Issue 21(2), 2021, pp. 120-136
https://doi.org/10.18757/ejtir.2021.21.2.5504

TU Delft OPEN

EJTIR

ISSN: 1567-7141
http://ejtir.tudelft.nl/

Curbing cost overruns in infrastructure investment: Has reference class forecasting delivered its promised success?

Jung Eun Park[1]
Urban Planning, Columbia University, USA

Infrastructure projects around the world have long been notorious for exceeding their budgets. To address persistent cost overruns, the American Planning Association urged planners to adopt reference class forecasting alongside traditional methods but the practice has not caught on in the U.S. Conversely, the U.K. adopted Kahneman's Nobel Prize-winning theory to challenge biases in human judgment and mandated reference class forecasting for major projects in 2003. Has reference class forecasting, originally developed to rectify honest mistakes, brought the promised success in the public sector wherein political pressure is significant? Through before-and-after and with-and-without comparisons of 107 major projects, this empirical study examines the practical relevance of reference class forecasting for infrastructure investments. A before-and-after comparison reveals that the average cost overrun declined from 38% to 5% following the introduction of reference class forecasting. A with-and-without comparison also demonstrates that the U.K. surpassed its targeted probability of completing projects within budget by 12% using reference class forecasting, whereas the U.S. underperformed by 17%. Thus, reference class forecasting has engendered notable improvements in estimation in the U.K. This empirical study demonstrates the benefits of supplementing or replacing the current forecasting method. The findings can be used to reduce substantial financial risks for the government as well as social and economic welfare losses for society.

Keywords: *cost overrun, major project, optimism, reference class forecasting, strategic misrepresentation*

---

[1] E: jep2104@columbia.edu / jep2104@gmail.com

1. Introduction

According to the World Bank, the world is experiencing the biggest infrastructure investment boom in human history and $9 trillion - or 14% of global GDP - is annually devoted to million and billion-dollar projects (McKinsey, 2017).[2] Infrastructure projects, however, often surprise the public by exceeding the budgets (Winch, 2012). Flyvbjerg, Garbuio, & Lovallo find that cost overruns of 50% are common and cost overruns above 100% are not uncommon (Flyvbjerg, Garbuio, & Lovallo, 2009). Her Majesty's Treasury (HM Treasury) also finds that the average level of underestimation of large public procurements in the U.K. is 38% for capital expenditures (Mott MacDonald, 2002). The discrepancy between the estimated and the observed has incurred severe criticism (Kitamura, Yoshii, & Yamamoto, 2009).

Most literature explains inaccuracies in the cost, schedule, and benefits forecasting of infrastructure investment relative to technological biases such as unreliable or outdated data and inappropriate models. Undoubtedly, huge investments have been made for many years in the development of sophisticated models and in the update of massive data. However, there is little evidence to indicate that the ability to forecast contingency values is getting any better (Batselier & Vanhoucke, 2016, 2017; Bordley, 2014; Chang et al., 2016; Flyvbjerg, 2008; Flyvbjerg, Holm, & Buhl, 2002; Horne, 2007; Kahneman, 2011). Attributing persistent cost overruns to psychological bias, the U.K. first adopted Kahneman's Nobel Prize-winning theory to challenge cognitive biases in infrastructure investment decisions. In 2003, the U.K. made reference class forecasting mandatory, followed by the European Union, Denmark, Germany, Norway, Sweden, Switzerland, and the Netherlands.[3] In 2005, the American Planning Association urged planners to use reference class forecasting in addition to conventional methods but the practice has yet to catch on in the U.S. (American Planning Association, 2005).

Reference class forecasting is a method that aims to predict the future by looking at similar situations in the past and their outcomes. Originally, reference class forecasting was developed to moderate the effects of optimism in human judgment. For this reason, Flyvbjerg, who introduced the idea in the practice of urban planning, claims that the potential for reference class forecasting is low where political-economic bias is the chief reason for the inaccuracy (Flyvbjerg, 2006). On the other hand, Kahneman, who is the originator of the theory, believes that reference class forecasting helps to mitigate not only psychological bias but also political-economic bias because the approach bypasses such biases by directly cutting to empirical results and constructing estimates based on those (Kahneman, Personal Communication, November 26th, 2015 and December 17th, 2015). The problem is whose argument should be trusted. Unfortunately, there is very limited information documenting the use and effectiveness of reference class forecasting in the public procurement at this time. Therefore, this study aims to answer whether the impact of political-economic bias can be efficiently minimized with the remedy to cure psychological bias and whether reference class forecasting helps to attain the promised success in assisting decisions in infrastructure investment.

2. Literature Review

2. 1. Psychological and Political Economic Biases as the Reasons for Forecasting Inaccuracy
What explains the occurrence of cost overruns during the delivery of infrastructure projects even after the application of various methods for handling uncertainties? There are sharp divergences between auditors prioritizing the technical and managerial explanations, such as inadequate data

---

[2] The World Bank's definition of infrastructure includes utilities, public works, and other transport sectors.
[3] HM Treasury has mandated reference class forecasting in relation to costs, duration, and benefits. However, the Department for Transport applies the method to costs only.

and bad forecasting techniques, and scholars prioritizing the psychological and political-economic explanations (Siemiatycki, 2009). Unlike the auditors, the scholars believe that technical or managerial explanations do not sufficiently explain the reasons for significant cost overruns and schedule delays in public procurement (Chapman, 2010). According to much of the academic literature, the underlying reasons for forecasting errors are broadly divided into two categories: (1) honest mistakes or delusions and (2) intentional manipulation or deception (Flyvbjerg, Garbuio, & Lovallo, 2009).

According to HM Treasury, there is a "high level of optimism in project estimates arising from underestimating project costs" (Mott MacDonald, 2002). Such cognitive biases can cause a tendency to underestimate risk and to be overconfident assuming that things will not go wrong (Kahneman & Tversky, 1979, Kitsos et al., 2013). One of the reasons for this poor decision stems from our tendency to focus too much on internal matters and to ignore what is outside the circle. As a result, we are unable to think beyond what is directly in front of us.

While other dynamics such as scope creep, ground conditions, technical and managerial difficulties, material or labor price changes, and improperly managed risk and uncertainty are also at work, political-economic explanations best account for the available evidence (Ahiaga-Dagbui & Smith, 2014; Gil & Lundrigan, 2012; Love, Edwards, & Irani, 2011; Okmen & Oztas, 2010). Planners and local governments know that if the real estimates were to be published, the project would never get sanctioned. Therefore, they deliberately misrepresent costs, demand, and risks to increase the probability for their projects, and not those of their competitors, to gain approval and funding (Batselier & Vanhoucke, 2016, 2017; Chang et al., 2016; Flyvbjerg, 2007, 2008; Flyvbjerg, Bruzelius, & Rothengatter, 2003; Flyvbjerg, Holm, & Buhl, 2002, 2006; Horne, 2007; Kahneman, 2011). This not only explains why the forecasts of infrastructure projects consistently fail but also explains why people have not learned to make such forecasts any better (Winch, 2012).

The inherent uncertainties of decision making in infrastructure investment mean that optimism[4] in good faith and strategic misrepresentation in bad faith are pervasive in predictions about the future. The problem is not with the models and data but within the individuals themselves. Thus, estimates will remain inaccurate unless we address the biases associated with predicting future events (Flyvbjerg, 2008, Rizzi, 2003, Liu et al., 2018).

2. 2. Reference Class Forecasting as the Answer to Cost Overruns in Large Procurement Projects

Prospect theory, for which Kahneman received the Nobel Memorial Prize in Economic Sciences, provides insights into errors in infrastructure investment decision making and possible remedies (Rizzi, 2003). One interesting finding of Kahneman and Tversky's research is that reliance on human estimates tends to result in an excessively optimistic view of the risks (Kahneman & Tversky, 1979).

The outside view, also known as reference class forecasting, totally ignores the specifics of the current project. Instead, it looks at the history of a class of comparable projects, assesses the rough

[4] **Optimism vs. Optimism Bias**
It is important to note that the term "optimism bias" has different meanings from that of the authors. HM Treasury defines optimism bias as "a measure of the extent to which actual project costs and duration exceed those estimated. (Mott MacDonald, 2002)." Flyvbjerg, who implemented Kahneman's prospect theory in the practice of urban planning, defines optimism bias as "a cognitive predisposition found with most people to judge future events in a more positive light than is warranted by actual experience (Flyvbjerg, 2007)." Together with strategic misrepresentation, which is explained as a deliberate and strategic behavior to increase the likelihood if gaining approval and funding, optimism bias, according to Flyvbjerg, accounts for inaccurate forecasts. Kahneman, who is the originator of the idea, defines optimism bias as any deception that causes optimism regardless of whether it is an honest mistake or has intentional sources (Kahneman, Personal Communication, 2015). Hence, Flyvbjerg's optimism bias is equivalent to what Kahneman defines as optimism and Kahneman's optimism bias is equivalent to what Flyvbjerg defines as optimism bias and strategic misrepresentation. This study follows Kahneman's definition in further discussion.

distribution of the outcomes for the reference class, and then positions the project at hand in that distribution (Lovallo & Kahneman, 2003). Through examination of data from both successful and unsuccessful past projects, decision-makers are forced to consider the full range of potential outcomes and look for external evidence that offers a more accurate context. Moreover, it is not necessary to try to calculate the exact outcomes of the project at hand. This prevents decision-makers from focusing on initiatives that are successful and close to the decision at hand and further prevents the natural and often cognitively impaired decisions (Flyvbjerg, Garbuio, & Lovallo, 2009). As a result, the outside view is more likely to produce accurate estimates than the inside view and much less likely to produce highly unrealistic ones (Lovallo & Kahneman, 2003).

To mitigate optimism bias, reference class forecasting positions a proposed action in the statistical distribution of the outcomes of comparable but already concluded actions. Reference class forecasting requires the following steps for a specific project. First, a relevant reference class of previous similar projects is identified to provide a context for a project under consideration. To be meaningful but also narrow enough to be comparable with the target effort, the selected reference class should be large enough. In order to apply the correct adjustment factors, projects need to be classified. If it is innovative, has mostly unique elements, or involves a high degree of complexity and difficulty, a project is considered non-standard. Second, probability distribution for the chosen reference class is established for the outcome in question. Third, the specific project is compared with the distribution of the reference class in order to determine where the current project should sit on the distribution and establish the most possible outcome for the specific project (Davenport & Harding, 2010; Jenner, 2016; Winch, 2012).

In practice, however, there is no standard practice code for reference class forecasting. There are variations in how the approach is implemented (See Technical Appendix 1 for the methodology of traditional methods).

In an attempt to address cost overruns, HM Treasury mandated reference class forecasting for HM Treasury projects larger than £40 million and for Department for Transport projects larger than £5 million in 2003 (Flyvbjerg, 2006; Flyvbjerg, Bruzelius, & Rothengatter, 2003; Mott MacDonald, 2002). In 2005, the American Planning Association announced that "APA encourages planners to use reference class forecasting in addition to traditional methods as a way to improve accuracy……Planners should never rely solely on civil engineering technology as a way to generate project forecasts" (American Planning Association, 2005). Reference class forecasting is increasingly being adopted in Europe to mitigate two main factors, optimism and strategic misrepresentation, which cause persistent cost overruns and schedule delays in infrastructure projects (Liu, Wehbe, & Sisovic, 2010).

2. 3. Does Reference Class Forecasting Effectively Detect and Cure Biases in Infrastructure Investment Decision?

Unlike private sector projects, the funding system of public sector projects encourages strategic misrepresentation (Flyvbjerg & COWI, 2004). To promote central government-funded capital investment programs, planners and local governments intentionally misrepresent costs, benefits, and risks to increase the probability of their projects gaining approval and funding (Flyvbjerg, 2007). While this practice is "not unknown in private sector ventures, it is not very common, simply because there is usually no taxpayer available to foot the bill later" (Merrow, 2011). This raises the following question. Are forecasts for public sector projects fundamentally different from those for private sector projects from which Kahneman has developed new theoretical approaches to decision sciences?

Kahneman and Tversky originally developed reference class forecasting to curve cognitive bias (Flyvbjerg, 2006; Kahneman & Tversky, 1979). Kahneman believes that reference class forecasting tends to reduce all types of human biases, including strategic misrepresentation, as long as a reality check is carried out by independent agencies, since the outside view bypasses such biases by directly cutting to empirical results and building estimates on them (Kahneman, Personal Communication, November 26th, 2015 and December 17th, 2015). Although both optimism and strategic misrepresentation have similar results, strategic misrepresentation differs in that it results from intentional sources. For this reason, Flyvbjerg, who introduced the idea in the practice of urban planning, claims that the potential for reference class forecasting is low where political-economic bias is the chief reason for the inaccuracy (Flyvbjerg, 2006).

Can the impact of strategic misrepresentation be efficiently minimized with the remedy to cure honest mistakes? Does reference class forecasting work as well where strategic misrepresentation is the main reason for inaccuracy?

2. 4. Arguments Against Reference Class Forecasting

There is, however, no silver bullet solution and reference class forecasting poses a range of significant issues (Jenner, 2016; Love and Ahiaga-Dagbui, 2018; Love et al., 2019; Moschoulia, Soeciptob & Vanelslandera, 2019; Themsen, 2019; Winch, 2012). Probability based approaches rely on accurate estimates of probabilities, but evidence shows that even experts can be woefully inaccurate when it comes to estimating probabilities (Ayers, 2007; Kahneman and Tversky, 1979; Taleb, 2004). Another point to note about expected value approaches, where different potential outcomes are given probabilities and the expected value is then the sum of these outcomes multiplied by their probabilities, is that the result is the average value that would be expected if the project was replicated a large number of times. In the real world of urban planning, however, this assumption is rarely correct. Furthermore, adjusting costs and demand for optimism bias based on past project performance might not be acceptable if we do not have accurate data on past performance or if future circumstances are likely to be different from those in the past. There is also the chance that planners and local governments will become much more optimistic in their forecasting as a result of the fact that their forecasts will be adjusted (Jenner, 2016). Most importantly, the suggested adjustments are large contingencies and will result in project budgets that are relatively generous overall. The presence of contingencies in the budget may itself create a moral hazard issue. If the funds are available, they are more likely to be spent than if they are not (Winch, 2012).

Has reference class forecasting overcome those drawbacks and delivered what it has promised to the U.K. over the past 17 years? Does reference class forecasting lead to more accurate cost estimates by counteracting psychological and political-economic biases?

2. 5. Gaps in Knowledge

Due to limited data availability, few studies have been able to provide evidence on the effectiveness of reference class forecasting and fewer still on its applications to infrastructure investment (Batselier & Vanhoucke, 2016; Liu, Wehbe, & Sisovic, 2010; Makovsek, Tominc, & Logozar, 2012). Even those few studies involve retroactive forecasting. Using only data that would have been available before forecasting, retroactive forecasting, or hindcasting, evaluates how well these simulated reference class forecasts would have performed compared to forecasts using traditional methods and the actual observations. Retroactive forecasting helps to a limited extent in assessing the expected performance of reference class forecasting. Because of its own nature that forecasts are made for already approved, funded, and completed projects, however, the retroactive forecasting approach is not suitable to provide evidence on the effectiveness of reference class forecasting which was developed to address optimism and strategic misrepresentation about an uncertain future.

Based on genuine forecasts of future events, therefore, this empirical study provides a statistical view of the impact of reference class forecasting in cost overruns of large transportation projects. As the before-and-after approach offers better evidence on intervention effectiveness than the other non-experimental approaches, the study demonstrates the immediate impacts of reference class forecasting through the before-and-after comparison of projects in the U.K. To capture time-varying factors, such as improvements in technologies and construction techniques, which can distort the impact of reference class forecasting over a more extended period, it further measures the outcomes for projects in a comparable country that did not adopt reference class forecasting but was subject to the similar set of environmental conditions. Finally, classical hypothesis tests are conducted to assess whether the differences in cost overruns are statistically significant between before and after and with and without reference class forecasting.

3. Methods

3. 1. Project Selection

In 2003, the U.K. first adopted reference class forecasting to challenge psychological biases in infrastructure investment decisions and made reference class forecasting mandatory. Therefore the U.K. was chosen for before-and-after and with-and-without evaluations. The practice has been followed by many European countries, but it has yet to catch on in the U.S. and Asia.

In 2004, the U.K. Department for Transport commissioned a study to implement reference class forecasting in transport planning. In an attempt to identify a relevant reference class of past similar projects for each type of transport scheme and to use it as a basis for applying reference class forecasting, the U.K. Department for Transport found that similar regulatory and construction schemes applied to U.S. projects but not to Asian projects. And testing further showed no significant differences between U.K. and U.S. projects (Flyvbjerg & COWI, 2004; Flyvbjerg, 2007). Therefore, the U.S. was chosen for the with-and-without evaluation. Although there are difficulties in formulating hypotheses free of unwarranted assumptions, the international performance comparison is expected to offer a good way to assess performance outcomes of policy measures (Phillips, 2019; Smith, 2001).

Due to the U.K.'s record management and retention policy of six years plus current after the last entry, records and information of the pre-reference class forecasting projects are no longer held. However, HM Treasury reviewed the outcome of large public procurement projects in the U.K. over the last 20 years in the preparation for adopting reference class forecasting in its treasury guidance to central government (Mott MacDonald, 2002). Therefore, this study compares the outcomes after the implementation of reference class forecasting with the known values from the HM Treasury cost overrun analysis prior to the implementation of the method. To compare like-for-like, the current study collects the post-implementation data following the same approach used for the previous HM Treasury study.

For the before-and-after evaluation, this study gathers the data of 39 large U.K. transportation projects that were approved after 2003, when reference class forecasting was first adopted in the HM Treasury's binding guidance for appraisal and evaluation in central government, and completed before 2019. Due to the paucity of large public procurements, this portfolio includes virtually all transportation projects on the database of the Infrastructure and Projects Authority (under the Cabinet Office and HM Treasury) constructed during the study period, whose total cost was over $100 million, and for which complete information is available.

For the with-and-without evaluation, this study gathers the data of 68 large U.S. transportation projects whose costs were estimated without reference class forecasting. As described in 23 U.S. Code 106(h), two types of projects require a financial plan in the U.S.: (1) the recipients of Federal financial assistance with a minimum estimated total cost of $100 million and (2) Transportation Infrastructure Finance and Innovation Act projects even if the projects receive no Federal-aid grant assistance. Following the same approach used for the U.K. project database, the U.S. portfolio includes virtually all transportation projects that require a financial plan, were approved after 2003, were completed before 2019, had a total cost over $100 million, and for which complete information is available.

In total, the project list consists of 107 large transportation projects. The project portfolio is worth approximately $91 billion. The size of projects ranges from $100 million to $5 billion with an average project size of $854 million. There are 30 projects exceeding $1 billion in value, 37 projects exceeding $500 million, and 40 projects exceeding $100 million (based on U.K. pound to U.S. dollar exchange rate for December 31st, 2003 from the Bank of England). The portfolio consists of three types of transport schemes: road (highway, local road, bus lane, fixed link), rail (high-speed rail, conventional rail, light rail, subway), and building projects (intermodal transit hub, station, terminal) (See Technical Appendix 2 for the limitation related to data access).

3. 2. Data Collection

In the study, it was difficult to obtain the data. Although there is strong interest in the topic, scholars have noted that key methodological limitations related to data access have made it difficult to shed light on the prevalence and causes of cost overruns in infrastructure investment (Nijkamp & Ubbels, 1999). In the case of this study, adequate data were not readily available. Unlike scholars, however, the Department of Transportation has a unique position within the governing process, enabling it to have far better access to data. As part of the government, the department is tasked with monitoring the accountability, effectiveness, efficiency, and integrity of public spending. To carry out its duties, the department has the authority to solicit internal documents from a state's Department of Transportation, project sponsors, and station staff members within a department or agency. Department of Transportation officials can also interview people. In the present study, therefore, data collection required their assistance. The Office of the Secretary of Transportation has helped facilitate discussions with the states' Departments of Transportation for this study and the Office of the Chief Counsel of the Federal Highway Administration has reached out to project sponsors to encourage them to share what materials they are able to with this study.

Thus, the financial plans of major transportation projects are the primary source of information for this study. The financial plan is a comprehensive document reflecting the scope, cost estimates, and funding structures of the project in order to provide reasonable assurance that it will be implemented and completed as planned. The financial plan must be submitted and approved prior to the first authorization of federal construction funds, authorization of the construction project, or initial procurement requests (Federal Highway Administration, 2014). Financial plans are drawn up and submitted by the project sponsor who is any entity providing project funding and administering any construction, construction engineering, or inspection activities for the project. When formally deciding to launch a project, the appropriate level of cost estimate is set out in the initial financial plan. Preparing the annual financial plan ensures that the required financial resources are identified, made available, and handled during the project's lifecycle. The content of the financial plans obtained with the assistance of the Department of Transportation was certified by the project sponsor's chief executive officer as correct and reasonable to the best of his knowledge. The business case is, likewise, one of the most important information sets available for significant projects in the U.K. It not only underpins the final investment decision but also justifies the financial investment over the entire project life-cycle (Highways Agency, 2013). In this study, a project was only included in the database when financial plans were available for it. Thus, relevant information for this analysis is primarily extracted from business cases and financial plans and further supplemented with appropriate documents from the time of the investment decision for each project in the portfolio.[5] Financial plans and business cases

[5] Unlike best and final funding bids or applications for final approval, the quality of financial plans and business cases varies by project sponsors or senior responsible owners.

obtained from the archives of the approving agencies for this study enable consistency in evaluating the cost performance.

Both the reliability of the data and the quality of information are essential for this type of analysis (Den Heijer, 2011). This study therefore seeks information of the best possible quality and builds on detailed expenditure data from the relevant institutions and not aggregate reports. This enabled the study to present comparably more accurate results.

3. 3. Methodological Procedure
The current study defines actual costs as actual, accounted costs calculated at the time of project completion and estimated costs as budgeted, or planned, costs including contingency allowances at the time of the construction approval in compliance with the U.K. Department for Transport guidelines (Mott MacDonald, 2002; Flyvbjerg, 2006; Flyvbjerg & COWI, 2004). This study covers all the costs and values of all the resources required to carry out preliminary engineering, right-of-way, environmental mitigation, construction, project management, public outreach, and external third-party work costs such as utility and railway relocations for the project cost. After taking inflation into account, it then computes cost overruns. This study analyses the differences between the final actual cost of each project and its cost estimates in their local currencies, rather than converting to the US dollar or UK pound to ensure that the forecast accuracy is not in affected by the exchange rate.

The dataset was divided by its location into U.K. projects and U.S. projects. Reference class forecasting was used as an estimating method in all U.K. projects and Monte Carlo simulation[6] was used in all U.S. projects. The dataset consists of three types of transport schemes: road (highway, local road, bus lane, fixed link), rail (high-speed rail, conventional rail, light rail, subway), and building projects (intermodal transit hub, terminal).

Although forecasting models can be quite complicated, their validation is straightforward (Makovsek, Tominc, & Logozar, 2012). In line with the definition provided by HM Treasury, this study defines and calculates the cost overrun as the relative increase in capital expenditure from what was expected to the actual capital expenditure at the time of investment decision (i.e., Cost Overrun (%) = 100 (%) X (Actual – Estimated) / Estimated) (Mott MacDonald, 2002).

To ascertain whether the difference in cost overruns between before and after reference class forecasting is statistically significant, this study tests the null and alternative hypotheses on datasets using the cost overrun (%) of each project. It uses a 0.05 significance level to evaluate the null hypothesis.

The degree of willingness to accept risk differs by projects and by governments. In the U.K., the Department for Transport's investment decisions are based on the P50 level of costs, unlike in the U.S. where the P70 level of costs is used. This means that if everything goes according to plan, 50% of projects in the U.K. and 70% of projects in the U.S. should be completed within their original budgets as proposed at the time of investment decision. Since the U.K. and the U.S. take different degrees of risk about the possibility of cost overruns, this study adjusts the estimated budgets of U.K. projects to the P70 level of costs for the degree of cost overrun comparison following the procedures of the U.K. Department for Transport to deal with optimism bias in transport planning (See Technical Appendix 3 for the procedure).

Since the U.S. does not provide the P50 level of costs, this study was unable to adopt traditional error measures. Instead, the probability of cost overrun comparison determines whether the actual

[6] Monte Carlo simulation is a mathematical method used to estimate the potential outcomes of an uncertain event. During World War II, John von Neumann and Stanislaw Ulam developed the Monte Carlo simulation to improve decision making under uncertain conditions. Monte Carlo simulations have assessed the impact of risk in many real-life situations such as cost estimation since its introduction (IBM Cloud Education, 2020).

probability of completing the project within the budget has reached 0.5 in the U.K. and 0.7 in the U.S. as promised through binomial tests.

3. 4. Limitations
This study only included completed projects in its database, not those that were cancelled or that are still under construction. Thus, it did not include many of the worst projects that were terminated before completion. Excluding these types of projects means that this study does not account for major transportation infrastructure costs: spending on facilities that do not end up being constructed but can still have an impact on project sponsors and taxpayers.

Moreover, the data only measured the cost increases that occurred after the projects had been fully designed. Thus, cost rises that occurred during project planning, where much of the cost increase occurs, were not reflected in the study's data.

4. Findings

4. 1. Cost Performance Before and After Reference Class Forecasting

Findings from the Previous Research: Before Reference Class Forecasting
To evaluate the propensity to underestimate the expense of current and future projects, past projects needed to be reviewed. As part of this exercise, HM Treasury commissioned a study in 2002 to review the outcomes of 50 large public procurement projects in the U.K. over the last 20 years (Mott MacDonald, 2002). Based on the findings of this extensive investigation, the U.K. became the first government in the world to mandate reference class forecasting in 2003.

As shown in Table 1, the average capital expenditure underestimation reported by the HM Treasury study was 38% for all 50 projects and 47% for 39 traditional procurement projects. The findings clearly demonstrate a tendency for project forecasts to be highly optimistic prior to the implementation of reference class forecasting (Mott MacDonald, 2002).

Table 1. Average Tendency for a Project's Cost to Be Underestimated: Before vs. After Reference Class Forecasting

Project Type	**Cost Underestimation Tendency (%)**	
	Before-Reference Class Forecasting	After-Reference Class Forecasting
Traditional Procurement	47	4
PFI/PPP	1	9
Total Procurement	**38**	**5**

Findings from the Current Research: After Reference Class Forecasting
Because of the U.K. record management and retention policy and the resultant unavailability of the data before the implementation, this study compares the post-implementation outcomes with the pre-implementation outcomes from the HM Treasury study. The HM Treasury study reviewed the results using the procurement method. To allow for comparison of results with the previous study, the current study also divides projects into Private Financial Initiative / Public Private Partnership (PFI/PPP) and traditional (i.e., non-PFI/PPP).

Table 1 shows the average cost overruns before and after reference class forecasting. Compared with the historical average of 38% for total procurement and 47% for traditional procurement, the average cost overrun of 5% for total procurement and 4% for traditional procurement following the adoption of reference class forecasting shows that the method has significantly improved cost overruns on large public procurement projects. The one-sample t-tests for equality of means confirm that the averages of cost estimation errors are significantly different before and after reference class forecasting (p = 0.0000 for both total and traditional procurement). The hypothesis is supported by a substantially higher average estimation error before the introduction of reference class forecasting.

It is noteworthy that reference class forecasting has not produced more accurate estimates for PFI/PPP projects after its implementation. It also means that 31 traditional procurement projects outperformed 8 PFI/PPP projects after the implementation of reference class forecasting. This result is quite contrary to the previous study.

When HM Treasury commissioned Mott MacDonald to conduct a study, Mott MacDonald classified projects into traditional procurement and PFI/PPP. It found that the cost overruns were 46 % higher for traditional procurement than for PFI/PPP. Mott MacDonald attributed this difference "to the negotiated transfer of project risks from the public sector to the private sector" and to "the high level of diligence demanded by PFI procurement to establish the business case" (Mott MacDonald, 2002). On the contrary, the result of the current study shows that the cost overruns for traditionally procured projects are 5% lower compared with those for PFI/PPP projects.

What, then, accounts for the difference? Although PFI/PPP is equally subject to optimism, above all, PFI/PPP is less subject to strategic misrepresentation. As private sector investors were hardly incentivized to make biased forecasts and to place their projects in a more flattering light, there was little room for improvement from the beginning. In addition, the small sample sizes may have precluded one from drawing a firm conclusion. Since only 11 and 8 PFI/PPP projects are included in the previous and current studies, the risk that observation is greatly affected by outliers is high. It is, therefore, necessary to interpret the findings of the previous and current studies on PFI/PPP projects with caution and to test them in future studies using larger size samples.

In summary, this study found that the method under study applied in addition to traditional methods produces accurate estimates that are significantly better than historical results recorded in the HM Treasury analysis right before the implementation of reference class forecasting.

4. 2. Cost Performance with and Without Reference Class Forecasting

Degree of Cost Overrun

As technologies and construction techniques improve, it may be construed that cost estimates have become more accurate over time without the implementation of reference class forecasting. To take into account time-varying factors, this study further measures the outcomes for projects in a comparable country that did not adopt reference class forecasting but was subject to a similar set of environmental conditions. Based on the U.K. Department for Transport's appraisal that similar regulatory and construction regimes apply to projects in the U.S. and subsequent findings from scholars that there is no significantly different performance between the two countries' projects, the U.S. was chosen for with-and-without comparison (Flyvbjerg & COWI, 2004; Flyvbjerg, 2007).

All projects involve risk. However, degrees of willingness to accept risk may vary by organizational level or by the regulatory regime to incentivize cost efficiency (Flyvbjerg, 2006). The U.K. decided to deliver projects within a budget with at least a 50% probability. In contrast, the U.S. wanted a higher degree of certainty, so the U.S. Department of Transportation required at least 70% certainty that there would be no cost overrun. Theoretically, the U.S. projects as a class are thus expected to have a bigger budget with a large contingency, to experience a smaller degree of cost overruns, and to show a lower frequency of cost overruns.

The average cost overrun of the original U.K. P50 estimates, as stated in Table 2, is 5.28%. Although P70 estimates are more conservative and should thus experience less cost overrun, the average U.S. P70 estimate experiences a cost overrun of 7.72% which is even higher than the cost overrun of the U.K. projects with P50 cost estimates. Compared with the average cost overrun of the adjusted U.K. P70 estimates of -4.72%, the accuracy of the U.S. P70 estimates is significantly different. With a 95% confidence level, Welch's two-sample t-test with unequal variance further confirms that the statistical significance of the difference in outcomes (p = 0.0322). The outcome indicates that the U.S. projects with a large contingency face more significant cost overruns. In short, the degrees of cost overruns seem to differ meaningfully with and without reference class forecasting.

Table 2. Cost Overruns

Project Samples	N	Mean (%)	Median (%)	Standard Deviation (%)	Minimum (%)	Maximum (%)
All						
UK (P50)	39	5.28	-4.76	39.69	-58.14	176.09
UK (P70: Adjusted)	39	-4.72	-13.76	35.83	-62.67	150.00
US (P70)	68	7.72	0.00	27.16	-33.33	126.62
Traditional Procurement						
UK (P50)	31	4.43	-4.76	41.98	-58.14	176.09
UK (P70: Adjusted)	31	-5.47	-13.76	38.01	-62.67	150.00
US (P70)	57	9.00	0.00	29.35	-33.33	126.62
Road						
UK (P50)	35	4.66	-5.88	39.71	-31.05	176.09
UK (P70: Adjusted)	35	-5.16	-14.43	35.04	-37.57	150.00
US (P70)	59	6.81	0.00	26.26	-27.21	126.62

This study further explores how forecasting errors vary for traditional procurement or road projects to measure the average cost overruns for similar projects. 88 traditional procurement projects represent 82% of the total projects under study (79% of the U.K. projects and 84% of the U.S. projects). Table 2 presents the overall results for all projects. Traditional procurement projects are substantially similar, although the difference between the U.K. and the U.S. grows slightly bigger. The difference between the U.K. P70 estimates and the U.S. P70 estimates for traditional procurement also reaches statistical significance (p = 0.0357).

Focusing exclusively on 94 road projects (35 U.K. projects and 59 U.S. projects) representing 88% of the total projects under study (90% of the U.K. projects and 87% of the U.S. projects), outcomes in Table 2 slightly differ. In contrast to the traditional procurement, the difference between the U.K. and the U.S. becomes smaller than the total procurement. As a result, the gap between the U.K. P70 estimates and the U.S. P70 estimates appears to reach the lower level of statistical significance for road projects (p = 0.0854).

Because of the limited number of projects and resultant lack of statistical power, this study leaves the PFI/PPP, rail, and building projects out of the sectoral review.

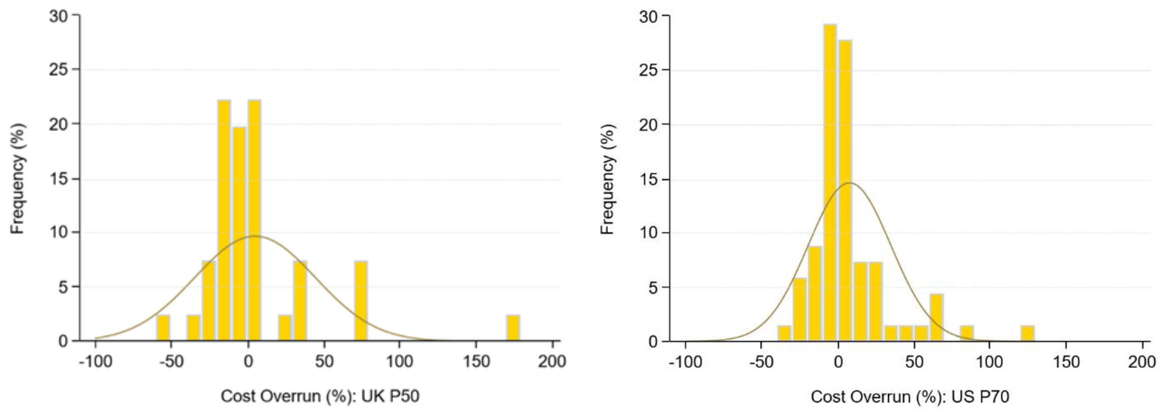

Figure 1. Cost Overrun Distribution: UK P50 vs. US P70

It is important to note that both distributions are skewed to the right and the tails of the distributions are longer on the right when measured from the estimates supporting investment decisions (i.e., from the U.K. P50 and U.S. P70 estimates). Although the P70 estimate should be skewed to the left (i.e., cost underrun) in theory, the frequency histogram and its normal density plot of the U.S. estimation errors are in fact skewed to the right (i.e., cost overrun). Although the U.K. P50 estimates should have similar dispersion around 0, the frequency histogram and its normal density plot are also skewed to the cost overrun to a lesser degree than the U.S. P70 estimates.

Given that reference class forecasting yields constant forecasts, it can be concluded that the method exhibits greater forecasting stability than the traditional method-only approach. The findings are consistent with Kahneman's theory that statistical distributional information help overcomes not only optimism but also strategic misrepresentation.

Probability of Cost Overrun

Since the U.K. and the U.S. require different degrees of certainty for their estimates, this study cannot use traditional error measures such as mean absolute percentage error and root mean square percentage error. Through binomial probability tests, this study instead tests whether the actual probability of completing the project within the budget has reached 0.5 in the U.K. and 0.7 in the U.S. as promised To see the relative frequency of success (i.e., completing a project within the budget) and deviation from their prediction, posterior predictive distributions are further presented.

As seen in Figure 2, the evidence-based assessment shows that 62% of the U.K. projects successfully completed within budget while 53% of the U.S. projects successfully completed within budget. Given that the U.K. and U.S. decided to remain within the budget with 50% and 70% certainty, the U.K. overachieved its expectations by 12%, while the U.S. fell short of its expectations by 17%. In short, the U.S. projects with a large contingency show a higher frequency of cost overruns. Consequently, the risks of cost overruns seem to differ meaningfully with and without reference class forecasting.

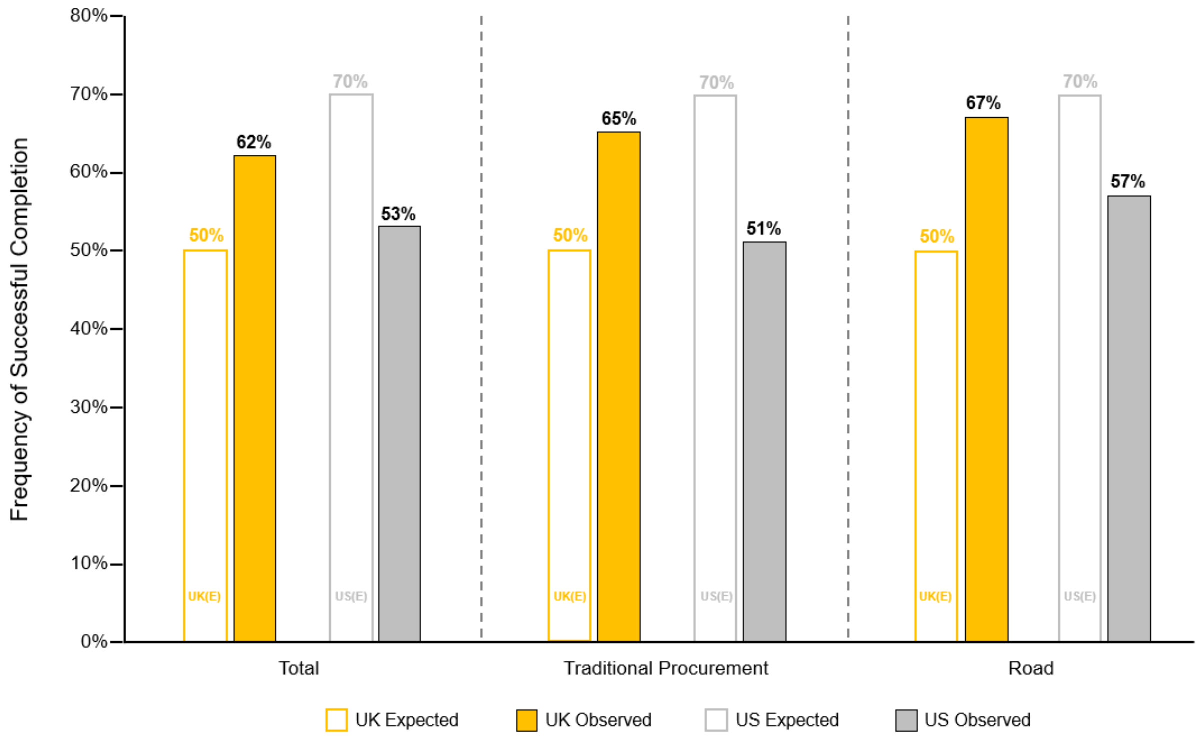

Figure 2. *Frequency of Successful Completion Without Cost Overrun*

To see if the observed results differ from what was expected in a statistically meaningful way, binomial probability tests are performed to the estimates which support investment decisions in each country: the U.K. P50 estimates and the U.S. P70 estimates. The null hypothesis that the probability of completing a project within the budget is at least 50% in the U.K. is accepted ($p = 0.999965$). On the contrary, the null hypothesis that the probability of completing a project within the budget is at least 70% in the U.S. is rejected at a 99% confidence level ($p = 0.002256$). This result confirms the earlier observation that the actual probability of success of the U.K. projects exceeds the assumed probability of success whereas that of the U.S. projects comes short of what the theory predicts. Consequently, the risks of cost overruns seem to differ meaningfully with and without reference class forecasting.

Figure 3 illustrates posterior predictive distribution of success to show an empirical distribution for success and its deviation from prediction. The vertical lines represent the expected probability of completing a project within the budget for each country. The posterior distribution for the U.K. projects (black curve) is positioned to the right of the expected P50 line (black line), while the posterior distribution for the U.S. projects (red curve) is positioned to the left of the expected P70 line (red line). Unlike the initial assumption, the U.K. P50 estimates are even positioned to the right of the U.S. P70 estimates. Ironically, it suggests that the U.K. P50 estimates are more likely to deliver projects within budget than the U.S. P70 estimates. From the posterior distribution, it is possible to predict the comparative accuracy of future observations through reference class forecasting.

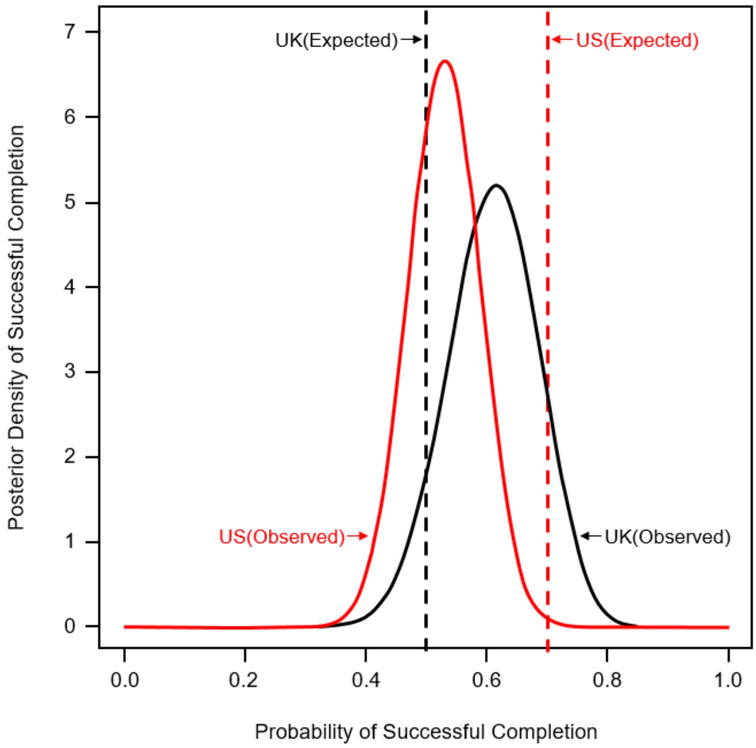

Figure 3. Posterior Distribution of Successful Completion Within Budget

In support of HM Treasury's view that there is a forecasting bias toward under-budgeting and that it can be mitigated by introducing reference class forecasting, the findings suggest that supplementing traditional methods with reference class forecasting is superior to using traditional methods alone. The results are consistent with Kahneman's theory that statistical distributional information help overcomes not only optimism but also strategic misrepresentation.

5. Conclusion

The main objective of this study is to examine the practical relevance of reference class forecasting to decision making in infrastructure investment where there are significant political pressures. By mean of a before and after comparison, the study finds that the average cost overrun is reduced from 38% to 5% for total procurement and from 47% to 4% for traditional procurement following the introduction of reference class forecasting. By means of a with and without comparison, the study also finds that the U.K. overachieved its targeted 50% probability of completing projects within budget by 12% with the benefit of reference class forecasting while the U.S. fell short of its targeted 70% probability by 17%. The results are consistent with Kahneman's theory that the impact of political-economic bias can be efficiently minimized with the remedy to cure honest mistakes. Reference class forecasting has attained the promised success in assisting decisions in infrastructure investment. As this is the case, we might not have to move beyond reference class forecasting just yet.

Studies on the accuracy of specific forecasting methods were called for as it would identify a fruitful direction in developing theories and practices for more accurate estimation (Batselier & Benhouke, 2017; Liu, Wehbe, & Sisovic, 2010). The U.S. Department of Transportation, for example, acknowledges that there is still room for improvement and therefore continues to work on enhancing its cost estimating procedures. It would like to consider supplementing its current processes with other industry-accepted cost estimating approaches after they have been sufficiently validated and proven to be effective. At the present time, however, there is very limited information documenting the use and effectiveness of reference class forecasting in the planning sector (The U.S. Department of Transportation, Personal Communication, February 8th, 2020). Consequently, the findings of this study contribute to the theory and practice by presenting empirical evidence on the accuracy of reference class forecasting. However, much further research is still needed in this area.

This study does not suggest that other conditions that contribute to cost overruns are not important. However, this study does suggest that optimism and strategic misrepresentation are the best explanations for cost overruns. In conclusion, reference class forecasting helped effective decision making wherein strategic misrepresentation is prevalent and has improved the quality of investment decisions on infrastructure.

References

Ahiaga-Dagbui, Dominic and Simon Smith. (2014). Rethinking construction cost overruns: cognition, learning and estimation. *Journal of Financial Management of Property and Construction* 19(1), 38-54.

American Planning Association. (2005). Planners called to help end inaccuracies in public project revenue. <http://www.planning.org/newsreleases/2005/ftp040705.htm>

Batselier, Jordy, and Mario Vanhoucke. (2016). Practical application and empirical evaluation of reference class forecasting for project management. *Project Management Journal* 47(5), 36–51.

________. (2017). Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. *International Journal of Project Management* 35(1), 28-43.

Bordley, Robert. (2014). Reference class forecasting: resolving its challenge to statistical modeling. *The American Statistician* 68(4), 221–229.

Chang, Welton, Eva Chen, Barbara Mellers, and Philip Tetlock. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making* 11(5), 509–526.

Chapman, Jeffery. (2010) Complex debt in financing infrastructure. In Ingram, Gregory and Yu-Hung Hong (ed) *Land Policy: Municipal Revenues and Land Policies*, Lincoln Institute of Land Policy, Cambridge, MA.

Davenport, Thomas, and Stephen Harding. (2010). *Manager Redifined: The Competitive Advantage in the Middle of Your Organization*. Jossey-Bass, San Francisco, CA.

Den Heijer, Alexandra. (2011). *Managing the University Campus: Information to Support Real Estate Decisions*. Eburon Uitgeverij, Delft.

Federal Highway Administration. (2014). *Financial Plans Major Project Financial Plan Guidance*. Federal Highway Administration, Washington, DC.

Flyvibjerg, Bent. (2006). From Nobel Prize to project management: Getting risks right. *Project Management Journal* 37(3), 5-15.

________. (2007). Cost overruns and demand shortfalls in urban rail and other infrastructure. *Transportation Planning and Technology* 30(1), 9-30.

________. (2008) Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice. *European Planning Studies* 16(1), 3–21

Flyvbjerg, Bent, and COWI. (2004) *Procedures for Dealing with Optimism Bias in Transport Planning: Guidance Document*. UK Department for Transport, London.

Flyvbjerg, Bent, Massimo Garbuio, and Dan Lovallo. (2009). Delusion and deception in large infrastructure projects: Two models for explaining and preventing executive disaster. *California Management Review* 51(2), 170-193.

Flyvbjerg, Bent, Mette Holm, and Soren Buhl. (2002). Understand costs in public works projects: Error or lie? *Journal of the American Planning Association* 68(3), 279-295.

Flyvbjerg, Bent, Nils Bruzelius, and Werner Rothengatter. (2003). *Megaprojects and Risk: An Anatomy of Ambition*. Cambridge University Press, Cambridge.

Gil, Nuno, and Colm Lundrigan. (2012) *The Leadership and Governance of Megaprojects*. The University of Manchester, Manchester.

Highways Agency. (2013). *The Project Control Framework Handbook*. Highway Agency, London.

Horne, John. (2007). The four 'knowns' of sports mega-events. *Leisure Studies 26(1),* 81–96.

Jenner, Stephen. (2016). *Transforming Government and Public Services: Realising Benefits Through Project Portfolio Management*. Gower, Farnham.

Kahneman, Daniel. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York, NY.

Kahneman, Daniel, and Amos Tversky. (1979). Prospect Theory: Analysis of Decisions Under Uncertainty. *Econometrica* 47, 263-291.

Kitamura, Ryuichi, Toshio Yoshii, and Toshiyuki Yamamoto, ed. (2009). *Expanding Sphere of Travel Behavior Research: Selected Papers from the 11th Conference of the International Association for Travel Behavior Research*. Bingley: Elsevier.

Kittos, Christos, Teresa Oliveira, Alexandros Rigas, and Sneh Gulati, ed. (2013). *Theory and Practice of Risk Assessment*. ICRA 5, Tomar.

Liu, Huimin, Canhui Jiang, Yan Liu, Marcel Hertogh, and Xue Lyu. (2018). Optimism bias evaluation and decision-making risk forecast on bridge project cost based on reference class forecasting: Evidence from China. *Sustainability* 10, 3981-4010.

Liu, Li, George Wehbe, and Jonathan Sisovic. (2010). The Accuracy of hybrid estimating approaches: A case study of an Australian State Road and Traffic Authority. *The Engineering Economist* 55, 225-245.

Lovallo, Dan, and Daniel Kahneman. (2003). Delusions of success: How optimism undermines executives' decisions. *Harvard Business Review* 81(7), 56-63

Love, Peter, Dominic Ahiaga-Dagbui. (2018). De-bunking 'fake news' in a post-truth era: The plausible untruths of cost underestimation in transport infrastructure projects. *Transportation Research Part A* 113, 357–368.

Love, Peter, David Edwards, and Jahir Irani. (2012). Moving beyond optimism bias and strategic misrepresentation: An explanation for social infrastructure project cost overruns. *IEEE Transportation Engineering Management* 59(4), 560-561.

Love, Peter, Michael Sing, Lavagnon Ika, and Sidney Newton. (2019). The cost performance of transportation projects: The fallacy of the planning fallacy account. *Transportation Research Part A* 122 , 1-20.

Makovšek, Dejan, Polona Tominc, and Klavdij Logožar. (2012). A cost performance analysis of transport infrastructure construction in Slovenia. *Transportation* 39(1), 197-214.

McKinsey Global Institute. (2017). *Bridging Infrastructure Gaps: Has the World Made Progress?* McKinsey & Company.

Merrow, Edward. *Industrial Megaprojects: Concepts, Strategies, and Practices for Success*. John Wiley and Sons, Hoboken, NJ.

Moschouli, Eleni, Raden Soecipto, and Thierry Vanelslander. (2019). Cost performance of transport infrastructure projects before and after the global financial crisis: Are differences observed in the conditions of project performance? *Research in Transportation Economics* 75, 21-35.

Mott MacDonald. (2002). *Review of Large Public Procurement in the UK*. Mott MacDonald, Surrey.

Nijkamp, Peter, and Barry Ubbels. (1999). How reliable are estimates of infrastructure costs?: A comparative analysis. *International Journal of Transport Economics* 26(1), 23-53.

Okmen, Onder, and Ahmet Oztas. (2010). Construction cost analysis under uncertainty with correlated cost risk analysis model. *Construction Management and Economics* 28(2), 203-212.

Park, Jung Eun. (2021). Schedule delays of major projects: What should we do about it? *Transport Reviews*.

Phillips, Mark. (2019). *The Practitioner's Handbook of Project Performance*. Routledge, London.

Rizzi, J. (2003). Behavioral bias: The hidden risk in risk management. *Commercial Lending Review* 18(6), 2-8.

Siemiatycki, Matti. (2009). Academics and auditors: Comparing perspectives on transportation project cost overruns. *Journal of Planning Education and Research* 29(2), 142-156.

Smith, Keith. (2001). Comparing economic performance in the presence of diversity. *Science and Public Policy* 28(4), 267-276.

Themsen, Tim. (2019). The processes of public megaproject cost estimation: The inaccuracy of reference class forecasting. *Financial Accountability and Management* 35(4), 337-352.

Winch, Graham. (2012). *Managing Construction Projects: An Information Processing Approach*. Wiley-Blackwell, Somerset.