Department for Transport

OXFORD GLOBAL PROJECTS

UPDATING THE EVIDENCE BEHIND THE OPTIMISM BIAS UPLIFTS FOR TRANSPORT APPRAISALS

2020 data update to the 2004 Guidance Document *Procedures for Dealing with Optimism Bias in Transport Planning*

Department for Transport

OXFORD GLOBAL PROJECTS

TABLE OF CONTENTS

0. Executive summary

Reference Class Forecasting (RCF) is an established method for accounting for the systematic underestimation of cost and schedule overrun in projects. The underlying causes of this underestimation can include optimism bias (OB), strategic misrepresentation and economic incentives to see projects progress. RCF was first introduced in UK transport projects in the 2004 report *Appraisal Guidance for Optimism Bias* as the standard method to adjust estimates to account for biases in project cost estimates. This document provides an update to the 2004 report, bringing much more data into the analysis as evidence for optimism bias uplifts. Drawing on a much larger sample size, this report considers further dimensions of OB in addition to capital costs, e.g. OB on benefits, operational costs and project delivery schedules.

In addition, this report reviews recent methodological developments in RCF, finding that RCF has a track record of providing more accurate forecasts than conventional cost estimation methods. Furthermore, using RCF is found to increase the probability of delivering a project on time and on budget. The treatment of cost inflation is also explored in depth and an RCF for real cost inflation is developed.

Finally, data from 2,522 rail, road, bridge and tunnel new build projects show that risks are even larger at earlier project stages such as at the Outline Business Case and Strategic Outline Business Case stages. The report demonstrates that optimism exists for both cost, schedule, benefit and operational cost forecasts throughout all project stages. A novel insight is that risks mainly diminish in the tail of the distributions of overruns as projects progress, rather than steadily throughout the distribution. This has material implications for recommended OB adjustments for different business case stages.

1. Introduction

1.1 Background and objective

Since the first report *Appraisal Guidance for Optimism Bias* was published in 2004, the sources of evidence have greatly expanded. Considerable effort has been undertaken by academics and consultancies around the world to better understand the issues of Optimism Bias and to improve the available data. Further, new approaches and tools for Quantitative Cost Risk Analyses have been developed and practice has evolved on how to better integrate and combine other methods with Optimism Bias.

In response to this, The Department for Transport has contracted Oxford Global Project to undertake the consultancy assignment "Updating the evidence behind the optimism bias uplift for transport appraisals". The present research report is the result of this assignment.

First, this assignment has, increased the number of reference cases for each class of project types.

Second, this assignment has expanded the reference classes to include not only estimates from the stage of a full business case (FBC). The increase in available data makes it possible to track the projects through the lifecycle and covers the stages SOBC and OBC and is as such more informative for earlier baseline estimates.

Third, whereas the 2004 report only included cost estimates, Optimism Bias is equally an issue in estimates of schedule and benefits. The current report includes schedule and benefit estimates. Furthermore, data on operational costs (OPEX) have been collected and are included.

In addition, this assignment also investigates the extent of Optimism Bias in the UK and compares data from UK projects with international data.

2. Review of Reference Class Forecasting Methodology

This section provides an overview of developments and research that has taken place since 2004 to increase the accuracy of forecasts.

Common traditional project forecasting methods include three-point estimates, Monte Carlo simulations and Earned Value Management (EVM), once project work has started. The use of these methods has led projects to experience large cost overruns and schedule delays. The use of these methods has led projects to estimate median (P50) or mode (most likely) accurately, however they also lead to some projects experiencing large cost overruns and schedule delays. One of the main explanations for this is optimism bias, the tendency to be overly optimistic about future actions, resulting in underestimation of cost and schedule. Due to optimism bias project owners may be ignorant or underestimate the risk/uncertainties in estimates. Optimism bias is the result of taking an 'inside view', focusing on the project at hand and estimating costs and duration of activities bottom-up.

Traditional forecasting techniques typically take an 'inside view', they include a fixed contingency to the project cost estimate to account for risk and uncertainty in cost estimation, often 10% of the estimated cost. However, this method is considered to be biased because of the arbitrary way of deciding on the contingency amount (Liu et al., 2010).

Instead, Reference Class Forecasting (RCF) is an estimating approach that deals with optimism bias by taking an 'outside view' in determining the contingency amount that is based on statistical modelling of similar projects. Monte Carlo simulation can be considered as a 'semi outside view' because even though it makes use of historical data, it still relies on assumptions from the project manager to construct the distributional information (Batselier and Vanhoucke, 2016).

Since the Optimism Bias Guidance was published in 2004 the risk management profession has improved their approaches and tools for Quantitative Cost Risk Analyses (QRA).

A review of developments in RCF and its use in providing accurate forecast was made in 2019 (Oxford Global Projects, 2019). Based on several studies using RCF in various industries including hydropower dam projects, building projects, chemical projects and wind farms, RCF has shown to result in more accurate estimates than using conventional methods. The key findings are as follows:

- Application of RCF to Bujagali hydropower dam project resulted in a more reliable cost estimate and increased the accuracy of the cost-benefit analysis (Awojobi and Jenkins, 2016).
- A study of 420 building projects in Turkey revealed improved cost forecast accuracy when using RCF (Bayram and Al-Jabouri, 2016a).
- The same study of 420 building projects in Turkey showed RCF provided the most accurate forecasts in the early stages of the project (Bayram and Al-Jabouri, 2016b).
- Based on samples of nine and ten offshore wind farms in the United Kingdom respectively, using RCF increases the probability of delivering a project on time and on budget (Koch and Sondergaard, 2010; Koch, 2012).

The effectiveness of RCF depends on the similarity of the reference class. If the project fits well into the reference class, the resulting uplift from the RCF will provide a more reliable estimate of the cost of the project (Awojobi and Jenkins, 2016; Batselier and Vanhoucke, 2016). Moreover, the effectiveness of RCF is influenced by the size of the projects and the size of the reference class (Batselier and Vanhoucke, 2016; Walczak and Majchrzak, 2018); projects need to be sufficiently large and the reference class should include enough projects. Only if these criteria, similarity, project size, reference class size, are met will RCF outperform other methods.

In practical terms, any data is better than no data and a reference class comprising 20-30 past, similar projects is robust to derive meaningful insights. Moreover, as with the RCF analysis below, once data are pooled, they can be analysed to statistically test for similarities between subtypes of projects in the reference class or other characteristics, e.g. size, cost, timelines, location which might show statistically significantly different risk profiles.

Based on the review of methodological developments of RCF, we conclude RCF is still a valid and best practice approach in forecasting particularly in the early stages of project development. Best results in forecasting accuracy overall will be achieved by combining the bottom-up and top-down methods, particularly RBE and Bayesian forecasting, combining the ‘outside view’ with ‘inside view’ approaches (Kim and Reinschmidt, 2011; Koch, 2012; Leleur et al., 2015).

A more detailed review of reference class forecasting as a project forecasting method is found in Appendix B.

3. UPDATING CURRENT REFERENCE CLASSES

RCF uses historical project data as a predictor of the uncertainty and risk of future projects, including the risk of optimism bias. The first step in our analyses was to identify the historical projects which are relevant for updating the reference classes for the most common transport projects under the UK Department for Transport.

We reviewed and analysed cost and schedule data provided by Department for Transport from 5,294 completed Network Rail projects, 33 programmes from the latest Department for Transport GMPP return, a large dataset from Highways England, and data from the Department for Transport local projects evaluation. Further, we reviewed the methodology behind the datasets and ensured that the data were prepared in a way that would allow it to be integrated with the Oxford Global Projects dataset on project performance, which covers international transport and other infrastructure projects.

Compared to the Optimism Bias Guidance from 2004, the cost overrun benchmarking is now based on larger datasets of reference data as the Flyvbjerg database has expanded and is now integrated into the Oxford Global Projects database. In addition, DfT's own data added to the pool of reference data. Table 1 below shows an overview of the cost data basis for the 2004 OB guidance document and this present 2020 data update.

TABLE 1: OVERVIEW OF DATA IN 2004 GUIDANCE AND 2020 DATA UPDATE

Categories	**Example of project subtypes**	***2004 guidance document***	***2020 data update***
Rail	Light rail, conventional rail, urban rail, high-speed rail	46 (3 UK)	355 (18 UK)
Roads	Trunk roads, motorways, highways	172 (128 UK)	977 (202 UK)
Fixed links	Bridge and tunnels	34 (4 UK)	117 (6 UK)
Buildings	Stations, depots, concert halls, office buildings, museums	Mott Macdonald - Non-standard Buildings Capital Expenditures	149 (25 UK)
IT	IT system development	Mott Macdonald - Non-standard Buildings Capital Expenditures	5303 (171 UK)

Land & Property	Property purchases	-	48 (0 UK)
Rolling Stock	Powered and unpowered vehicles	-	20 (0 UK)

3.1. DATA

Projects under the Department for Transport are generally divided into three main categories; roads, rail and fixed links. In addition, the department is also either directly or indirectly responsible for a handful of other project types, of which Oxford Global Projects is able to provide data for reference class construction for building projects, IT projects, and the cost of acquiring land and property, which is applicable to many project types. Previously, the reference classes only consisted of data on capital cost risk, however this current report additionally incorporates data on schedule, benefit and OPEX risk. Similarly, the first report only consisted of estimated at the Full Business Case (FBC) stage, while this report additionally incorporates data at the Strategic Outline Business Case (SOBC) and Outline Business Case (OBC) stages. In the presentation of the results, we refer to any percentile values taken from the distribution of reference class curves as *RCF X.* RCF50 is the median and RCF80 is the 80th percentile. For instance, as shown in Table 2 below, 80% of rail projects in the reference class had a cost overrun of 60% or less compared to the base cost estimate. We adopted this language to help projects clearly differentiate between bottom-up risk estimates, which refer to e.g. P50 and P80, and the results of the reference class analyses to avoid confusion. Finally, all means provided in this report are adjusted means based on data from the 5th to 95th percentiles. Table 2 below shows a descriptive overview for the data basis of the data update.

TABLE 2: HIGH LEVEL OVERVIEW OF DATA IN 2020 OB DATA UPDATE (FBC)

	Sample size	***Mean deviation***	***Frequency***	***Median (RCF 50)***	***RCF 80***	***Historical range***
Rail						
Cost overrun	355	30%	7 out of 10	19%	60%	1964-2011
Schedule overrun	133	28%	6 out of 10	20%	61%	1964-2011
Benefit shortfall	132	-25%	7 out of 10	-30%	-64%	1970-2011
Roads						

Cost overrun	977	22%	8 out of 10	16%	47%	1954-2019
Schedule overrun	340	20%	7 out of 10	11%	57%	1965-2019
Benefit shortfall	973	-5%	6 out of 10	-7%	-35%	1954-2017
Fixed Links						
Cost overrun	117	28%	7 out of 10	20%	62%	1927-2016
Schedule overrun	54	17%	6 out of 10	4%	40%	1931-2016
Benefit shortfall	53	-13%	4 out of 10	-14%	-47%	1970-2010
Buildings						
Cost overrun	149	44%	7 out of 10	13%	84%	1851-2017
Schedule overrun	112	32%	6 out of 10	5%	68%	1851-2018
Benefit shortfall	21	-4%	6 out of 10	-4%	-48%	1993-2005
IT						
Cost overrun	5303	42%	4 out of 10	0%	50%	2002-2017
Schedule overrun	1318	28%	5 out of 10	0%	58%	2001-2017
Benefit shortfall	211	21%	5 out of 10	0%	-85%	2006-2017
Land and property						
Cost overrun	48	-4%	4 out of 10	-4%	11%	1994-2019
Rolling Stock						
Cost overrun	20	35%	9 out of 10	30%	64%	NA
Schedule overrun	20	4%	2 out of 10	0%	0%	NA
OPEX						
Roads	23	70%	7 out of 10	21%	185%	1998-2007
Rail	49	1%	4 out of 10	-10%	40%	1985-2007
Pooled*	74	23%	5 out of 10	-2%	77%	1981-2007

*Pooled OPEX includes two bridge projects.

Table 3 below shows a descriptive overview for the UK data basis of the data update.

TABLE 3: HIGH LEVEL OVERVIEW OF UK DATA IN 2020 OB DATA UPDATE (FBC)

	Sample size	*Mean*	*Frequency*	*Median (RCF 50)*	*RCF 80*
Rail					
Cost overrun	18	39%	6 out of 10	12%	68%
Schedule overrun	8	9%	8 out of 10	9%	17%
Benefit shortfall	14	-10%	6 out of 10	-4%	-36%
Roads					
Cost overrun	202	20%	9 out of 10	18%	37%
Schedule overrun	7	-2%	3 out of 10	0%	5%
Benefit shortfall	219	-1%	5 out of 10	-1%	-22%
Fixed Links					
Cost overrun	6	60%	10 out of 10	51%	107%
Benefit shortfall	7	-24%	9 out of 10	-16%	-63%
Buildings					
Cost overrun	25	26%	7 out of 10	11%	42%
Schedule overrun	12	11%	5 out of 10	0%	29%
Benefit shortfall	17	-7%	6 out of 10	-4%	-19%
IT					
Cost overrun	171	9%	4 out of 10	0%	30%
Schedule overrun	36	78%	8 out of 10	51%	140%
Benefit shortfall	16	-29%	6 out of 10	-14%	-75%

Table 4 below gives an overview of the geographic distribution of the cost overrun data basis for the data update.

TABLE 4: HIGH LEVEL OVERVIEW OF GEOGRAPHIC DATA FOR PROJECT COST OVERRUNS IN 2020 OB DATA UPDATE

	Africa (53)	*Asia (501)*	*South America (214)*	*North America (3501)*	*Oceania (214)*	*Europe (2379)**	*UK only (423)*
Rail (n=355)	1	69	2	83	12	176	18
Roads (n=977)	11	255	43	30	33	605	202
Fixed links (n=117)	0	8	0	24	9	76	6

Buildings (n=149)	0	34	0	40	5	70	25
IT (n=5303)	41	143	169	3348	74	1528	171
Land and Property (n=88)	0	0	0	0	0	88	0
Rolling Stock (n=20)	0	0	0	11	3	6	0
* Includes UK							

3.1.1. Network rail data

We cleaned and analysed a total of 5,294 completed projects in the Network Rail database. Our initial analysis showed that the projects in the dataset were small with a median size of £650.000. This is due to most projects in the dataset being maintenance and enhancement projects, which are very different from new build projects. Therefore, we recommend using broad rail RCFs based on data from the Oxford Global Projects (OGP) database.

Furthermore, our analysis led to the same results as those of Bert De Reyck and his team in a 2015 report on optimism bias in rail infrastructure[1]. Therefore, we recommend using the existing guidelines found in the report for future maintenance and enhancement projects. For practical appraisal purposes, we suggest defining rail maintenance and enhancement projects as projects with a base cost less than £7 million. In thread with this, we suggest defining rail new builds as projects with a base cost estimate of more than £7 million.

3.1.2. Collecting additional data

As a part of the data update, Oxford Global Projects collected additional data OPEX. In total, we were able to collect OPEX data from 74 projects; 23 roads, 49 rail, and 2 bridge projects. The projects originated from three geographical regions with 21 of the projects from China, seven from France, two from the Philippines, one from Sweden, one from the UK and 42 from the United States.

3.1.3. Data preparation

Reference Class Forecasting requires a like-for-like comparison. For cost reference class forecasts, specifically, projects need to compare outturn cost with cost estimates at the same price level and

[1]De Reyck, Bert, et al. (2015) "Optimism Bias Study: Recommended Adjustments to Optimism Bias Uplifts." *UK Department for Transport*. United Kingdom.

comparing real-term estimates with real-term outturn cost. The price level of a cost figure is adjusted to the year of the estimate and to the currency of the estimate. Similarly, cost forecasts have been stripped off contingency.

In accordance with the HMT Green Book 2018, we used country-specific GDP implicit deflators from the World Bank[2] to adjust price levels of cost estimates to the same year for the projects in the data sample. Note that the World Bank GDP implicit deflator for UK slightly deviates from the ONS GDP deflator recommended by HMT. The deflation data from World Bank was chosen to ensure comparability to international cost figures by using the same deflation index for all the projects in the data set. Deflation was done for the minority of the projects, since many of the OGP data are derived from research, in which case costs are commonly listed in real terms. The GDP implicit deflator is the ratio of GDP in current local currency to GDP in constant local currency. For more in-depth analyses of different inflation measures, see Appendix C.

For currency exchange rates, we used official exchange rates from the World Bank[3] to standardise cost estimates of the projects in the data sample. Official exchange rate refers to the exchange rate determined by national authorities or to the rate determined in the legally sanctioned exchange market. It is calculated as an annual average based on monthly averages (local currency units relative to the U.S. dollar).

For the statistical analysis, international projects were furthermore categorised into geographic regions using the *UN M49* nomenclature for area codes developed and maintained by the United Nations Statistics Division[4].

[2] GDP implicit deflators are based on World Bank national accounts data and OECD National Accounts data files. Source: "GDP Deflator (Base Year Varies by Country)." *The World Bank Data*, World Bank, https://data.worldbank.org/indicator/NY.GDP.DEFL.ZS.

[3] Official exchange rates are based on data from the International Financial Statistics database under the International Monetary Fund (IMF). Source: "Official exchange rate (LCU per US$, period average)." *The World Bank Data*, World Bank, https://data.worldbank.org/indicator/PA.NUS.FCRF.

[4] "Standard Country or Area Codes for Statistical Use (M49)." *United Nations Statistics Division*, United Nations, https://unstats.un.org/unsd/methodology/m49/.

Cost overrun is calculated as *Actual Cost / Estimated Cost – 1*, where estimated cost is measured at the relevant business case stage and actual cost at project completion. The estimated cost is the base cost, i.e. the estimated cost excluding provisions for risk or optimism bias.

Schedule overrun is calculated as *Actual Schedule / Estimated Schedule – 1*, where estimated schedule is measured from the approval of the relevant business case, i.e. the date of decision to build for FBC, date of SOBC and OBC, to the planned date of completion. Actual schedule is measured as the time passed from the date of SOBC/OBC/FBC approval to project completion as the date of actual opening.

Benefit shortfall is generally calculated as *Actual benefits / Estimated benefits -1*. Except in the few cases where the estimated benefits are negative (which is a phenomenon primarily found in IT projects), in which the benefit shortfall is calculated as *(Actual benefits – Estimated benefits) / Estimated benefits -1*. Benefit data is estimated for the first years of operation. If data is not available at first year, then the first reported year of operation within the first 5 years after opening is used as a proxy of the benefits achieved. Benefits are measured as number of passengers for rail, traffic counts for roads and fixed links, usage figures for buildings (i.e. visitor numbers for concert halls), and cashable benefits for IT projects.

OPEX overrun is measured as *Actual operational expenditure / Estimated operational expenditure - 1*. OPEX overrun data include costs for salary, wages and benefits, maintenance and repair, materials/outsourcing transport cost, leasing charges, fuel, materials and administration cost/loading cost. The data do not include capital upgrades. Actual operational expenditure is the expenses on the first year after opening. If data is not available the first year, then the first reported year within 5 years after opening is used. Estimated operational expenditure is measured from the approval of the relevant business case.

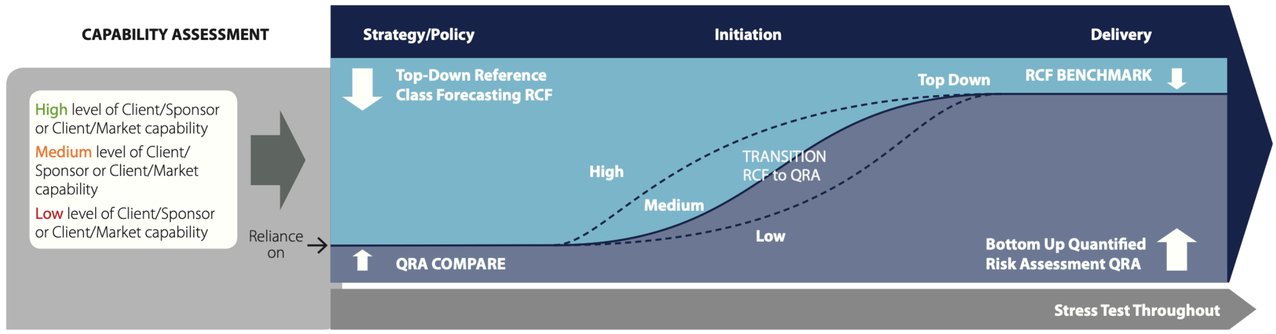
3.1.4. QA PROCEDURES

The Oxford Global Projects data used for the update have undergone strict quality assurance procedures. The OGP team collects data, which is then reviewed by a separate person on the team. Statistical analysis including inspection of histograms, various hypothesis tests and outlier analysis is then used to identify any anomalies, which are inspected and corrected if necessary. In addition, most of the transport data in the OGP database have been used in academic research and have thus been peer reviewed. However, this is not the case for the OPEX data that were collected specifically for the Department for Transport Optimism Bias update, which instead have been subject to the standard QA procedure.

3.1.5. Base cost estimates

All the cost uplift estimates in this report should be applied to the base cost estimate for projects and not to any estimates that include contingency such as the QRA Pmean. RCF (top-down) and QRA (bottom-up) estimates both apply to the base cost estimate. Typically, RCF is more reliable in earlier stages of the project, while QRA is more informative in later stages. The point of changing from RCF to QRA is determined by the degree of definition of the project and the organizational maturity (process and capability). The more mature an organization, the earlier the transition point. The Infrastructure and Projects Authority suggests that that the transition as minimum should not be earlier than optioneering, when there are still significant options open for a meaningful bottom up risk model[5].

FIGURE SHOWING THE TRANSITION PROCESS FROM TOP-DOWN RCF TO BOTTOM-UP QRA



SOURCE: INFRASTRUCTURE AND PROJECTS AUTHORITY, 2016

Additionally, the two assessments can be carried out in conjunction and used to inform the other and the project's approach to risk management. A gap between the top-down and bottom-up estimates could be an indication about the assurance validation of the QRA and that the project might not have understood risks correctly.

5 Infrastructure Projects Authority (2016). "Improving Infrastructure Delivery: Project Initiation Routemap - Risk Management Module". *Infrastructure Projects Authority,* United Kingdom.

3.1.6. Estimating at different stage gates

As outlined above, this assignment expands the calculation for the Optimism Bias Uplift to now include SOBC and OBC estimates. Contrary to expectations, the analysis of the different stages found that the RCF 50 (median) uplifts remain constant throughout the front-end of projects, i.e. when the project is developed from SOBC to FBC. The difference in uplift figures throughout the project life cycle changes only in the tails. Further analysis and previous analysis of the GMPP data across the UK government led to a few possible explanations for this unexpected finding.

First, key risks that are subject to optimism bias besides design risks are construction risks. Risks reduce as construction progresses and assumptions in estimates are proven by actual completion of the work.

Second, early estimates have an anchoring and lock-in effect[6], another form of well-documented bias in planners. This means that, for a portion of the projects, early estimates are only adjusted slightly during the planning process, therefore optimism bias on the typical project remains at similar levels throughout planning. Consequently, across the GMPP not only for transport projects, baselines of the typical projects (P50) remain similar during the front-end with only small adjustments between the different treasury approval points.

Third, the Optimism Bias changes only in the tails. A key explanation for cost overruns is scope change. The typical project, however, will not see major scope changes that fundamentally change baseline estimates and the risk a project is exposed to during construction. The impacts of this might explain the behavior of the tails of the distribution, where changes in the risk exposure and therefore changes to the Optimism Bias is present.

---

[6] Cantarelli, C.C., Flyvbjerg, B., van Wee, B. and Molin, E.J., 2010. Lock-in and its influence on the project performance of large-scale transportation infrastructure projects: investigating the way in which lock-in can emerge and affect cost overruns. *Environment and Planning B: Planning and Design*, *37*(5), pp.792-807.

Terrill, M. and Danks, L., 2016. *Cost overruns in transport infrastructure* (No. 2016-13). Carlton, Victoria, Australia: Grattan Institute.

3.2. Reference class construction

In this section, we test whether there is any indication in the data that the updated reference classes should be constructed or split on the basis of specific factors. First, we test whether bridges and tunnels should continue to be pooled into a single fixed links reference class. Then, we test other influential background variables. Finally, we compare the international data against the UK data to ensure that geographically pooled reference classes are meaningful for estimating risk in UK projects. Note that all the p-values listed in this report are outputs from non-parametric Wilcoxon rank-sum tests[7], which are preferable to classic t-tests when the data do not follow a normal distribution. The Wilcoxon rank-sum test is used to test whether two samples are likely to derive from the same population (i.e., that the two populations have similarly shaped distributions). This test is sometimes interpreted as a test of the null hypothesis that the medians of two distributions are equal. The tests were adjusted using Holm-adjustments to control for family-wise error rates. In addition to the Wilcoxon rank-sum tests, we additionally conducted all hypothesis tests using Welch's t-tests as well as non-parametric Kolmogorov-Smirnov, Anderson-Darling k-sample and Kruskal-Wallis rank-sum tests.

3.2.1. Statistical validation of reference classes

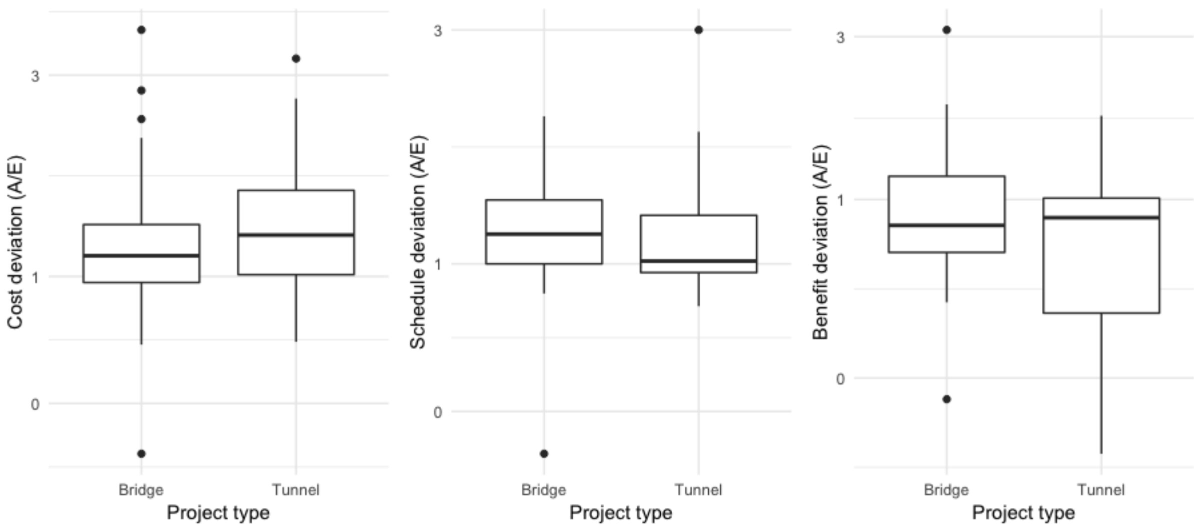
The 94 bridge projects were compared with the 75 tunnel projects in the sample. Here, we found that bridges were not statistically significantly different from tunnels in terms of neither cost overruns (p = 0.40), schedule overruns (p = 1.00) nor benefit shortfalls (p = 0.62). Bridges and tunnels are therefore continuously pooled in a single fixed link reference class, since the statistical tests showed no statistically significant difference with respect to cost overruns, schedule overruns and benefit shortfalls.

Figure 1 below show the distribution of these cost, schedule and benefit deviations for bridge and tunnel projects though box plots. The boxes show the middle portion of the data: the inter-quartile range (IQR). The bottom and top of the boxes mark the first quartile (the 25% mark – or P25) and the third quartile (the 75% mark – or P75). The thick bars in the middle of the boxes are the medians (P50). The upper whisker extends from the top of the box to the largest value no further than 1.5 *

[7] Except for the tests of the relationship between the estimated project duration/budget size and project cost overruns, which were conducted using simple linear regression.

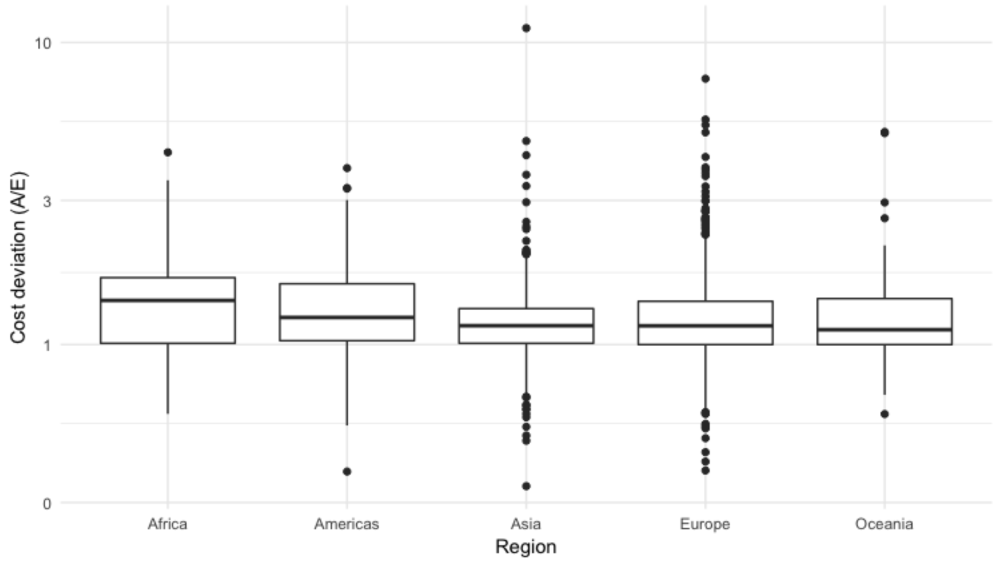
IQR from the top. The lower whisker extends from the bottom of the box to the smallest value at most 1.5 * IQR of the bottom. Data beyond the end of the whiskers are called "outlying" points and are plotted individually.

FIGURE 1: COMPARISON OF COST, SCHEDULE AND BENEFIT DEVIATION BETWEEN BRIDGES AND TUNNELS



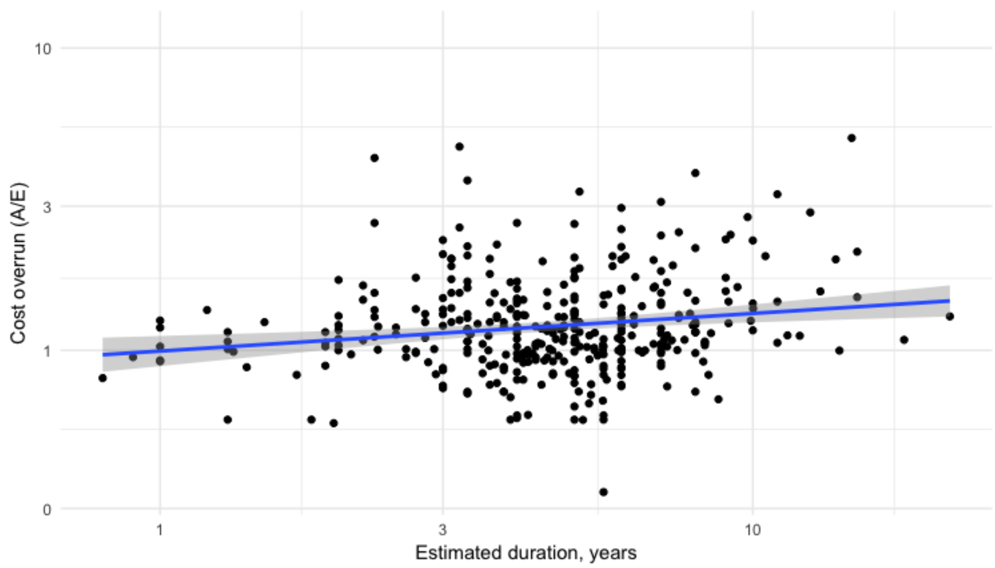
When comparing the performance of transport projects between major geographical regions, we found no association in the dataset between geographical regions and cost overruns (all $p \geq 0.09$), which indicates that the reference classes overall should not be split by geographical region. The distribution of capital cost overruns by geographical region is shown in figure 2 below.

FIGURE 2: TRANSPORT PROJECT COST OVERRUNS BY GEOGRAPHICAL REGION



The data show a statistically significant trend between the estimated duration of projects and cost overrun ($p < 0.001$). The trend is modest with 3-year projects having an estimated 13% cost overrun, 5-year projects having a 20% cost overrun, and 10-year projects having a 30% cost overrun. The trend is shown in figure 3 below, in which the blue line is the projected correlation and the shaded bands are 95% confidence intervals.

FIGURE 3: RELATIONSHIP BETWEEN ESTIMATED TRANSPORT PROJECT DURATION AND PROJECT COST OVERRUNS



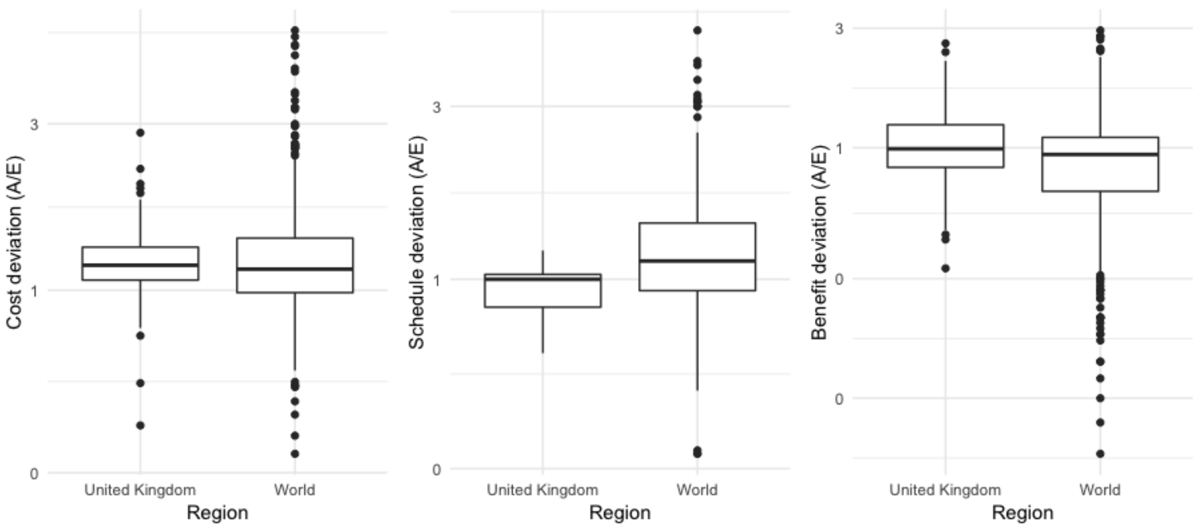
However, once the interaction between project type and estimated duration is accounted for, the trend is no longer statistically significant ($p = 0.282$). This is a sound argument for not splitting the reference classes by estimated project duration.

In addition, we tested the relationship between overruns and budget size ($p = 0.50$) but found no indication that the reference classes should split by the estimated project cost.

3.2.2. UK VS WORLD

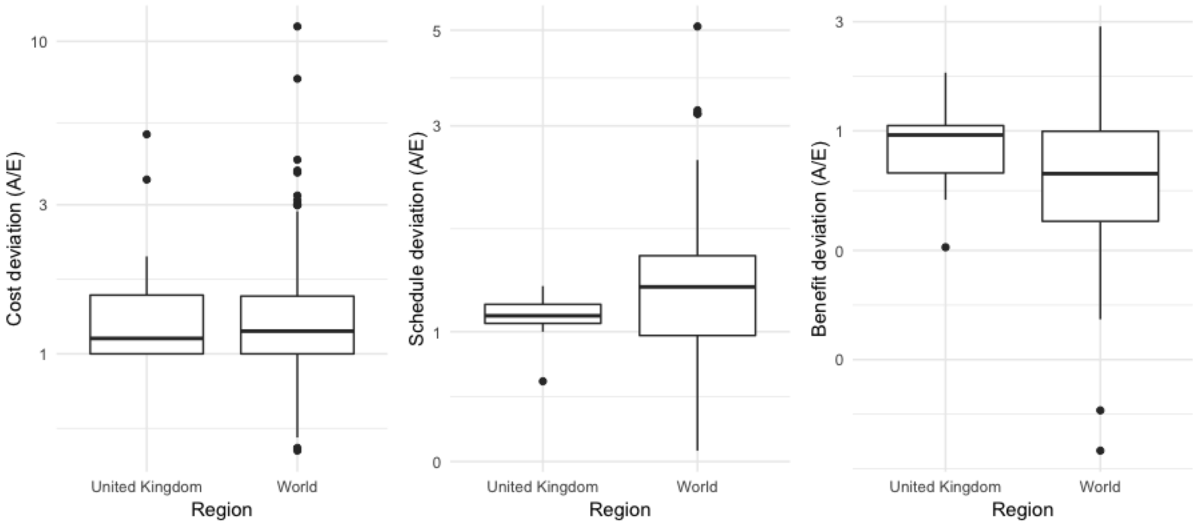
We find that UK road projects are not significantly different from international projects in terms of cost overruns ($p = 0.08$) when tested using Wilcoxon rank-sum tests. However, the Kolmogorov-Smirnov, Anderson-Darling k-sample, and Kruskal-Wallis rank-sum tests all find a statistically significant difference between UK and international road projects ($p \leq 0.001$). The tests did not find a significant difference in terms of schedule overruns ($p = 0.08$). Finally, all tests showed a statistically significant difference between UK and international road projects in terms of benefit shortfalls ($p \leq 0.001$). This indicates that DfT should use geographically pooled reference classes for schedule overruns but use UK-specific reference classes for road costs and benefits, especially considering the large UK-specific samples for both road cost overruns ($n = 203$) and benefit shortfalls ($n = 219$). Figure 4 below shows the UK vs world distribution of road project performance.

FIGURE 4: UK VS WORLD COMPARISON OF ROAD COST, SCHEDULE AND BENEFIT DEVIATION



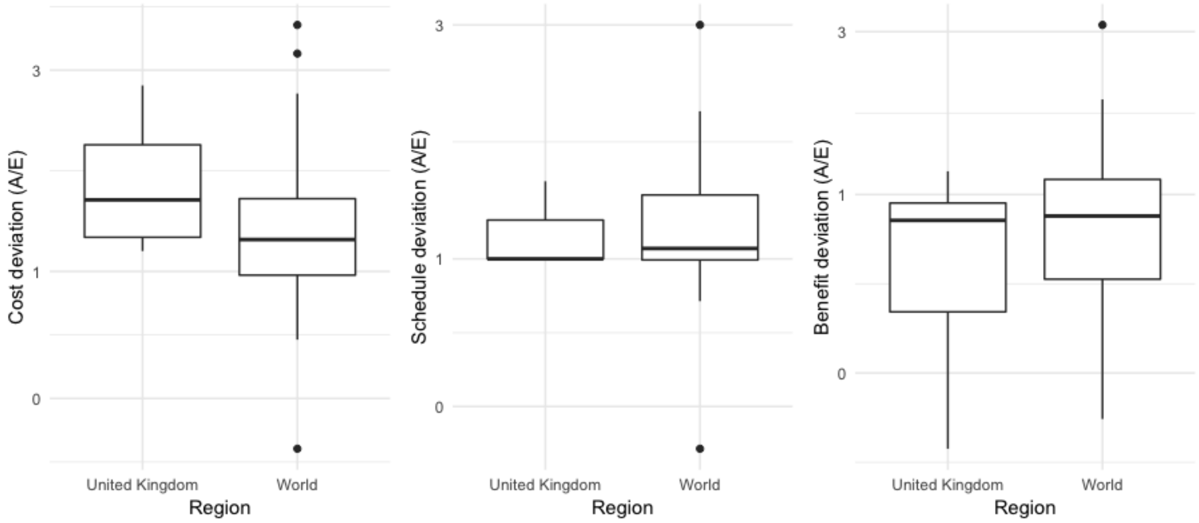
In terms of rail, the data show no significant difference between UK and international projects for cost overruns (p = 0.85), schedule overruns (p = 0.24), or benefit shortfalls (p = 0.053). The almost-significant p-value for might indicate that the UK projects are different from the international projects in terms of benefit performance, however the UK data for the statistical comparison is limited to 14 projects. Therefore, we recommend using an internationally pooled rail benefit reference class, which even though it shows a larger benefits shortfall than the UK specific one, and then testing the difference again once more data become available. Figure 5 below shows the UK vs world distribution of rail project performance.

FIGURE 5: UK VS WORLD COMPARISON OF RAIL COST, SCHEDULE AND BENEFIT DEVIATION



In terms of fixed links, the data show no statistically significant differences for cost overruns (p = 0.09), schedule overruns (p = 0.94) or benefit overruns (p = 0.48). This indicates that DfT should use pooled reference classes for fixed link projects. Note that this analysis is affected by data scarcity, since the dataset only includes UK fixed links data for 6 cost overruns, 3 schedule overruns and 7 benefit overruns. Thus, we also recommend DfT to use reference classes based on international data. Figure 6 below shows the UK vs world distribution of fixed link project performance.

FIGURE 6: UK VS WORLD COMPARISON OF FIXED LINKS COST, SCHEDULE AND BENEFIT DEVIATION



3.3 Results: Picking the right level of contingency

The risk appetite differs among decision makers and thus the required contingency commonly also varies. However, as a rule of thumb we would suggest picking RCF50 as the low contingency bound, RCF65 as the mid-point and RCF80 as the high contingency bound. For economic appraisals, we suggest using the trimmed means (i.e. the mean value within P5-P95) listed in this report due to the highly skewed distributions of project performance, which otherwise skew the means towards extremely large contingency uplifts.

3.4 Results: Updated reference classes

With the available data, we constructed six project specific international (including UK) reference classes for cost overruns. For each of the six project types, three separate reference classes were produced with estimates for the SOBC, OBC and FBC business case stages. Table 5 below displays an overview of the 18 total international capital cost reference classes. RCF curves can be found in Appendix A.

TABLE 5: OVERVIEW OF INTERNATIONAL (INCL. UK) CAPITAL COST REFERENCE CLASSES

	Cost overrun (mean)	*50% certainty of the estimate (RCF50)*	*80% certainty of the estimate (RCF80)*
Rail (n=355)			
FBC	30%	19%	60%
OBC	33%	19%	77%
SOBC	56%	19%	121%
Road (n=977)			
FBC	22%	16%	47%
OBC	25%	16%	64%
SOBC	48%	16%	108%
Fixed links (n=117)			
FBC	28%	20%	62%
OBC	32%	20%	79%
SOBC	55%	20%	123%
Building (n=149)			
FBC	44%	13%	84%
OBC	48%	13%	101%
SOBC	70%	13%	145%
IT (n=5303)			
FBC	42%	0%	50%
OBC	50%	0%	60%
SOBC	69%	0%	111%
Land and Property (n=88)			
FBC	-4%	4%	11%
OBC	14%	0%	62%
SOBC	33%	3%	116%
Rolling Stock (n=20)			
FBC	35%	30%	64%
OBC	38%	30%	81%
SOBC	61%	30%	125%

We constructed five international reference classes for schedule overruns. For each of the five project types, three separate reference classes were produced with estimates for the SOBC, OBC and FBC business case stages. Table 6 below displays an overview of the 15 total international schedule reference classes. RCF curves can be found in Appendix A.

TABLE 6: OVERVIEW OF INTERNATIONAL SCHEDULE REFERENCE CLASSES

	Schedule overrun (mean)	*50% certainty of the schedule estimate (RCF50)*	*80% certainty of the schedule estimate (RCF80)*
Rail (n=133)			
FBC	28%	20%	61%
OBC	33%	20%	78%

SOBC	31%	20%	73%
Road (n=340)			
FBC	20%	11%	57%
OBC	25%	11%	74%
SOBC	23%	11%	69%
Fixed links (n=54)			
FBC	17%	4%	40%
OBC	22%	4%	57%
SOBC	20%	4%	52%
Building (n=112)			
FBC	32%	5%	68%
OBC	37%	5%	85%
SOBC	35%	5%	80%
IT (n=1318)			
FBC	28%	0%	58%
OBC	32%	0%	70%
SOBC	31%	0%	70%
Rolling Stock (n=20)			
FBC	4%	0%	0%
OBC	9%	0%	17%
SOBC	7%	0%	12%

We constructed five international benefit reference classes at the FBC stage. Table 7 below displays an overview of the five reference classes. RCF curves can be found in Appendix A.

TABLE 7: OVERVIEW OF INTERNATIONAL BENEFIT REFERENCE CLASSES AT FBC STAGE

	Benefit shortfall (mean)	***50% certainty of the benefit estimate (RCF50)***	***80% certainty of the benefit estimate (RCF80)***
Rail (n=132)	-25%	-30%	-64%
Road (n=793)	-5%	-7%	-35%
Fixed links (n=53)	-13%	-14%	-47%
Building (n=21)	-4%	-4%	-48%
IT (n=211)	-21%	0%	-85%

Finally, constructed three international OPEX reference classes at the FBC stage. Table 7 below displays an overview of the three reference classes. RCF curves can be found in Appendix A.

TABEL 8: OVERVIEW OF INTERNATIONAL OPEX REFERENCE CLASSES AT FBC STAGE

	OPEX overrun (mean)	***50% certainty of the OPEX estimate (RCF50)***	***80% certainty of the OPEX estimate (RCF80)***
Road (n=23)	23%	-2%	77%
Rail (n=49)	70%	21%	185%
Pooled* (n=74)	1%	-10%	40%

*The pooled reference class includes OPEX data from two bridge projects

In addition, UK specific reference classes were constructed as well. In total, we constructed five UK project-specific reference classes for cost overruns. For each of the five project types, three separate reference classes were produced with estimates for the SOBC, OBC and FBC business case stages. We suggest using international data for any UK reference class with a sample size below 20. Table 9 below displays an overview of the 15 total UK cost reference classes. RCF curves can be found in Appendix A.

TABLE 9: OVERVIEW OF UK COST REFERENCE CLASSES

	Cost overrun (mean)	***50% certainty of the cost estimate (RCF50)***	***80% certainty of the cost estimate (RCF80)***
Rail (n=18)			
FBC	39%	12%	68%
OBC	42%	12%	85%
SOBC	65%	12%	130%
Road (n=202)			
FBC	20%	18%	37%
OBC	23%	18%	54%
SOBC	46%	18%	98%
Fixed links (n=6)			
FBC	60%	51%	107%
OBC	62%	51%	124%
SOBC	85%	51%	168%
Building (n=25)			
FBC	26%	11%	42%
OBC	29%	11%	59%
SOBC	51%	11%	103%
IT (n=171)			
FBC	9%	0%	30%
OBC	12%	0%	47%
SOBC	35%	0%	91%

Note that we consider a sample size of under 20 projects to be too small for proper reference class forecasting.

Similarly, we constructed five UK project-specific reference classes for schedule overruns. For each of the five project types, three separate reference classes were produced with estimates for the SOBC, OBC and FBC business case stages. We suggest using international data for any UK reference class with a sample size below 20. Table 10 below displays an overview of the 15 total UK schedule reference classes. RCF curves can be found in Appendix A.

TABLE 10: OVERVIEW OF UK SCHEDULE REFERENCE CLASSES

	Schedule overrun (mean)	***50% certainty of the schedule estimate (RCF50)***	***80% certainty of the schedule estimate (RCF80)***
Rail (n=8)			
FBC	9%	9%	17%
OBC	13%	9%	34%
SOBC	11%	9%	29%
Road (n=7)			
FBC	-2%	0%	5%
OBC	0%	0%	22%
SOBC	-2%	0%	17%
Fixed links (n=3)			
FBC	NA	NA	NA
OBC	NA	NA	NA
SOBC	NA	NA	NA
Building (n=12)			
FBC	11%	0%	29%
OBC	16%	0%	46%
SOBC	14%	0%	41%
IT (n=36)			
FBC	78%	51%	140%
OBC	83%	51%	157%
SOBC	81%	51%	152%

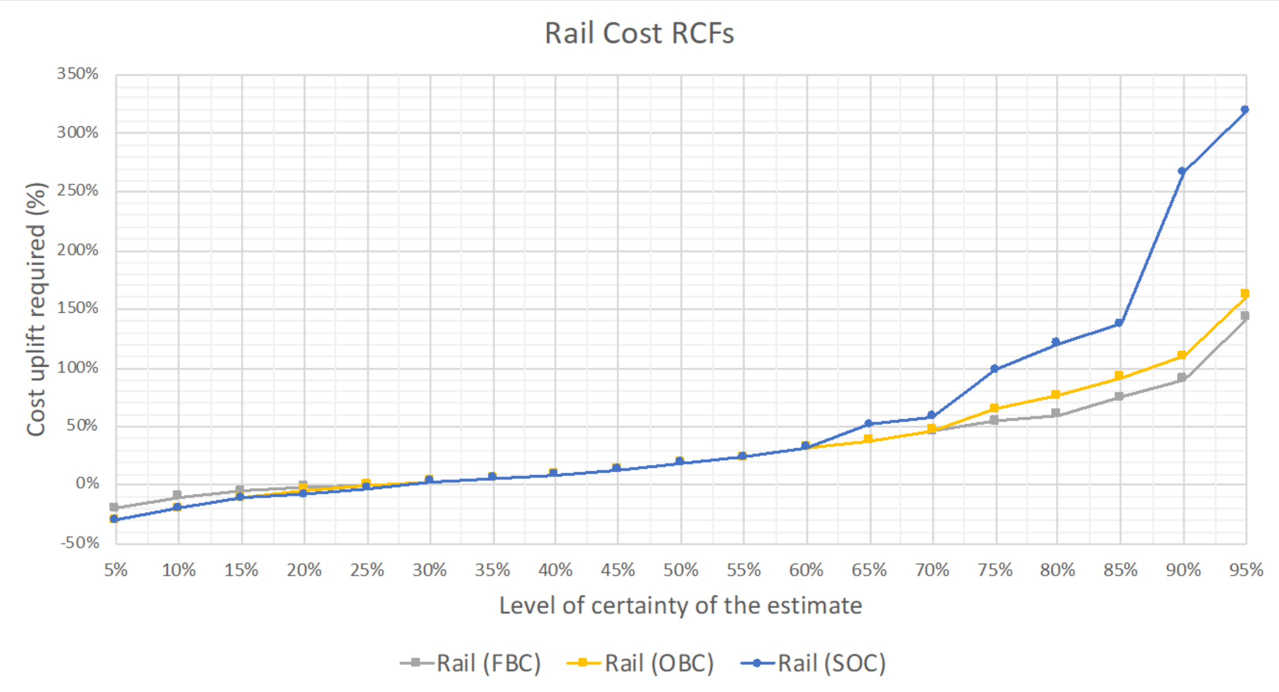
Note that we consider a sample size of under 20 projects to be too small for proper reference class forecasting.

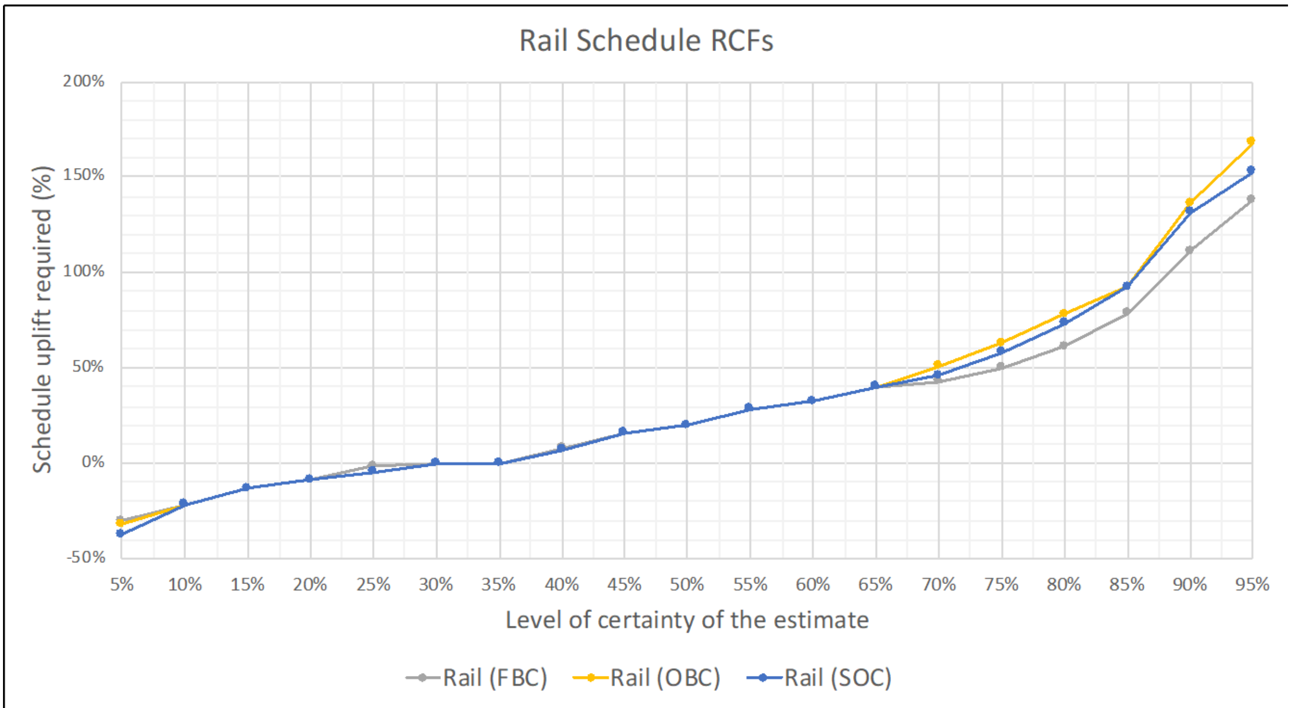
Finally, we constructed five UK-specific benefit reference classes at the FBC stage. Table 11 below displays an overview of the five reference classes. RCF curves can be found in Appendix A.

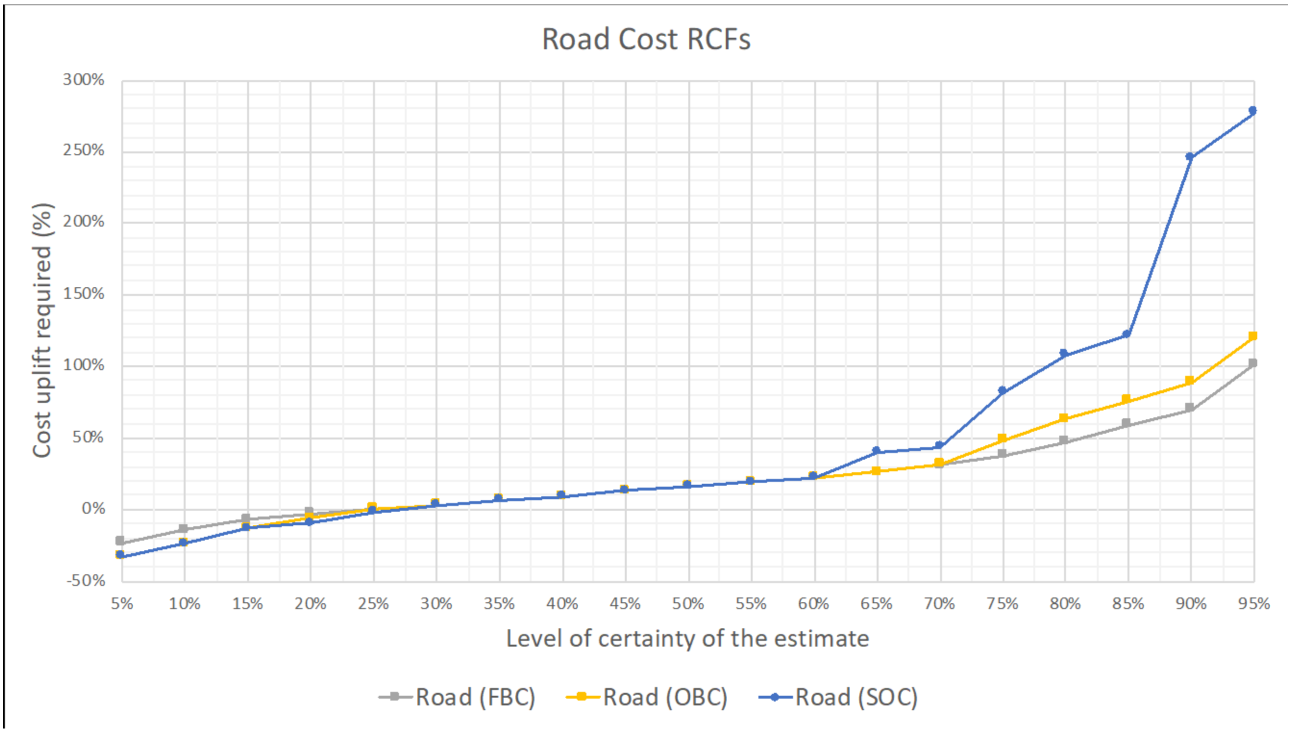
TABLE 11: OVERVIEW OF UK BENEFIT REFERENCE CLASSES AT FBC STAGE

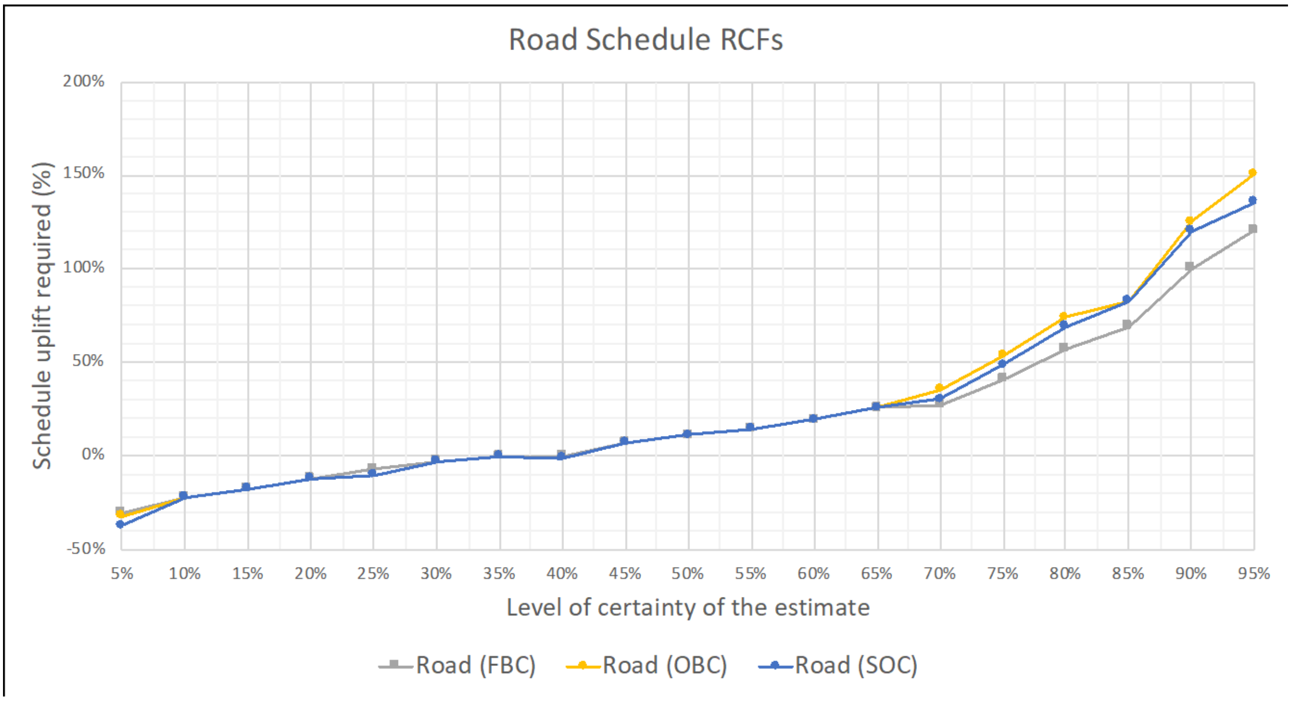
	Benefit shortfall (mean)	*50% certainty of the benefit estimate (RCF50)*	*80% certainty of the benefit estimate (RCF80)*
Rail (n=14)	-10%	-4%	-36%
Road (n=219)	-1%	-1%	-22%
Fixed links (n=7)	-24%	-16%	-63%
Buildings (n=17)	-7%	-4%	-19%
IT (n=16)	-29%	-14%	-75%

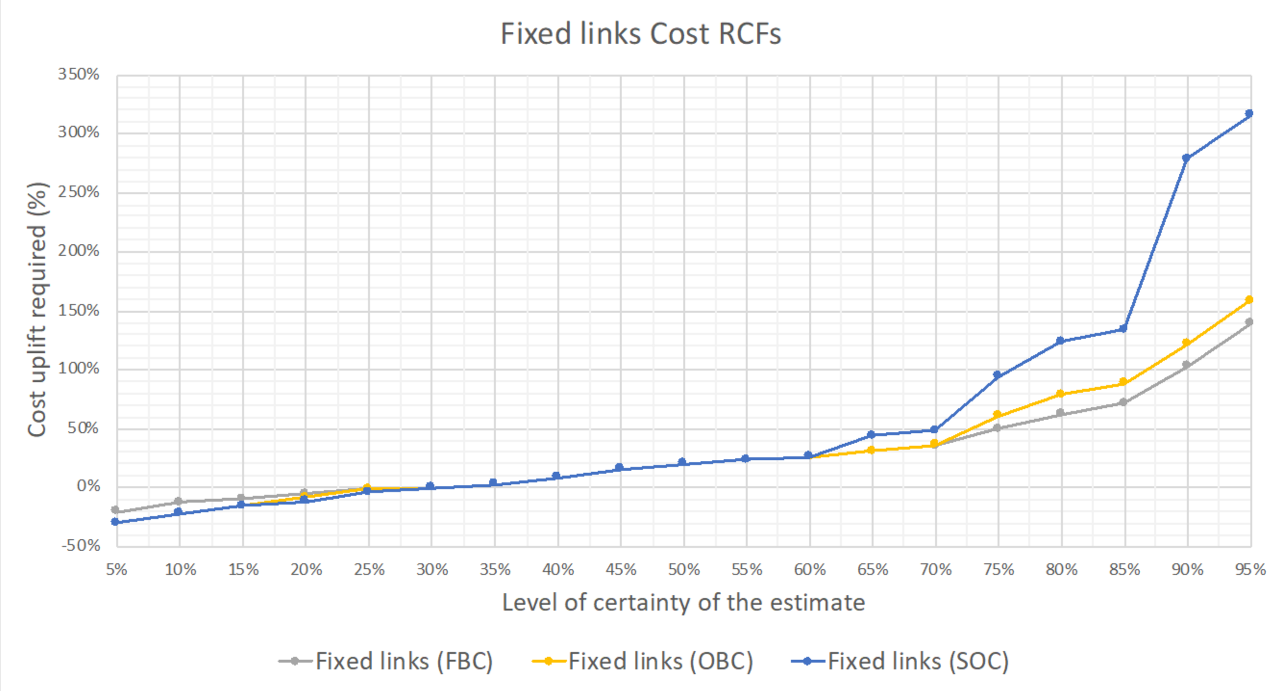
APPENDIX A: UPDATED REFERENCE CLASSES

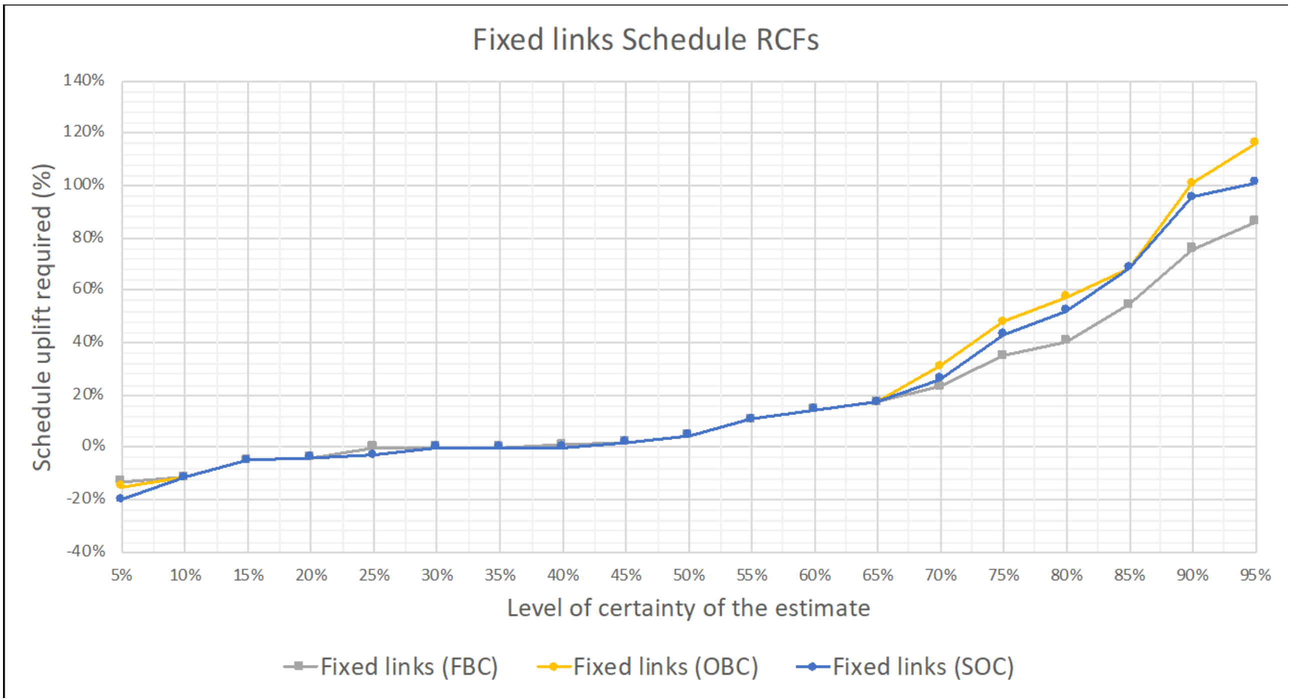


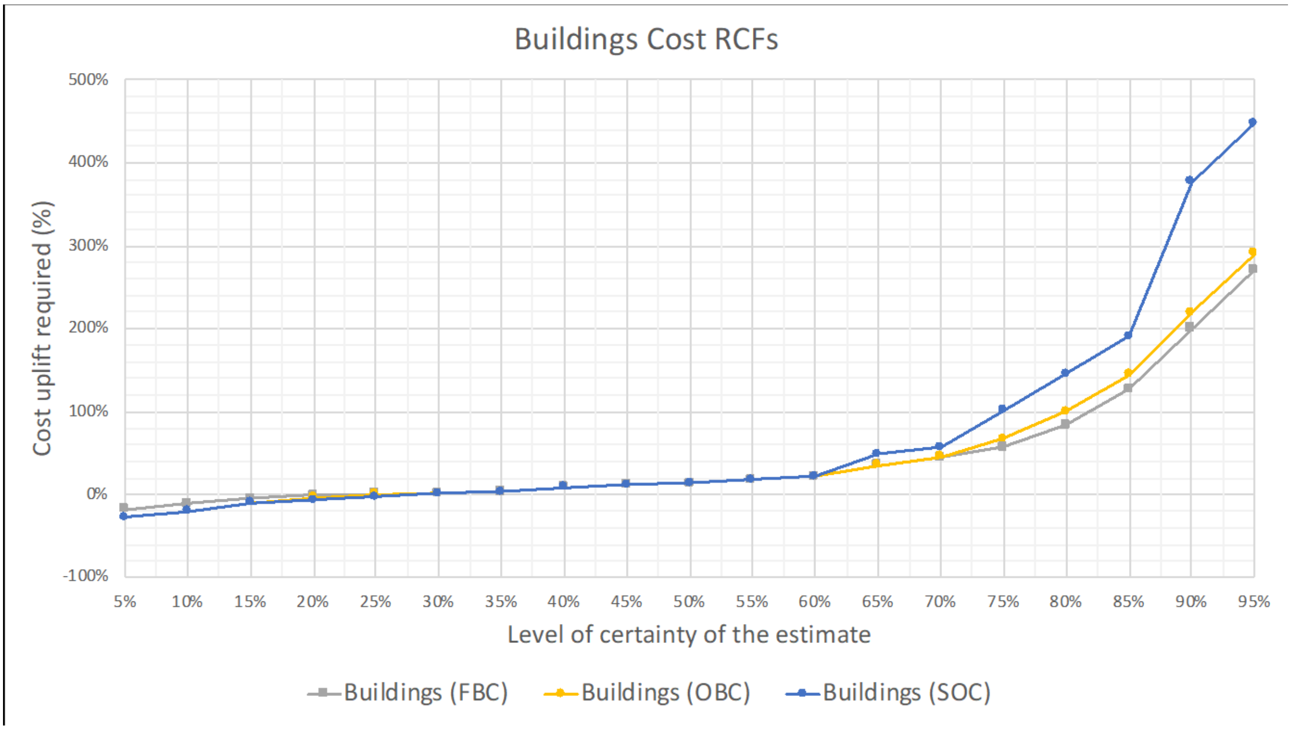


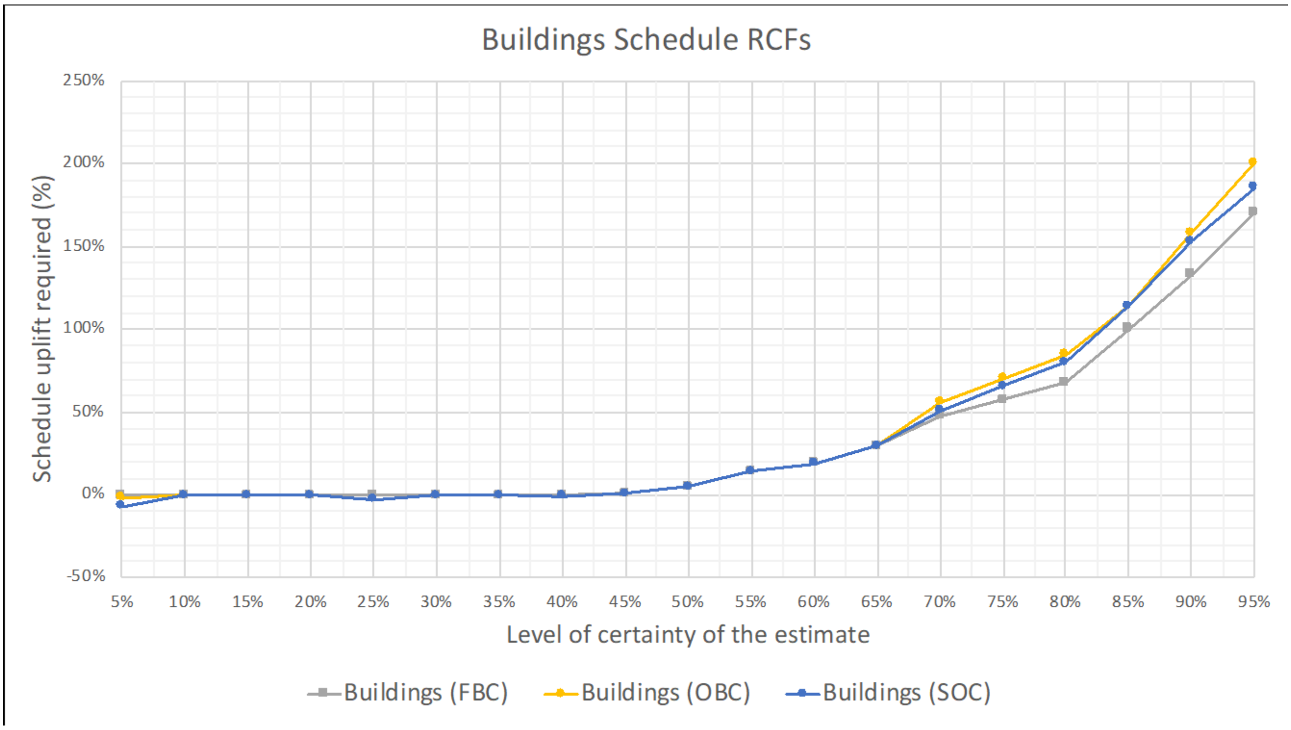

Road Cost RCFs
300%
250%
200%
150%
100%
50%
0%
-50%
Cost uplift required (%)
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Road (FBC)
Road (OBC)
Road (SOC)

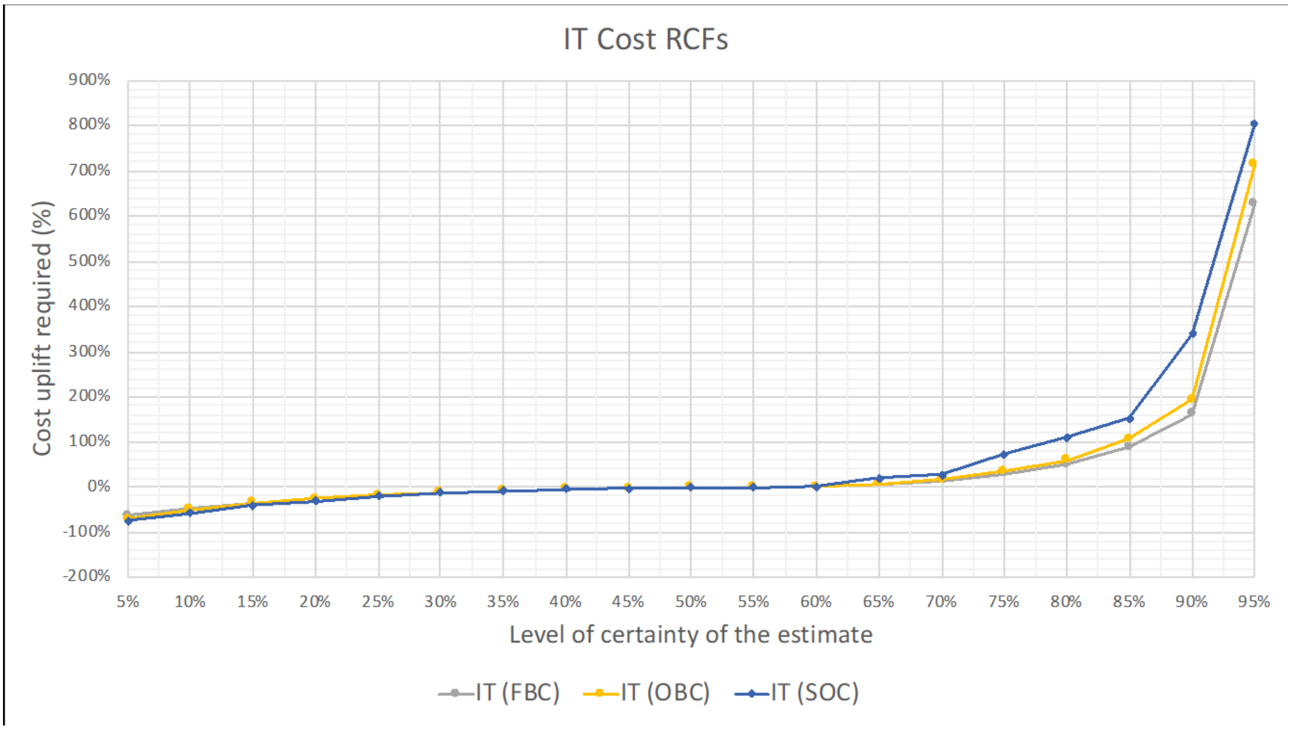

Road Schedule RCFs
200%
150%
100%
50%
0%
-50%
Schedule uplift required (%)
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Road (FBC)
Road (OBC)
Road (SOC)

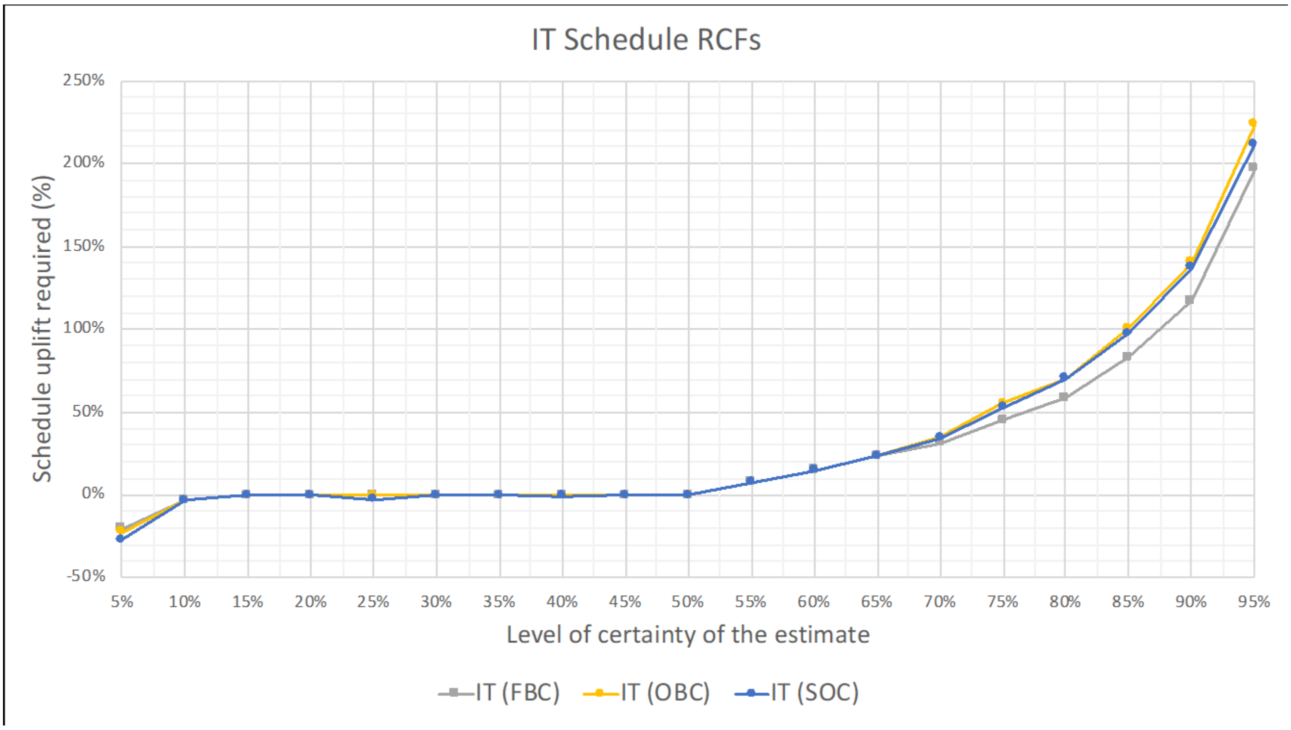

Fixed links Cost RCFs
Cost uplift required (%)
350%
300%
250%
200%
150%
100%
50%
0%
-50%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Fixed links (FBC)
Fixed links (OBC)
Fixed links (SOC)

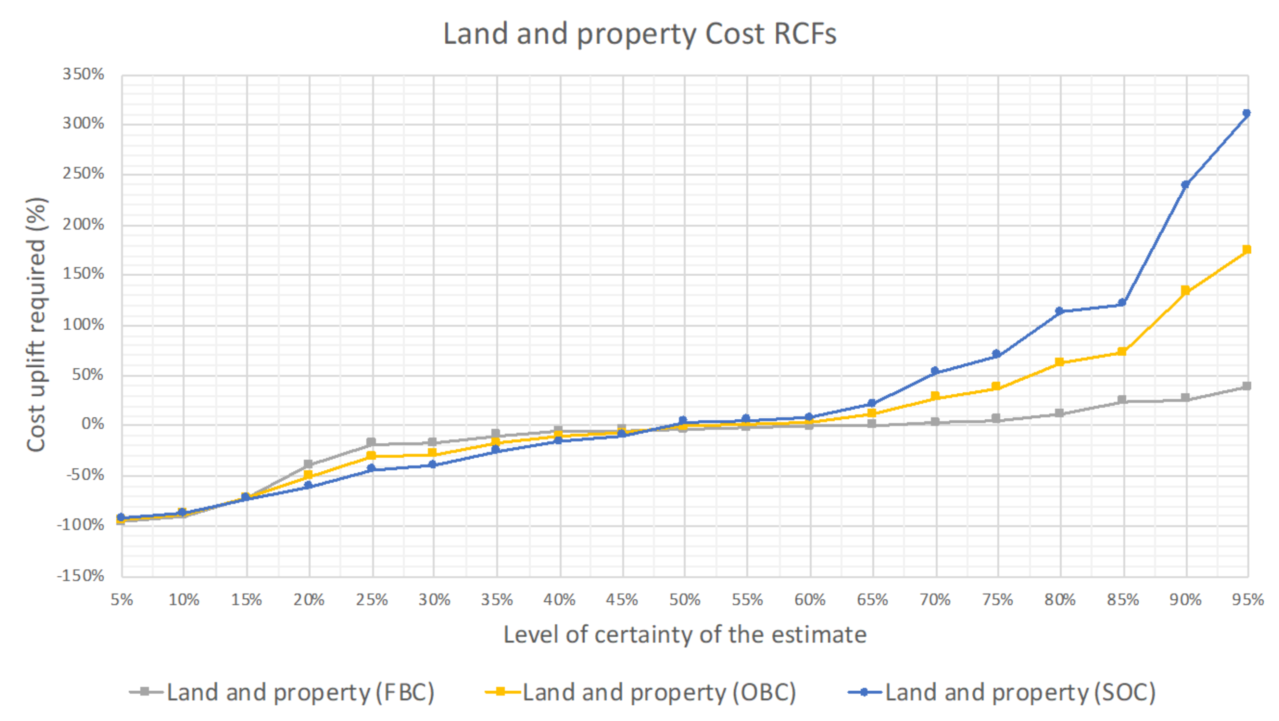

Fixed links Schedule RCFs
Schedule uplift required (%)
140%
120%
100%
80%
60%
40%
20%
0%
-20%
-40%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Fixed links (FBC)
Fixed links (OBC)
Fixed links (SOC)

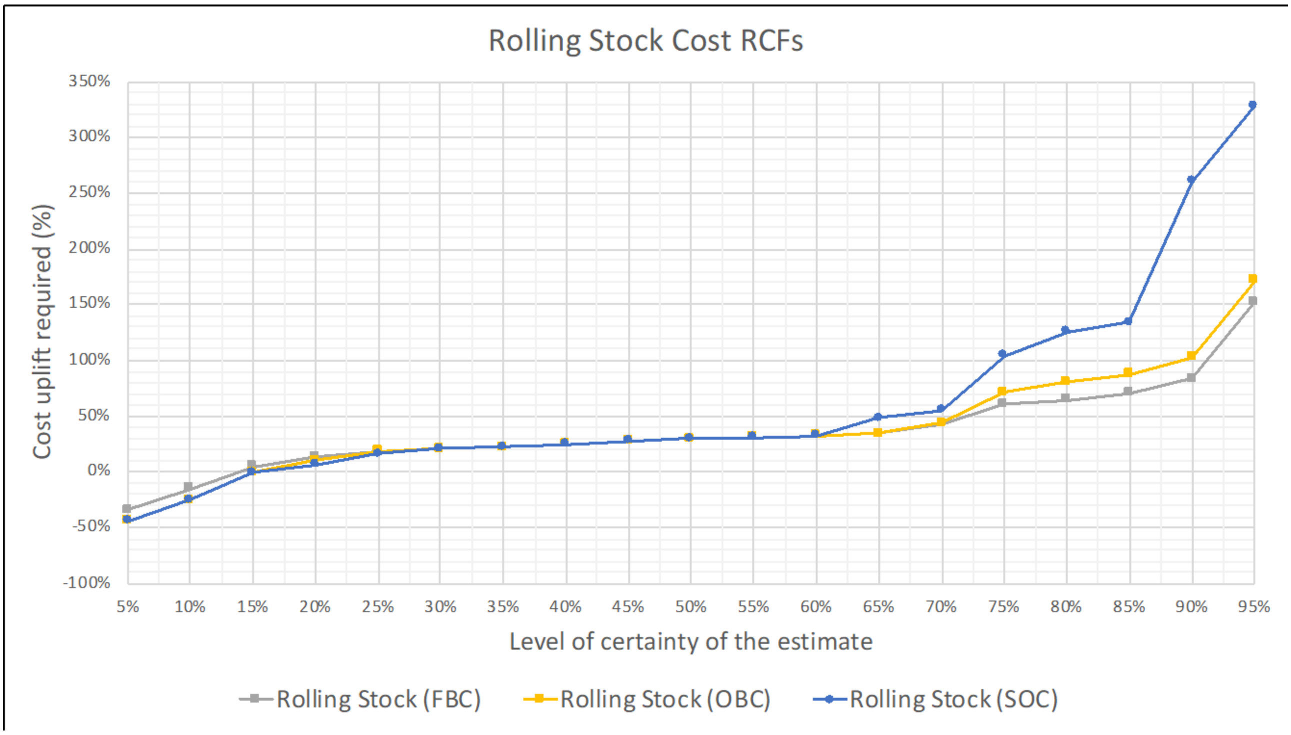

Buildings Cost RCFs
Cost uplift required (%)
500%
400%
300%
200%
100%
0%
-100%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Buildings (FBC)
Buildings (OBC)
Buildings (SOC)

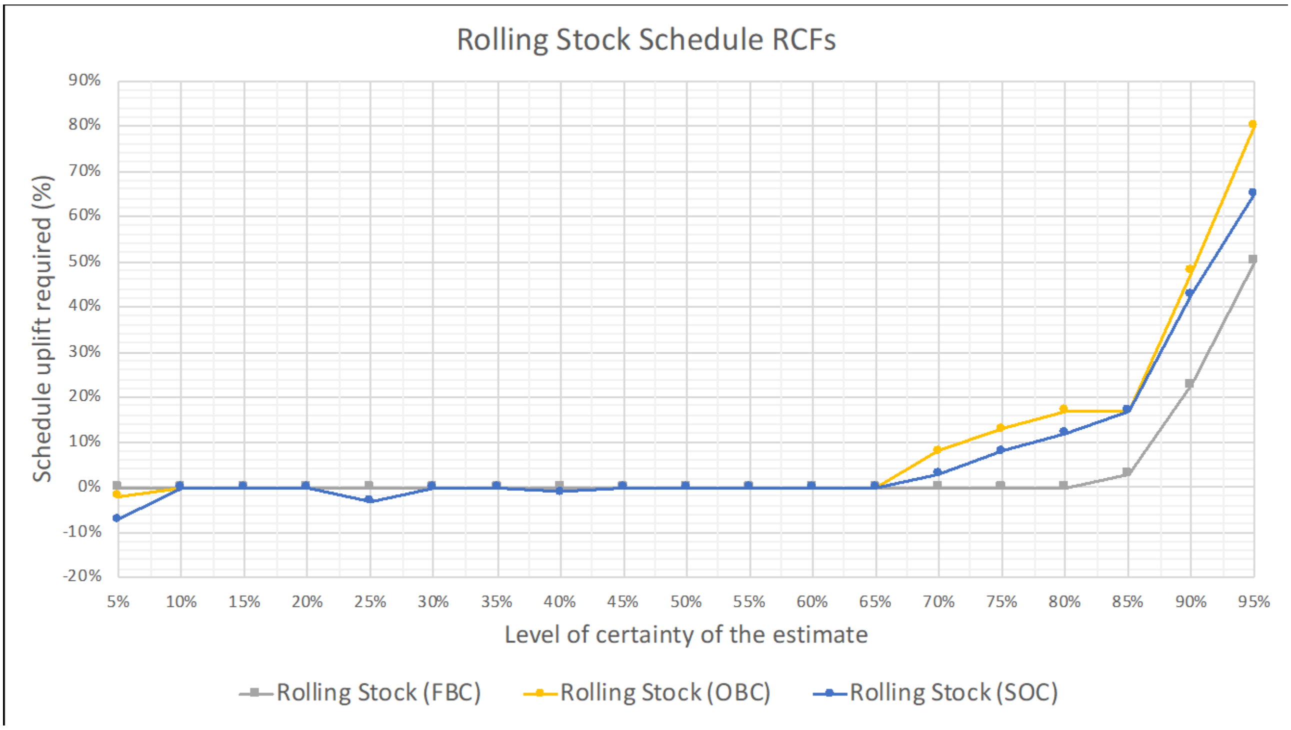

Buildings Schedule RCFs
Schedule uplift required (%)
250%
200%
150%
100%
50%
0%
-50%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Buildings (FBC)
Buildings (OBC)
Buildings (SOC)

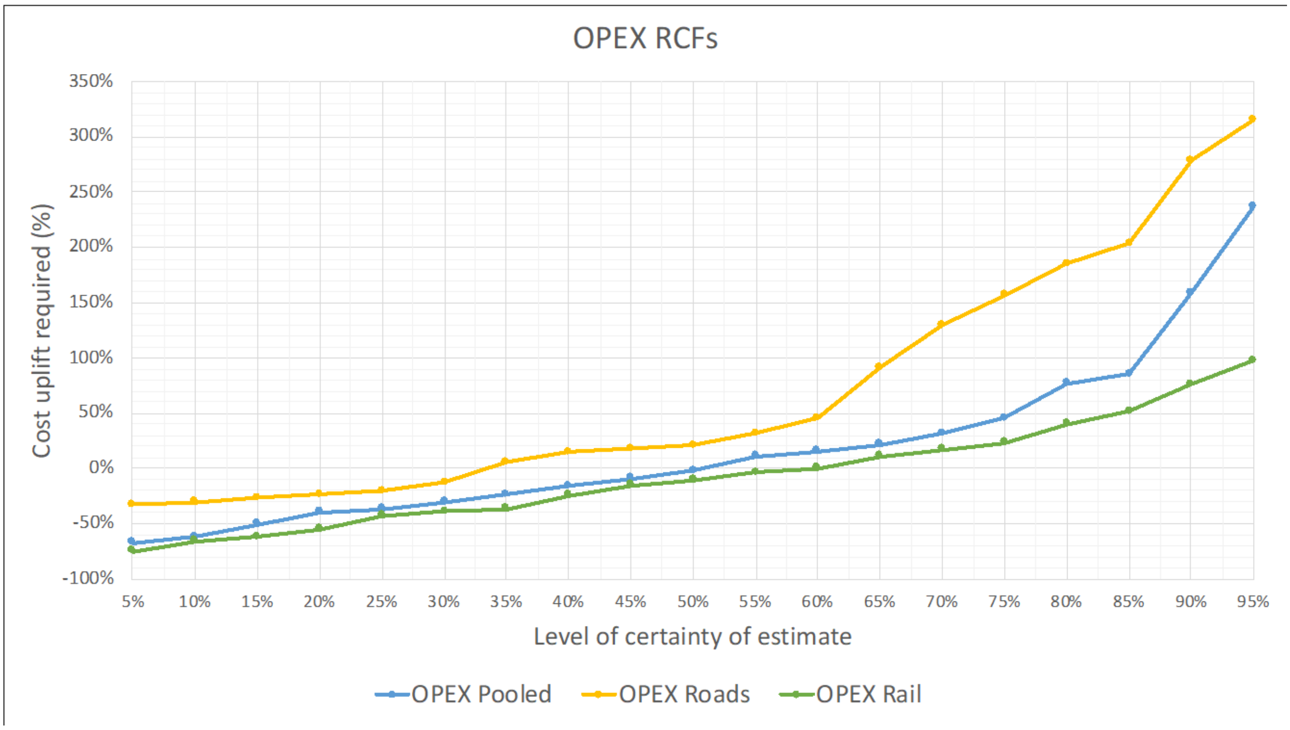

IT Cost RCFs
Cost uplift required (%)
900%
800%
700%
600%
500%
400%
300%
200%
100%
0%
-100%
-200%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
IT (FBC)
IT (OBC)
IT (SOC)

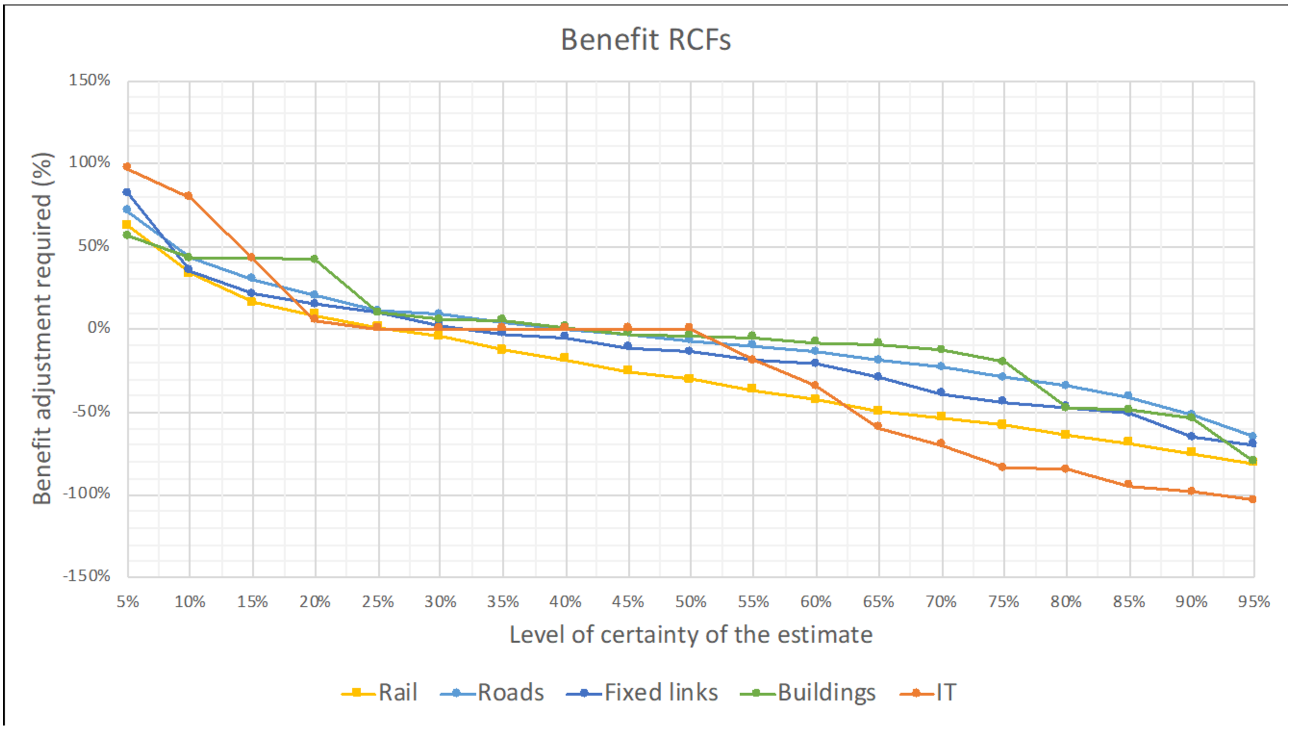

IT Schedule RCFs
Schedule uplift required (%)
250%
200%
150%
100%
50%
0%
-50%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
IT (FBC)
IT (OBC)
IT (SOC)

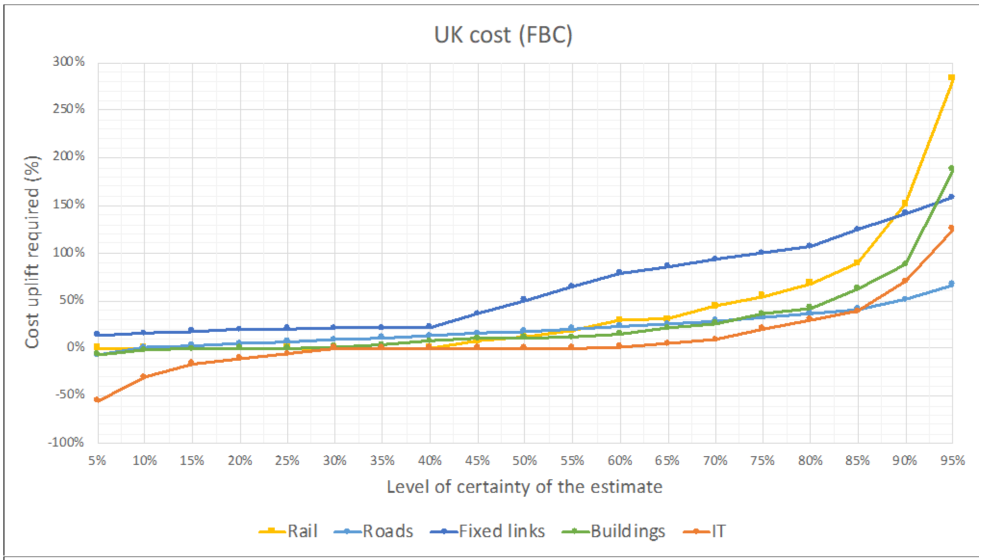

Land and property Cost RCFs
Cost uplift required (%)
350%
300%
250%
200%
150%
100%
50%
0%
-50%
-100%
-150%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Land and property (FBC)
Land and property (OBC)
Land and property (SOC)

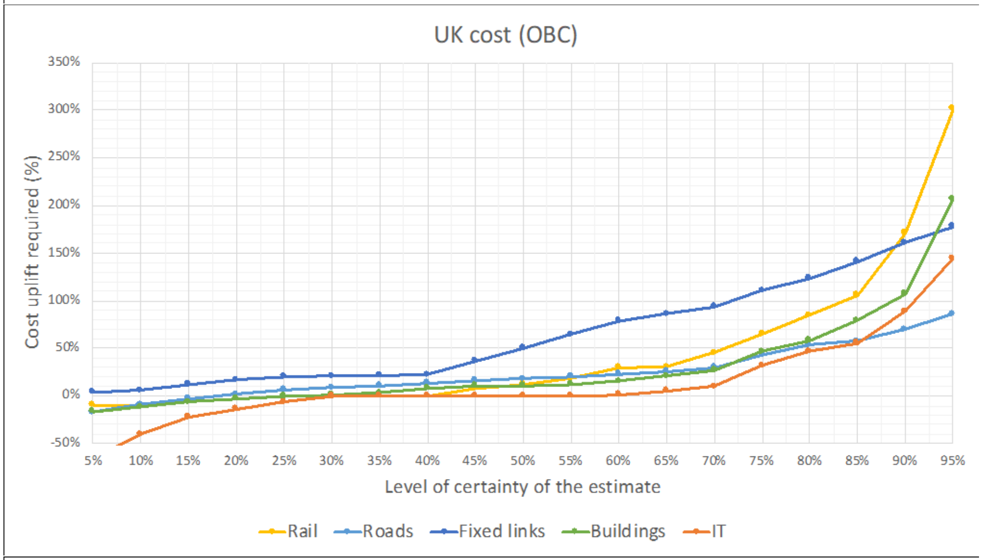

Rolling Stock Cost RCFs
Cost uplift required (%)
350%
300%
250%
200%
150%
100%
50%
0%
-50%
-100%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rolling Stock (FBC)
Rolling Stock (OBC)
Rolling Stock (SOC)

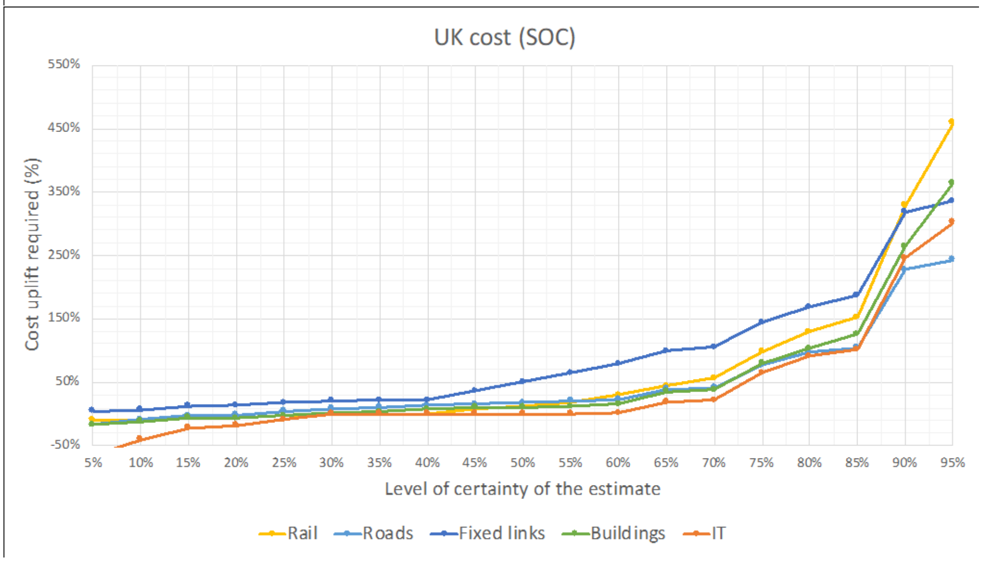

Rolling Stock Schedule RCFs
Schedule uplift required (%)
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
-10%
-20%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rolling Stock (FBC)
Rolling Stock (OBC)
Rolling Stock (SOC)

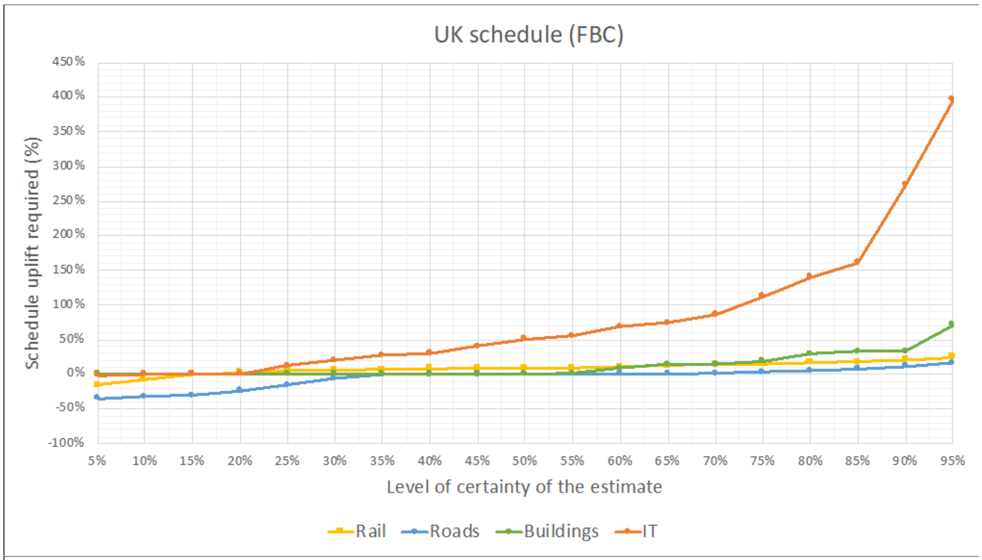

OPEX RCFs
Cost uplift required (%)
350%
300%
250%
200%
150%
100%
50%
0%
-50%
-100%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of estimate
OPEX Pooled
OPEX Roads
OPEX Rail

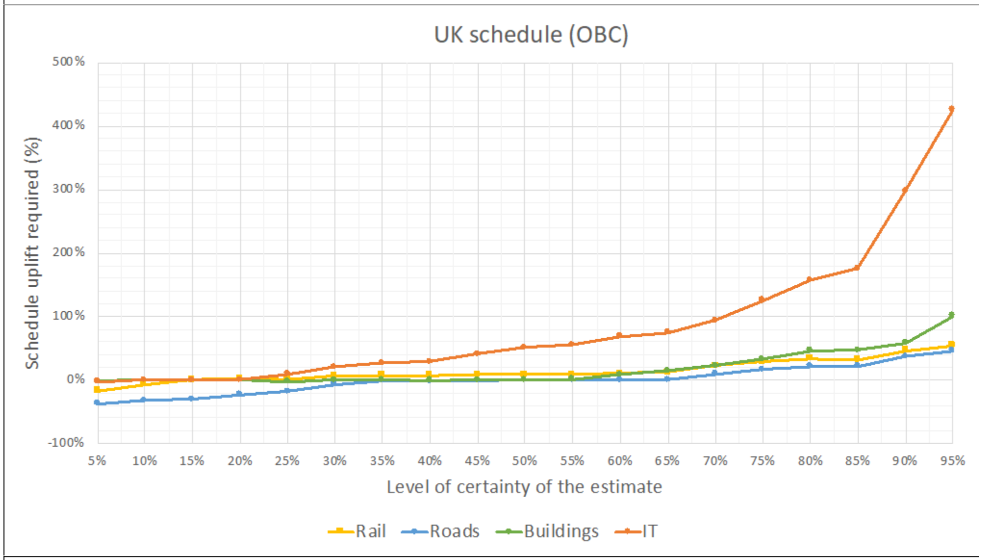

Benefit RCFs
Benefit adjustment required (%)
150%
100%
50%
0%
-50%
-100%
-150%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rail
Roads
Fixed links
Buildings
IT

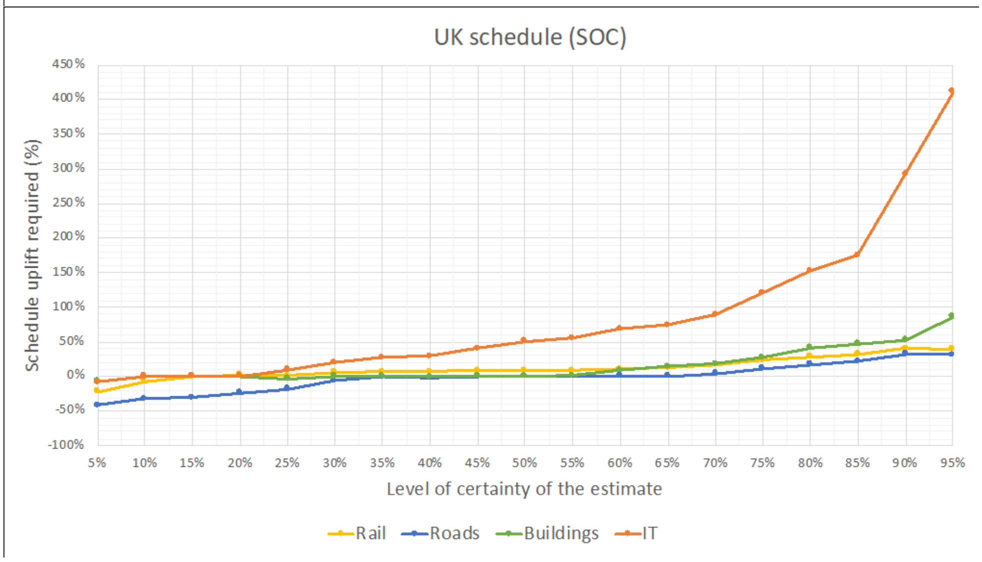

UK cost (FBC)
Cost uplift required (%)
300%
250%
200%
150%
100%
50%
0%
-50%
-100%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rail Roads Fixed links Buildings IT

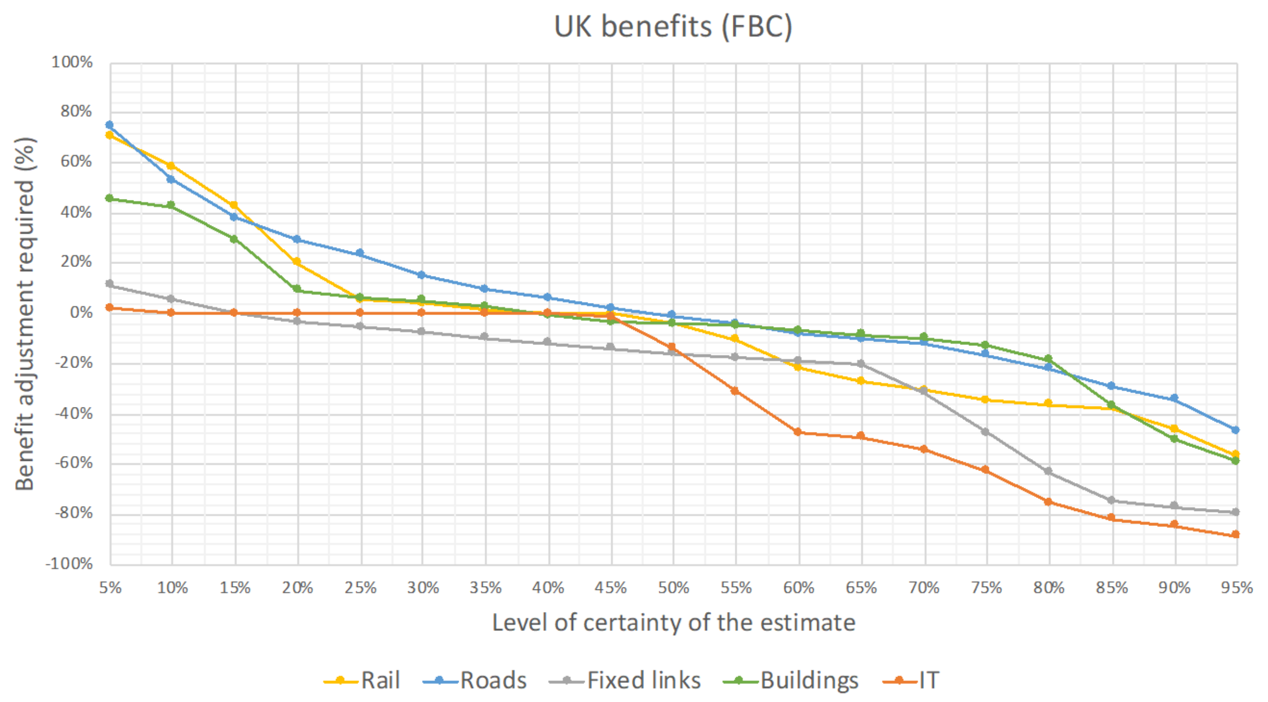

UK cost (OBC)
Cost uplift required (%)
350%
300%
250%
200%
150%
100%
50%
0%
-50%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rail Roads Fixed links Buildings IT


UK cost (SOC)
Cost uplift required (%)
550%
450%
350%
250%
150%
50%
-50%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rail Roads Fixed links Buildings IT


UK schedule (FBC)
Schedule uplift required (%)
450%
400%
350%
300%
250%
200%
150%
100%
50%
0%
-50%
-100%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rail Roads Buildings IT


UK schedule (OBC)
Schedule uplift required (%)
500%
400%
300%
200%
100%
0%
-100%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rail Roads Buildings IT


UK schedule (SOC)
Schedule uplift required (%)
450%
400%
350%
300%
250%
200%
150%
100%
50%
0%
-50%
-100%
5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95%
Level of certainty of the estimate
Rail Roads Buildings IT


UK benefits (FBC)
Benefit adjustment required (%)
100%
80%
60%
40%
20%
0%
-20%
-40%
-60%
-80%
-100%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
55%
60%
65%
70%
75%
80%
85%
90%
95%
Level of certainty of the estimate
Rail
Roads
Fixed links
Buildings
IT

APPENDIX B: OGP DATA IN THE OPEX REFERENCE CLASS

Project name	Type	Country
A14 Orgeval-La Defense, 1988-1996	Road	France
A16 L'Isle Adam-Boulogne, 1992-1994,1997,1998	Road	France
A19 Sens-Courtenay, 1993-1997	Road	France
A26 Chalons-Troyes, 1989-1992	Road	France
A29 Saint Saens-A13, excluding Pont de Normandie, 1991-1998	Road	France
A39 Dijon-Dole et Dole-Bourg en Bresse	Road	France
A837 Saintes-Rochefort, 1991-1997	Road	France
Atlanta MARTA metro, 1971-1986	Road	United States
Atlanta North Line Extension	Rail	United States
Baltimore BWI, Hunt Valley, Penn Station LRT extensions, 1991-1997	Rail	United States
Baltimore Central LRT Double Tracking, 2000-2006	Rail	United States
Beijing-Tianjin-Tanggu Expressway, 1987-1993	Rail	China
Buffalo LRT, 1977-1986	Rail	United States
Chengdu-Nanchong Expressway, 1998-2002	Rail	China
Chicago Metra North Central, 1998-2006	Rail	United States
Chicago Metra Southwest Corridor, 1998-2006	Rail	United States
Chicago Metra UP West, 1998-2006	Rail	United States
Chicago Orange Line SW Transitway	Rail	United States
Chingquing-Guizhou Leichong Expressway, 2000-2005	Road	China
Dallas North Central LRT Extension, 1996-2002	Rail	United States
Dallas South Oak Cliff LRT, 1990-1996	Rail	United States
Denver Southeast LRT Corridor, 1997-2006	Rail	United States
Denver Southwest LRT, 1995-2000	Rail	United States
Detroit DPM, 1980-1987	Rail	United States
Guang-Mei-Shan Railway, 1992-1995	Road	China
Hailar-Manzhouli Highway, 2005-2007	Road	China
Hudson-Bergen light rail MOS-1 & MOS-2, New Jersey	Rail	United States
Humber Bridge, 1967-1981	Road	United Kingdom
Jacksonville Skyway Express ASE AGT, 1982-2000	Rail	United States
Ji-Tong Inner Mongolia Line, 1989-1995	Road	China
Jinan-Qingdao Highway, 1989-1993	Rail	China
Jiujiang-Jingdezhen. Jiangxi Province Expressway, 1996-2002	Road	China
Laoyemiao-Jining Highway	Road	China
Liaoning Expressway, 1995-1998	Road	China
Liupanshui-Shuicheng Guizhou Shuibai Line, 1998-2004	Rail	China
Longgang-Tanxi Highway, 1992-1996	Rail	China

Los Angeles Red Line, 1983-2000	Rail	United States
Manila Metrostar Express, 1996-1999/2000 (MRT-3)	Bridge	Philippines
Manila MRT 2, 1995-2004	Rail	Philippines
Memphis Medical Center LRT Extension, 1997-2004	Rail	United States
Miami DPM, 1980-1986	Rail	United States
Miami metro, 1978-1985; correct: 1977-1985	Rail	United States
Miami Omni and Brickell AGT extensions, 1987-1994	Rail	United States
Miami South Florida Tri-Rail Double Tracking Segment 5, 1998-2007	Rail	United States
Minneapolis, Twin Cities Hiawatha LRT Corridor	Rail	United States
Newark Elizabeth Rail Link MOS-1, 1997-2006	Rail	United States
Portland Interstate MAX LRT Extension, 1998-2004	Rail	United States
Portland LRT, 1978-1987	Rail	United States
Qingxian-Wuqiao Expressway, 1998-2001	Rail	China
Sacramento LRT, 1981-1987	Rail	United States
Sacramento South LRT Phase 1, 1996-2003	Rail	United States
Salt Lake City I-15/State St South LRT, 1990-1999	Rail	United States
Salt Lake City Medical Center LRT extension	Rail	United States
Salt Lake City University & Medical Center LRT Extensions1997-2003	Rail	United States
San Diego El Cajon LRT extension, 1985-1989	Rail	United States
San Diego Mission Valley East LRT Extension, 1997-2005	Rail	United States
San Francisco BART to SFO Extension, FTA 2008, 1995-2003	Rail	United States
San Francisco Colma BART station, 1988-1996	Rail	United States
San Jose Guadalupe corridor LRT, 1981-1991	Rail	United States
San Jose-Tasman West LRT, 1991-1999	Rail	United States
San Juan Tren Urbano, Puerto Rico, FTA 2008, 1995-2005,	Rail	United States
Sanyuan-Tongchuan Highway, 1987-1993	Road	China
Shanghai-Nanpu Bridge, 1989-1991	Road	China
SHANGHAI-ZHEJIANG HIGHWAY PROJECT	Road	China
St, Louis Metrolink initial LRT system, 1984-1993	Rail	United States
St, Louis St, Clair Metrolink LRT extension, GAO 98, 1995-2001	Rail	United States
Stockholm Arlanda airport link (1995- 1999)	Road	Sweden
Urumqi-Kuitun Highway, 1996-2001	Road	China
Washington Largo Metrorail Extension, 1996-2004	Rail	United States
Washington metro, na-1985	Rail	United States
Xian-Hefei Railways, 2000-2004	Road	China
XINJIANG HIGHWAY PROJECT	Bridge	China
Yuanjiang-Mohei Expressway, 1998-2003	Road	China
Zhangzhou-Zhaoan Expressway	Rail	China

APPENDIX C: METHODOLOGY REVIEW OF REFERENCE CLASS FORECASTING

METHODOLOGY REVIEW

REFERENCE CLASS FORECASTING (RCF)

October 2019

SUMMARY

This report has reviewed the developments in RCF and its use in providing accurate forecasts. Based on several studies using RCF in various industries including hydropower dam projects, building projects, chemical projects and wind farms, RCF has shown to result in more accurate estimates than using conventional methods. The key findings are as follows:

- Application of RCF to Bujagali hydropower dam project resulted in a more reliable cost estimate and increased the accuracy of the cost-benefit analysis (Awojobi and Jenkins, 2016).
- A study of 420 building projects in Turkey revealed improved cost forecast accuracy when using RCF (Bayram and Al-Jabouri, 2016a).
- A study of 420 building projects in Turkey showed RCF provided the most accurate forecasts in the early stages of the project (Bayram and Al-Jabouri, 2016b).
- Based on samples of nine and ten offshore wind farms in the United Kingdom respectively, using RCF increases the probability of delivering a project on time and on budget (Koch and Sondergaard, 2010; Koch, 2012).

The effectiveness of RCF depends on the similarity of the reference class. If the project fits well into the reference class, the resulting uplift from the RCF will provide a more reliable estimate of the cost of the project (Awojobi and Jenkins, 2016; Batselier and Vanhoucke, 2016). Moreover, the effectiveness of RCF is influenced by the size of the projects and the size of the reference class (Batselier and Vanhoucke, 2016; Walczak and Majchrzak, 2018); projects need to be sufficiently large and the reference class should include enough projects. *Only if these criteria, similarity, project size, reference class size, are met will RCF outperform other methods.*

RCF provides more accurate forecasts than conventional methods, particularly in the early stages of project development. In the later stages other forecasting methods such as regression analysis could provide increased accuracy in forecasting compared to conventional methods as well (Batselier and Vanhoucke, 2016; Bayram and Al-Jibouri, 2016b). *Using a combination of forecasting methods throughout the life cycle of the project development is therefore recommended.*

Besides using RCF approaches in forecasting cost and schedule, several studies have compared the forecast accuracy of RCF with other methods:

- A study of 56 construction projects shows that RCF outperforms traditional methods of forecasting Earned-Value Management and Monte Carlo Simulation on three criteria: accuracy, stability and timeliness. (Batselier and Vanhoucke, 2016, 2017)
- At Sydney Water Corporation, based on 11 water infrastructure projects, risk-based estimation (RBE) *with* RCF resulted in more accurate cost estimates than projects that used conventional methods for cost estimation, and RBE *with* RCF significantly increased the likelihood of projects completing under budget (Liu and Napier, 2009)
- For infrastructure projects in Saudi-Arabia a risk-based cost contingency estimation model (RBCCEM) regressing project cost overrun on clusters of causes of cost overrun of past similar projects results in better forecast accuracy than conventional forecasting and RCF (Allaheim et al., 2015).

The latter method, RBCCEM can be considered hybrid estimation method as it uses the concept of RCF when examining the causes of cost overrun and adjusting the cost estimate accordingly.
Several other hybrid estimation approaches have been considered, combining RCF with conventional contingency based forecasting, combing RCF with QRA and Monte Carlo simulation, combining RCF with Baysian forecasting, and combining RCF with EVM and exponential smoothing approach. The main findings of the studies using a hybrid forecasting approach are as follows:

- At Australian State Road and Traffic Authority, based on a sample of 44 projects, the hybrid approach to forecasting (combining RCF with traditional contingency based forecasting) represents significant improvement in forecast accuracy against the conventional contingency approach but it reduces forecast accuracy compared to RBE (yet the sample characteristics are very different) (Liu, Wehbe and Sisovic, 2010).
- A bridge construction project using cost Bayesian adaptive forecasting produced more accurate forecasts than using the 'inside view' or 'outside view' methods separately (Kim and Reinschmidt, 2011).

- Based on a set of 8 car manufacturing plans, Bayesian forecasting with RCF produced more accurate forecasts than forecasts using RCF or statistical modeling separately. (Bordley, 2014)
- Based on a sample of 23 construction projects, integrating RCF with EVM and exponential smoothing forecasting approach (eXponential Smoothening-based method XSM) provides considerable more accurate predictions of cost and schedule performance than other EVM forecasting methods (Batselier and Vanhoucke, 2016, 2017)

Another methodological development was made by Salling and Leleur (2012, 2017) who proposes the method Reference *Scenario* Forecasting. This method combines RCF with Quantitative Risk Analysis (QRA) based on Monte Carlo simulation and exploratory scenarios. It provides probabilities of achieving certain BC ratios and as such can support investment appraisal decisions.

Besides methodological advancements, the experience with the forecasting method may impact its results. A study of 399 political forecasters shows that those with probabilistic-reasoning training and practice in using RCF had higher forecasting accuracy than those who did not (Chang, Chen, Mellers, Tetlock, 2016).

To conclude, based on the review of methodological developments of RCF we conclude *RCF is still a valid and best practice approach in forecasting particularly in the early stages of project development.* Best results in forecasting accuracy overall will be achieved by *combining the bottom-up and top-down methods*, particularly RBE and Bayesian forecasting, combining the 'outside view' with 'inside view' approaches (Kim and Reinschmidt, 2011; Koch, 2012; Leleur et al., 2015).

1. INTRODUCTION

Common traditional project forecasting methods include Monte Carlo simulation and Earned Value Management (EVM). The use of these methods have led projects to experience large cost overruns and schedule delays. One of the main explanations for this is optimism bias, the tendency to be overly optimistic about future actions, resulting in underestimation of cost and schedule. Due to optimism bias project owners may be ignorant or underestimate the risk/uncertainties in estimates. Optimism bias is the result of taking an 'inside view', focusing on the project at hand and estimating costs and duration of activities bottom-up. Traditional forecasting techniques typically take an 'inside view', they include a fixed contingency to the project cost estimate to account for risk and uncertainty in cost estimation, often 10% of the estimated cost. However, this method is considered to be biased because of the arbitrary way of deciding on the contingency amount (Liu et al., 2010). Instead, Reference Class Forecasting is an estimating approach that deals with optimism bias by taking an 'outside view' in determining the contingency amount that is based on statistical modelling of similar projects. Monte Carlo simulation can be considered as a 'semi outside view' because even though it makes use of historical data, it still relies on assumptions from the project manager to construct the distributional information (Batselier and Vanhoucke, 2016).

Since the Optimism Bias Guidance was published in 2004 the risk management profession has improved their approaches and tools for Quantitative Cost Risk Analyses (QRA).

This report aims to answer the following three main questions:

1. What have been the developments of Reference Class Forecasting since the publication in 2004 of "Procedures for Dealing with Optimism Bias in Transport Planning. Guidance Document" by The British Department for Transport?
2. How are bottom-up (QRA) and top-down (RCF) estimation working together?
3. How can we handle inflation above GDP deflators?

Chapter 2 will provide a review of the methodological developments in RCF addressing the first two questions. Chapter 3 will address how inflation is being handled in the construction sector.

2. Review of Methodological Developments in RCF

Use of Reference Class Forecasting

Reference Class Forecasting (RCF) has been applied in various industries including hydroelectric dams, buildings, chemical projects, and wind farms. This section discusses the main findings of the studies that have used RCF.

A study by Awojobi and Jenkins (2016) apply RCF to estimate cost risks of hydroelectric dams. They use maximum likelihood estimators to generate three different probability distributions covering three main development regions. They find the resulting cumulative probability distribution function fits better than those used in other studies (e.g. Sovacool et al., Ansar et al.).

Application of RCF to Bujagali hydropower dam project resulted in a more reliable cost estimate and it increased the accuracy of the cost-benefit analysis.

Based on a databased of 420 completed building projects in Turkey, Bayram and Al-Jobouri (2016a) compare cost estimates (based on contract sums) produced using the conventional method against cost forecasts using RCF. They use 75% of the total data to determine the probability distribution, establish an optimism bias curve, and develop the reference class. The other 25% is used to test the RCF method to predict the projects' final actual costs.

To compare the accuracy of the RCF with the conventional method three measures are used, root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The magnitude of error differs depending on which of these measures was used but overall 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, 2016). Bayram and Al-Jibouri (2018) find similar results.

A study of 420 building projects in Turkey revealed improved cost forecast accuracy when using RCF.

Moreover, Bayram and Al-Jibouri (2016b) compare the forecasts of traditional and non-traditional approaches in building projects in Turkey. The three traditional methods include unit area costs (UAC), client detailed costs (UPA), and contract sums (CS). The five non-traditional methods include multilayer perceptron (MLP), radial basis function (RBF), grid partitioning algorithm (GPA), reference class forecasting (RCF), and regression analysis (RA).

Again using 75% of the data traditional forecasting models were established, while the remaining 25% of the data were used to test the forecast accuracy of the five non-traditional approaches. Regarding RCF, the client detailed costs required higher uplift values compared with unit area costs indicating a higher forecast accuracy for unit area costs. Considering all error measures, in general, the study shows that based on unit area costs (carried out in the predesign stage), RCF produced more accurate forecasts. 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 postdesign stage).

A study of 420 building projects in Turkey showed RCF provided the most accurate forecasts in the early stages of the project.

Based on a sample of 65 chemical projects in one company completed between 2012-2014 Walczak and Majchrzak (2018) use reference class forecasting for small projects whereas typically RCF is used for large projects. They conclude that RCF is only suitable for large homogenous projects where it is possible to separate uniform groups of projects, while the usability of the method is limited in real business environment implementing a limited number of project types.

The effectiveness of RCF is influenced by the size of the projects and the reference class; projects need to be sufficiently large and the reference class should include a sufficient number of projects.

Based on a sample of nine offshore wind farm projects in the United Kingdom, Koch and Sondergaard (2010) use reference class forecasting to estimate the cost and duration for the London Array wind farm. They conclude that there is a much higher probability of delivering the project on time and budget if the budget is increased with 15% and the schedule extended with 30%. Moreover Koch (2012) argue that 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, p618). Using this socio-technical perspective may result in smaller more narrow reference classes. Based on a sample of ten offshore wind farms in the United Kingdom, Koch (2012) recommend to increase the budget by 35-40% and extend the schedule by 30% to obtain higher level of probabilities of meeting budget and schedule of 70% and 60% respectively.

Based on samples of nine and ten offshore wind farms in the United Kingdom respectively, using reference class forecasting increases the probability of delivering a project on time and on budget. Koch and Sondergaard (2010) and Koch (2012)

Using a probabilistic-reasoning training module Chang et al. (2016) investigate debiasing to improve probability judgements. They investigate four factors as possible explanations for the probabilistic-reasoning training effects, i.e. targeted practice, cognitive ability, cognitive motivation, and training. They found that subjects trained to apply probabilistic reasoning principles will be more accurate than subjects who do not. Training had the highest explanatory power for accuracy, followed by targeted practice, and cognitive ability.

Probabilistic reasoning training boosted forecasting accuracy, with the use of comparison classes and base-rates having the highest contribution highlighting the benefits of adopting an outside view.

A study of 399 political forecasters shows that those with probabilistic-reasoning training and practice in using RCF had higher forecasting accuracy than those who did not.

Zarikas and Kitsos (2015) suggest two improvements to the current RCF method. First, instead of using confidence interval to estimate the percentile that corresponds to a risk of cost overrun, they recommend the use of tolerance intervals. Secondly, instead of using conventional simple linear regression they recommend the use of the proposed best fitting polynomial according to the criterion of best prediction. The new model used in RCF is then “the one that best predicts on an average the future value, which lie in a certain interval with some probability. This is achieved using beta expected tolerance regions. So, while the regression oriented prediction is based on the extrapolation or interpolation of the best model fitting the data, the proposed method is based on a probabilistic reasoning and provides the model which best predicts the next value within experimental region” (Zarikas and Kitsos, 2015, p6).

Earned Value Management (EVM)

Batselier and Vanhoucke (2016) compare the traditional project forecasting methods Monte Carlo simulation and Earned Value Management (EVM) with Reference Class Forecasting based on a set of 56 construction projects.

For the Earned Value Management technique, two methods were used, EAC-1 (estimated cost at completion) with a project performance factor (PF) assuming the future cost performance will be according to plan, and EAC-CPI with a PF assuming the future cost performance will be equal to current cost performance. Similarly two methods were used for time forecasting, ESM-1 (earned schedule method) with a PF assuming the future time performance will be according to plan, and ESM-SPI(t) with a PF assuming the future time performance will be equal to the current time performance (Batselier and Vanhoucke, 2016).

For the Reference Class Forecasting technique, they use four different RC compositions with different levels of similarity (broad sector level to specific company level RC). The comparison of methods is based on the mean absolute percentage error (MAPE) measure, with lower MAPE values indicating higher levels of accuracy. Forecasting quality is assessed on accuracy, stability and timeliness.

On all three of these criterion RCF method outperforms the other two traditional methods. Specifically, RCF with the highest level of similarity (the most specific reference class of projects from the same company) is the most accurate cost forecasting method. It is also showing the greatest forecasting stability due to the use of one fixed constant pre-project forecast throughout the entire project, rather than using updated forecasts based on actual progress data. RCF also clearly outperforms EVM on timeliness criterion; RCF estimates are more accurate in the early stages compared to other techniques, and this is important as they allow adequate corrective actions to be taken in a timely manner.

Batselier and Vanhoucke (2016) conclude that for both cost and time forecasting RCF has a strong performance. However, the reference class level of similarity and size are important factors in the performance of RCF. Only if the reference class is of sufficient size and similarity will RCF outperform other techniques.

A study of 56 construction projects shows that RCF outperforms traditional methods of forecasting EVM and Monte Carlo Simulation on all three accuracy criterion of accuracy, stability and timeliness.

Risk-based Estimation (RBE)

One of the drawbacks for using RCF is the requirement of a sufficiently large dataset of similar projects. An alternative forecasting method that is able to adopt an outside view without the requirement of large dataset of historical data is Risk-based Estimation. RBE is a method that has been increasingly adopted for forecasting in large projects.

RBE is said to be able to adopt the outside view through a collective decision-making process. RBE models a probabilistic distribution of *component costs* of a project. It quantifies the base estimates and risk contingencies for each project component (rather than for the total project cost as in the conventional contingency approach and RCF). This quantification is based on the reference class concept because it uses previous projects' performance data. After quantification RBE runs a Monte Carlo simulation to produce the probability distribution function of the entire project costs by aggregating the component probability distribution functions (Chau, 1995 in Liu and Napier, 2006).

Liu and Napier (2009) test the effectiveness of RBE by comparing the cost estimates of a set of 11 water infrastructure projects estimated using RBE with RCF with a sample of 30 construction projects estimated using the conventional approach. Cost estimation accuracy was measured in two dimensions: i) *closeness*: percentage variation from the estimate measured as the percentage difference between the final contract amount (construction costs plus variation) and the contract award amount (base estimate plus contingency, and ii) *consistency*: the degree of dispersion in the sample around the mean measures by the variance or standard deviation.

The mean percentage variation is -3.5% and 5.2% for the RBE with RCF and conventional samples respectively. Regarding the dispersion of variation the standard deviations are 5.9% and 9.6% for the RBE with RCF and conventional samples respectively. This shows that on both dimensions of closeness and consistency RBE with RCF performs better than the conventional method, thus resulting in higher cost estimate accuracy. Moreover, Liu and Napier (2009) found that projects using RBE with RCF appear more likely to be under-budget while the ones using the conventional approach appear more likely to be over-budget.

At Sydney Water Corporation, based on 11 water infrastructure projects, RBE with RCF resulted in more accurate cost estimates than projects that used conventional methods for cost estimation, and RBE with RCF significantly increased the likelihood of projects completing under budget.

Another study that uses RBE method in forecasting is by Allaheim et al. (2015). They argue that the actual impact of cost risk of past similar projects could produce more accurate cost contingency estimates. They identify four clusters or groups of causes of cost overruns and suggest to include these in the forecasting method similar to the reference class concept. Specifically they develop a risk-based cost contingency estimation model (RBCCEM) which regresses the project cost overrun on the four clusters of causes. A bootstrapping variant of RBCCEM was also estimated. Comparing these models with the conventional forecasting using a fixed contingency of 10% and the method of reference class forecasting, and using standard measures of forecast accuracy (error indices MAPE, MAE, MSE, and RMSE), Allaheim et al. (2015) conclude RBCCEM outperforms the conventional and RCF methods on dimensions of error indices, adjusted cost overrun, and variance.

Considering the adjusted cost overrun both RBCCEM and RCF reduce overrun significantly. However, one of the main differences between the forecasting techniques is that "RCF estimates are subject to the acceptable risk of cost overrun, whereas RBCCEM are not." (Allaheim et al., 2015, 1299)

Similar to the findings of Liu and Napier, projects using RBCCEM tends to underrun budget while projects using the RCF model tends to overrun budget. (Allaheim et al., 2015)

For infrastructure projects in Saudi-Arabia a risk-based cost contingency estimation model (RBCCEM) regressing project cost overrun on clusters of causes of cost overrun of past similar projects results in better forecast accuracy than conventional forecasting and RCF (Allaheim et al., 2015).

The RBCCEM can be considered a hybrid approach as it uses the concept of reference class forecasting when examining the causes of cost overrun and adjusting the cost estimate accordingly. Risk-based Estimation can also be considered a hybrid approach because it combines probabilistic models with reference class forecasting. Several other studies have proposed hybrid estimation approaches which will be discussed in the next section.

Hybrid Estimation Approaches

Liu, Wehbe, and Sisovic (2010) argue that organisations develop these hybrid approaches to tailor generic estimating approaches to their operating environment. These hybrid approaches often combine reference class forecasting with conventional approaches.

Liu et al. (2010) compare the accuracy of the hybrid estimation approach (combining RCF with traditional contingency based forecasting) with i) historical data reported by other studies (which typically use the fixed contingency approach), ii) accuracies in the conventional approach, and iii) accuracies in the RBE approach. They find that the hybrid approach represents significant improvement against the conventional contingency approach (mean estimation error is lower) but reduces accuracy compared to RBE (mean estimation error is higher). Moreover, the variance of the estimation error is higher in the hybrid sample again suggesting RBE is more accurate.

The results may be explained by the large variation in the project type that is used in the different samples. The RBE sample only contains 11 water infrastructure projects, the hybrid sample includes 44 road and traffic projects conducted by one organization, whereas the historical data (the conventional sample) includes road, toll roads, mining, rail, bridges, tunnels, World Bank, metro, and 'various' type of projects.

At Australian State Road and Traffic Authority, based on a sample of 44 projects, the hybrid approach to forecasting (combining RCF with traditional contingency based forecasting) represents significant improvement in forecast accuracy against the conventional contingency approach but it reduces forecast accuracy compared to RBE.

Salling and Leleur (2012, 2017) propose a new technique to forecasting called Reference Scenario Forecasting. This method combines RCF with Quantitative Risk Analysis (QRA) based on Monte Carlo simulation and makes use of a set of exploratory scenarios. The approach first performs a Cost Benefit Analysis (CBA) which provides a deterministic point estimate. Then this point estimated is transformed into an interval result through the Monte Carlo simulation, using random input parameters based on reference class forecasting. The result are scenario-based graphs that show the risk probability of achieving specific BC ratios. The advantage of this method is thus the ability to consider the probability of implementing a non-feasible project or not implementing a feasible one. Salling and Pryn (2015) further expand this method by including sustainable planning criteria in the assessment of projects. However, none of these three studies show whether this combined CBA-RCF method is superior to RCF in itself.

Reference Scenario Forecasting, which combines RCF with QRA based on Monte Carlo simulation and exploratory scenarios, provides probabilities of achieving certain BC ratios and can support investment appraisal decisions.

Another method that combines the QRA and CBA with RCF is proposed by Leleur et al. (2015). The method combines RCF with expert judgements based on overconfidence theory interpretation. The main advancement is the combination of outside view and inside view approaches in contrast to solely using the inside view as RCF does. The method replaces the point estimate benefit-cost ratio with a probability-based interval (referred to as certainty graph). The average or agreed 'consensus' values by experts on the provided Min and Max cost forecasts represents the 'inside view' forecast. However, overconfidence theory states that the ranges indicated by the experts are often too narrow and Leleur et al. (2015) therefore propose to calibrate the interplay between RCF and expert judgements, by modifying the probability distribution produced by the RCF.

Bayesian Updating

Kim and Reinschmidt (2011) suggest combining the inside and outside view of cost forecasting by using Bayesian inference and the Bayesian model averaging technique. They propose a probabilistic cost forecasting framework for Bayesian adaptive forecasting to incorporate the actual performance data from earned value management into the predictions and revise the pre-project cost estimates.

The cost Bayesian adaptive forecasting (cost-BAF) method is used to update the current estimate of cost at completion throughout the execution phase. Kim and Reinschmidt (2011, p960) explain "The inside-view prior estimate and the outside-view prior estimate are applied separately, in conjunction with the reported performance data. The predictions from the cost-BAF model using different pre-project estimates are combined to compute a combined posterior distribution using the Bayesian model averaging technique."

A prior estimate of project cost is based on pre-project planning or historical data (this can be both the inside-view estimate and the outside-view estimate), whereas the posterior estimate reflects the actual performance of the project. As the project progresses the impact from a prior estimate diminishes because more data on actual performance has become available. Earned value method may be able to revise predictions using actual performance data. This method to systematically update pre-project estimates with actual project performance data differentiates the cost-BAF method from conventional cost forecasting methods.

Based on a hypothetical bridge construction project, Kim and Reinschmidt (2011) test the forecasting accuracy of the cost-BAF with other methods. They compare the cost predictions from the inside-view, outside-view and cost-BAF and conclude that the combined cost-BAF predictions are more sensitive to actual performance data, that is, they provide more precise predictions.

A bridge construction project based on cost Bayesian adaptive forecasting produced more accurate forecasts than using the 'inside view' or 'outside view' methods individually.

Similar to Kim and Reinschmidt (2011) Bordley uses a Bayesian approach to forecasting in healthcare. The forecasting method combines reference class forecasting with model-based forecasting. The reference class forecast information is used to specify the Bayesian prior. The prior was then updated with model-based forecasts to generate a posterior probability. The Bayesian cost forecast is higher than the model-based forecast because higher actual costs are provided by the reference class forecasts and model-based forecasts are typically systematically underestimated. Bordley (2014, p221) found that the "Bayesian posterior forecast had lower variance (and lower forecast error) than either the model-based forecast or the reference-class forecast."

Based on a set of 8 car manufacturing plans, Bayesian forecasting with RCF produced more accurate forecasts than forecasts using RCF or statistical modeling separately.

eXponential Smoothening-Based Method (XSM)

With the traditional EVM forecasting technique one can either use an unweighted method (EAC-1 and ESM-1 for estimated cost at completion and earned schedule method with performance factor assuming the future cost/schedule performance will be according to plan) or a weighted method (CPI and SPI(t) for estimated cost at completion and earned schedule method with performance factor assuming the future cost/schedule performance will be equal to current performance). However, the unweighted method does not accurately account for two possible influences, i.e. the occurrence of natural performance improvement and the effect of corrective management actions during the course of the project.

Batselier and Vanhoucke (2017) propose a new technique, eXponential Smoothening-based Method (XSM) which deals with these limitations. It integrates the earned value management approach with the exponential smoothing forecasting approach. The exponential smoothening forecast approach makes it possible to assign higher weighting to more recent results. XSM uses one smoothing parameter which allows for anticipated changes in management. To set the smoothening parameter there are two approaches, the static and dynamic approach. In the static approach, the value of the smoothing parameter is chosen before the project starts and remains constant throughout the entire project. In the dynamic approach, the smoothing parameter can be adjusted every tracking period. The XSM technique allows the integration with RCF by using the data from the reference class of projects to inform the value of the smoothing parameter (static approach).

Batselier and Vanhoucke (2017) investigated four different XSM approaches, varying in the smoothing parameter β, using three static and one dynamic version and compared this with other forecasting methods. For the static approach, β is chosen before the project starts and remains constant throughout the entire project.

- β_{opt}: constant β value that produces the most accurate (optimal) forecast
- $\beta_{opt, oa}$: constant optimal β value over all projects in the database
- $\beta_{opt, oa, rc}$: constant optimal β value over all historical projects with the same characteristics as the considered project (this incorporates reference class forecasting concept into XSM)
- β_{dyn}: β is based on progress of the considered project itself. It is based on a quantitative analysis whereby β_{dyn} value is "calculated for every tracking period based on the performance of the past tracking periods"[8] (Vatselier and Vanhoucke, 2017, p42). A qualitative analysis can also be used to calculate the β_{dyn}.

The results show that based on MAPE comparison, XSM(t)-β_{opt} is 14.8% and 31.8% more accurate than forecasts obtained from ESM-1 and ESM-SPI(t) respectively. XSM(t)-β_{opt} is however difficult to apply in practice and

[8] Or the value for a certain tracking period is calculated as the β that would have produced the most accurate forecasts over all of the preceding tracking periods" (Batselier and Vanhoucke, 2017, p42).

therefore SXM-$\beta_{opt, oa}$ is preferred. Yet XSM(t)-$\beta_{opt, oa}$ results in a considerable reduction in accuracy but it is still more accurate than ESM-1. The quantitative dynamic approach does not improve accuracy; XSM(t)-β_{dyn} is similar to ESM-1 and while it is considerably more accurate than ESM-SPI(t) (19.6%), it is worse compared to XSM(t)-β_{opt}. The static approach can enhance performance of accuracy by considering reference classes. XSM(t)-$\beta_{opt, rc}$ obtains the highest forecast accuracy, with 13.9% improvement over the best EVM forecasting method ESM-1.

Cost forecasting show similar results to time forecasting when comparing XSM versions. XSM($)-$\beta_{opt}$ shows a substantial potential accuracy gain of 25.1% over the best EVM approach for cost forecasting EAC-CPI. XSM($)-$\beta_{opt, oa}$ again reduces accuracy compared to XSM($)-$\beta_{opt}$ but still remains more accurate than EAC-1 and EAC-CPI although the accuracy gain is low. Again XSM($)-$\beta_{opt, rc}$ has the largest forecast accuracy, with relative improvements of 22.2% and 22.6% over EAC-1 and EAC-CPI respectively.

The above comparisons clearly show that the XSM has great potential improving the accuracy of both cost and schedule forecast, especially when it incorporates the reference class component. XSM-$\beta_{opt, rc}$ outperforms other forecasting methods, with performance accuracy for costs being even more significant than for time.

Based on a sample of 23 construction projects, integrating RCF with EVM and exponential smoothing forecasting approach (eXponential Smoothening-based method XSM) provides considerable more accurate predictions of cost and schedule performance than other EVM forecasting methods (Batselier and Vanhoucke, 2016, 2017)

Conclusion

1. What have been the developments of Reference Class Forecasting since the publication in 2004 of "Procedures for Dealing with Optimism Bias in Transport Planning. Guidance Document" by The British Department for Transport?

Studies have provided more evidence of improved forecasting accuracy by using RCF over conventional forecasting techniques. The method has been recognized more widely with increased adoption of RCF in various industries and countries.

Most methodological developments are focused on combing RCF with other estimation techniques rather than making adjustments in RCF itself.

2. How are bottom-up (QRA) and top-down (RCF) estimation working together?

RCF has its strengths in the early stages of the project development, outperforming other methods, while other estimating methods also provide accurate forecasting results in the later stages.

The review shows combining both bottom-up and top-down estimation techniques provides the best results in forecasting accuracy. Current advancements combine RCF with Risk-based Estimation, Bayesian forecasting, and EVM with exponential smoothening forecasting. The review suggests that a first step is to conduct RCF and QRA side by side each based on the base cost estimate. This encourages projects and planners to investigate the gap between the two analyses and get an understanding of the completeness and ranges of their risk assessments. A second step might be to expand RCF and integrate hard distributional data in the bottom-up analysis. In practical terms, if all items on a risk register, including an item for unknowns and correlations between risks, are evaluated with unbiased and accurate data, i.e. RCF curves for each risk, the top-down RCF approach and the bottom-up QRA approach should reach similar conclusions.

Appendix D: Suggestions for Inflation treatment

Suggestions for Inflation Treatment

The reference class forecasts in this report have removed all inflation, i.e. used real-term cost estimates and real-term outturns to calculate overruns.

In some cases, projects will have to produce estimates that include inflation and therefore additional consideration needs to be given to any optimism bias included in inflation forecasts.

We took two different approaches to identify the optimism uplift required for estimates including inflation, often referred to as nominal estimates or year-of-expenditure estimates. The first approach applies the Reference Class Forecast logic to historical construction inflation in the UK. The second approach uses historic data on real-terms cost overruns and nominal cost overruns in past, completed UK projects.

Reference Class Forecast Based on Historical Inflation Indeces

In the UK several different measures are available for the forecast of inflation:

- Consumer price indices: CPI and CPIH issued by the Office of National Statistics (note: RPI is not considered to be a national statistic any longer);
- Construction Price Indices: in this context most notably the producer price index for infrastructure (previously issued by BEIS now ONS), general civil engineering cost index (issued by RICS), construction cost index (issued by Eurostat), tender price indices (issued by various entities e.g. RICS); and
- Government deflator: GDP deflator issued by the Office of National Statistics and prescribed HMT Greenbook.

The *consumer price indices* are not relevant for capital investment project appraisals.

Construction price indices exist for key industries in the UK. In the UK, the Building Cost Information Service (BCIS, n.d.) of RICS prepares four price adjustment indices:

1. Building;
2. Specialist engineering;
3. Civil engineering; and

4. Highways maintenance.

In addition, to the general civil engineering cost price index RICS also issues a *Tender Price Index*. Tender price indices suffer from the typical problem that tender prices are lowballed to win bids when markets are contracting. In theory tender prices should be a suitable measure, however, this theory rests on the assumption that markets are rational (i.e. that tenderers would not bid at loss-making prices).

The Office of National Statistics issues a *Construction Output Price Index* (OPI). Since 2015, this index has replaced the Construction Price and Cost Index (CPCI). Further changes to the OPI were made in 2017 to improve the quality of the measure. Unfortunately, this means that little historic data is available, the oldest quarterly measure of OPI is October 2014.

Eurostat issues a *Construction Cost Index*, which has the advantage that it is available for the EU28 and has a relatively long history (earliest 1980, earliest UK reporting period 1993). This measure follows the same methodology as the OPI.

The *government deflator* is a broad measure for the prices of all goods and services sold in an economy. The GDP deflator also includes the prices of investment goods, government services and exports, and excludes the price of UK imports. The wider coverage of the GDP deflator makes it more appropriate for deflating and inflating public expenditure series. The HMT Greenbook 2018 states that projects should use the GDP deflator to strip out inflation.

The comparison of the three key indices discussed above (General Civil Engineering Cost, GDP Deflator, Construction Cost) is shown in Figure 1. The historic data show a gap between the three series, where the GDP deflator shows less inflation than the other indices.

Department for Transport

OXFORD GLOBAL PROJECTS

FIGURE 1 COMPARISON OF THE BCIS GENERAL CIVIL ENGINEERING COST INDEX (GCECI), THE GDP DEFLATOR, AND THE EUROSTAT CONSTRUCTION COST INDEX (STS_COPI_Q), 1993=100

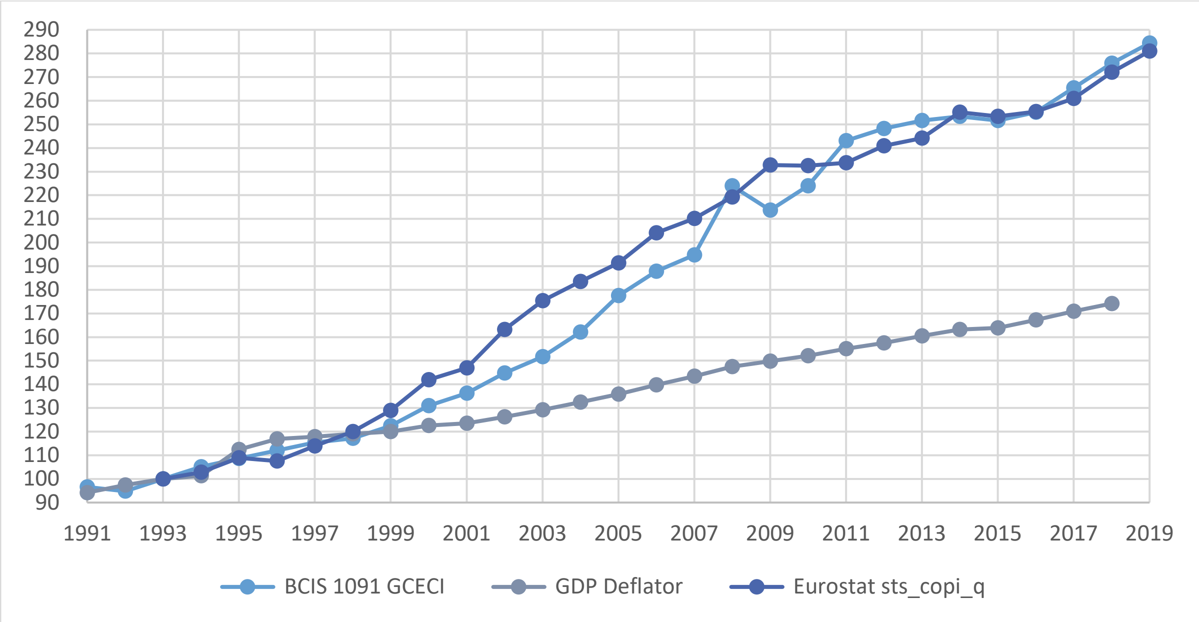


Figure 1 illustrates the timeseries of the different deflators. The gap between the different deflators is widening, because compound effects of small differences in the growth rates of inflation are amplified when the data are viewed with the reference index being this far in history.

FIGURE 2 COMPARISON OF THE BCIS GENERAL CIVIL ENGINEERING COST INDEX (GCECI), THE GDP DEFLATOR, AND THE EUROSTAT CONSTRUCTION COST INDEX (STS_COPI_Q), AND THE ONS CONSTRUCTION OUTPUT PRICE INDEX, 2018=100

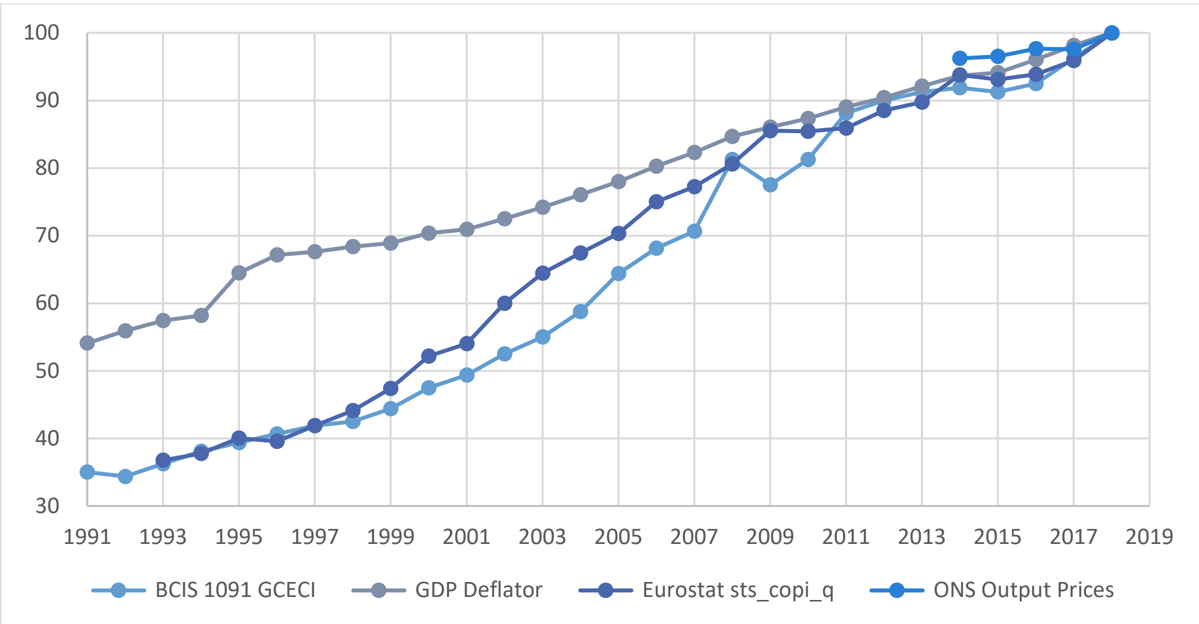


Figure 2 shows the same time series as Figure 1 but baselined at 2018, i.e. 2018=100. The figure shows that the GDP deflator and the industry specific construction cost indices trail more closely between 2008-2018.

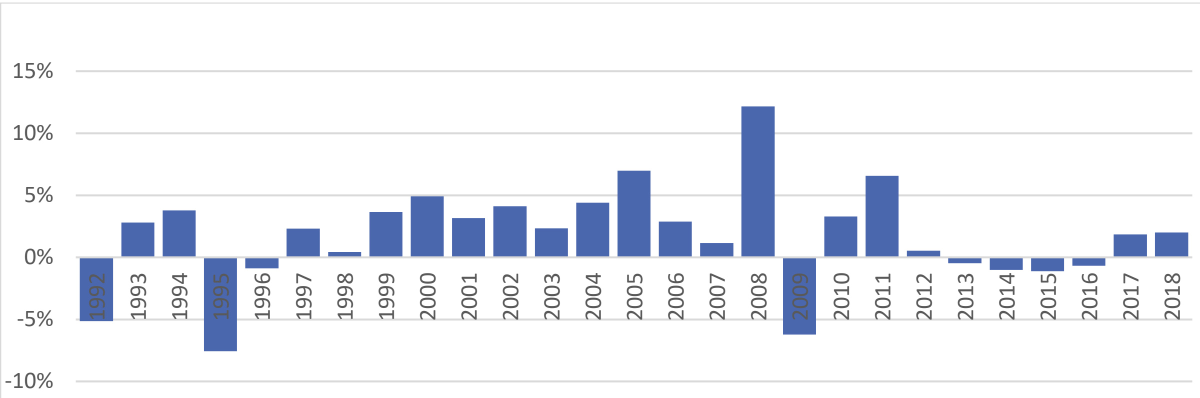
The long running mean of the annual GDP deflator growth in the UK is around 2% p.a.; for construction cost this is the case between 2008-2018, while for the full data series, i.e. 1993-2018, the inflation is about 4% p.a.

TABLE 1 COMPOUND ANNUAL GROWTH RATE OF GENERAL CIVIL ENGINEERING COST, CONSTRUCTION COST, AND GENERAL PRICE INFLATION (GDP DEFLATOR) FOR THE YEARS 1993-2018, 2008-2018, AND 2013-2018

Mean annual change (CAGR)	General Civil Engineering Cost Index	Construction Cost Index	GDP Deflator
1993-2018	4.0%	3.9%	1.9%
2008-2018	1.9%	2.0%	1.5%
2013-2018	1.5%	1.8%	1.4%

Figure 3 shows the difference of the inflation in civil engineering cost above and beyond the GDP deflator for the years 1992-2018. The difference does not follow any clear patterns. For example, economic recessions in the UK only happened between Q2 2008 and Q2 2009 during this period and in these two years civils cost inflation was higher and lower than the GDP deflator.

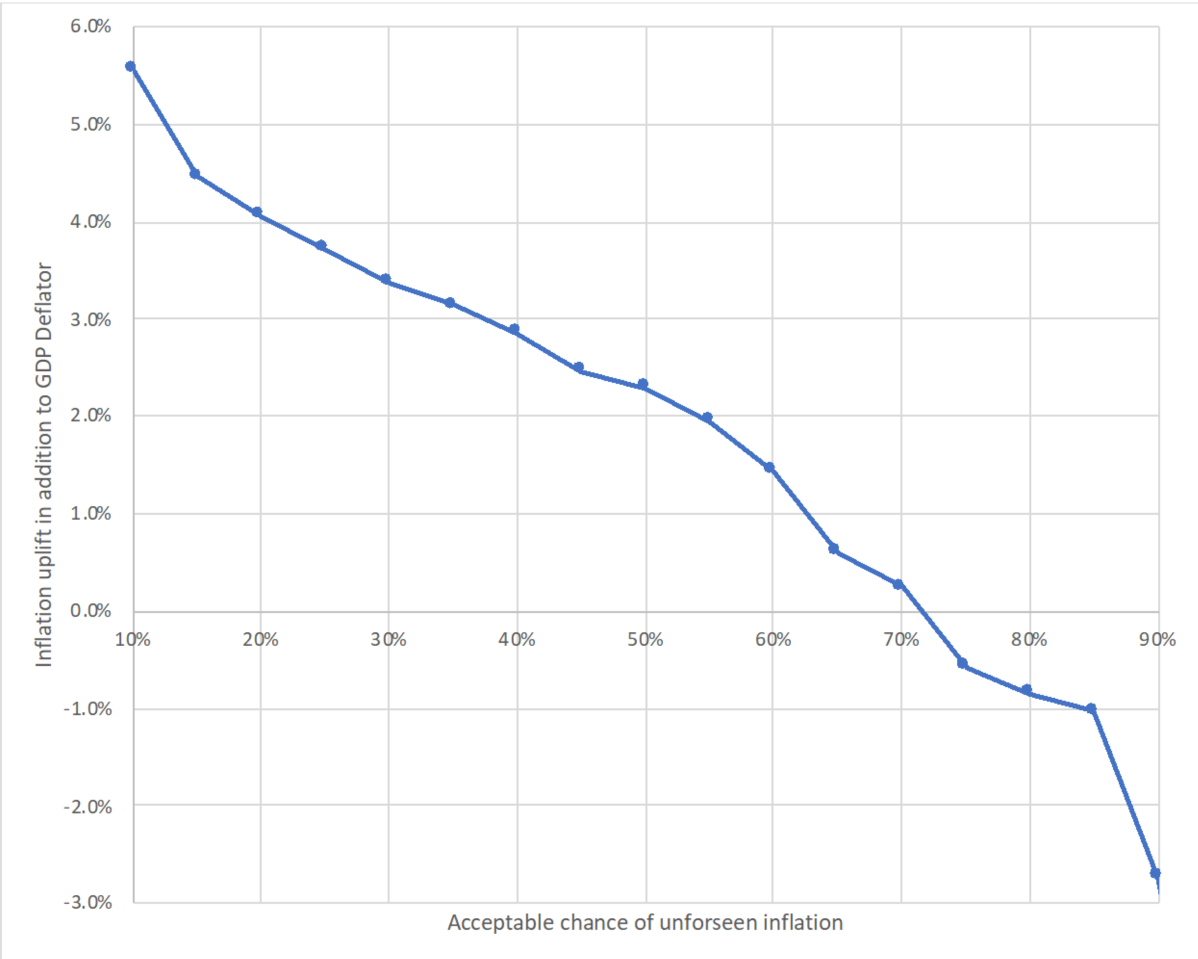
FIGURE 3 DIFFERENCE BETWEEN CIVIL ENGINEERING COST INFLATION AND GDP DEFLATOR



Thus, for the construction of the Reference Classes the GDP Deflator has been used to turn nominal outlays and nominal forecasts into real-term forecasts. This is in line with HMT Greenbook suggests that projects use GDP Deflators to express nominal values in real-term values (the HMT Greenbook allows projects to plan with inflation above and beyond GDP deflator), which is between 2% and 3.5% depending on the length of the forecast period.

A Reference Class Forecast can be used if a project wants to forecast additional inflation above and beyond GDP deflators for the full range of available data 1992-2018). This Reference Class Curve is shown in Figure 4. The most likely scenario (P50 = 50% acceptable chance of unforeseen inflation) is an uplift of 2.3pp on top of the GDP deflator and at P80 the uplift is approximately 4pp.

FIGURE 4 REFERENCE CLASS FORECAST FOR INFLATION UPLIFT NEEDED FOR THE GDP DEFLATOR FOR A GIVEN ACCEPTABLE CHANCE OF UNFORSEEN INFLATION



Reference Class Forecast of Inflation Based on Past UK Projects

The second approach to formulate a reference class forecast for the unanticipated inflation is based on 116 projects from the UK. This reference class is based on the difference of

Cost overrun in nominal terms – Cost overrun in real terms.

The cost overrun in nominal terms is calculated as

Actual cost in nominal terms / Estimated cost in nominal terms.

The cost overrun in real terms is calculated as

Actual cost in real terms / Estimated cost in real terms.

The difference in the overruns represents the additional optimism bias that stems from producing estimates in nominal terms above and beyond the optimism bias that is in the base cost estimate.

For 116 UK projects the both nominal and real-terms cost overruns were available. Figure 5 shows the difference between the nominal and real-terms overruns for these projects. The P50 estimate is a 4pp additional uplift; the P80 estimate is a 10pp uplift.

FIGURE 5 OPTIMISM BIAS OF INFLATION ESTIMATES IN UK PROJECTS (N=116)

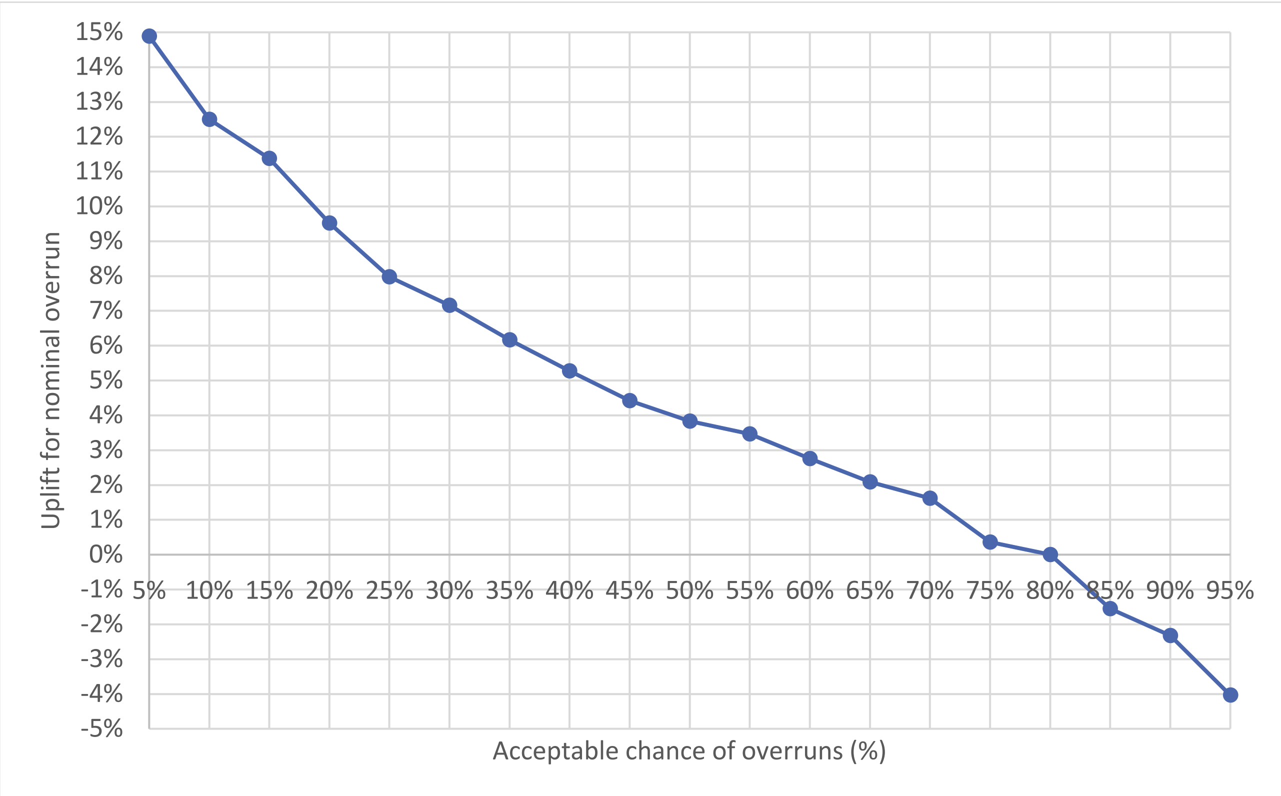


TABLE 2 REFERENCE CLASS FORECAST FOR INFLATION

Acceptable chance of overrun (%)	Provision for unexpected inflation
95%	-4%
90%	-2%
85%	-2%
80%	0%

75%	0%
70%	2%
65%	2%
60%	3%
55%	3%
50%	4%
45%	4%
40%	5%
35%	6%
30%	7%
25%	8%
20%	10%
15%	11%
10%	12%
5%	15%

To illustrate how these figures could be used the following provides a worked example. A rail project estimates the base cost at GBP 250m in nominal terms, i.e. including inflation as per HMT Greenbook. Then on top of the base cost the project needs to add an uplift for cost overruns from the reference classes (see the main part of the report). At FBC the uplift at P50 is 19% and for P80 the uplift is 60%. Thus, additional GBP 47.5m are required as Optimism Bias Uplift at P50 and GBP 150m are required at P80. The additional uplift for inflation is 4% at P50 and 10% at P80, thus GBP 10m and GBP 25m respectively (calculated on the base cost). In sum, the total uplift at P50 is GBP 57.5m and GBP 175m at P80, which brings the total adjusted estimate to 307.5m and 425m respectively.

Most major construction projects limit the exposure to inflation, e.g. through contractual clauses that regulate price adjustments by specifying a specific index or negotiating a specific annual percentage adjustment. Therefore, most projects are only openly exposed to inflation prior to contracting, which explains also the difference in the two RCF approaches. Thus, we recommend the following:

(1) If a project forecasts inflation only on the GDP deflators recommended in the Greenbook to establish a budget in nominal terms the risk of unforeseen inflation should be derived from the first RCF curve, which compare actual construction inflation against the GDP deflator.

(2) If a project has limited exposure to inflation through specialist technical forecasts or commercial strategies the second RCF curve should be used, which compares real-terms and nominal forecasts.

REFERENCES

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

Awojobi, O. and Jenkins, G.P. (2016), "Managing the cost overrun risks of hydroelectric dams: An application of reference class forecasting techniques", *Renewable and Sustainable Energy Reviews*, Elsevier, Vol. 63, pp. 19–32.

Batselier, J. and Vanhoucke, M. (2016), "Practical Application and Empirical Evaluation of Reference Class Forecasting for Project Management", *Project Management Journal*, Vol. 47 No. 5, pp. 36–51.

Batselier, J. and Vanhoucke, M. (2017), "Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting", *International Journal of Project Management*, Elsevier Ltd and Association for Project Management and the International Project Management Association, Vol. 35 No. 1, pp. 28–43.

Bayram, S. and Al-Jibouri, S. (2016a), "Application of Reference Class Forecasting in Turkish Public Construction Projects: Contractor Perspective", *Journal of Management in Engineering*, Vol. 32 No. 3.

Bayram, S. and Al-Jibouri, S. (2016b), "Efficacy of Estimation Methods in Forecasting Building Projects' Costs", *Journal of Construction Engineering and Management*, Vol. 142 No. 11.

Bayram, S. and Al-Jibouri, S. (2018), "Cost forecasting using RCF: a case study for planning public building projects costs in Turkey", *International Journal of Construction Management*, Taylor & Francis, Vol. 18 No. 5, pp. 405–417.

BCIS (n.d.), Inflation adjustment clauses.

Bordley, R.F. (2014), "Reference Class Forecasting: Resolving Its Challenge to Statistical Modeling", *American Statistician*, Vol. 68 No. 4, pp. 221–229.

Chang, W., Chen, E. and Mellers, B. (2016), "Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments", *Judgment and Decision Making*, Vol. 11 No. 5, pp. 509–526.

El Yamami, A., Ahriz, S., Mansouri, K., Qbadou, M. and Illoussamen, E.H. (2018), "Rethinking IT project financial risk prediction using reference class forecasting technique", *Proceedings of the 2018 International Conference on Optimization and Applications, ICOA 2018*, Mohammedia, Morocco.

Kim, B.S. and Reinschmidt, K.F. (2011), "Combination of Project Cost Forecasts in Earned Value Management", *Journal of Construction Engineering and Management*, Vol. 137 No. 11, pp. 958–966.

Koch, C. (2012), "Contested overruns and performance of offshore wind power plants", *Construction Management and Economics*, Vol. 30 No. 8, pp. 609–622.

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

Leleur, S., Salling, K.B., Pilkauskiene, I. and Nicolaisen, M.S. (2015), "Combining reference class forecasting with overconfidence theory for better risk assessment of transport infrastructure investments", *European Journal of Transport and Infrastructure Research*, Vol. 15 No. 3, pp. 362–375.

Liu, L. and Napier, Z. (2010), "The accuracy of risk-based cost estimation for water infrastructure projects: Preliminary evidence from australian projects", *Construction Management and Economics*, Vol. 28 No. 1, pp. 89–100.

Liu, L., Wehbe, G. and Sisovic, J. (2010), "The accuracy of hybrid estimating approaches: A case study of an Australian state road traffic authority", *Engineering Economist*, Vol. 55 No. 3, pp. 225–245.

Martin, J. (2018), "Giving a good account", *Construction Journal,* February/March 2018, pp. 14-15.

Salling, K.B. and Leleur, S. (2012), "Modelling of transport project uncertainties: Feasibility risk assessment and scenario analysis", *European Journal of Transport and Infrastructure Research*, Vol. 12 No. 1, pp. 21–38.

Salling, K.B. and Leleur, S. (2015), "Transport Project Evaluation: Feasibility Risk Assessment and Scenario Forecasting", *Transport*, Vol. 32 No. 2, pp. 180–191.

Salling, K.B. and Pryn, M.R. (2015), "Sustainable transport project evaluation and decision support: Indicators and planning criteria for sustainable development", *International Journal of Sustainable Development and World Ecology*, Taylor & Francis, Vol. 22 No. 4, pp. 346–357.

Walczak, R. and Majchrzak, T. (2018), "Implementation of the Reference Class Forecasting Method for Projects Implemented in a Chemical Industry Company", *Acta Oeconomica Pragensia*, Vol. 26 No. 1, pp. 25–33.

Walczak, R., Majchrzak, T., Osiecka, K. and Neumann, P. (2016), "Stochastic Risk Assessment of Projects Carried Out in a Company of Chemical Industry", *The 10th International Days of Statistics and Economics*, Prague, pp. 2046–2055.

Zarikas, V., and Kitsos, C.P. (2015), "Risk Analysis with Reference Class Forecasting Adopting Tolerance Regions", in Kitsos, C.P., Oliveira, T.A., Rigas, A. and Gulati, S. (Eds.), *Theory and Practice of Risk Assessment*, Springer International Publishing.