Prepared for the UK Department for Transport

Optimism Bias Study

Recommended Adjustments to Optimism Bias Uplifts

Final Report

Authors

Bert De Reyck, Professor and Director, UCL School of Management, University College London

Yael Grushka-Cockayne, Associate Professor, Darden School of Business, University of Virginia

Ioannis Fragkos, Assistant Professor, Rotterdam School of Management, Erasmus University Rotterdam

Jeremy Harrison, Honorary Research Associate, University College London and Director, Faithfields

Daniel Read, Professor, Warwick Business School

The authors would like to thank **Mike Bartlett**, Principal Risk & Value Manager, Network Rail, for his contributions to this report.

Table of Contents

1. Executive Summary

This study investigates the phenomenon of *optimism bias* in UK rail infrastructure projects. Optimism bias is the tendency of individuals to expect better than average outcomes from their actions. In the context of rail infrastructure projects, optimism bias can lead to underestimation of project duration, overestimation of its benefits and underestimation of its total cost.

Network Rail, the operator of the UK rail network, is well aware of the impact of optimism bias when planning and executing rail projects, and follows procedures that aim to remedy its impact. Guidelines have been published by the UK Department for Transport and are updated periodically to adjust to the ever-changing business environment and to changes in Network Rail's forecasting procedures. In the context of updating the optimism bias guidelines, the Department for Transport has commissioned the authoring team of this report to investigate the existence and magnitude of optimism bias in cost forecasts of rail infrastructure projects, and to recommend adjustments to the guidelines so that they correspond to the current environment under which forecasts are made.

The authoring team has cooperated closely with Network Rail and has collected data from a large sample of recent projects, including projects of various types, size and complexity, and from different geographical territories. We collected capital cost forecasts at different stages of the projects' development and actual realized costs. Using statistical analysis, we provide recommendations on how to update the existing optimism bias guidelines. Concretely, current practice recommends that an optimism bias uplift, expressed as a percentage increase of the original cost forecast, should be applied for the purpose of project appraisals, throughout the stages of project development. Our recommended optimism bias uplifts are summarised on the next page.

Our report also provides recommendations for the Department for Transport and Network Rail on how to further improve their investment appraisal and risk management framework.

(i) Our analysis reveals that optimism bias exists throughout all the stages of a project's development. We therefore present updated optimism bias uplifts for every stage. We recommend, however, that ideally, optimism bias uplifts only be used in the early stages of project appraisals, with quantitative risk assessment replacing optimism bias uplifts in the later stages. A proper quantitative risk assessment eliminates the need for optimism bias adjustments, and the evidence from Network Rail shows that when risk assessment is carried out, cost forecasts are indeed quite accurate. For early stages, however, limited available detail on the project scope makes a robust risk assessment difficult to implement, making an optimism bias adjustment more practical. Our analysis shows, however, that the current risk assessment procedures used by Network Rail are not yet sufficiently compensating for optimism bias, so in the short term we present optimism bias uplifts for every project development stage, until Network Rail's quantitative risk assessment is improved and can replace optimism bias uplifts for later project stages.

(ii) We recommend that optimism bias uplifts should not be added to a project's budget, as this might lead to overspending. A more suitable use for optimism bias adjustment values is to confirm that a business case remains robust if costs rise to this level, and to create a contingency held at the portfolio level.

(iii) We recommend that Network Rail improves its data capture for future projects, creating standardised data entry procedures, and in particular, records any applied adjustments separately (whether based on optimism bias or risk assessment), and provides cost estimates for the entire project, even if approval is only sought for one stage.

Table E-1 below shows the recommended optimism bias uplifts. We recommend a different uplift for different project types, i.e. for renewals (projects that maintain and renew the existing network) versus enhancements (improvement and extension projects), and for each GRIP stage, i.e. the different stages in the investment cycle of a project as defined in the Governance to Railway Investment Projects:

- GRIP 1 — Output Definition
- GRIP 2 — Feasibility
- GRIP 3 — Option Selection
- GRIP 4 — Single Option Development
- GRIP 5 — Detailed Design

As the recommend uplifts are not very different for the different project types, they could also be combined in a single recommendation (Table E-2), although this could lead to a slight underestimation or overestimation of cost.

Table E-1. Recommended OB uplifts based on GRIP stage and project type[1]

Project Type	GRIP 1/2	GRIP 3	GRIP 4	GRIP 5
Renewals	66%	18%	13%	3%
Enhancements	60%	17%	0%	8%

Table E-2. Recommended OB uplifts based on GRIP stage[1]

	GRIP 1/2	GRIP 3	GRIP 4	GRIP 5
All Projects	64%	18%	9%	4%

In line with the current procedures adopted by Network Rail, the OB uplifts for GRIP stages 1 and 2 are to be applied to the cost point estimate, whereas for GRIP stages 3, 4 and 5, they are to be applied to the mean of the Quantitative Risk Assessment (QRA) distribution (obtained using Monte Carlo simulation), as shown in Figure E-1. Clearly, the uplifts for stages 3, 4 and 5 are significantly smaller than the uplifts for stages 1 and 2, supporting the claim that a QRA could eliminate the need for optimism bias uplifts. However, a slight optimism bias is still present, possibly due to overconfidence, i.e. the ranges of cost estimates specified during QRA being too small. Therefore, before replacing optimism bias uplifts with QRA, a modification to the current risk assessment procedures would be required in order to reduce overconfidence as much as possible[2].

[1]The recommended OB uplift for GRIP stages 1 and 2 are not significantly different from each other, and are therefore combined in a single recommendation.

[2]Forecasts can be recalibrated to reduce overconfidence by training estimators when providing worst-case and best-case estimates or by adjusting the estimates themselves (allowing for the possibility of actual costs to fall outside of the specified boundaries).

Figure E-1. Application of recommended OB uplifts across the GRIP stages[3]

[3] The QRA mean is the mean (average) of the project cost probability distribution, determined using a Monte Carlo simulation analysis. It is typically higher than the (most likely) point estimate because of the right-skewed nature of the cost distribution. Therefore, using the QRA mean instead of the point estimate typically results in an increased cost estimate, and reduced optimism bias. At NR, the term "QRA" is often used to indicate the difference between the QRA mean and the point estimate, i.e. as an addition to the point estimate. We do not recommend this, but instead recommend using the term "QRA mean" as the overall average cost.

Alternatively, OB uplifts in GRIP stages 1 and 2 could be applied to the AFC, the *Anticipated Final Cost*, rather than the cost point estimate. For budgeting purposes NR determines an AFC for GRIP stages 1 and 2 by increasing the cost point estimate typically by 60% and 40%[4], respectively, in order to account for risk. Our analysis reveals that on average, the AFC estimates are reasonably accurate, with cost escalations above the AFC of only 3% and 17% for GRIP 1 and 2, respectively. Therefore, one could also apply (smaller) OB uplifts to the AFC instead of the cost point estimate, as in Tables E-3 and E-4.

Table E-3. Recommended OB uplifts based on AFC for GRIP 1 and 2

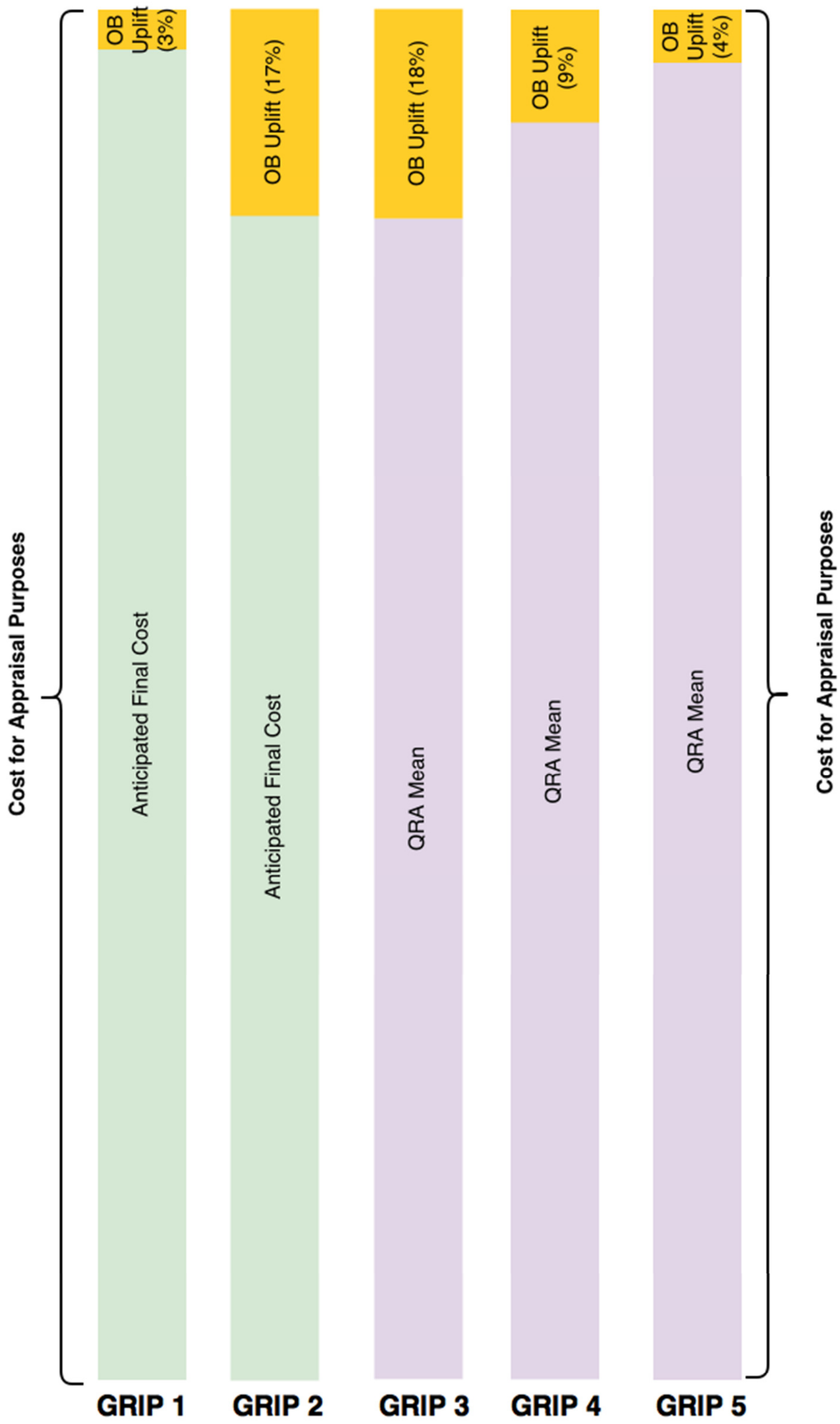
Project Type	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
Renewals	3%	28%	18%	13%	3%
Enhancements	1%	0%	17%	0%	8%
Applied to	AFC	AFC	QRA Mean	QRA Mean	QRA Mean

Table E-4. Recommended OB uplifts based on AFC for GRIP 1 and 2

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
All Projects	3%	17%	18%	9%	4%
Applied to	AFC	AFC	QRA Mean	QRA Mean	QRA Mean

[4] Although 60% and 40% are the general guidelines adopted by NR, it appears that the uplifts applied vary from project to project, although details of such variations are not recorded.

Figure E-2. An alternative application of OB uplifts based on AFC in GRIP stages 1 and 2



2. Introduction

2.1 Optimism Bias and its Impact on Rail Projects

Optimism bias (OB) is defined as the tendency of individuals to expect better than average outcomes from their actions. In a managerial context, OB leads to underestimation of the cost and completion time of planned tasks and overestimation of their benefits. A specific form of OB is the *Planning Fallacy* (Kahneman and Tversky 1979), which is a phenomenon of individuals making optimistic predictions regarding the amount of time needed to complete future tasks. There is a growing body of evidence that OB is one of the most important biases when it comes to factors that impact the quality of forecasts in project planning (Flyvbjerg 2011).

Nobel laureate Daniel Kahneman, (Kahneman and Tversky 1979; Kahneman and Lovallo 1993; Kahneman 2011), suggests that taking an *outside view* to forecasting helps reduce the errors that results from OB. Taking the outside view involves considering previous forecasts and actual realisations in similar forecasting environments and utilising this evidence when making new forecasts. In the context of project planning, this suggests that when forecasting the cost of a certain project, it is beneficial to consider outcomes of previous projects of similar nature. Specifically, historic forecasting errors and their distribution can be valuable guides towards calibrating a future forecast, if a set of similar projects, also called a *reference class*, can be utilised.

2.2 Scope and Content of this Study

The aim of this study is to investigate the degree of capital cost escalation in rail projects in the United Kingdom, and to recommend appropriate OB uplifts of cost forecasts, based on the nature and attributes of individual projects. To this end, the authors worked closely with Network Rail UK (NR) on curating data from recent rail projects, and performed an analysis of cost escalations. The analysis considered how the escalations depend on a variety of project characteristics.

The study is based on a large set of diverse projects, including different territories, asset and sponsoring groups and national and local programmes. Excluded projects included mergers of projects[5] and projects of certain categories that NR deemed not relevant to the scope of this report[6], such as operational expenses projects, generic works and annual programmes of work.

It should be noted that the application of OB as described in this report was deemed appropriate for NR. These recommendations, however, might not be suitable for other organisations with different estimating procedures, or without data to demonstrate how actual costs fluctuate from earlier estimates.

[5] The cost of a merged project cannot be allocated to individual projects, and neither can the costs of individual projects be added so that they accurately reflect the equivalent cost of a merged project.

[6] This report is concerned with capital costs and therefore excludes operational expenses. Annual programmes and generic works have specific cost structures, which require different cost estimation procedures.

3. The Governance to Railway Investment Projects

In this section, we provide an overview of how NR manages its investment projects. The G*overnance to Railway Investment Projects* (GRIP) breaks down the investment cycle of a project into eight distinct stages; in each stage a review takes place that ensures that the project can successfully progress to the next stage:

1. **GRIP 1 — Output Definition**
   Strategic requirements are set, and project goals and scope are defined.
2. **GRIP 2 — Feasibility**
   An initial feasibility study defines the scope of the investment, identifies constraints and assesses if the outputs can be delivered.
3. **GRIP 3 — Option Selection**
   Several options that address the constraints of stage 2 are developed. A single preferred option is selected at the end of this stage, which should be approved by the main stakeholders.
4. **GRIP 4 — Single Option Development**
   Initiation of the development and of the design specifications of the selected option.
5. **GRIP 5 — Detailed Design**
   Specifies the full design to which the project will be built.
6. **GRIP 6 — Construction, Test and Commission**
   The main implementation part of the project, which should ultimately be delivered according to the specifications, tested and commissioned into use.
7. **GRIP 7 — Scheme Hand Back**
   The project is handed over to its operator and maintainer.
8. **GRIP 8 — Project Closeout**
   Assessment of benefits is carried out, and any contingencies and guarantees are put into place.

Figure 1 gives an overview of the GRIP process. Note that small projects might group a number of GRIP stages together and only report at the end of the main ones, i.e. stages 1, 4 and 8, in order to expedite their execution. However, all projects have at minimum GRIP stages 1, 4 and 8. As a project progresses through the different GRIP stages, its scope becomes more specific and uncertainty is reduced. Therefore, the amount of OB in each GRIP stage might be different. To address this possibility, we collected data from projects in different GRIP stages.

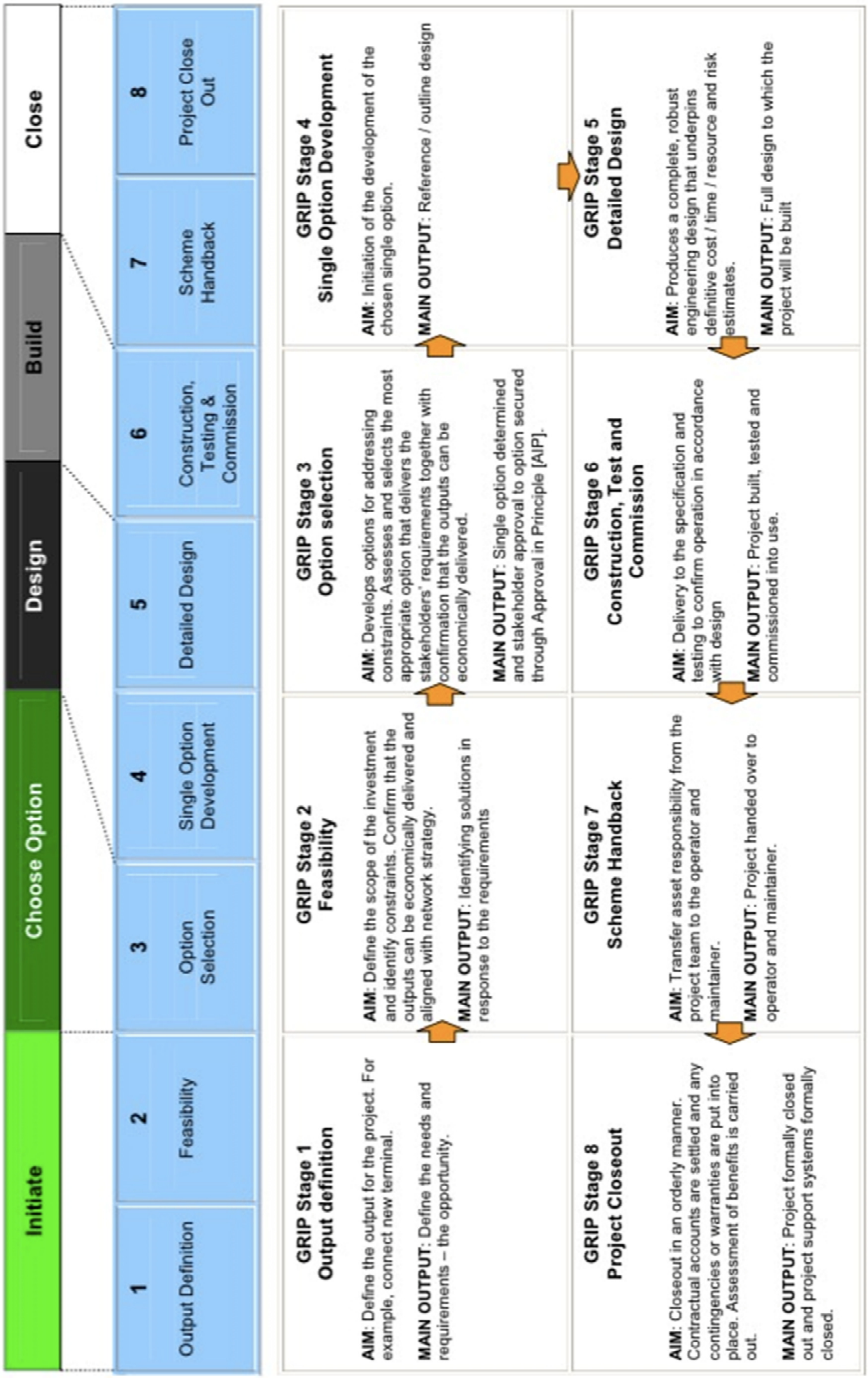


Figure 1. Network Rail's Governance to Railway Investment projects (GRIP)

4. Previous Optimism Bias Studies and Current Guidelines

In this section, we provide an overview of the most recent OB studies carried out by the Department for Transport (DfT) and NR, from 2002 onwards.

4.1 Mott MacDonald Study

In 2002, Mott MacDonald (2002) studied cost escalations of six distinct project types, with the results in Table 1. Rail projects were classified as either standard or non-standard civil engineering projects, and showed early-stage cost escalations of 44% and 66%, respectively, with corresponding recommended OB uplifts. For rail projects that have both standard and non-standard elements, the study recommends that uplifts should be weighted appropriately. As the project evolves and the amount of uncertainty and optimism is reduced, the recommended uplifts are reduced accordingly, down to 3% and 6% for rail projects upon contract award.

Table 1. Mott MacDonald cost escalations per project type

Project Type	Early-stage OB uplift	Late-stage OB uplift (before contract award)
Standard Buildings	24%	2%
Non-standard Buildings	51%	4%
Standard Civil Engineering	44%	3%
Non-standard Civil Engineering	66%	6%
Equipment/Development	200%	10%
Outsourcing	41%	0%

Based on these recommendations, the UK Treasury developed the *Supplementary Guidance on OB*, which was adopted by the Department for Transport and published in the Transport Analysis Guidance (DfT, WebTAG A5.3, 2014), hereafter referred to as "WebTAG guidance". The WebTAG guidance provide guidelines for adjusting estimates of capital and operating costs, benefits and time profiles, using the OB uplifts in Table 2[7]. The uplifts for GRIP 1, 2 and 3 are to be applied to the point estimate, i.e. the most likely base cost estimate, whereas the uplifts for GRIP 4 and 5 are to be applied to the mean of the distribution that results from a *Quantitative Risk Assessment* (QRA)[8], which already incorporates some adjustment for risk.

Table 2. WebTAG OB guidance

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
	66%	50%	40%	18%	6%
Applied to	Point Estimate	Point Estimate	Point Estimate	QRA Mean	QRA Mean

[7] See also WebTAG Unit A5.3 Table 3.

[8] *Quantitative Risk Assessment* refers to a risk analysis approach using Monte Carlo simulation, in which activity duration and/or costs are assessed for probability of occurrence and range of impact and aggregated on a scenario basis.

The WebTAG guidance, however, has some shortcomings. First, it classifies all rail projects as non-standard projects, without differentiating between projects of different types, scale and complexity. Second, the OB uplifts for intermediate stages were derived by extrapolating the Mott MacDonald findings for GRIP 1 and 5. To remedy these shortcomings, our study provides evidence-based uplifts for separate stages in the project's development, with different recommendations for different types of projects, thereby refining the recommendations given by the WebTAG guidance.

4.2 Halcrow Group OB Study

The Halcrow Group was commissioned by the DfT in 2008 to analyse OB for *Renewals* and *Enhancements* in rail projects. Renewals are projects that maintain and renew the existing network, while enhancements are improvement and extension projects. Halcrow investigated 187 renewal and 58 enhancement projects, and recommended the uplifts in Table 3.

Table 3. OB uplifts (%) recommended by the Halcrow Group report on OB (2009)

Stage	Renewals	Enhancements
GRIP 1	66%	55%
GRIP 2	51%	46%
GRIP 3	22%	27%
GRIP 4	8%	18%

The Halcrow study provides more finely grained guidelines than the Mott MacDonald study. However, a limitation of the Halcrow study is the small set of projects being analysed. Moreover, the recommended OB uplifts for stages 2 and 3 are still extrapolations from the OB uplift for stages 1 and 4, rather than being independently derived from project data.

4.3 Sweett Group Benchmarking Study

The Sweett Group attempted to identify efficiency targets for NR's Strategic Business Plan. Although the initial scope involved utilising cost data from 96 projects (for producing unit cost rate comparisons between territories, assess the use of the newly introduced CAF and provide recommendations), the lack of data led them to provide only general recommendations. As such, their recommendations are of strategic nature and are not directly comparable with our report.

4.4 Infrastructure Risk Group Report

A relevant report that addresses implementation issues of OB was published in 2013 by the Infrastructure Risk Group (IRG). The IRG has members that are leading managers and risk analysts in a variety of organisations, such as NR, London Underground, Crossrail, Heathrow, HS2 and Infrastructure UK. Although the report does not provide a quantitative assessment of OB, it emphasizes that an erroneous application of OB uplifts can result in overspending. In particular, the reports warns against adding OB uplifts to individual project budgets, and instead using OB estimates to verify that a project business case remains robust even if costs rise to this level. We have adopted and underlined this recommendation in our report as well.

4.5 Current Study

Our study is based on a sample of 2,050 projects, and provides results for all GRIP stages. Using statistical analysis and data cleansing tools, we provide deeper insight into the cost escalations of UK rail projects, with an in-depth analysis of project characteristics and their impact on cost escalations. Compared to earlier studies, our analysis is based on a larger and more recent set of projects, thereby incorporating the impact of recent changes implemented by NR when forecasting project costs, such as the recently introduced *Cost Analysis Framework* (CAF)[9].

[9] The *Cost Assessment Framework* refers to a standardized process and a database of standardized tasks that are incorporated in cost forecasts of capital investment projects.

5. Description of Project Characteristics

5.1 Project Data Sources

The data used in this study was collected between September 2013 and May 2014 from the following sources at NR:

- NR Oracle Projects Database
- Author2k3 File
- Internal NR reports on OB

For each project entry in these databases, the following information is available:

- Title and number
- Funder
- Current authority (amount of funding that was requested when the entry was recorded)
- Anticipated Final Cost (AFC) of approved stages and of all stages
- Cost of Work Done (COWD)
- Delivery group, sponsoring group, asset category and asset group
- Indication of whether the project is funded from multiple resources (multi-funded)
- Indication of whether the project involves multiple assets (multi-asset)
- Territory
- Names of programme managers, projects managers and sponsors

Using this database, we were also able to classify the projects as renewals and enhancements[10].

The AFC for each project consists of a point estimate, i.e. an estimate of the most likely cost of the project, and an allowance for risk. Although these allowances are not recorded separately, but instead included in the overall estimate, NR guidelines (Network Rail, 2014) provide the following recommendations[11]:

- GRIP 1: AFC = point estimate plus 60%
- GRIP 2: AFC = point estimate plus 40%
- GRIP 3, 4 and 5: AFC = mean of QRA distribution

As we were unable to determine the individual uplifts applied to each project, we assume that the NR guidelines were followed when determining the AFC for each project, which allowed us to calculate the underlying point estimates in GRIP 1 and 2[12]. We assume that no uplifts were applied in GRIP 3, 4 and 5 as per the guidelines.

Note that the allowances above differ from the WebTAG OB guidance (Table 2), which at NR is used for project appraisals, but not for calculating project AFCs. This is in line with the WebTAG guidance

[10] The classification was made based on whether the string "renewal" or "enhancement" was found within the funding category field entry of each project.

[11] Although 60% and 40% are the general guidelines adopted by NR, it appears that the uplifts applied vary from project to project.

[12] Naturally, if the guidelines were not adhered to, then our estimates of the cost point estimates could be incorrect. However, as long as the guidelines were followed on average, i.e. with some uplifts higher and others lower than what the guidelines prescribe, then our recommendations remain valid.

(paragraph 2.4.8), which states that "optimism bias values are only required for appraisal purposes. It may be appropriate to use a different way of taking into account potential cost overruns for financial or accounting purposes, such as the use of contingency".

Between 2006 and 2008, NR introduced the *Cost Analysis Framework* (CAF), which represented a major improvement in recording and standardising repeatable work items, which in turn may have led to improved project cost estimates. We expect that the introduction of the CAF, as well as the adoption of the WebTAG guidance on OB uplifts, will have improved the accuracy of forecasts. Therefore, we excluded any pre-2009 projects to ensure that our results accurately reflect current practice.

5.2 Project Data Cleansing

All previous studies mention data reliability issues and ours is no exception. We found errors in project stages, dates, and cost estimates in the datasets provided by NR. From our investigation, it seems that these errors stem from the manual data entry procedure, and from the fact that the data is recorded by several managers, with little standardisation. As a result, many cost estimates are missing or entered erroneously, or the same information is represented in multiple ways, making analysis cumbersome. Also, a lack of consistency in scales and formatting presented a major challenge.

Another common shortcoming in the data is that the recorded cost estimates sometimes refer to the cost of the development phase only, and not to the cost of the entire project. This is due to commitment only being requested for the development phase, rather than the entire project. As a result, it does not allow for monitoring total project costs against a baseline. Also, the lack of consistency in recording can result in inaccuracies, as it is sometimes not clear which phase(s) the cost estimates refer to. Other information, such as asset categories or territories, is also inconsistent and required much cleansing for the data to be made usable for analysis.

Another issue with the data is that some projects experience significant changes in scope throughout their development, sometimes resulting in significant cost escalations. It is still open for debate whether such significant changes are part of optimism bias, or whether they should be analysed separately.

To address these problems, we took the following steps to cleanse the data. First, we removed project entries with corrupted or missing data, for instance where the GRIP stage was not recorded. Second, we analysed all project entries with extremely low or high cost escalations, and either corrected the data after discussing with NR, or removed the entries altogether. We also removed project entries with cost escalations exceeding 500% or below -50%, as these were considered to be either incorrect due to data entry errors, or due to significant scope changes.

Note that despite these correction and cleansing procedures, issues will still remain with some of the project data, which may distort the recommendation in this report[13].

[13] A further potential problem with the data is that some project costs use different price bases during the GRIP process, so some increases in cost may reflect these price base changes rather than optimism bias. However, given that inflation has been relatively low since 2009, this is not believed to be a significant issue.

5.3 Project Data Statistics

After cleansing, we retained 2,050 projects that contained seemingly reliable information on project costs and a range of project characteristics, on which we elaborate in this section.

5.3.1 Project Type

Our dataset contains approximately 75% renewals (1,555) and 25% enhancements (495).

5.3.2 GRIP Stages

Table 4 shows the number of project entries in each of the five GRIP stages. In total, we have 2,460 entries, relating to 2,050 projects. Most entries relate to GRIP 1, 4 and 5, but we also retained a substantial number of projects with data on GRIP stages 2 and 3, thereby allowing a robust analysis of the cost escalations on these stages as well. However, note that our analysis is not based on projects with data recorded for every GRIP stage, but rather on projects that have at least one GRIP stage recorded, in order to retain a larger and more diverse dataset.[14]

Table 4. Project entries across each of the five GRIP stages

Project Type	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
Renewals	346	34	210	587	649
Enhancements	198	17	54	196	169
All projects	544	51	264	783	818

5.3.3 Asset Categories

Figure 2 shows for the largest 12 project asset categories, accounting for 80% of all projects, the number of projects in each category. The largest asset category for enhancement projects is the *National Stations Improvement Programme* (NSIP), a major programme that aims to improve the quality of facilities and services provided to passengers in train stations nationwide. The second largest enhancement asset category is *Access for All*, which aims to improve accessibility at train stations. Renewal projects relate to maintain the quality of the existing rail network, and exist across a wide variety of asset categories.

[14] 1,685 projects only have an entry for 1 GRIP stage, 325 projects have 2 entries, 36 projects have 3 entries, 3 projects have 4 entries, and only 1 project has an entry for every GRIP stage.

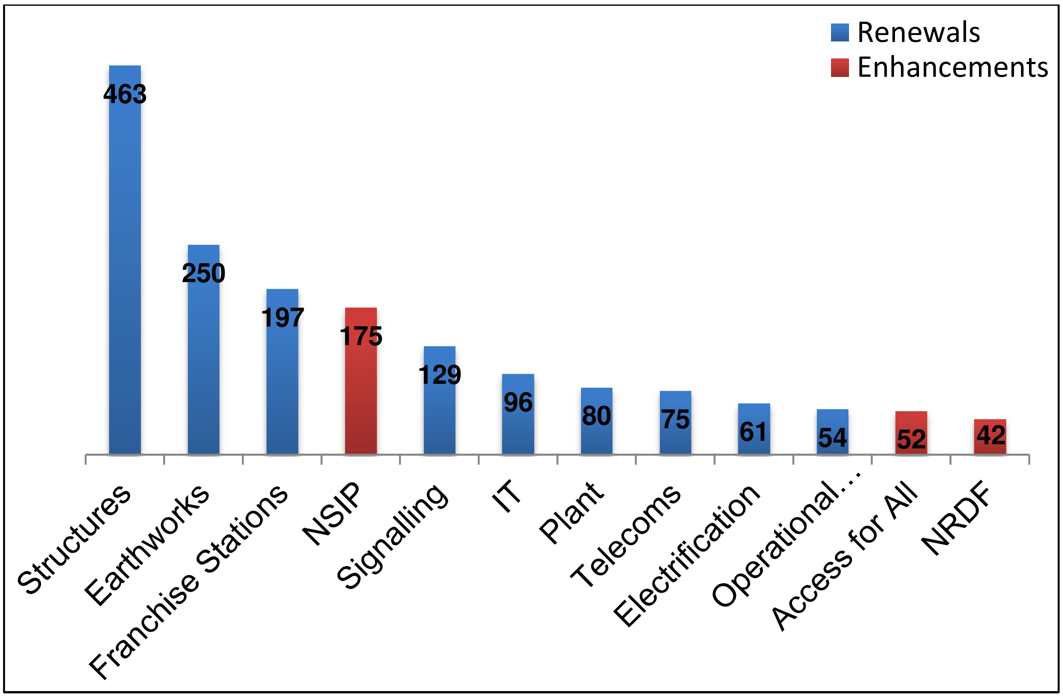


Figure 2. Number of projects in the 12 largest asset categories

5.3.4 Project Size

Figure 3 shows the varying project sizes in the dataset, as measured by their final cost. The smallest and largest project final cost values are £100,000 and £254 million, respectively. (The figure only shows entries with a final cost of less than £10 million.) The median and average final costs are £509,000 and £1.9 million, respectively. Renewal projects are typically smaller than enhancements, with an average final cost of £1.2 million versus £3.9 million.

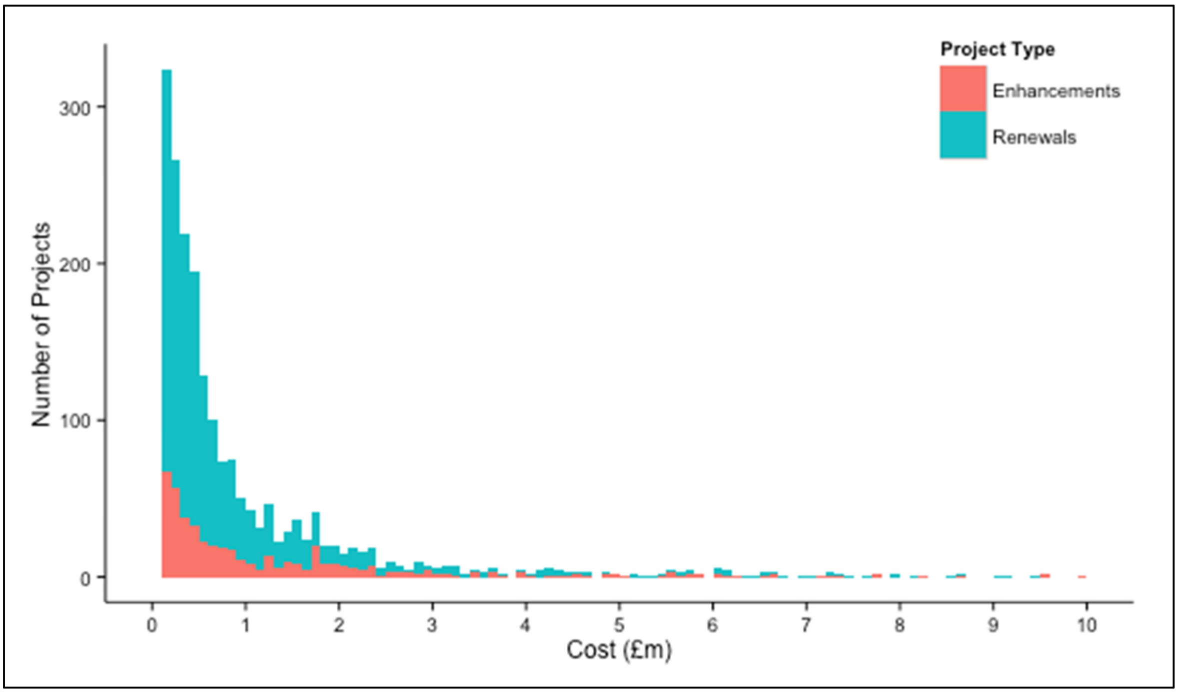


Figure 3. Project sizes as measured by final cost (shown up to £10 million)

5.3.5 Project Age

Figure 4 shows the number of projects initiated or in progress in each year[15].

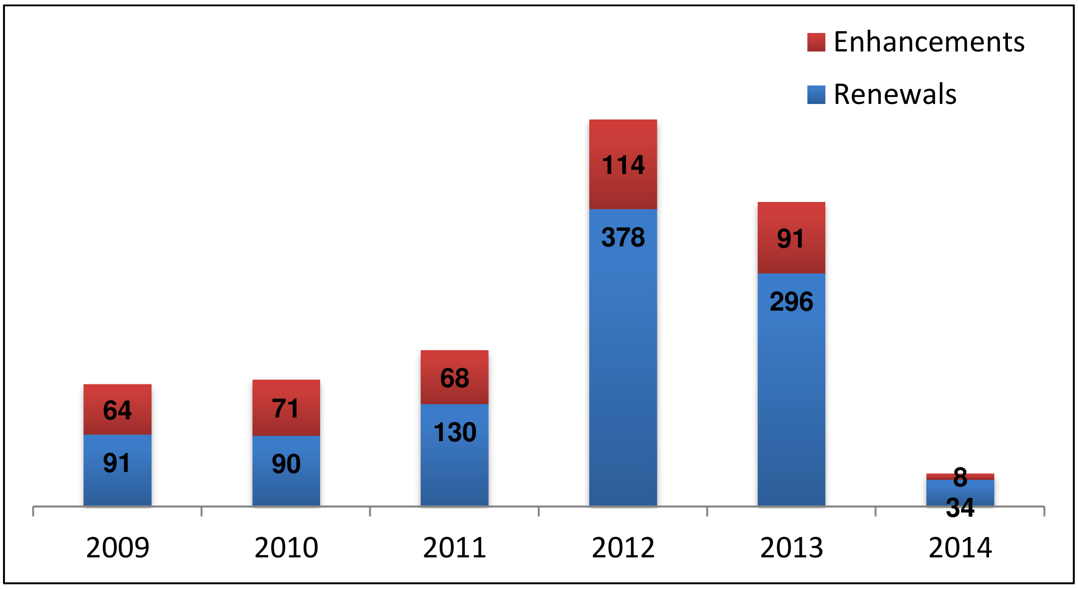


Figure 4. Number of projects by year of first recorded cost estimate

5.3.6 Territories

Figure 5 shows the distribution of projects across territories. The majority of projects are carried out in London NW and NE, and Scotland. The 75/25 proportion of renewals versus enhancements is rather constant in each territory, except for Scotland, where renewals account for the large majority of projects.

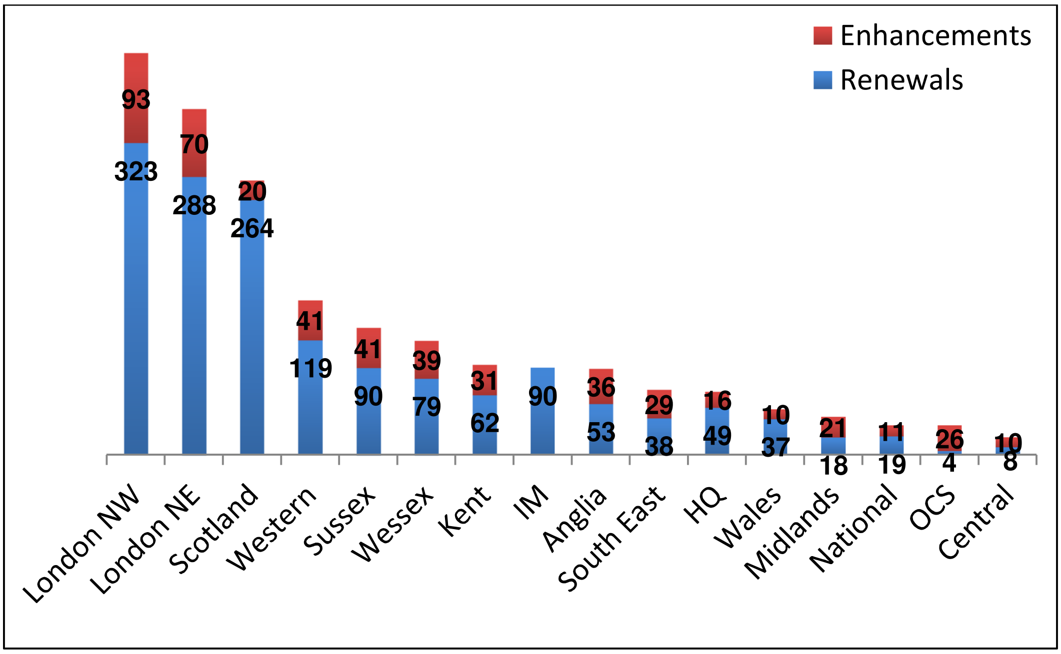


Figure 5. Number of projects in each territory and project category

[15] The classification is based on the earliest date for which a cost estimate was available. The age of some projects could not be determined due to data entry errors.

5.3.7 NR versus Consortium Projects

814 (40%) projects are carried out by one dominant partner, while 1,236 (60%) were carried out by a consortium, such as EDF Energy, Costain and Carillion Construction[16]. Enhancement projects have a larger proportion (68%) of consortium projects compared to renewals (58%).

5.3.8 Multi-Funded and Multi-Asset Projects

A small number of projects, namely 111 (5%), involve multiple assets and multiple funding sources, 58 of which are renewals and 53 enhancements. These projects typically involve a higher degree of complexity.

5.3.9 Cost Escalations

For each project, we measure the relative difference between the estimated final cost (AFC) and the actual final cost as follows, where i stands for each GRIP stage:

$$Escalation_i = \frac{FinalCost - AFC_i}{AFC_i}$$

By so doing we obtain a cost escalation for each GRIP stage. In other words, the derived cost escalation is the amount by which the AFC estimate of each GRIP stage needs to be uplifted to match the final cost:

$$AFC_i(1 + Escalation_i) = FinalCost$$

A positive cost escalation implies that the final cost was higher than the estimated cost, and therefore that the forecast (which might already incorporate an allowance for risk) was optimistic, while a negative escalation implies that the final cost was lower than the estimated cost, showing that the forecast was too conservative.

Figure 6 shows a histogram of the cost escalations from the GRIP 1 AFC to the final cost. As the figure suggests, most projects have cost escalations close to zero, indicating reasonably accurate forecasts, with a majority of projects coming in below the AFC. We also see projects with extreme escalations, either close to -100% or with a final cost being a large multiple of the AFC (not shown in Figure 6). As mentioned earlier, we removed such project entries with extreme cost escalations (exceeding 500% or below -50%), as these were considered to be either incorrect due to data entry errors, or due to significant scope changes. Overall, we found that:

- 78% of projects have zero or negative GRIP-1 escalations, i.e. with actual total cost equal to or lower than the AFC
- 52% of projects have negative GRIP-1 escalations, i.e. with actual total cost lower than the AFC
- 26% have zero GRIP-1 escalations, i.e. with actual total cost equal to the AFC
- 22% have positive GRIP-1 escalations, i.e. with actual total cost higher than the AFC
- 5% of projects have a GRIP-1 escalation larger than 50%
- Average GRIP-1 escalation: 3%, with 81% of projects below average

[16] Consortium projects can be identified as projects for which an "Industry AFC" figure is reported.

Overall, we find that the large majority of projects are delivered with a total cost lower than the AFC, and that on average, the AFC accurately reflects the actual total project costs. However, in the next section we will investigate whether specific project characteristics can help predict future cost escalations more accurately.

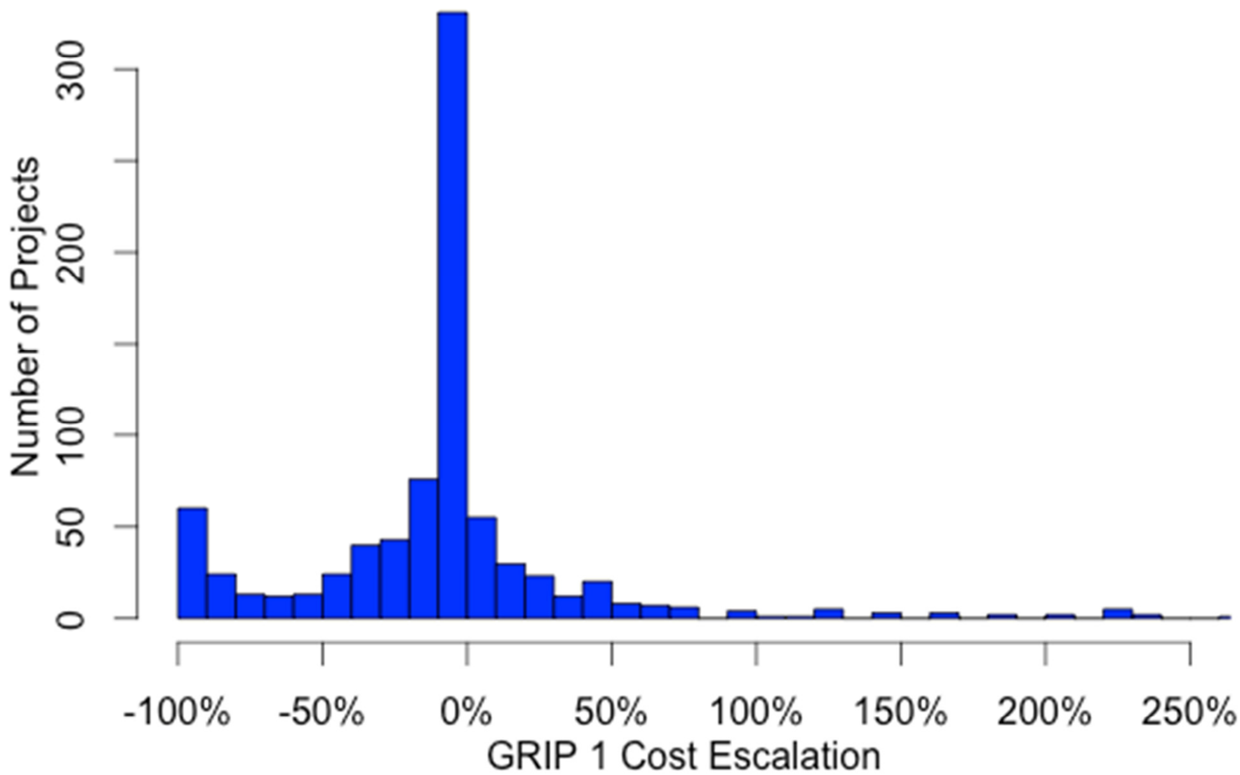


Figure 6. Cost Escalations from GRIP 1 (only shown up to 250%)

6. Project Characteristics that Influence Cost Estimates

In this section, we investigate a wide range of project characteristics and their impact on cost escalations. The aim is to identify a set of characteristics than can help predict future cost escalations more accurately, and provide an OB uplift for each individual project based on its underlying characteristics.

1. *Project GRIP stage*
   We expect cost escalations to be greater for projects that are in early GRIP versus late GRIP stages, as more accurate information about the project is available at later GRIP stages, and as uncertainty is reduced. Therefore, we analyse cost escalations for each GRIP stage separately.
2. *Project type* (enhancements versus renewals)
   Generally speaking, enhancement projects are more risky than renewal projects. Therefore, we expect the greater cost escalations for enhancements.
3. *Project size* (measured by the anticipated final cost AFC)
   We expect relative cost escalations, as a percentage of AFC, to be different between small and large projects.
4. *Project complexity*
   We define project complexity as the number of interfaces and parties that a project involves, which can be measured by proxy characteristics such as whether a project is multi-funded or multi-asset (MFMA), or whether an Industry AFC figure is reported (indicating that it is a consortium project). We expect that the more complex the project, the greater the cost escalation.
5. *Territory*
   The area in which a project takes place could have an effect on cost escalation, perhaps due to geographical features, complexity of interaction with existing infrastructure, or the nature of infrastructure that is being developed in each territory. Therefore, we expect that there might be different cost escalations for each territory.
6. *Project asset category*
   Projects that belong to different assets, such as signalling, structures and earthworks, differ greatly and therefore the degree of cost escalation might vary across asset categories.

Using a statistical analysis, we determined which of these characteristics have a significant[17], sizeable[18] and reliable[19] impact on past cost escalations. Table 5 provides an overview of our findings.

[17] An impact that is statistically different from zero.
[18] An impact that is statistically different from zero, and not small.
[19] Some of the data available for a project characteristic is unreliable because of data errors or unreliable data recording processes.

Table 5. Impact of project characteristics

Project characteristic	Significant impact?	Sizeable impact?	Reliable?
GRIP stage	Yes	Yes	Yes
Project type	Yes	Yes	Yes
Project size	No	No	Yes (cleansing required)
Project complexity	Yes	Yes	Yes
Territories	No	No	Yes
Asset category	Yes (1 category)	No	Yes

In the following sections, we provide detailed results for the impact of each project characteristic on cost escalations, measured by the relative deviations between the final cost and the AFC in each GRIP stage. Note that the AFC already includes a contingency for risk and OB (in the form of a percentage uplift for GRIP 1 and 2) or a risk allowance (based on QRA for GRIP 3, 4, and 5)[20], so our reported deviations should be interpreted as required modifications to the currently applied contingencies and allowances.

6.1 Cost Escalations across GRIP Stages

Table 6 shows the average cost escalation for each GRIP stage, for all projects. The observed escalations suggest that overall the currently applied uplifts and contingencies result in reasonably accurate cost estimates, with only a 7% average escalation, and near-zero escalations for GRIP stages 1 and 5. However, all observed escalations are positive, implying that a bias still remains, and that the contingencies currently applied need to be increased, especially for stages 2 and 3.

Table 6. Cost escalations per GRIP stage

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Cost Escalations	3%	17%	18%	9%	4%	**7%**

6.2 Renewals versus Enhancements

Table 7 provides a separate breakdown of cost escalations for renewals and enhancements. Overall, the escalations are slightly higher for renewals. We again see that cost escalations in GRIP 1 and 5 are relatively low, meaning that the forecasts are reasonably accurate, with higher escalations for GRIP 2, 3 and 4, although GRIP 2 and 4 forecasts for enhancements are very much on target. These results indicate that the contingencies for renewals need to be adjusted upwards, at least for stages 2, 3 and 4, but that for enhancements, a major adjustment is only required for stage 3.

[20] We were unable to verify the contingencies or QRA allowances included in the AFC for individual projects, as these are not separately recorded. However, NR confirmed that the typical uplifts / allowances included in the AFCs (see Section 5.1) were consistent with NR's "Cost estimating, cost analysis and benchmarking" guidelines (NR, 2014). At GRIP 1 and 2, the uplifts are sometimes modified based on expert judgement, and in GRIP 3, 4 and 5, the QRA allowance is sometimes based on the P80 rather than mean, or a QRA may not have been applied at all, but a contingency included instead. However, as long as the applied uplifts or contingencies conform to the guidelines *on average*, our conclusions will still be valid.

Table 7. Average cost escalations per GRIP stage and project category

Type	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Renewals	3%	28%	18%	13%	3%	**9%**
Enhancements	1%	-5%	17%	-2%	8%	**3%**

6.3 Project Size

There is some evidence (Flyvbjerg, Bruzelius and Rothengatter 2003) that large projects tend to overspend proportionally more than small projects. Tables 8 and 9 show average cost escalations for small (AFC<£5m), medium (£5m<AFC<£50m) and large projects (AFC>£50m).

Table 8. Average cost escalations per GRIP stage and size. A dash (-) indicates no data. An asterisk (*) shows that the average escalation is based on less than 30 projects, and therefore less reliable.

Size	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Small	3%	20%	19%	11%	5%	**8%**
Medium	2%*	4%*	12%*	-7%*	-7%*	**-2%***
Large	-	-18%*	9%*	-11%*	-6%*	**-9%***

Table 9. Average cost escalations per GRIP stage, project type and size. A dash (-) indicates no data. An asterisk (*) shows that the average escalation is based on less than 30 projects, and therefore less reliable.

Type	Project Size	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Renewals	*Small*	3%	30%	19%	14%	4%	**9%**
	Medium	8%*	-4%*	11%*	0%*	-6%*	**1%***
	Large	-	-	-	-	-3%*	**-3%***
Enhancements	*Small*	2%	-5%*	18%	-1%	11%	**5%**
	Medium	-8%*	13%*	13%*	-15%*	-8%*	**-7%***
	Large	-*	-18%*	9%*	-11%*	-9%*	**-11%***

From Tables 8 and 9, we can see some evidence that typically, larger projects actually have smaller relative cost escalations compared to smaller ones. The analysis, however, is hampered by the fact that the dataset only contains a small number of large projects, which can skew the results. For example, the cost escalation for GRIP 3 large enhancements (9%) is based on only one project, and therefore it is not reliable for future projects. Therefore, we do not provide recommendations that depend on the size of each project[21].

6.4 Project Complexity

Projects with multiple interfaces and involved parties are not only harder to forecast, but also to plan, manage and execute. In Tables 10 and 11 we classify projects as simple or complex based on whether an Industry AFC figure is reported or not, as the reporting of an Industry AFC is compulsory in cases where a consortium of contractors take part in the project.

[21] We also explain in the Appendix that we have performed further analysis (hypothesis testing and linear regression) that confirms the decision not to use project size to predict future cost escalations.

Table 10. Average cost escalations per GRIP stage and project complexity (NR versus consortium projects). An asterisk (*) shows that the average escalation is based on less than 30 projects.

Complexity	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Simple	-5%	-7%*	7%	1%	-2%	**0%**
Complex	4%	26%	27%	17%	9%	**12%**

Table 11. Average cost escalations per GRIP stage, project type and project complexity (NR versus consortium projects). An asterisk (*) shows that the average escalation is based on less than 30 projects.

Type	Complexity	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Renewals	*Simple*	-3%	-8%*	-1%	3%	-2%	**0%**
	Complex	4%	37%*	33%	24%	7%	**14%**
Enhancements	*Simple*	-12%*	-6%*	37%*	-8%	-4%	**-1%**
	Complex	3%	-4%*	-1%*	2%	14%	**5%**

Tables 10 and 11 reveal that project complexity has a large influence on cost escalations. Overall, complex projects exhibit larger cost escalations, and therefore require higher OB uplifts[22]. In fact, simple projects are, on average, forecasted perfectly on target, with complex projects requiring an additional 12% uplift, or 14% for renewals and 5% for enhancements.

A second feature that captures project complexity is whether a project is multi-funded or involves multiple assets. Tables 12 and 13 show that projects that involve multiple assets and are funded by multiple resources overall are indeed more likely to have higher cost escalations. However, very few projects are multi-funded and multi-asset, making the analysis unreliable. Therefore, we did not use this characteristic in our recommendations.[23]

Table 12. Average cost escalations per GRIP stage and project complexity (multi-funded/multi-asset projects). An asterisk (*) shows that the average escalation is based on less than 30 projects.

Complexity	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Simple	1%	19%	16%	9%	4%	**7%**
Complex	31%*	-10%*	53%*	14%	10%	**17%**

Table 13. Average cost escalations per GRIP stage, project type, further classified by project complexity (multi-funded/multi-asset projects). An asterisk (*) shows that the average escalation is based on less than 30 projects.

Type	Complexity	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Renewals	*Simple*	2%	29%	17%	13%	4%	**8%**
	Complex	31%*	-22%*	54%*	22%	-11%*	**14%**
Enhancements	*Simple*	-1%	-5%*	13%	-3%	3%	**0%**
	Complex	31%*	-6%*	52%*	5%*	27%	**21%**

[22] This statement is also supported by statistical analysis.

[23] A statistical analysis also reveals that this dimension of project complexity does not add much in addition to the consortium-measure of complexity.

6.5 Geographical Territories

We also examined average cost escalations for different territories. The results do not reveal a systematic pattern of area-related deviations, in the sense that no particular area seems to perform consistently better or worse than others. After confirming with further statistical testing, we decided not to use territory as a differentiating factor for OB uplift recommendations.

6.6 Asset Categories

NR allocates each project to an asset category, which gives an indication of the different nature of these projects. Signalling projects, for example, are different from plant projects or structures, not only because of different technical challenges but also because of different management teams. Our dataset contains 64 separate asset categories. We found no significant differences, however, between the escalations in these asset categories. Also, a further potential issue with using asset categories is that the classification to assets depends on NR's organisational structure, and therefore any future changes might render this classification unusable. We thus decided to not utilise asset categories in our recommendations.

6.7 Summary of influential project characteristics

In summary, our results show that the following characteristics have a significant influence on the cost escalations of projects, and we therefore proceed to provide OB uplift recommendations by utilising the aforementioned characteristics:

- GRIP stage
- Project type (renewals versus enhancements)
- Complexity (NR versus consortium)

7. Recommendations

7.1 Recommended OB uplifts

Having examined various characteristics that could influence the accuracy of cost forecasts, we now provide recommendations on how the existing guidelines should be updated.

NR uses the latest WebTAG guidelines (Table 14) for project appraisals, but for estimating a project's AFC, NR uses different guidelines (Table 15). Compared to the latest WebTAG guidelines, the NR contingencies are lower for every GRIP stage, with a marked difference for GRIP 3, which can be explained by the fact that for that stage NR uses the QRA mean as the AFC estimate instead of the point estimate inflated by 40%. This is the result of QRA now also becoming the norm in GRIP 3, rather than just for GRIP 4 and 5.

Table 14. Current guidelines on OB uplifts as suggested by the latest WebTAG

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
Cost Escalations	66%	50%	40%	18%	6%
Applied to	Point estimate	Point estimate	Point estimate	QRA mean	QRA mean

Table 15. NR risk allowance estimates

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
Cost Escalations	60%	40%	0%	0%	0%
Applied to	Point estimate	Point estimate	QRA mean	QRA mean	QRA mean

Table 16 shows the changes to the currently applied contingencies, based on the GRIP stage, category and complexity of the project, defined using the consortium categorisation. The table also shows a confidence range for each reported deviation[24]. Cost escalations highlighted with an asterisk (*) have wide confidence ranges because they are based on a small number of projects. Bold numbers indicate deviations that are found to be significantly different from zero, and would therefore benefit from an adjustment, whereas the non-bold numbers indicate that current practice could result in accurate forecasts.

Overall, the results show that complex renewals and simple enhancements have rather large escalations, thereby implying reduced forecasting accuracy.

[24] In cases where only a sample of projects is available, the actual average cost escalation might be different from the one calculated using the sample, and lie in a range around the calculated average. This range depends on how large the sample is and on how different the projects in the sample are.

Table 16. Recommended changes to OB uplifts (with confidence ranges). An asterisk (*) shows that the average escalation is based on less than 30 projects.

Type	Complexity (Consortium)	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
Renewals	*Simple*	-3% [-10%, 4%]	-8%* [-15%, 0%]	-1% [-5%, 3%]	3% [-1%, 7%]	-2% [-5%,1%]
	Complex	**4%** [1%, 7%]	**37%*** [12%, 59%]	**33%** [29%, 47%]	**24%** [19%, 29%]	**7%** [4%, 10%]
Enhancements	*Simple*	**-12%*** [-16%, -8%]	**-6%*** [-10%, -2%]	**37%*** [10%, 64%]	**-8%** [-13%, -3%]	-4% [-13%, 5%]
	Complex	3% [-1%, 7%]	-4%* [-14%, 6%]	-1%* [-10%, 8%]	2% [-7%, 3%]	**14%** [5%, 23%]

In table 17, we provide updated OB uplift recommendations[25], calculated as:

$$New\ OB\ Uplift = (1 + Current\ Contingency) * (1 + Cost\ Escalation) - 1$$

Table 17.Recommended OB uplifts.

Project category	Complexity (Consortium)	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
Renewals	*Simple*	56%	29%	-1%	3%	-2%
	Complex	67%	91%	33%	24%	7%
Enhancements	*Simple*	41%	32%	37%	-8%	-4%
	Complex	65%	34%	-1%	2%	14%

7.2 Recommended OB uplifts with varying levels of disaggregation

The current WebTAG guidance recommends OB uplifts based on GRIP stage only. In Table 18 and 19, we provide recommended uplifts in case one would prefer to retain the current structure of the OB uplifts, i.e. based on GRIP stage only (Table 18), or based on GRIP stage and project type (Table 19) only.

Table 18. Recommended OB uplifts based on GRIP stage only

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
OB Uplifts	64%	63%	18%	9%	4%

Table 19. Recommended OB uplifts based on GRIP stage and project type only

Type	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
Renewals	65%	79%	18%	13%	3%
Enhancements	62%	33%	17%	0%[26]	8%

[25] Although the revised uplifts are not always significantly different from the contingencies currently in use, the revised numbers represent the best (unbiased) estimates of the escalations, and therefore the best choice of recommended OB uplift.

We also performed an analysis (see Appendix) on whether the recommended OB uplifts for each GRIP stage are significantly difference from each other. If they are not, then some stages could be combined, with a single OB uplift recommendation for several stages. The results show that the required OB uplifts in Tables 18 and 19 for GRIP stages 1 and 2 are not significantly different from each other, except for enhancements stages 1 and 2. Therefore, the more simplified recommendations for OB uplifts given in Tables 20 and 21 would be appropriate as well.

Table 20. Recommended OB uplifts based on aggregated GRIP stages

	GRIP 1/2	GRIP 3	GRIP4	GRIP 5
OB Uplifts	64%	18%	9%	4%

Table 21. Recommended OB uplifts based on aggregated GRIP stages and project type

Type	GRIP 1/2	GRIP 3	GRIP 4	GRIP 5
Renewals	66%	18%	13%	3%
Enhancements	60%	17%	0%	8%

The final choice of the level of granularity for making OB uplift recommendations depends on the trade-off between ease of implementation versus the benefit provided by the enhanced accuracy. However, we would recommend using either Table 20 or 21.

7.3 Robustness

Tables 22 and 23 show the confidence intervals around the estimated required OB uplifts.

Table 22. 90% confidence interval on recommended OB uplifts per GRIP stage

	GRIP 1/2	GRIP 3	GRIP4	GRIP 5
OB Uplifts	[62%, 66%]	[13%, 23%]	[6%, 12%]	[2%, 6%]

Table 23. 90% confidence interval on recommended OB uplifts per GRIP stage and project type

Type	GRIP 1/2	GRIP 3	GRIP 4	GRIP 5
Renewals	[63%, 69%]	[13%, 23%]	[10%, 16%]	[1%, 5%]
Enhancements	[57%, 63%]	[3%, 31%]	[-5%, 1%]	[1%, 15%]

[26] The average required uplift for GRIP 4 enhancements is -2%. However, we do not recommend using negative uplifts, and use 0% instead, which is within the confidence interval for this estimate, [-5%, 1%].

7.4 Implementation of recommended uplifts

For GRIP stages 1 and 2, the OB uplift should be applied to the point estimate, i.e. the base cost estimate, as follows:

$$Point\ Estimate * (1 + OB\ uplift\ \%)$$

In GRIP stages 3, 4 and 5, the OB uplift should be applied to the cost estimate based on the QRA mean[27], as follows:

$$QRA\ Mean * \ (1 + OB\ uplift\ \%)$$

Note that this differs from the current WebTAG guidance, which for GRIP stage 3 recommends applying an uplift to the point estimate rather than the QRA mean. We recommend changing this recommendation to reflect the fact that QRA has now become standard at NR for project appraisals, and that the QRA results can therefore be used in conjunction with a lower uplift value.

If we compare the recommendations in Table 21 with the current WebTAG guidelines, the differences are rather minor:

- GRIP 1: The recommended 66% uplift is confirmed for renewals, but needs to be lowered slightly for enhancements, to 60%.
- GRIP 2: The recommended 50% uplift needs to be increased, to 66% for renewals, and 60% for enhancements;
- GRIP 3: The recommended 40% uplift based on the cost point estimate needs to be replaced by a 18% or 17% uplift applied to the QRA mean for renewals and enhancements, respectively.
- GRIP 4: The recommended 18% uplift needs to be lowered to 13% for renewals, and no uplift is required for enhancements;
- GRIP 5: The recommended 6% uplift needs to be modified to 3% and 8% for renewals and enhancements, respectively.

Alternatively, for GRIP stages 1 and 2, OB uplifts could be applied to the AFC estimates instead of the cost point estimates, resulting in the OB uplifts in tables 24-25.

Table 24. Recommended OB uplifts based on AFC for GRIP stages 1 and 2

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
OB Uplifts	3%	17%	18%	9%	4%
Applied to	AFC	AFC	QRA mean	QRA mean	QRA mean

[27] The QRA mean is the mean (average) of the project cost probability distribution, determined using a Monte Carlo simulation analysis. It is typically higher than the (most likely) point estimate because of the right-skewed nature of the cost distribution. Therefore, using the QRA mean instead of the point estimate typically results in an increased cost estimate, and reduced optimism bias. At NR, the term "QRA" is often used to indicate the difference between the QRA mean and the point estimate, i.e. as an addition to the point estimate. We do not recommend this, but instead recommend using the term "QRA mean" as the overall average cost.

Table 25. Recommended OB uplifts based on AFC for GRIP stages 1 and 2

Type	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5
Renewals	3%	28%	18%	13%	3%
Enhancements	1%	0%[28]	17%	0%	8%
Applied to	AFC	AFC	QRA mean	QRA mean	QRA mean

We also recommend the following actions for DfT and NR on how to further improve their investment appraisal and risk management framework.

- First, we would ideally recommend that OB uplifts are used in the early stages of project appraisals, with robust risk assessment replacing OB uplifts in the later stages. We therefore commend NR's approach of using risk uplifts for GRIP 1 and 2 estimates only, and relying on the QRA for providing robust estimates for GRIP 3, 4 and 5. The analysis clearly shows that cost escalations in later GRIP stages are small, proving that the QRA procedures currently in place sufficiently cover for the risk of cost escalations in later stages. The remaining small cost escalations that still exist in late-stage appraisals indicate, however, that the risk assessment procedures should be slightly adjusted, taking into account bigger ranges (eliminating the effect of so-called overconfidence). This could eliminate the need for OB uplifts in stages 3, 4 and 5 in the longer term.
- Second, OB uplifts should not be added to a project's budget, as this might lead to overspending. We found that for more than a quarter of projects, the final cost was identical to the GRIP-1 AFC, meaning that the 60% contingency was fully used, but nothing more. We would expect projects to come in below and above the AFC, but actual total costs being exactly equal to the AFC might indicate that money was unnecessarily spent[29]. Note that the recommended OB uplifts presented above are averages, with most projects being delivered with a cost escalation below the average. Adding the average escalation to every project's budget will therefore overestimate the cost for the majority of projects. Instead of adding OB to budgets, one should use the OB uplifts to confirm that a business case remains robust if the costs do rise to this level. Additionally, other than confirming viability of a business case, a good use of OB uplifts could also be in the form of a contingency held at the portfolio level.
- Third, we recommend that NR improves its data capture for future projects, creating standardised data entry procedures, unified templates, and in particular, records any applied OB uplift and risk assessment separately, and provides cost estimates for the entire project, even if approval is sought for one stage only. Also, we recommend that when a QRA is carried out, the 80% and 90% values of the QRA distribution should be recorded as well.

[28] The average required AFC uplift for GRIP 2 enhancements is -4%. However, we do not recommend using negative uplifts, and use 0% instead, which is within the confidence interval for this estimate, [-14%, 4%].

[29] An alternative explanation could be that when a project reaches its allocated budget, its scope is cut in order to contain the costs within the budget.

7.5 Quality Assurance

In this section we briefly describe the quality assurance (QA) steps we undertook to ensure the accuracy of the results reported in this report. The QA process consisted of two parts, namely verifying the accuracy of the data, and of the analysis.

To validate the data, project entries were checked line-by-line, and project entries with extreme reported cost escalations were sent to NR in order to obtain feedback on the validity of the recorded values. Some of these entries were corrected by NR while others were eliminated from the dataset altogether, as they could not be verified. No project entries with extreme escalations were retained in the final dataset. In addition to manual inspection, automatic cleansing operations were also used, removing invalid project entries and outliers. All automated operations were performed in the R programming language, with *RStudio* as the Integrated Development Editor. Although it is not possible to guarantee that no errors remain in the data, the removal of extreme observations, the line-by-line validation performed by NR and the large amount of project data gives us the confidence that our recommendations are unlikely to be altered significantly should further data errors be detected.

As mentioned earlier, a limitation of our analysis is the fact that allowances for risk and optimism bias are not recorded separately in the databases, leaving us unable to determine the individual uplifts present in each project's AFC. Therefore, we had to make certain assumptions. We chose to assume that the NR guidelines were followed when determining the AFC for each project, which allowed us to calculate the underlying cost point estimates. If, however, these guidelines were not adhered to, then our estimates of the cost point estimates could be incorrect. We note, however that as long as the guidelines were followed on average, i.e. with some uplifts higher and others lower than what the guidelines prescribe, then our recommendations remain valid.

The main parts of the analysis that involved generating the tables, the figures, the hypothesis tests and the regression models, were implemented both in R and in Excel by three different teams, in order to ensure the validity of the conclusions. In particular, the dataset was first analysed in R, using standard functions for calculating the mean, sample size, performing hypothesis tests and regression models. The analysis was then re-done in Excel, first using manual excel formulas, and then using pivot tables.

8. Bibliography

Ashley, M., W. Boxall, R. Halstead, J. Harrison, M. Bartlett, T. Rose, I. Murphy, G. Thomas, S. Longstaffe, A. Garlick, D. Penhallurick. *Managing Cost Risk & Uncertainty in Infrastructure Projects.* Infrastructure Risk Group Report. 2013.

Cost Engineering Terminology, *Recommended Practice 10S-90*, AACE International, WV. 2007.

Department for Transport. *Transport Analysis Guidance (TAG) Unit A5.3 - Rail Appraisal*. November 2014. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/381943/TAG_Unit_A5.3_-_Rail_Appraisal_November_2014.pdf. Accessed March 2015.

Flyvbjerg, B. *The Oxford Handbook of Project Management.* Oxford: Oxford University Press. 2011.

John Filicetti, *PMO and Project Management Dictionary*. 2007.

Halcrow Group. *Optimist Bias Study.* Halcrow Group. 2009.

HM Treasury guidance. *Early financial cost estimates of infrastructure programmes and projects and the treatment of uncertainty and risk*. March 2015. https://www.gov.uk/government/publications/green-book-supplementary-guidance-valuing-infrastructure-spend/early-financial-cost-estimates-of-infrastructure-programmes-and-projects-and-the-treatment-of-uncertainty-and-risk#the-nature-and-structure-of-early-cost-estimates. Accessed July 2015.

Kahneman, D. “New challenges to the rationality assumption.” *Journal of Institutional and Theoretical Economics* 150, 18-36. 1994.

Kahneman, D. *Thinking fast and slow.* New York: Farrar, Straus and Giroux. 2011.

Kahneman, D. and D. Lovallo. "Timid choices and bold forecasts: A cognitive perspective on risk taking", *Management Science* 39. 1993.

Kahneman, D. and A. Tversky. “Intuitive prediction: biases and corrective procedures.” *TIMS Studies in Management Science* 12, 313-327. 1979.

Mott MacDonald. *Review of Large Public Procurement in the UK*. 2002.

Network Rail. *Cost estimating, cost analysis and benchmarking*, Issue 2.0. 2014.

Sweett Group. *Benchmarking Report*. 2012.

A. Appendix

In this appendix, we elaborate on the technical analyses that we performed for determining which project characteristics have a significant effect on cost escalations, and could therefore be used in fine-tuning OB uplifts based on a project's characteristics.

A1 Hypothesis Testing

We examined the impact of several project characteristics using hypothesis testing. Here we report on some of our most important results. Additional experiments that check the robustness of certain threshold values are omitted for the sake of brevity.

A1.1 GRIP Stages

Tables A1, A2 and A3 show the result of a t-test[30] that checks the hypothesis that the different GRIP stages require different OB uplifts, for all projects (table A1), renewals (Table A2) and enhancements (Table A3). The results show that the required OB uplifts for stages 1, 2 and 3 do not necessarily need to be different, which means that a single recommendation for all three stages would also be appropriate.

Table A1. t-test results for differences in GRIP stage OB uplift recommendations

All Projects	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
GRIP 1		Inconclusive	**Significant**	**Significant**	**Significant**	**Significant**
GRIP 2			**Significant**	**Significant**	**Significant**	Inconclusive
GRIP 3				Inconclusive	**Significant**	**Significant**
GRIP 4					**Significant**	Inconclusive
GRIP 5						**Significant**

Table A2. t-test results for differences in GRIP stage OB uplift recommendations – renewals only

Renewals	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
GRIP 1		Inconclusive	**Significant**	**Significant**	**Significant**	**Significant**
GRIP 2			**Significant**	**Significant**	**Significant**	Inconclusive
GRIP 3				Inconclusive	**Significant**	**Significant**
GRIP 4					**Significant**	**Significant**
GRIP 5						**Significant**

Table A3. t-test results for differences in GRIP stage OB uplift recommendations – enhancements only

Enhancements	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
GRIP 1		**Significant**	**Significant**	**Significant**	**Significant**	Inconclusive
GRIP 2			Inconclusive	**Significant**	**Significant**	Inconclusive
GRIP 3				Inconclusive	Inconclusive	Inconclusive
GRIP 4					Inconclusive	**Significant**
GRIP 5						Inconclusive

[30] Assuming unequal variances and applying the Welsh modification.

A similar analysis, but taking into account project complexity, yields the results in Table A4. The results indicate that GRIP stages 2 and 3 could be combined, but there are some significant differences between stages 1 on the one hand, and stages 2 and 3 on the other.

Table A4. t-test results for differences in GRIP stage OB uplift recommendations

GRIP Stage Comparison	Renewals Simple	Renewals Complex	Enhancements Simple	Enhancements Complex
GRIP 1 - 2	**Significant**	Inconclusive	Inconclusive	**Significant**
GRIP 1 - 3	**Significant**	**Significant**	Inconclusive	**Significant**
GRIP 1 - 4	**Significant**	**Significant**	**Significant**	**Significant**
GRIP 1 - 5	**Significant**	**Significant**	**Significant**	**Significant**
GRIP 2 - 3	**Significant**	**Significant**	Inconclusive	**Significant**
GRIP 2 - 4	**Significant**	**Significant**	**Significant**	**Significant**
GRIP 2 - 5	**Significant**	**Significant**	**Significant**	Inconclusive
GRIP 3 - 4	Inconclusive	Inconclusive	**Significant**	Inconclusive
GRIP 3 - 5	Inconclusive	**Significant**	Inconclusive	Inconclusive
GRIP 4 - 5	Inconclusive	**Significant**	Inconclusive	Inconclusive

A1.2 Renewals vs Enhancements

Table A5 shows the result of a t-test[31] that checks the hypothesis that renewals and enhancements require different OB uplifts per GRIP stage. Although the results are not conclusive for every stage, overall there seems to be a significant difference.

Table A5. t-test results for renewals versus enhancements

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Renewals ≠ Enhancements?	Inconclusive	**Yes**	Inconclusive	**Yes**	Inconclusive	**Yes**

A1.3 Project Size

Table A6 shows the result of a t-test that checks if projects classified as large require significantly higher OB uplifts compared to small projects. Clearly, the data does not support this hypothesis.

Table A6. t-test results for Small and Large projects across the GRIP stages

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Small ≠ Large?	Not enough observations	Inconclusive	Not enough observations	Inconclusive	Inconclusive	Inconclusive

A1.4 Project Complexity

Table A7 shows the result of a t-test that checks if projects of high complexity, i.e. consortium projects, require significantly higher OB uplifts compared to NR-only projects. The test provides very strong evidence that there is indeed a difference. Further, Table A8 shows the result of a t-test that checks the

[31] Assuming unequal variances and applying the Welsh modification.

same hypothesis, but with complexity defined as projects that are multi-funded or multi-asset. Clearly, this classification does not provide conclusive results.

Table A7. t-test results for complex (consortium) versus simple (NR-only) projects

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Complex ≠ Simple?	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Table A8. t-test results for complex (multi-funded or multi-asset) versus simple projects

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Complex ≠ Simple?	Inconclusive	**Yes**	Inconclusive	Inconclusive	Inconclusive	Inconclusive

A1.5 Project Age

Table A9 shows that overall, there is a significant difference between the OB uplift required for older versus more recent projects.

Table A9. t-test results for old (≤2008) versus new (>2008) projects across the GRIP stages

	GRIP 1	GRIP 2	GRIP 3	GRIP 4	GRIP 5	Overall
Old ≠ New?	**Yes**	Inconclusive	**Yes**	**Yes**	Inconclusive	**Yes**

A2 Regression Modelling

We have run several regression models in an attempt to assess the predictive power of each of the project characteristics. Here we report on the results of a generic regression model that takes into account the following characteristics:

- GRIP stage
- Project type
- Complexity (measured both by "Consortium" and "multi-funded, multi-asset")
- Asset category
- territory
- Project size
- AFC (logarithmic model)

After using stepwise regression, the optimal forecasting model contains the following variables:

- Complexity (both measures)
- Asset category
- Log(AFC)

According to the results in Table A10, the vast majority of asset categories have non-significant coefficients. After eliminating these variables, the optimal model contains the following variables (See Table A11):

- Complexity (both measures)
- Log(AFC)
- GRIP stage
- Project type

We therefore decided to use these variables in our further analysis.

Table A10. Regression model with the cost escalation as dependent variable. The model has an R^2 adjusted of 8%. An asterisk (*) denotes a significant variable at the 95% level.

Variable	Estimate	Std. error	t-value	p-value
(Intercept)	0.78	0.16	4.84	0.00*
log(AFC)	-0.07	0.01	-6.69	0.00*
GRIP == 2	0.15	0.08	1.88	0.06
GRIP == 3	0.18	0.04	4.41	0.00*
GRIP == 4	0.10	0.03	3.07	0.00*
GRIP == 5	0.05	0.03	1.68	0.09
MF.MA	0.20	0.05	4.54	0.00*
Consortium	0.17	0.02	7.54	0.00*
Category - Ayrshire Inverclyde	3.69	0.50	7.33	0.00*
Category - Civils CP4 Enhd Spend	0.06	0.11	0.56	0.58
Category - Commercial Property	-0.15	0.29	-0.50	0.62
Category - Corporate Offices	-0.16	0.36	-0.43	0.67
Category - CP5 Development Fund	-0.18	0.26	-0.68	0.50
Category - Cross Rail	-0.12	0.20	-0.59	0.56
Category - Depots MDU	-0.02	0.36	-0.06	0.95
Category - Discretionary Investment	0.22	0.23	0.97	0.33
Category - Earthworks	0.02	0.06	0.24	0.81
Category - ECML Improvements	-0.18	0.26	-0.72	0.47
Category - Electrification	0.00	0.08	0.00	1.00
Category - Enhancements	-0.01	0.36	-0.03	0.98
Category - Franchise Stations	0.07	0.07	0.99	0.32
Category - HLOS Performance Fund	-0.08	0.11	-0.76	0.45
Category - IM TMS	0.23	0.51	0.45	0.65
Category - IT	0.00	0.08	0.01	0.99
Category - Kings Cross	0.15	0.36	0.43	0.67
Category - LMD	-0.09	0.12	-0.80	0.42
Category - Lvl Crossing Risk Red Fund	0.09	0.12	0.76	0.45
Category - MDU	-0.12	0.13	-0.92	0.36
Category - Mid Tier Accessibility	-0.03	0.21	-0.14	0.89
Category - Midlands Improvement Prog	-0.10	0.26	-0.39	0.69
Category - Northern Urban Centre Leeds	-0.04	0.29	-0.14	0.89
Category - Northern Urban Centre Mcr	0.10	0.51	0.20	0.85
Category - NRDF	-0.05	0.09	-0.62	0.53
Category - NSIP	0.02	0.07	0.27	0.79
Category - NW Elec	-0.05	0.26	-0.20	0.84
Category - Operational Property	0.05	0.09	0.57	0.57
Category - Orbis	0.02	0.36	0.06	0.96
Category - Out Performance	-0.05	0.26	-0.20	0.84
Category - Performance Recovery	-0.08	0.13	-0.60	0.55
Category - Plant	-0.04	0.08	-0.46	0.65
Category - Platform Lengthening South	0.01	0.14	0.09	0.93
Category - PR08 Enhancement Other	0.18	0.50	0.36	0.72

Category - PSU	-0.31	0.23	-1.36	0.18
Category - RAB	-0.02	0.12	-0.20	0.84
Category - Reading	0.43	0.29	1.46	0.14
Category - Retail	0.23	0.23	0.98	0.33
Category - S&C	0.14	0.19	0.74	0.46
Category - S&E Plan	0.16	0.11	1.45	0.15
Category - Scotland Projects	0.14	0.30	0.47	0.64
Category - Scottish Small Proj Fund	1.14	0.30	3.84	0.00*
Category - Self Financing (Cost) Oth	0.10	0.16	0.65	0.52
Category - Self Financing (Income) Oth	0.03	0.51	0.05	0.96
Category - Seven Day Railway	0.17	0.10	1.70	0.09
Category - SFN	-0.07	0.23	-0.30	0.76
Category - Signalling	0.12	0.07	1.72	0.09
Category - Southern Capacity	-0.07	0.20	-0.35	0.72
Category - Spacia	-0.12	0.17	-0.69	0.49
Category - Station Security (E & W)	-0.36	0.50	-0.71	0.48
Category - Stations	0.10	0.26	0.38	0.71
Category - Structures	0.16	0.06	2.54	0.01*
Category - Telecoms	-0.00	0.07	-0.04	0.97
Category - Thameslink	0.50	0.18	2.82	0.00*
Category - Tier 3 Project Development	0.06	0.30	0.20	0.84
Category - Track	0.16	0.10	1.68	0.09
Category - WCRM Renewals	-0.08	0.30	-0.29	0.77
Category - Western Improvements Prog	-0.32	0.29	-1.07	0.28

Table A11. Regression model with the cost escalation as dependent variable. The model has an R^2 adjusted of 5%. An asterisk (*) denotes a significant variable at the 95% level.

Variable	Estimate	Std. Error	t-value	p-value
(Intercept)	0.74	0.13	5.82	0.00*
log(AFC)	-0.07	0.01	-7.21	0.00*
GRIP == 2	0.21	0.04	5.40	0.00*
GRIP == 3	0.13	0.03	4.30	0.00*
GRIP == 4	0.07	0.03	2.45	0.01*
GRIP == 5	0.04	0.02	1.60	0.11*
Type Renewals?	0.21	0.04	4.75	0.00*
MF.MA	0.17	0.02	7.49	0.00*
Consortium	0.74	0.13	5.82	0.00*

B. Glossary

In this glossary, we give brief definitions of terms that are used throughout the report. To align with terminology currently used by NR, we employ some of the definitions given in NR's cost estimating report (NR 2014) and in the relevant HM Treasury guidance (HM Treasury 2015).

Allowance

An allowance is a percentage add-on for *foreseen* cost elements, which might relate to:

- Contractor preliminaries
- Feasibility
- Design
- Project management
- Sponsor-related costs
- Environmental mitigation
- Safety management

This list is not exhaustive. NR utilises allowances for cost estimates in GRIP stages 1 and 2.

Anticipated Final Cost (AFC)

An AFC is a valuation (or cost forecast) of work completed (invoiced) on a project at a specific time in its lifecycle (normally applicable from the commencement of the construction phase – or contract award). It also includes a valuation of outstanding work to complete the project and usually includes a cost of agreed variations, claims and residual risk. The AFC is typically comprised of a point estimate and an allowance (for optimism bias and risk) in GRIP stages 1 and 2, and a QRA mean estimate with added risk contingency in GRIP stages 3, 4 and 5.

Budget

The total amount of financial resources authorized or allocated for the particular purpose of carrying out a specific project scope in a specific period of time.

Contingency

An amount added to an estimate to allow for items, conditions, or events for which the state, occurrence, or effect is uncertain and that experience shows will likely result, in aggregate, in additional costs. Contingencies are typically estimated using statistical analysis or based on judgment based on past experience. A contingency usually excludes:

- Major scope changes, such as changes in end product specification, capacities, building sizes, and location of the asset or project.
- Extraordinary events such as major strikes and natural disasters.
- Management reserves.
- Escalation and currency effects.

Some of the items, conditions, or events for which the state, occurrence, and/or effect is uncertain include, but are not limited to, planning and estimating errors and omissions, minor price fluctuations (other than general escalation), design developments and changes within the scope, and variations in market and environmental conditions. Contingency is generally included in most estimates, and is expected to be expended (Cost Engineering Terminology, 2007).

Point estimate

A point estimate is the total capital cost estimate excluding risk allowance and represents the most likely estimate. It provides an estimate of the base construction works (direct construction costs and indirect construction costs) and project / design team Fees and other project costs.

Project appraisal

The process of calculating a project's viability (Filicetti, 2007).

Risk

An uncertain event that, when materialised, has a tangible impact on a project's objectives (duration, scope or budget).

Risk Allowance

A risk allowance needs to account for the cost consequences of (HM Treasury, 2015):

- The development and refinement of the design.
- The greater understanding of the solution's interfaces with its physical environment.
- Legitimate changes in requirement scope[32].
- A reducing provision for other areas of uncertainty which are not addressed by the above 3 bullets (as provided for through Optimism Bias in earlier high level assessments).
- Specific risks e.g. changes in key personnel during the project, pending legislative changes which would impact on the project.

[32] If within the functional output commitment of the project.