Mixed-method development of a Bayesian Network model for cost overruns, integrating RCF and three rounds of expert interviews and surveys.	A Hybrid method that integrates traditional RCF with Object-Oriented Bayesian Network (OOBN)) demonstrated an improvement over using traditional RCF means alone.
Zani, Adey, and Carroll (2024)	Quantitative method development and validation using a simulated application of adjusted RCF weights on empirical reference data from previous studies.	The study finds that Weighted RCF provides more accurate cost deviation forecasts than the traditional RCF method.

2016; Liu, Wehbe, and Sisovic 2010). Estimators and planners attempt to foresee potential issues in scenario planning but may underestimate the probability of unfavourable events or risks in projects (Lovallo and Kahneman 2003).

In response to limitations of RCF, alternative forecasting approaches have been proposed to enhance the accuracy of cost contingency estimation. Notably, RCF and RBE (Liu, Wehbe, and Sisovic 2010) represent outside-view models designed to validate single-point estimates and inform the determination of appropriate contingency levels (Koch 2012; Oberlender and Trost 2001).

However, the availability of more accurate forecasting methods does not eliminate the need for contingency governance. As Lovallo, Cristofaro, and Flyvbjerg (2023) argue, while reference class data enhance forecasting accuracy, it must be embedded within a broader strategy that resists opportunistic use of contingencies. Similarly, Flyvbjerg, Hon, and Fok (2016) propose a multi-layered framework for contingency management, emphasizing that contingencies should not be readily accessible, but preserved for genuine risks as they materialize.

4.3. RCF is relatively easy to use compared to alternative methods

RBE has been found to outperform RCF in terms of forecasting accuracy, according to the only comparative study identified in the literature (Liu, Wehbe, and Sisovic 2010). However, the evidence remains insufficient to decisively favour RBE over RCF, particularly considering the methodological complexity and intensive data requirements RBE imposes, constraints that are especially demanding for large and complex projects (Allahaim, Liu, and Kong 2016). In contrast, RCF offers a pragmatic advantage through its ability to link contingency levels to the statistical probability of cost and schedule overruns (Awojobi and Jenkins 2016). This addresses one of the most critical limitations in traditional approaches; the lack of empirical grounding in risk probability. Moreover, our literature review has identified various other forecasting methods that are often combined with RCF providing improved forecasts, including exponential smoothing (Batselier and Vanhoucke 2017), Bayesian methods (Bordley 2014), and SBF (Lovallo, Clarke, and Camerer 2012). However, while promising, there is no concrete evidence of their efficacy.

Considering the challenges with RCF in the first place, most notably establishing a relevant reference class, combining them with other forecasting methods will exacerbate data requirements demands. Besides data collection, the RCF method is relatively easy to use, with a limited number of parameters needed to perform the method (Bayram and Al-Jibouri 2016a; Chadee et al. 2023). Moreover, RCF has shown greater forecasting stability, because it uses one fixed constant pre-project forecast throughout the entire project, while other projects use updated forecasts based on actual progress data (Batselier and Vanhoucke 2016).

5. Problems with RCF

Advocates for RCF have argued the necessity of applying the "outside view" because, as Kahneman and Tversky (1979b) noted, planners are inherently optimistic – prone to minimizing problems while exaggerating benefits – so their estimates simply cannot be trusted to be accurate. Therefore, applying a multiplier benchmarked against a large database of similar projects (the reference class) is the only way to generate sensible estimates of the likely cost of these projects. Based on a review of the literature noted and summarized above, there are some patterns of potential problems with the use of RCF. We suggest that these problems are not, of themselves, fatal flaws with the theory, but rather questions for which deeper investigation is warranted. In summarizing these issues, we hope to offer some avenues for additional research on this topic.

5.1. Biases are collective, often socio-political in nature, and extremely hard to model

The arguments for bias in decision making of the sort that Kahneman and Tversky modelled in laboratory studies formed the basis for their promotion of the outside view and statistical modelling to ostensibly account for and correct these implicit flaws. Flyvbjerg's analyses of the planning fallacy and strategic misrepresentation are designed to support these contentions by applying RCF as a post hoc smoothing, or budget uplift that more accurately reflects the "long-tailed" statistical probabilities of cost overruns for entire classes of project. However, even within the studies we examined, clear evidence for the existence of optimism bias, particularly as a precursor to cost underestimation and benefit overestimation, was mixed. While the majority of studies found it a plausible explanation, based on evidence from their research, others either found no evidence or, in fact, determined that the opposite was the case; that is, evidence of pessimism bias. Moreover, there is some suggestion that planning fallacy explanations are deeply rooted in our preconceptions, as in the case of Batselier and Vanhoucke (2016), who, while failing to find that optimism bias was operating in their cost estimation process, simply ascribed the alternative cause as strategic misrepresentation. In other words, since the one cause was not found, their assumption was that the under-performance must be due to the other – again, without clear evidence of any causal link.

Estimates are further complicated by the existence of collective estimation. Original research on optimism bias was based on individual results in laboratory experiments. As Love et al. (2022) have noted, for large projects, it is far more common for teams of estimators to work together to generate cost estimates, suggesting that optimism bias may be less common than we assume, especially in megaprojects or those that employ multiple people to develop estimates. Employing an argument based on actor-network theory, Themsen (2019, 338), suggested that cost estimation is "a continually generated effect of the web of relations that brings the practices about". In his research, he undertook a longitudinal case-study of a multi-billion Kroner Danish public megaproject and found that, contrary to other studies, the use of RCF did not improve the accuracy of cost estimates. Themsen (2019, 337) theorized that this problem was because, despite our best assumptions, 'estimates are always a relational network effect of human and nonhuman actors' "biased" efforts to establish them' and concluded that it is naïve to expect a debiasing technique to yield meaningful outcomes when the estimation process itself is embedded in politicized and subjective practices. In short, the psychology, motivations, and interactions (intertwining) of multiple stakeholders who can affect final project bids is extremely complex and resists the over-simplification of cost estimation to exclusively planning fallacy causes, of which optimism bias is perhaps the best known.

5.2. It is hard to determine what is a reasonable and attainable uplift (what project sponsors can accept) and what quantifiable role risk mitigation plays

The compromise between "likely" overruns, according to Flyvbjerg, and what a funding agency may be willing to provide is a critical issue. Flyvbjerg's data across multiple project classes supports the need for quite large uplifts. For example, Flyvbjerg (2008) notes that for rail projects, past data indicate that to achieve a 50/50 risk of cost overrun, that is, a 50% likelihood that actual costs will not exceed the estimate, the initial cost estimate would have to be adjusted upwards by 40%. To lower the likelihood of an overrun to 10%, clients would have to accept a budget uplift of 68%. For public projects, at least, the practical question to be considered is the potential for politicians or public officials convincing the holders of the purse strings to change, after the fact, a project's budget from, say, $1 billion USD to $1.68 billion *after the estimate was ostensibly completed*. In short, it is hard to credit most organizations and funding units with accepting such a massive post hoc readjustment, suggesting that some form of widespread stakeholder negotiations is needed. So, in the compromise between what the confidence interval data show and the art of the possible, would a modest uplift of 20–30% just be the worst of both worlds – not enough to be meaningful but a potential trigger for injecting too much slack into the project (see Pinto 2023)?

Flyvbjerg and Gardner (2023) note that using data from the extreme end of the distribution (what they refer to as a "fat-tailed distribution") is a practical impossibility, as it is prohibitively expensive to simply add a 200% or 300% uplift to a cost estimate, and instead suggests finding risk mitigation strategies that are designed to "cut off the fat tail" of the distribution. By this, they are advocating for focusing on key questions – where are the likely most expensive causes of delays or overruns coming from and what can we do to mitigate this risk in advance? This is reasonable advice in theory but is heavily compromised by the challenge of determining how much (and what sort) of risk mitigation is required and then monetizing it (assigning a value) that allows for a concomitant shrinkage of the initial inflator figure. Salling and Banister (2010, 118) refer to a "scaling exercise", while Love et al. (2024, 209) describe RCF as a "contingency on a contingency" or an inflator device; it inflates costs by applying uplifts on top of existing contingencies. In other words, if HM Treasury's Green Book (2022) suggests that risk management can lower the required percentage multiplier for cost estimates, it begs the follow-on questions: how much risk mitigation is necessary, what types are preferred (risk management and mitigation processes come in all shapes and sizes), and how much will these actions then accordingly adjust the uplift multiplier? That is, risk mitigation to offset cost multiplier effects leads to some important questions:

- Where do you stop; in other words, if a project team knows that they need to mitigate likely risks, how many do they address? Using Pareto's Law, for example, can we determine which 20% of project risk mitigation actions will lead to an 80% decrease in uplift?
- Statistically, what does this mitigation actually look like? Can we state with confidence that our 20% uplift is now accurate (within a confidence interval of, say, 80%) because we have successfully mitigated "enough" risks? How do we know?

In short, Flyvbjerg and Gardner's guidance makes sense until we follow the logic to its eventual destination. It still does not offer statistical justification for an uplift value that includes risk mitigation. As was recently noted, "calculating anything like a cost/benefit trade-off for adding both budget uplift and the cost of mitigation gets very murky, very quickly" (Pinto 2023, 6).

5.3. Identifying the actual reference class is extremely complex and the accuracy of RCF depends on doing it correctly

What *is* the reference class? One of the main challenges with the RCF method is finding a suitable representative database that is large enough for statistical analysis (Awojobi and Jenkins 2016; Baerenbold 2023; Chadee et al. 2023; Leleur et al. 2015; Liu, Wehbe, and Sisovic 2010; Liu and Napier 2010; Love et al. 2024). A large sample of similar projects with accurate information needs to be collected. Developing such a database takes effort and time (Garbuio and Gheno 2023; Liu and Napier 2010; Lovallo, Clarke, and Camerer 2012). If the reference class is too broadly constituted, it will not allow for accurate comparisons, while too narrowly selected and the resulting class would be too small to permit statistical smoothing (Leleur et al. 2015). Batselier and Vanhoucke (2016) found that RCF only outperforms EVM and Monte Carlo simulation when the degree of similarity between the project and the projects from the reference class is sufficiently high.

Liu, Wehbe, and Sisovic (2010) further point to two other problems that can make it hard to develop a representative reference class. First, for project types that are relatively rare (e.g. nuclear power plants), achieving a sample size sufficient for statistical analysis may be unattainable. Brown, Lux, and Cowan (2024) argue that for fusion power plants, a reference class that contains non-fusion projects with similar novelty and complexity as comparators must be used, given that no fusion power plants have previously been built. Second, private companies may be reluctant to disclose sensitive information to competitors and government agencies. Besides, Lovallo, Clarke, and Camerer (2012) also argue that constructing the reference class requires careful and structured consideration to avoid bias in the sample (e.g. selecting too few cases or cases that are too favorable). In practice, this would mean first identifying a large pool of projects, after which based on specified criteria the reference class is selected. Zarghami (2023) stresses that without a systematic way to assess similarity, reference classes may be biased, undermining the accuracy of RCF. To mitigate this risk, they propose a quantitative similarity measure to guide the selection of more appropriate reference projects.

This problem leads to the challenge of creating subclasses that more accurately reflect the nature of historical overruns. Thus, the data may cloak real differences within these referent classes, rendering any post hoc smoothing inaccurate at best, or even misleading for cost estimation. An argument can be made that RCF and post hoc budget uplifts are only as useful as they reflect very specific sets of projects within the classes (Awojobi and Jenkins 2016; Leleur et al. 2015). Broad categorization simply will not work if, say, the actual range (+/− 2 SDs) of IT project overruns is so broad that when an average value is employed, it badly misses the historical data for a specific set of projects within the larger class. As Servranckx, Vanhoucke, and Aouam (2021, 1161) noted: "An important aspect of RCF is the construction of the different reference classes. The reference classes cannot be too large as this might increase the diversity within the reference classes, but they cannot be too small either because this reduces the reliability of the results".

Shaping future discussion of appropriate reference classes is the recent argument of Winch (2025a). Reflecting on the challenge of applying classification schemes to megaprojects and referring to the use of decision support tools such as RCF, Winch (2025a, 7) observes:

> [T]he UK's Building Cost Information Service (https://bcis.co.uk/) has been providing such a service for six decades, but its applicability with any rigour depends on relatively low levels of uniqueness of the project being evaluated. This suggests that we need a measure of uniqueness … (Shenhar and Dvir 2007; Winch, Maytorena-Sanchez and Sergeeva 2022) to give a pretty good sense of how unique the project being evaluated actually is. … Megaprojects are indeed relatively unique, even if their promotors sometimes make them more unique than they need to be (Winch 2025b), which is arguably another form of uniqueness bias.

The critical point argued in the previous passage suggests the need to consider the potential advantages of reframing our original ideas of what constitutes a recognizable and distinct "class" of project. Is it truly the industrial classification, such as rail infrastructure, hydroelectric dams, or Olympic Games, or is the alternative we should consider predicated on some combination of project type as well as the more unique properties of structural complexity, which would require an alternate set of variables in establishing internal reliability and divergent validity of various project classes? At a minimum, future research should consider reformulating the way we arrive at our initial classification schemes to determine their predictive utility.

5.4. It is important not to over-promise the advantages offered by RCF

In arguing for the use of RCF as a means to employ the outside view, remove embedded bias, and account for planning fallacy, it has been suggested that Flyvbjerg oversteps by positing corrective benefits that either may not be clearly demonstrated or for which contradictory evidence can be found (Love et al., 2023). For example, one of the proposed advantages of RCF is the contention that through collecting a large data sample of past projects, it is possible to account for uncertainty (which is assumed to have been found within the past project histories) (Flyvbjerg and Gardner 2023). However, this argument has been refuted as conflating risk and uncertainty. That is, project risk involves the identification of known risk factors for which statistical probability can be accounted. While it may be statistically possible to quantify risks based on a sufficiently large data set, uncertainty ("unknown-unknowns") cannot be modelled a priori, suggesting that post-hoc smoothing cannot address the effect of uncertainty on projects (Love et al. 2024; Pinto 2023). However, several studies (e.g. Asadabadi and Zwikael 2024; Love et al. 2023, 2024) have highlighted that RCF can be open to manipulation, as project sponsors may intentionally set lower initial estimates knowing that later uplifts will increase the budget. Thus, it is necessary to consider that even post-hoc data smoothing techniques such as RCF need to be carefully reviewed, as those applying seemingly "objective" devices for adjusting project estimates are themselves often subject to the vagaries of supporting a pre-established position, either in favour or opposing a proposed project investment.

Both the benefits and challenges of RCF noted above lead us invariably to positing some follow-on research questions. As noted, the goal here is not to suggest that these issues are insurmountable, but that they offer some useful avenues for additional research on RCF. As our literature review demonstrated, despite its wide-spread acceptance and use, particularly in the UK and Eurozone, studies of RCF continue to point to some inconsistencies and competing results in various settings, suggesting that the "jury" may still be out regarding the efficacy of the method. In the following section we will offer some thoughts on directions for additional research.

6. Propositions and directions for future research

RCF is increasingly being used as a cost estimation correction technique. Given the list of strengths and weaknesses that we have identified, it is necessary to consider some next steps in how we can understand and better employ the methodology; first, by offering some directions for future research on RCF. While not an exhaustive list, these arguments allow us to frame its utility and areas where caution may be warranted.

Proposition 1: RCF is a corrective device that is subject to the way it is employed by its users, influenced by myriad behavioural motivations and forces.

As some critics of RCF are quick to note, there is a philosophical disconnect between the argument criticizing standard deterministic cost estimation techniques as being influenced by planning fallacy "noise" and human bias and their subsequent correction through other deterministic methods (Pinto 2023). That is, if we think of the event chain for the use of RCF, it is predicated on a series of steps that operate as follows: (1) recognize that attempts to use some estimation technique to create initial cost number (deterministic methods, usually) are subject to a variety of human biases (Flyvbjerg 2021) that will artificially lower our initial estimates, and (2) revert to correcting these planning fallacy behaviours through alternative probabilistic methods (RCF) to account for the planning fallacy behaviours that arose at the outset. The challenge, of course, is minimizing the impact of human behaviour (which led to the initial estimate flaws) on our subsequent corrective actions. In effect, RCF suggests using a calculated methodology to correct the problems caused by people when originally employing … a calculated methodology! A key limitation is that many initial estimates remain deterministic, expressed as single-point figures that fail to capture uncertainty. Moreover, because human judgement is involved both in producing the initial estimate and in applying the corrective RCF procedure, bias may continue to shape outcomes even when probabilistic adjustments are introduced.

It is important to reflect that planners, employing data to advocate a position (e.g. supporting a new project), are themselves subject to a variety of external pressures to gather or promote analysis that supports a position for which they are either personally invested or are being encouraged to support (Wachs 1989). In short, as Wachs (1989) noted, planning is not just analytical; planners are constantly trapped between two competing models of their role: the scientist and the advocate. Or, as he suggests: "We work in the fishbowl of politics and public-policy making. Our agencies, employers, and clients favour particular policies or programs for reasons that may be derived more directly from ideology, political commitments, or economic self-interest than from the results of analytical studies" (Wachs 1989, 476). Researching the potential for this behaviour becomes a key step in verifying the efficacy of RCF. Are the same calculations that led to the supposed initial downplaying of costs resolved by adding this uplift, despite of more behavioural theory (e.g. Goldratt 2017) that suggests that once people know their initial estimates are going to be subjected to an inflator effect, they will factor this expected adjustment into initial estimates to create the number they were originally aiming for? Put another way, Goldratt (2017) and Graham (1989) would argue that human ingenuity in the face of statistical corrections impels them to reconfigure the initial data to generate the approximate cost (for purposes of winning the bid) that was originally sought (an argument also advanced by Wachs (1989)). While it could be validly argued that this effect would only happen once and then that project would become part of the reference class for future calculations, we can see that a steady, repetitive cycle of "adjusted" estimates would do little to improve estimation accuracy through RCF, if the initial data is first modified to account for expected budget uplifts. We note also the incentive problem that the sales manager bidding for work will likely have a different view than the project manager tasked with delivering it, further complicating the search for meaningful numbers.

It is possible that ethical issues may also arise as a consequence. This points to the need for better incentive alignment. RCF is more likely to be effective when combined with accountability mechanisms that reward accurate forecasts and discourage strategic misrepresentation (Flyvbjerg 2008). Forcing unrealistic (but notionally 'correct') goals on teams may lead to excessive stress and a corresponding drop in performance – the exact opposite of the intention. Future research needs to carefully weigh the manner in which initial cost estimates are created and the role that expected RCF adjustments might play in their creation. An adjacent and separate, but valuable, line of enquiry might investigate how pre-agreed "quit/kill criteria" (Duke 2022) might be added such that projects that deviate negatively from expectations may be terminated swiftly, freeing up resources for alternative work.

Finally, it is necessary to better understand the nature of the relationship between RCF and risk mitigation methods. For example, it has been suggested that it is advisable to test the sensitivities of a BC analysis by considering a forecast cost from the tail. On the other hand, Flyvbjerg and Gardner (2023) argue that the first step to best use of risk mitigation techniques is to "cut off the fat tail" of the distribution to better focus on most likely disruptions. Thus, while RCF and risk management share some common characteristics, future research needs to better understand the implicit relationship between risk and RCF as a corrective device.

Proposition 2: The key to successful use of the technique appears to be an "enhanced RCF method".

Based on our literature review, it appears that while the use of RCF holds promise, it demonstrates more accurate results when combined with other cost estimation methods, as a hybrid form. Recent methodological innovations in RCF, including SBF, RSF, and WRCF, propose enhancements aimed at improving forecast accuracy through a better reference class project selection process. While these approaches, such as weighting similar cases more heavily (Lovallo, Clarke, and Camerer 2012), integrating scenario analysis (Salling and Leleur 2017) or applying empirical weightings at the work-package level (Zani and Adey 2025) are promising from a theoretical perspective, their empirical validation remains limited. Studies reporting improved performance of RCF over other methods are typically based on a small number of projects or a narrowly defined scope, making it difficult to determine whether these hybrid or enhanced models consistently outperform traditional RCF models.

To advance the empirical foundation of hybrid RCF, future research should undertake two distinct but complementary research directions. First, a meta-analysis of (hybrid) RCF performance using data from published studies should be conducted. While heterogeneity in datasets, project types, and evaluation methods poses challenges, a systematic synthesis of reported (or calculated) performance metrics, such as Mean Absolute Error or Mean Absolute Percentage Error can provide valuable insights into the comparative performance of different RCF variants.

Second, research should conduct structured benchmarking studies using controlled experimental designs and comparable performance metrics. These could include retrospective forecasting using real-world datasets or simulation-based experiments with synthetic project data. Such experiments allow for a systematic comparison of RCF variants under identical conditions, enabling researchers to isolate model-specific effects and identify trade-offs between accuracy, interpretability, and robustness across different forecasting methods. Due to the limited availability of real-world data, this approach is less feasible than approaches using generated data.

Hybrid RCF models often include additional components, such as Bayesian updating (Bordley 2014; Zangeneh and McCabe 2022), expert weighting (Leleur et al. 2015), or machine learning techniques (Natarajan 2022), which rely on assumptions about data availability, project structure, and risk typologies that are context specific. For example, adaptive Bayesian models that use real-time project data (Kim and Reinschmidt 2011) may not be useful in settings like IT or healthcare, where such data are harder to collect or interpret. Similarly, clustering-based hybrids, which group projects by shared characteristics (Garbuio and Gheno 2023), may not work well when project types are very diverse or poorly defined. These issues mean that hybrid RCF models may perform well in one context, such as infrastructure, but poorly in others. Future research should test how well different hybrid RCF models work across sectors with different project types and data environments. It is especially important to study how the structure of each model, such as the use of priors, data clustering, or weighting, affects its accuracy and usefulness in different settings.

Proposition 3: The utility of RCF will strongly depend on the project selection process to construct a suitable reference class.

Our findings from these studies and previous theoretical analysis show a strong link between establishing a suitable reference class (not so broad as to result in excessive within-group variance, nor so narrow as to restrict sufficient data points for creating a comparative class) and the utility of RCF as a guidance device for accurate cost adjustment. Indeed, as we have noted, a key challenge with RCF is the establishment of the comparative class (Baerenbold 2023). Future research needs to investigate the way to optimally identify and employ this outside view database. Even if accurate data on previous projects can be obtained (itself a substantial challenge), the decisions on which represent the accurate reference class is not straightforward and will in reality involve managerial judgement as to the relevance of each and appropriate inclusion/exclusion criteria. The RCF concept is far less neat in practice than in theory. Notwithstanding different actors' incentive problems, mentioned above, selecting which data to draw upon will necessitate an element of subjectivity, and this appears to be an important area to test empirically. Given historical data, how much variability is experienced if different managers are asked to determine the appropriate reference class?

This also raises the question of whether organizations can get better at using RCF to improve their estimation accuracy through learning. This has two distinct challenges. First, can experience of RCF over time enhance the ability to identify the 'correct' reference class to use? Second, historical data give the 'what', not the 'why' of previous outcomes. A better understanding of the details of past work will allow more nuanced decision-making. A large body of work on organizational learning indicates this is particularly difficult, but it is nonetheless a valuable aspiration.

Proposition 4: Additional research is needed to support the accuracy of claims that RCF helps projects complete within the newly created budget (post uplift).

We have seen that prescribed uplifts to support 50% or 90% likelihoods of staying "on cost" can inflate the initial budget substantially. This raises the spectre of situations in which political and budgetary realities force compromise – say, for example, an uplift of 20%. Because of the need to find an economically feasible compromise position, does the injection of some alternative amount simply offer the worst of both worlds: not enough to meaningfully affect the final cost performance but enough to inject additional slack into the budget, leading to the potential for scope and configuration control challenges, as more interested parties push for project enhancements? To prevent the risk that uplifted budgets lead to inefficiency or overspending, RCF should be implemented alongside tight contractual arrangements and clear accountability mechanisms to maintain cost discipline (Flyvbjerg 2008; Flyvbjerg, Glenting, and Rønnest 2004). In short, additional research needs to track not just the uplift decision, but the dynamics of the project post-uplift, to observe how the extra funds added to the budget affect development. It is disingenuous to claim that with sufficient budget uplift, a project is bound to come in closer to its "true" cost when in fact, the opposite may be the case – where the problem is being exacerbated by the introduction of extra budget slack. Does it do any good? This is an important problem, although it will likely be difficult to ascertain given the "messiness" of the real-world environments in which it would be evaluated. Purely internal projects are hard enough to evaluate, but the added challenge of working with a wider supply chain provides significant extra complexity to attempt to unpack. However, valuable behavioural insights could initially be obtained with controlled experiments such as classroom exercises.

Proposition 5: The variability of the methods being employed to study RCF is both a challenge and strength of the current state of research.

Our literature review found a widely diverse set of scholarly articles and research methods used in analysing the effects of RCF as a useful cost estimation function, including single-case studies, longitudinal assessments, maths modelling using a variety of techniques, and other empirical research methods. The advantage of these multiple approaches lies in their ability to shed light on the subject from a variety of perspectives; however, they also complicate our ability to offer defensible conclusions across the breadth of the research, as these multiple methods make intra-study comparisons (using methodologies such as meta-analysis) quite difficult. It is not our purpose to propose a standard for future research, which can continue to effectively assess the efficacy of RCF from these multiple perspectives, but instead to note that these mixed-method approaches are both advantageous and challenging for arriving at synthesis.

Based on the findings of the SLR and propositions, the following recommendations are proposed for practitioners:

- Train project teams to recognize cognitive biases and develop skills in probabilistic reasoning, ensuring that RCF results are interpreted and applied correctly without reverting to deterministic judgments.
- Adopt probabilistic estimation methods from the outset of project planning to avoid reliance on single-point figures and reduce subjective judgement. Ideally, replace single-point estimates and use intervals to communicate uncertainty to decision-makers.
- Increase accountability by separating the teams responsible for producing initial estimates from those applying RCF adjustments. An independent RCF unit can verify underlying assumptions, apply uplifts proportionately, and thereby maintain cost discipline while ensuring that

adjustments improve forecast accuracy rather than inflate budgets.

- Use independent reviews to reduce the influence of personal judgement as well as strategic misrepresentation and ensure that RCF produces more objective and reliable forecasts.
- Test hybrid or enhanced RCF models carefully before adopting them widely. Running pilot studies and comparing results with traditional RCF will help determine which approaches work best for different project contexts.
- Use clear criteria for constructing reference classes, defining inclusion and exclusion criteria based on comparable scale, context, and governance conditions, ensuring the selected reference class is statistically valid and comparable.
- Start developing a database of completed projects that can be used for RCF if no such database exists. Keep this database up to date and apply clear selection criteria to ensure RCF forecasts are more accurate and reliable over time.

7. Conclusions

The use of projects for realizing a wide variety of social and economic benefits continues to expand at a dramatic rate, as do unfortunately, numerous examples of their poor performance through cost overruns and unrealized goals. As scholars and public policy professionals alike struggle to identify the causes of these poor results, they have embraced the use of RCF as a means for correcting some of the human biases that are assumed to cause projects to continue to fail to live up to their promise. The purpose of this paper was to examine the current state of RCF, as it has become a widely used method for bias correction and identifying cost underestimation, to the point where its use is mandated for government projects in many European countries. While we found substantial evidence to support the efficacy of RCF as a more accurate means to assess the likely "true" costs of projects, its use is not without controversy and a set of as yet unresolved questions remains. The popularity and future use of this technique demands an objective reflection on what it can and, to date, cannot offer in terms of addressing the recurrent problem of project cost under-performance.

Our paper has shown how RCF has gradually undergone a conceptual shift. Originally, RCF functioned as an empirical non-parametric forecasting technique that relied on the observed distribution of outcomes from comparable past projects. However, recently, the method has moved to more analytical forecasting, integrating RCF with probabilistic models (hybrid RCF models) that are used within risk frameworks.

This shift has also changed the type of concerns raised in recent studies. Problems such as the risk of manipulating results by adjusting baselines or applying RCF too rigidly are mostly derived from how the method is used within risk management rather than from weaknesses in the method itself, in other words, most current critiques focus on practical implementation or governance, not on the core logic of RCF.

The recent developments of RCF are also linked to wider improvements in data quality and forecasting methods. As project datasets have become more detailed and reliable, the way RCF is applied has also matured. Early studies were limited by small or aggregated samples, but with the availability of richer and better-structured data, researchers can now apply RCF directly to baseline estimates and combine it with parametric or probabilistic models. In this context, RCF functions both as a benchmarking tool, grounded in historical evidence, and as a complement to more advanced risk modelling, thereby strengthening the role of RCF within project governance.

RCF is no doubt an important addition to managers' toolboxes. However, any tool needs skilled use in its application, and due care and attention is essential. Scholarly work to date has shown its benefits as well as its challenges and limitations. We have identified that there are significant opportunities for further valuable research here, including the difficulty of determining which previous projects should constitute the reference class, how data should be interpreted, and by whom, and how any budget uplift actually plays out in practice. Issues such as incorporating the wider supply chain further complicate the technique but offer the opportunity for valuable discussions at the estimation phase. Further empirical work should study the practical effects and outcome of RCF in organizations, and this can also be augmented by controlled experiments to examine any variability in managerial responses. We believe that the next steps in the RCF journey will be enlightening and allow more take-up of this system and, ultimately, better projects.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Chantal C. Cantarelli is an Associate Professor in Project and Programme Management at Cranfield University, Faculty of Business and Management. Prior to this, she was Assistant Professor at Sheffield University Management School and a Research Fellow at Oxford University's Saïd Business School. She gained her PhD in Systems Engineering, Policy Analyses, and Management from Delft University of Technology, The Netherlands. Her research focuses on the performance and management of major programs, with an emphasis on the unique challenges of megaprojects. Her primary interests include project complexity, innovation, resilience and decision-making, with the aim of improving the delivery of large-scale projects in uncertain environments.

Kate Davis is Reader in Project Management at Cranfield University, bringing over two decades of leadership in higher education. She has directed business management programmes and pioneered innovative employability initiatives that bridge academic learning with industry needs. Her research focuses on the dynamics of stakeholder impact in projects, advancing both theory and practice in complex organizational contexts. Beyond her academic contributions, she is committed to mentoring, supporting the professional growth of students, early career academics, and practitioners in the field of project management.

Jeffrey K. Pinto is the Andrew Morrow and Elizabeth Lee Black Chair in Management Technology and Professor with the Black School of Business, Penn State–Erie, The Behrend College, Erie, PA, USA. He is the author or editor of more than 25 books and over 200 scientific articles. His research has been published in the *Management Science*, *Research Policy*, *Journal of Management*, *Production Planning & Control*, *Expert Systems with Applications*, *Sloan Management Review*, *Journal of Management Studies*, *Journal of Product Innovation Management* and *IEEE Transactions on Engineering Management*. He is the author of the textbook, *Project Management: Achieving Competitive Advantage* with Pearson, now in revision for a sixth edition.

Neil Turner is a Professor of Project Management at Cranfield University, Faculty of Business and Management, having joined from a previous career as an engineering manager in a major telecommunications firm. His current research activities involve organizational knowledge and learning in the context of complex projects (the subject of his PhD), and how strategic operational choices can improve both delivery performance and resilience. His interests lie primarily in how managers deal with the organizational realities they face, including supply chain challenges in both the public and private sectors. He has published extensively in major academic journals, and his latest book is *Project Management, 5th edition*, with Dr. Harvey Maylor. He is the Course Director for the MSc by Research in Project Management at Cranfield.

ORCID

Chantal C. Cantarelli http://orcid.org/0000-0003-1376-941X

Data availability statement

No new data were created or analyzed during this study. Data sharing is not applicable to this article.

References

Allahaim, F. S., L. Liu, and X. Kong. 2016. "A Risk-Based Cost Contingency Estimation Model for Infrastructure Projects." In *International Conference on Construction and Real Estate Management (ICCREM)*, 1293–1303.

Ansar, A., B. Flyvbjerg, A. Budzier, and D. Lunn. 2014. "Should we Build More Large Dams? The Actual Costs of Hydropower Megaproject Development." *Energy Policy* 69: 43–56. https://doi.org/10.1016/j.enpol.2013.10.069.

Asadabadi, M. R., and O. Zwikael. 2024. "Unrealistic Project Goals: Detection and Modification." *Journal of Construction Engineering Management* 150 (3): 04023166.

Atapattu, C. N., N. Domingo, and M. Sutrisna. 2023. "A Bibliometric Review of the Statistical Modelling Techniques for Cost Estimation of Infrastructure Projects." *Smart and Sustainable Built Environment* (online ahead of print). https://doi.org/10.1108/SASBE-01-2023-0005.

Awojobi, O., and G. P. Jenkins. 2016."Managing the Cost Overrun Risks of Hydroelectric Dams: An Application of Reference Class Forecasting Techniques." *Renewable and Sustainable Energy Reviews* 63: 19–32. https://doi.org/10.1016/j.rser.2016.05.006.

Baerenbold, R. 2023. "Reducing Risks in Megaprojects: The Potential of Reference Class Forecasting." *Project Leadership and Society* 4: 100103. https://doi.org/10.1016/j.plas.2023.100103.

Batselier, J., and M. Vanhoucke. 2016. "Practical Application and Empirical Evaluation of Reference Class Forecasting for Project Management." *Project Management Journal* 47 (5): 36–51. https://doi.org/10.1177/875697281604700504.

Batselier, J., and M. Vanhoucke. 2017. "Improving Project Forecast Accuracy by Integrating Earned Value Management with Exponential Smoothing and Reference Class Forecasting." *International Journal of Project Management* 35 (1): 28–43. https://doi.org/10.1016/j.ijproman.2016.10.003.

Bayram, S., and S. Al-Jibouri. 2016a. "Application of Reference Class Forecasting in Turkish Public Construction Projects: Contractor Perspective." *Journal of Management in Engineering* 32 (3): 05016002. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000421.

Bayram, S., and S. Al-Jibouri. 2016b. "Efficacy of Estimation Methods in Forecasting Building Projects' Costs." *Journal of Construction Engineering and Management* 142 (11): 05016012. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001183.

Bayram, S., and S. Al-Jibouri. 2018. "Cost Forecasting Using RCF: A Case Study for Planning Public Building Projects Costs in Turkey." *International Journal of Construction Management* 18 (5): 405–417. https://doi.org/10.1080/15623599.2017.1333399.

Bordley, R. F. 2014. "Reference Class Forecasting: resolving Its Challenge to Statistical Modeling." *The American Statistician* 68 (4): 221–229. https://doi.org/10.1080/00031305.2014.937544.

Bracha, A., and D. J. Brown. 2012. "Affective Decision Making: A Theory of Optimism Bias." *Games and Economic Behavior* 75 (1): 67–80. https://doi.org/10.1016/j.geb.2011.11.004.

Brown, C., H. Lux, and J. R. Cowan. 2024. "Reference Class Forecasting and Its Application to Fusion Power Plant Cost Estimates." *IEEE Transactions on Plasma Science* 52 (9): 3628–3633. https://doi.org/10.1109/TPS.2024.3405631.

Caffieri, J. J., P. E. Love, A. Whyte, and D. D. Ahiaga-Dagbui. 2018. "Planning for Production in Construction: Controlling Costs in Major Capital Projects." *Production Planning & Control* 29 (1): 41–50. https://doi.org/10.1080/09537287.2017.1376258.

Callegari, C., A. Szklo, and R. Schaeffer. 2018. "Cost Overruns and Delays in Energy Megaprojects: How Big is Big Enough?" *Energy Policy* 114: 211–220. https://doi.org/10.1016/j.enpol.2017.11.059.

Cantarelli, C. C., B. Flyvbjerg, and S. L. Buhl. 2012. "Geographical Variation in Project Cost Performance: The Netherlands versus Worldwide." *Journal of Transport Geography* 24: 324–331. https://doi.org/10.1016/j.jtrangeo.2012.03.014.

Chadee, A., S. R. Hernandez, and H. Martin. 2021. "The Influence of Optimism Bias on Time and Cost on Construction Projects." *Emerging Science Journal* 5 (4): 429–442. https://doi.org/10.28991/esj-2021-01287.

Chadee, A., H. Martin, S. Gallage, and U. Rathnayake. 2023. "Reducing Cost Overrun in Public Housing Projects: A Simplified Reference Class Forecast for Small Island Developing States." *Buildings* 13 (4): 998. https://doi.org/10.3390/buildings13040998.

Duke, A. 2022. *Quit: The Power of Knowing When to Walk Away*. Portfolio/Pengui.

El Yamami, A., S. Ahriz, K. Mansouri, M. Qbadou, and E. Illoussamen. 2018. "Rethinking IT Project Financial Risk Prediction Using Reference Class Forecasting Technique." In *4th International Conference on Optimization and Applications (ICOA). Mohammedia, Morocco*, 1–6.

Flyvbjerg, B., and D. Gardner. 2023. *How Big Things Get Done:The Surprising Factors that Determine the Fate of Every Project, from Home Renovations to Space Exploration and Everything in Between*. New York: Crown Currency.

Flyvbjerg, B. 2021. "Top Ten Behavioral Biases in Project Management: An Overview." *Project Management Journal* 52 (6): 531–546. https://doi.org/10.1177/87569728211049046.

Flyvbjerg, B. 2014. "What You Should Know about Megaprojects and Why: An Overview." *Project Management Journal* 45 (2): 6–19. https://doi.org/10.1002/pmj.21409.

Flyvbjerg, B. 2011. “Over Budget, over Time, over and over Again: Managing Major Projects.” In *The Oxford Handbook of Project Management*, edited by P. W. G. Morris, J. K. Pinto, and J. Soderlund, 321–344. Oxford, UK: Oxford University Press.

Flyvbjerg, B. 2008. “Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice.” *European Planning Studies* 16 (1): 3–21. https://doi.org/10.1080/09654310701747936.

Flyvbjerg, B. 2006. “From Nobel Prize to Project Management: Getting Risks Right.” *Project Management Journal* 37 (3): 5–15. https://doi.org/10.1177/875697280603700302.

Flyvbjerg, B., C. Glenting, and A. Rønnest. 2004. “Procedures for Dealing with Optimism Bias.” In *Transport Planning Guidance Document*. London, UK: Department for Transport.

Flyvbjerg, B., and D. W. Bester. 2021. “The Cost-Benefit Fallacy: Why Cost-Benefit Analysis Is Broken and How to Fix It.” *Journal of Benefit-Cost Analysis* 12 (3): 395–419. https://doi.org/10.1017/bca.2021.9.

Flyvbjerg, B., A. Budzier, M. D. Christodoulou, and M. Zottoli. 2025. “The Uniqueness Trap.” *Harvard Business Review* 103 (2): 118–125.

Flyvbjerg, B., M. S. Holm, and S. Buhl. 2002. “Underestimating Costs in Public Works Projects: Error or Lie?” *Journal of the American Planning Association* 68 (3): 279–295. https://doi.org/10.1080/01944360208976273.

Flyvbjerg, B., C. K. Hon, and W. H. Fok. 2016. “Reference Class Forecasting for Hong Kong’s Major Roadworks Projects.” *Proceedings of the Institution of Civil Engineers - Civil Engineering* 169 (6): 17–24. https://doi.org/10.1680/jcien.15.00075.

Flyvbjerg, B., A. Stewart, and A. Budzier. 2016. “The Oxford Olympics Study 2016: Cost and Cost Overrun at the Games.” arXiv preprint arXiv:1607.04484.

Flyvbjerg, B., and C. R. Sunstein. 2016. “The Principle of the Malevolent Hiding Hand; or, the Planning Fallacy Writ Large.” *Social Research: An International Quarterly* 83 (4): 979–1004. https://doi.org/10.1353/sor.2016.0063.

Flyvbjerg, B., A. Budzier, J. S. Lee, M. Keil, D. Lunn, and D. W. Bester. 2022. “The Empirical Reality of IT Project Cost Overruns: Discovering a Power-Law Distribution.” *Journal of Management Information Systems* 39 (3): 607–639. https://doi.org/10.1080/07421222.2022.2096544.

Fridgeirsson, T. V. 2016. “Reference Class Forecasting In Icelandic Transport Infrastructure Projects. *Transport Problems*.” *Silesian University of Technology* 11 (2): 103–115.

Garbuio, M., and G. Gheno. 2023. “An Algorithm for Designing Value Propositions in the IoT Space: Addressing the Challenges of Selecting the Initial Class in Reference Class Forecasting.” *IEEE Transactions on Engineering Management* 70 (9): 3171–3182. https://doi.org/10.1109/TEM.2021.3077625.

Gold, S., S. Seuring, and P. Beske. 2010. “Sustainable Supply Chain Management and Inter-Organizational Resources: A Literature Review.” *Corporate Social Responsibility and Environmental Management* 17 (4): 230–245. https://doi.org/10.1002/csr.207.

Goldratt, E. M. 2017. *Critical Chain: A Business Novel*. UK: Routledge.

Graham, R. J. 1989. *Project Management as If People Mattered*. Bala Cynwyd, PA: Primavera Press.

Treasury, H. M. 2013. *Green Book Supplemental Guidance: Optimism Bias*. London: The Stationary Office (TSO).

Treasury, H. M. 2022. *The Green Book: Central Government Guidance on Appraisal and Evaluation*. London: The Stationary Office (TSO).

Ika, L. A., and L. T. Munro. 2022. “Tackling Grand Challenges with Projects: Five Insights and a Research Agenda for Project Management Theory and Practice.” *International Journal of Project Management* 40 (6): 601–607. https://doi.org/10.1016/j.ijproman.2022.05.008.

Ika, L. A., P. E. Love, and J. K. Pinto. 2022. “Moving beyond the Planning Fallacy: The Emergence of a New Principle of Project Behavior.” *IEEE Transactions on Engineering Management* 69 (6): 3310–3325. https://doi.org/10.1109/TEM.2020.3040526.

Jenkins, G., G. Olasehinde-Williams, and S. Baurzhan. 2022. “Is There a Net Economic Loss from Employing Reference Class Forecasting in the Appraisal of Hydropower Projects?” *Renewable and Sustainable Energy Reviews* 159: 112218. https://doi.org/10.1016/j.rser.2022.112218.

Kahneman, D., and A. Tversky. 1979b. “Prospect Theory: Analysis of Decisions Under Uncertainty.” *Econometrica* 47 (2): 263–291. https://doi.org/10.2307/1914185.

Kahneman, D., and A. Tversky. 1979a. “Intuitive Prediction: Biases and Corrective Procedures.” In *Studies in Management Sciences: Forecasting*, edited by S. Makridakis and S. C. Wheelwright, 313–327. Amsterdam: North Holland.

Kaiser, M. J., and B. Snyder. 2012. “Offshore Wind Capital Cost Estimation in the U.S. Outer Continental Shelf-A Reference Class Approach.” *Marine Policy* 36 (5): 1112–1122. https://doi.org/10.1016/j.marpol.2012.02.001.

Kim, B.-C., and K. F. Reinschmidt. 2011. “Combination of Project Cost Forecasts in Earned Value Management.” *Journal of Construction Engineering and Management* 137 (11): 958–966. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000352.

Koch, C. 2012. “Contested Overruns and Performance of Offshore Wind Power Plants.” *Construction Management and Economics* 30 (8): 609–622. https://doi.org/10.1080/01446193.2012.687830.

Koch, C., and J. Søndergaard. 2010. “Don Quixote Reborn? - Managing Offshore Wind Turbine Projects.” In *Portland International Conference on Management of Engineering and Technology (PICMET), Phuket, Thailand*, 1562–1570.

Leleur, S., K. B. Salling, I. Pilkauskiene, and M. S. Nicolaisen. 2015. “Combining Reference Class Forecasting with Overconfidence Theory for Better Risk Assessment of Transport Infrastructure Investments.” *European Journal of Transport and Infrastructure Research* 15 (3): 362–375. https://doi.org/10.18757/EJTIR.2015.15.3.3083.

Liu, H., C. Jiang, Y. Liu, M. Hertogh, and X. Lyu. 2018. “Optimism Bias Evaluation and Decision-Making Risk Forecast on Bridge Project Cost Based on Reference Class Forecasting: evidence from China.” *Sustainability* 10 (11): 3981–4010. https://doi.org/10.3390/su10113981.

Liu, L., and Z. Napier. 2010. “The Accuracy of Risk-Based Cost Estimation for Water Infrastructure Projects: Preliminary Evidence from Australian Projects.” *Construction Management and Economics* 28 (1): 89–100. https://doi.org/10.1080/01446190903431525.

Liu, L., G. Wehbe, and J. Sisovic. 2010. “The Accuracy of Hybrid Estimating Approaches: A Case Study of an Australian State Road Traffic Authority.” *The Engineering Economist* 55 (3): 225–245. https://doi.org/10.1080/0013791X.2010.502962.

Lordan-Perret, R., R. Bärenbold, H. Weigt, and R. Rosner. 2023. “An Ex-Ante Method to Verify Commercial U.S. nuclear Power Plant Decommissioning Cost Estimates.” *Energy Economics* 127: 107023. https://doi.org/10.1016/j.eneco.2023.107023.

Lovallo, D., C. Clarke, and C. Camerer. 2012. “Robust Analogizing and the outside View: Two Empirical Tests of Case-Based Decision Making.” *Strategic Management Journal* 33 (5): 496–512. https://doi.org/10.1002/smj.962.

Lovallo, D., M. Cristofaro, and B. Flyvbjerg. 2023. “Governing Large Projects: A Three-Stage Process toTo Get It Right.” *Academy of Management Perspectives* 37 (2): 138–156. https://doi.org/10.5465/amp.2021.0129.

Lovallo, D., and D. Kahneman. 2003. “Delusions of Success. How Optimism Undermines Executives’ Decisions.” *Harvard Business Review* 81 (7): 56–63, 117.

Love, P. E. D., and D. D. Ahiaga-Dagbui. 2018. “Debunking Fake News in a Post-Truth Era: The Plausible Untruths of Cost Underestimation in Transport Infrastructure Projects.” *Transportation Research Part A: Policy and Practice* 113: 357–368. https://doi.org/10.1016/j.tra.2018.04.019.

Love, P. E. D., L. Ika, and D. D. Ahiaga-Dagbui. 2019. “On de-Bunking Fake News in a Post-Truth Era: Why Does the Planning Fallacy Explanation for Cost Overruns and Benefit Shortfalls Fall Short?” *Transportation Research Part A: Policy and Practice* 126: 397–408. https://doi.org/10.1016/j.tra.2019.06.013.

Love, P. E. D., L. A. Ika, J. Matthews, and W. Fang. 2024. “Risk and Uncertainty in the Cost Contingency of Transport Projects: Accommodating Bias or Heuristics, or Both?” *IEEE Transactions on Engineering Management* 71: 205-–219. https://doi.org/10.1109/TEM.2021.3119064.

Love, P. E. D., L. A. Ika, and J. K. Pinto. 2023."Smart Heuristics for Decision-Making in the 'Wild': Navigating Cost Uncertainty in the Construction of Large-Scale Transport Projects." *Production Planning & Control* 35 (16): 2286–2303. https://doi.org/10.1080/09537287.2023.2248942.

Love, P. E., L. A. Ika, and J. K. Pinto. 2022. "Homo Heuristicus: From Risk Management to Managing Uncertainty in Large-Scale Infrastructure Projects." *IEEE Transactions on Engineering Management* 71: 1940–1949. https://doi.org/10.1109/TEM.2022.3170474.

Love, P. E. D., L. Ika, M. C. P. Sing, and S. Newton. 2019. "Cost Performance of Transportation Infrastructure Projects: The Fallacy of the Planning Fallacy account." *Transportation Research Part A: Policy and Practice* 122: 1–20. https://doi.org/10.1016/j.tra.2019.02.004.

Mak, S., and J. Raftery. 1992. "Risk Attitude and Systematic Bias in Estimating and Forecasting." *Construction Management and Economics* 10 (4): 303––320. https://doi.org/10.1080/01446199200000028.

Mayring, P. 2000. "Qualitative Content Analysis." *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 1: 20.

Natarajan, A. 2022. "Reference Class Forecasting and Machine Learning for Improved Offshore Oil and Gas Megaproject Planning: Methods and Application." *Project Management Journal* 53 (5): 456–-484. https://doi.org/10.1177/87569728211045889.

Nesvold, E., and R. B. Bratvold. 2022. "Debiasing Probabilistic Oil Production Forecasts." *Energy* 258: 124744. https://doi.org/10.1016/j.energy.2022.124744.

Oberlender, G. D., and S. M. Trost. 2001. "Predicting Accuracy of Early Cost Estimates Based on Estimate Quality." *Journal of Construction Engineering and Management* 127 (3): 173–182. https://doi.org/10.1061/(ASCE)0733-9364(2001)127:3(173).

Park, J. E. 2021a. "Schedule Delays of Major Projects: What Should we Do about It?" *Transport Reviews* 41 (6): 814–832. https://doi.org/10.1080/01441647.2021.1915897.

Park, J. E. 2021b. "Curbing Cost Overruns in Infrastructure Investment: Has Reference Class Forecasting Delivered Its Promised Success?" *European Journal of Transport and Infrastructure Research* 21 (2): 120–136. https://doi.org/10.18757/ejtir.2021.21.2.5504.

Pinto, J. K. 2023. "Is This How Big Things Get Done?" *International Journal of Project Management* 41 (5): 102484. https://doi.org/10.1016/j.ijproman.2023.102484.

Prater, J., K. Kirytopoulos, and T. Ma. 2017. "Optimism Bias within the Project Management Context: A Systematic Quantitative Literature Review." *International Journal of Managing Projects in Business* 10 (2): 370–385. https://doi.org/10.1108/IJMPB-07-2016-0063.

Salling, K. B., and D. Banister. 2010. "Feasibility Risk Assessment of Transport Infrastructure Projects: The CBA-DK Decision Support Model." *European Journal of Transport and Infrastructure Research* 10 (1): 103–120. https://doi.org/10.18757/EJTIR.2010.10.1.2871.

Salling, K. B., and S. Leleur. 2015. "Accounting for the Inaccuracies in Demand Forecasts and Construction Cost Estimations in Transport Project Evaluation." *Transport Policy* 38: 8–18. https://doi.org/10.1016/j.tranpol.2014.11.006.

Salling, K. B., and S. Leleur. 2012. "Modelling of Transport Project Uncertainties: Feasibility Risk Assessment and Scenario Analysis." *European Journal of Transport and Infrastructure Research* 12 (1): 21–38. https://doi.org/10.18757/EJTIR.2012.12.1.2947.

Salling, K. B., and S. Leleur. 2015. "Transport Project Evaluation: Feasibility Risk Assessment and Scenario Forecasting." *Transport* 32 (2): 180–191. https://doi.org/10.3846/16484142.2015.1063003.

Salling, K. B., and M. R. Pryn. 2015. "Sustainable Transport Project Evaluation and Decision Support: Indicators and Planning Criteria for Sustainable Development." *International Journal of Sustainable Development & World Ecology* 22 (4): 346–357. https://doi.org/10.1080/13504509.2015.1051497.

Servranckx, T., M. Vanhoucke, and T. Aouam. 2021. "Practical Application of Reference Class Forecasting for Cost and Time Estimations: Identifying the Properties of Similarity." *European Journal of Operational Research* 295 (3): 1161–1179. https://doi.org/10.1016/j.ejor.2021.03.063.

Seuring, S., and S. Gold. 2012. "Conducting Content-Analysis Based Literature Reviews in Supply Chain Management." *Supply Chain Management* 17 (5): 544–555.

Shenhar, A. J., and Dvir, D. 2007. *Reinventing Project Management: The Diamond Approach to Successful Growth and Innovation*. Harvard Business Review Press.

Sovacool, B. K., D. Nugent, and A. Gilbert. 2014. "Construction Cost Overruns and Electricity Infrastructure: An Unavoidable Risk?" *The Electricity Journal* 27 (4): 112–120. https://doi.org/10.1016/j.tej.2014.03.015.

Steininger, B. I., M. Groth, and B. L. Weber. 2020. "Cost Overruns and Delays in Infrastructure Projects: The Case of Stuttgart 21." *Journal of Property Investment & Finance* 39 (3): 256–282. https://doi.org/10.1108/JPIF-11-2019-0144.

Themsen, T. N. 2019. "The Processes of Public Megaproject Cost Estimation: The Inaccuracy of Reference Class Forecasting." *Financial Accountability & Management* 35 (4): 337–352. https://doi.org/10.1111/faam.12210.

Thomé, A. M. T., L. F. Scavarda, and A. J. Scavarda. 2016. "Conducting Systematic Literature Review in Operations Management." *Production Planning & Control* 27 (5): 408–420. https://doi.org/10.1080/09537287.2015.1129464.

Tranfield, D., D. Denyer, and P. Smart. 2003. "Towards a Methodology for developing evidence-Informed Management Knowledge by Means of Systematic Review." *British Journal of Management* 14 (3): 207–222. https://doi.org/10.1111/1467-8551.00375.

Wachs, M. 1989. "When Planners Lie with Numbers." *Journal of the American Planning Association* 55 (4): 476–479.

Walczak, R., and T. Majchrzak. 2018. "Implementation of the Reference Class Forecasting Method for Projects Implemented in a Chemical Industry Company." *Acta Oeconomica Pragensia* 26 (1): 25–33. https://doi.org/10.18267/j.aop.593.

Winch, G. M. 2025a. "Can Behavioral Decision Theory Provide the Microfoundations of Project Contingency?" *International Journal of Project Management* 43 (4): 102707. https://doi.org/10.1016/j.ijproman.2025.102707.

Winch, G. M. 2025b. *So, What Went Wrong With HS2?* Manchester: The Productivity Institute.

Winch, G. M., E. Maytorena-Sanchez, and N. Sergeeva. 2022. *Strategic Project Organizing*. Oxford: Oxford University Press.

Wohlin, C. 2014. "Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering." In *ACM International Conference Proceeding Series, London, United Kingdom*.

Wohlin, C., M. Kalinowski, K. Romero Felizardo, and E. Mendes. 2022. "Successful Combination of Database Search and Snowballing for Identification of Primary Studies in Systematic Literature Studies." *Information and Software Technology* 147 (1): 106908. https://doi.org/10.1016/j.infsof.2022.106908.

Zangeneh, P., and B. McCabe. 2022. "Modelling Socio-Technical Risks of Industrial Megaprojects Using Bayesian Networks and Reference Classes." *Resources Policy* 79: 103071. https://doi.org/10.1016/j.resourpol.2022.103071.

Zani, D., and B. T. Adey. 2025. "Empirical Comparison of Weighted and Non-Weighted Reference Class Forecasting." *Infrastructure Asset Management* 12 (3): 147–158. https://doi.org/10.1680/jinam.24.00024.

Zani, D., B. T. Adey, and S. Carroll. 2024. "An Approach to Support Reference Class Forecasting When Adequate Project Data Are Unavailable." *Results in Engineering* 22: 102333. https://doi.org/10.1016/j.rineng.2024.102333.

Zarghami, S. A. 2023. "Accommodating Uncertainty in Forecast Generation of Artificial Intelligence Tools in Construction Projects." In *2023 IEEE 5th International Conference on Architecture, Construction, Environment and Hydraulics, Taichung, Taiwan*, 8–11.

Zarikas, V., and C. P. Kitsos. 2015. "Risk Analysis with Reference Class Forecasting Adopting Tolerance Regions." In *Theory and Practice of Risk Assessment*, edited by C. P. Kitsos, T. A. Oliveira, A. Rigas, and S. Gulati, 235–247. New York: Springer International Publishing.

Reference class forecasting: promises, problems, and a research agenda moving forward

Cantarelli, Chantal C.

2026-05-19

Cantarelli CC, Davis K, Pinto JK, Turner N. (2026) Reference class forecasting: promises, problems, and a research agenda moving forward. Production Planning & Control, Volume 37, Issue 7, May 2026, pp. 691-709

https://doi.org/10.1080/09537287.2025.2578708