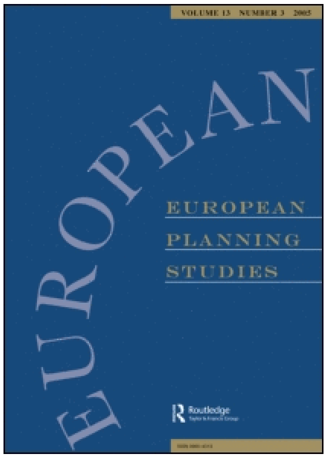

European Planning Studies
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713417253

Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice

Bent Flyvbjerg [a]
[a] Aalborg University, Denmark and Delft University of Technology, The Netherlands

Online Publication Date: 01 January 2008
To cite this Article: Flyvbjerg, Bent (2008) 'Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice', European Planning Studies, 16:1, 3 - 21
To link to this article: DOI: 10.1080/09654310701747936
URL: http://dx.doi.org/10.1080/09654310701747936

European Planning Studies Vol. 16, No. 1, January 2008

Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice

BENT FLYVBJERG

Aalborg University, Denmark and Delft University of Technology, The Netherlands

(Received May 2006; accepted August 2006)

ABSTRACT *The American Planning Association recently endorsed a new forecasting method called reference class forecasting, which is based on theories of planning and decision-making that won the 2002 Nobel prize in economics. This paper details the method and describes the first instance of reference class forecasting in planning practice. First, the paper documents that inaccurate projections of costs, demand, and other impacts of plans are a major problem in planning. Second, the paper explains inaccuracy in terms of optimism bias and strategic misrepresentation. Third, the theoretical basis is presented for reference class forecasting, which achieves accuracy in projections by basing them on actual performance in a reference class of comparable actions and thereby bypassing both optimism bias and strategic misrepresentation. Fourth, the paper presents the first case of practical reference class forecasting, which concerns cost projections for planning of large transportation infrastructure investments in the UK, including the Edinburgh Tram and London's £15 billion Crossrail project. Finally, potentials for and barriers to reference class forecasting are assessed.*

Introduction

In April 2005, based on a study of inaccuracy in demand forecasts for public works projects by Flyvbjerg *et al.* (2005), the American Planning Association (APA) endorsed a new forecasting method called "reference class forecasting". The new method achieves accuracy in projections by basing them on actual performance in a reference class of comparable actions. APA made the recommendation that planners employ reference class forecasting and never rely solely on conventional forecasting techniques when making forecasts:

Correspondence Address: Bent Flyvbjerg, Department of Development and Planning, 13 Fibigerstraede, Aalborg University, 9220 Aalborg, Denmark. Fax: +45 9815 3788; Tel.: +45 9635 7216; Email: flyvbjerg@plan.aau.dk

ISSN 0965-4313 print/ISSN 1469-5944 online/08/010003–19
DOI: 10.1080/09654310701747936

> APA encourages planners to use reference class forecasting in addition to traditional methods as a way to improve accuracy. The reference class forecasting method is beneficial for non-routine projects such as stadiums, museums, exhibit centers, and other local one-off projects. Planners should never rely solely on civil engineering technology as a way to generate project forecasts. (APA, 2005)

Reference class forecasting is based on theories of planning and decision-making under uncertainty that won Princeton psychologist Daniel Kahneman the Nobel prize in economics in 2002 (Kahneman & Tversky, 1979a, 1979b; Kahneman, 1994). As part of their work, Kahneman and Tversky uncovered a systematic fallacy in planning and decision-making under which people underestimate the costs, completion times, and risks of planned actions, whereas they overestimate the benefits of the same actions. This would later be known as "the planning fallacy", and Kahneman argued that this fallacy stems from actors taking an "inside view" focusing on the constituents of the specific planned action rather than on the outcomes of similar actions already completed. Kahneman also identified a cure to the fallacy, namely taking an "outside view" on planned actions using distributional information from previous, similar ventures. Distributional information is here understood as data on variation from the expected outcome for instance as expressed in common statistical measures such as standard deviation and variance. Doing so in a systematic fashion is called "reference class forecasting". A reference class forecast of a given planned action is based on knowledge about actual performance in a reference class of comparable actions already carried out.

Kahneman has never developed his innovative ideas about the planning fallacy and outside view for use in practical planning and decision-making, and neither has anyone else to the knowledge of the present author. In what follows, the focus will be on how this may be done. Where Flyvbjerg *et al.* (2005) briefly outlined the idea of reference class forecasting, this paper details the method and presents the first instance of reference class forecasting in practical policy and planning. The emphasis will be on transportation policy and planning, because this is where the first instance of reference class forecasting took place. It should be mentioned at the outset, however, that comparative research shows that the problems, causes, and cures identified for transportation apply to a wide range of other project types including sports arenas, exhibit and convention centres, concert halls, museums, urban renewal, power plants, dams, water projects, information technology (IT) systems, oil and gas extraction projects, and aerospace projects (Altshuler & Luberoff, 2003; Flyvbjerg *et al.*, 2003, pp. 18–19; Flyvbjerg *et al.*, 2002, p. 286; Flyvbjerg, 2005a, 2005b).

Inaccuracy in Forecasts

Two previous papers in the *Journal of the American Planning Association* (JAPA) document that forecasts of cost, demand, and other impacts of major plans and projects have remained constantly and remarkably inaccurate for decades (Flyvbjerg *et al.*, 2002, 2005; see also Ascher, 1979; Flyvbjerg *et al.*, 2003). No improvement in forecasting accuracy seems to have taken place, despite all claims of improved forecasting models, better data, etc. For example, inaccuracy in cost forecasts for transportation infrastructure projects was found to be on average 44.7% for rail, 33.8% for bridges and tunnels, and 20.4% for roads (constant prices, see Table 1).[1] For the 70-year period for which cost

Table 1. Inaccuracy in cost forecasts for rail, bridges, tunnels, and roads, respectively (construction costs, constant prices). For all project types inaccuracy is different from zero with extremely high significance

Type of project	Average inaccuracy (%)	Standard deviation (SD)	Level of significance, *p*
Rail	44.7	38.4	<0.001
Bridges and tunnels	33.8	62.4	0.004
Road	20.4	29.9	<0.001

Source: Flyvbjerg database on large-scale infrastructure projects.

data was available, accuracy in cost forecasts has not improved. Average inaccuracy for rail passenger forecasts was found to be –51.4%, with 84% of all rail projects being wrong by more than ±20%; this is equivalent to an average overestimate in rail passenger forecasts of 106%. For roads, average inaccuracy in traffic forecasts is 9.5%, with half of all road forecasts being wrong by more than ±20% (see Table 2). For the 30-year period for which demand data was available, accuracy in rail and road traffic forecasts has not improved.

When cost and demand forecasts are combined, for instance in the cost-benefit analyses that are typically used to justify large infrastructure investments, the consequence is inaccuracy to the second degree. Benefit-cost ratios are often wrong, not only by a few percent but by several factors (Flyvbjerg *et al.*, 2003, pp. 37–41). As a consequence, estimates of viability are often misleading, as are socio-economic and environmental appraisals, the accuracy of which are heavily dependent on demand and cost forecasts. These results point to a significant problem in policy and planning: More often than not the information that promoters and planners use to decide whether to invest in new projects is highly inaccurate and biased making plans and projects very risky. Comparative studies show that transportation projects are no worse than other project types in this respect (Flyvbjerg *et al.*, 2003).

Explaining Inaccuracy

Flyvbjerg *et al.* (2002, 2004, 2005) and Flyvbjerg and Cowi (2004) tested technical, psychological, and political-economic explanations for inaccuracy in forecasting. Technical

Table 2. Inaccuracy in forecasts of rail passenger and road vehicle traffic

	Rail	Road
Average inaccuracy (%)	−51.4 (SD = 28.1)	9.5 (SD = 44.3)
Percentage of projects with inaccuracies larger than ±20%	84	50
Percentage of projects with inaccuracies larger than ±40%	72	25
Percentage of projects with inaccuracies larger than ±60%	40	13

Source: Flyvbjerg database on large-scale infrastructure projects.

explanations are most common in the literature and they explain inaccuracy in terms of unreliable or outdated data and the use of inappropriate forecasting models (Vanston & Vanston, 2004, p. 33). However, when such explanations are put to empirical test they do not account well for the available data. First, if technical explanations were valid one would expect the distribution of inaccuracies to be normal or near-normal with an average near zero. Actual distributions of inaccuracies are consistently and significantly non-normal with averages that are significantly different from zero. Thus the problem is bias and not inaccuracy as such. Second, if imperfect data and models were main explanations of inaccuracies, one would expect an improvement in accuracy over time, since in a professional setting errors and their sources would be recognized and addressed, for instance through referee processes with scholarly journals and similar critical expert reviews. Undoubtedly, substantial resources have been spent over several decades on improving data and forecasting models. Nevertheless, this has had no effect on the accuracy of forecasts. This indicates that something other than poor data and models is at play in generating inaccurate forecasts. This finding has been corroborated by interviews with forecasters (Flyvbjerg & Cowi, 2004; Wachs, 1990).

Psychological and political-economic explanations better account for inaccurate forecasts. Psychological explanations account for inaccuracy in terms of optimism bias, that is, a cognitive predisposition found with most people to judge future events in a more positive light than is warranted by actual experience. Political-economic explanations, however, explain inaccuracy in terms of strategic misrepresentation. Here, when forecasting the outcomes of projects, forecasters and planners deliberately and strategically overestimate benefits and underestimate costs in order to increase the likelihood that it is their projects, and not the competition's, that gain approval and funding. Strategic misrepresentation can be traced to political and organizational pressures, for instance competition for scarce funds or jockeying for position, and to lack of incentive alignment. Optimism bias and strategic misrepresentation are both deception, but where the latter is intentional, the first is not, optimism bias is self-deception. Although the two types of explanation are different, the result is the same: inaccurate forecasts and inflated benefit-cost ratios. However, the cures to optimism bias are different from the cures to strategic misrepresentation, as will be shown later.

Explanations of inaccuracy in terms of optimism bias have been developed by Kahneman and Tversky (1979a) and Lovallo and Kahneman (2003). Explanations in terms of strategic misrepresentation have been set forth by Wachs (1989, 1990) and Flyvbjerg *et al.* (2002, 2005). As illustrated schematically in Figure 1, explanations in terms of optimism bias have their relative merit in situations where political and organizational pressures are absent or low, whereas such explanations hold less power in situations where political pressures are high. Conversely, explanations in terms of strategic misrepresentation have their relative merit where political and organizational pressures are high—this being the situation for the projects described in Flyvbjerg *et al.* (2002, 2005)—while they become immaterial when such pressures are not present. Thus, rather than compete, the two types of explanation complement each other: one is strong where the other is weak, and both explanations are necessary to understand the phenomenon at hand—the pervasiveness of inaccuracy and risk in decision-making—and how to curb it.

In what follows a forecasting method called "reference class forecasting" is presented, which bypasses human bias—including optimism bias and strategic misrepresentation—by cutting directly to outcomes. In experimental research carried out by Daniel Kahneman

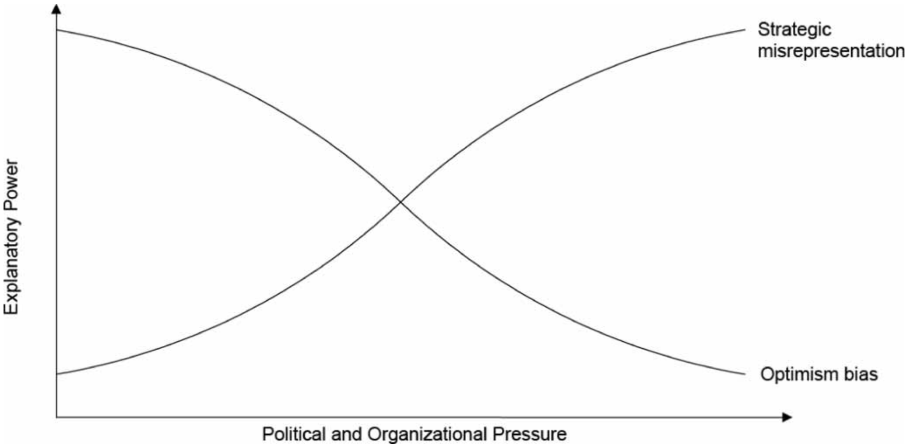

Figure 1. Explanatory power of optimism bias and strategic misrepresentation, respectively, in accounting for forecasting inaccuracy as a function of political and organizational pressure

and others, this method has been demonstrated to be more accurate than conventional forecasting methods (Kahneman & Tversky, 1979a, 1979b; Kahneman, 1994; Lovallo & Kahneman, 2003). First the theoretical and methodological foundations for reference class forecasting are explained, then the first instance of reference class forecasting in practical policy and planning is presented.

The Planning Fallacy and Reference Class Forecasting

The theoretical and methodological foundations of reference class forecasting were first described by Kahneman and Tversky (1979b) and later by Lovallo and Kahneman (2003). Reference class forecasting was originally developed to compensate for the type of cognitive bias that Kahneman and Tversky found in their work on decision-making under uncertainty, which won Kahneman the Nobel prize in economics 2002 (Kahneman, 1994; Kahneman & Tversky, 1979a). This work shows that errors of judgment are often systematic and predictable rather than random, manifesting bias rather than confusion, and that any corrective prescription should reflect this. It also shows that many errors of judgment are shared by experts and laypeople alike. Finally the work demonstrates that errors remain compelling even when one is fully aware of their nature. Thus awareness of a perceptual or cognitive illusion does not by itself produce a more accurate perception of reality, according to Kahneman and Tversky (1979b, p. 314). Awareness may, however, enable one to identify situations in which the normal faith in one's impressions must be suspended and in which judgment should be controlled by a more critical evaluation of the evidence. Reference class forecasting is a method for such critical evaluation. Human judgment, including forecasts, is biased. Reference class forecasting is a method for debiasing forecasts.

Kahneman and Tversky (1979a, 1979b) demonstrate that human judgment is generally optimistic due to overconfidence and insufficient regard to distributional information. Thus people will underestimate the costs, completion times, and risks of planned actions, whereas they will overestimate the benefits of the same actions. Following and expanding upon Buehler *et al.* (1994), Lovallo and Kahneman (2003, p. 58) call such common behaviour the "planning fallacy" and they argue that it stems from actors taking an "inside view" focusing on the constituents of the specific planned action rather than on the outcomes of similar actions that have already been completed. At the root of the planning fallacy is a tendency for actors to see each new venture as unique. In fact, ventures are typically more similar than actors assume, even ventures that on the surface of things may appear entirely different. For instance, planners may consider building a subway and building an opera house to be completely different undertakings with little to gain from each other. In fact the two may be—and often are—quite similar in statistical terms, for example as regards the size of cost overruns. And the lessons from one can be pooled with other similar projects and used as distributional information to statistically predict the outcome of the other.

Kahneman and Tversky (1979b) argue that the prevalent tendency to underweigh or ignore distributional information is perhaps the major source of error in forecasting. "The analysts should therefore make every effort to frame the forecasting problem so as to facilitate utilizing all the distributional information that is available" (Kahneman & Tversky, 1979b, p. 316). This may be considered the single most important piece of advice regarding how to increase accuracy in forecasting through improved methods. Using such distributional information from other ventures similar to that being forecasted is called taking an "outside view" and it is the cure to the planning fallacy. Reference class forecasting is a method for systematically taking an outside view on planned actions.

More specifically, reference class forecasting for a particular project requires the following three steps:

(1) Identification of a relevant reference class of past, similar projects. The class must be broad enough to be statistically meaningful but narrow enough to be truly comparable with the specific project.
(2) Establishing a probability distribution for the selected reference class. This requires access to credible, empirical data for a sufficient number of projects within the reference class to make statistically meaningful conclusions.
(3) Comparing the specific project with the reference class distribution, in order to establish the most likely outcome for the specific project.

Thus reference class forecasting does not try to forecast the specific uncertain events that will affect the particular project, but instead places the project in a statistical distribution of outcomes from the class of reference projects. In statisticians' vernacular, reference class forecasting consists of regressing forecasters' best guess toward the average of the reference class and expanding their estimate of credible interval toward the corresponding interval for the class (Kahneman & Tversky, 1979b, p. 326).

Daniel Kahneman relates the following story about curriculum planning to illustrate how reference class forecasting works (Lovallo & Kahneman, 2003, p. 61). Some years ago, Kahneman was involved in a project to develop a curriculum for a new subject area for high schools in Israel. The project was carried out by a team of academics and teachers. In time, the team began to discuss how long the project would take to complete. Everyone on the team was asked to write on a slip of paper the number of months needed to finish and report the project. The estimates ranged from 18 to 30 months. One of the team members—a distinguished expert in curriculum development—was then posed a challenge by another team member to recall as many projects similar to theirs as possible and to think of these projects as they were in a stage comparable to their project. "How long did it take them at that point to reach completion?", the expert was asked. After a while he answered, with some discomfort, that not all the comparable teams he could think of ever did complete their task. About 40% of them eventually gave up. Of those remaining, the expert could not think of any that completed their task in less than 7 years, nor of any that took more than 10. The expert was then asked if he had reason to believe that the present team was more skilled in curriculum development than the earlier ones had been. The expert said no, he did not see any relevant factor that distinguished this team favourably from the teams he had been thinking about. His impression was that the present team was slightly below average in terms of resources and potential. The wise decision at this point would probably have been for the team to break up, according to Kahneman. Instead, the members ignored the pessimistic information and proceeded with the project. They finally completed the project 8 years later, and their efforts went largely wasted—the resulting curriculum was rarely used.

In this example, the curriculum expert made two forecasts for the same problem and arrived at very different answers. The first forecast was the inside view; the second was the outside view, or the reference class forecast. The inside view is the one that the expert and the other team members adopted. They made forecasts by focusing tightly on the project at hand, considering its objective, the resources they brought to it, and the obstacles to its completion. They constructed in their minds scenarios of their coming progress and extrapolated current trends into the future. The resulting forecasts, even the most conservative ones, were overly optimistic. The outside view is the one provoked by the question to the curriculum expert. It completely ignored the details of the project at hand, and it involved no attempt at forecasting the events that would influence the project's future course. Instead, it examined the experiences of a class of similar projects, laid out a rough distribution of outcomes for this reference class, and then positioned the current project in that distribution. The resulting forecast, as it turned out, was much more accurate.

The contrast between inside and outside views has been confirmed by systematic research (Buehler *et al.*, 1994; Gilovich *et al.*, 2002). The research shows that when people are asked simple questions requiring them to take an outside view, their forecasts become significantly more accurate. For example, a group of students enrolling at a college were asked to rate their future academic performance relative to their peers in their major. On average, these students expected to perform better than 84% of their peers, which is logically impossible. The forecasts were biased by overconfidence. Another group of incoming students from the same major were asked about their entrance scores and their peers' scores before being asked about their expected performance. This simple diversion into relevant outside-view information, which both groups of subjects were aware of, reduced the second group's average expected performance ratings by 20%. That is still overconfident, but it is significantly more realistic than the forecast made by the first group (Lovallo & Kahneman, 2003, p. 61).

However, most individuals and organizations are inclined to adopt the inside view in planning new projects. This is the conventional and intuitive approach. The traditional way to think about a complex project is to focus on the project itself and its details, to bring to bear what one knows about it, paying special attention to its unique or unusual features, trying to predict the events that will influence its future. The thought of going out and gathering simple statistics about related projects seldom enters a planner's mind (Kahneman and Tversky use the term "planner" to denote individuals contemplating future actions). This is the case in general, according to Lovallo and Kahneman (2003, pp. 61–62). And it is certainly the case for cost and demand forecasting in transportation infrastructure planning. Of the several hundred forecasts reviewed in Flyvbjerg *et al.* (2003) and Flyvbjerg *et al.* (2002, 2005), not one was a reference class forecast.[2]

While understandable, planners' preference for the inside view over the outside view is unfortunate. When both forecasting methods are applied with equal skill, the outside view is much more likely to produce a realistic estimate, as shown by the simple experiments conducted by Kahneman and others. That is because the outside view bypasses cognitive and political biases such as optimism bias and strategic misrepresentation and cuts directly to outcomes. In the outside view project planners and forecasters are not required to make scenarios, imagine events, or gauge their own and others' levels of ability and control, so they cannot get all these things wrong. Human bias is bypassed. Surely the outside view, being based on historical precedent, may fail to predict extreme outcomes, that is, those that lie outside all historical precedents. But for most projects, the outside view will produce more accurate results. In contrast, a focus on inside details will be the road to inaccuracy.

The comparative advantage of the outside view will be most pronounced for non-routine projects, understood as projects that planners and decision-makers in a certain locale or organization have never attempted before—like building new infrastructure or catering to new types of demand. It is in the planning of such new efforts that the biases toward optimism and strategic misrepresentation are likely to be largest. To be sure, choosing the right reference class of comparative past projects would become more difficult when planners are forecasting initiatives for which precedents are not easily found, for instance the introduction of new and unfamiliar technologies. However, most projects are both non-routine locally and use well-known technologies. Such projects would, therefore, be particularly likely to benefit from the outside view and reference class forecasting.

Some people may find the method of reference class forecasting so obvious that it seems a puzzle why it has not been used before. It is like the Egg of Columbus: the solution to the problem is self-evident and anybody can do it, after they have been shown how. For reference class forecasting, it is not quite that simple, however. The real challenge in doing a reference class forecast lies in assembling a valid dataset that will allow a reliable forecast. Such datasets are rare in real-life policy-making and planning. Only because years of systematic research on cost overrun in large transport infrastructure projects had produced such a dataset was it possible to carry out the first instance of reference class forecasting in practical policy and planning (Flyvbjerg *et al.*, 2002).

First Instance of Reference Class Forecasting in Practice

The first instance of reference class forecasting in practice was carried out 2004–2005 under the auspices of HM Treasury and the UK Department of Transport with a view to curb optimism bias in transportation planning and other large public procurement.[3] Based on a study by Flyvbjerg and Cowi (2004), the Treasury and the Department for Transport decided to systematically employ reference class forecasting as part of project appraisal for large transportation projects in the UK.

The immediate background to this decision was the revision to "The Green Book" by HM Treasury in 2003, that identified for large public procurement a demonstrated, systematic tendency for project appraisers to be overly optimistic:

> There is a demonstrated, systematic, tendency for project appraisers to be overly optimistic. To redress this tendency appraisers should make explicit, empirically based adjustments to the estimates of a project's costs, benefits, and duration . . . [I]t is recommended that these adjustments be based on data from past projects or similar projects elsewhere. (HM Treasury, 2003b, p. 1)

Such optimism was seen as an impediment to prudent fiscal planning, for the government as a whole and for individual departments within government. To redress this tendency HM Treasury recommended that planners involved in large public procurement should make explicit, empirically based adjustments to the estimates of a project's costs, benefits, and duration. HM Treasury recommended that these adjustments be based on data from past projects or similar projects elsewhere, and adjusted for the unique characteristics of the project at hand. In the absence of a more specific evidence base, HM Treasury encouraged government departments to collect valid and reliable data to inform future estimates of optimism, and in the meantime use the best available data. The Treasury let it be understood that in future the allocation of funds for large public procurement would be dependent on valid adjustments of optimism in order to secure valid estimates of costs, benefits, and duration of large public procurement (HM Treasury, 2003a, 2003b).

In response to the Treasury's Green Book and its recommendations, the UK Department for Transport decided to collect the type of data, which the Treasury recommended, and on that basis to develop a methodology for dealing with optimism bias in the planning of transportation projects. The Department for Transport appointed Bent Flyvbjerg in association with Cowi to undertake this assignment as regards costing of large transportation procurement. The main aims of the assignment were two; first, to provide empirically based optimism bias uplifts for selected reference classes of transportation infrastructure projects, and, second, to provide guidance on using the established uplifts to produce more realistic forecasts of capital expenditures in individual projects (Flyvbjerg & Cowi 2004). Uplifts would be established for capital expenditures based on the full business case, i.e. time of decision to build.

The types of transportation schemes under the direct and indirect responsibility of the UK Department for Transport were divided into a number of distinct categories where statistical tests, benchmarkings, and other analyses showed that the risk of cost overruns within each category could be treated as statistically similar. For each category a reference class of projects was then established as the basis for reference class forecasting, as required by step 1 in the three-step procedure for reference class forecasting described earlier. The specific categories and the types of project allocated to each category are shown in Table 3. Projects and data for the reference classes were selected from the so-called "Flyvbjerg database", which contains high-quality data on actual performance regarding costs and benefits in several hundred large transport infrastructure projects. The database is described in detail in Flyvbjerg *et al.* (2002) and Flyvbjerg (2005b). Projects were placed in the same reference class if tests showed no significant differences between projects and if similar regulatory and construction regimes applied to projects. Thus projects in a reference class may be from different countries. For instance, following

Table 3. Categories and types of projects used as basis for reference class forecasting

Category	Types of projects
Roads	Motorway Trunk roads Local roads Bicycle facilities Pedestrian facilities Park and ride Bus lane schemes Guided buses on wheels
Rail	Metro Light rail Guided buses on tracks Conventional rail High speed rail
Fixed links	Bridges Tunnels
Building projects	Stations Terminal buildings
IT projects	IT system development
Standard civil engineering	Included for reference purposes only
Non-standard civil engineering	Included for reference purposes only

these criteria rail projects in Europe and North America were placed in the same reference class, whereas rail projects in developing nations and Japan were excluded from the reference class.

For each category of projects, a reference class of completed, comparable transportation projects was used to establish probability distributions for cost overruns for new projects similar in scope and risks to the projects in the reference class, as required by step 2 in reference class forecasting. For roads, for example, a class of 172 completed and comparable projects was used to establish the probability distribution of cost overruns shown in Figure 2. The share of projects with a given maximum cost overrun is also shown in Figure 2. For instance, 40% of projects have a maximum cost overrun of 10%; 80% of projects a maximum overrun of 32%, etc. For rail, the probability distribution is shown in Figure 3, and for bridges and tunnels in Figure 4. Figures 2–4 show that the risk of cost overrun is substantial for all three project types, but highest for rail, followed by bridges and tunnels, and with the lowest risk for roads.

Based on the probability distributions described earlier the required uplifts needed to carry out step 3 in a reference class forecast may be calculated as shown in Figures 5–7. The uplifts refer to cost overrun calculated in constant prices. The lower the acceptable risk for cost overrun, the higher the uplift. For instance, with a willingness to accept a 50% risk for cost overrun in a road project, the required uplift for this project would be 15%. If the investor were willing to accept only a 10% risk for cost overrun, then the required uplift would be 45%. In comparison, for rail with a willingness to accept a 50% risk for cost overrun, the required uplift would be 40%. If the investor were willing to

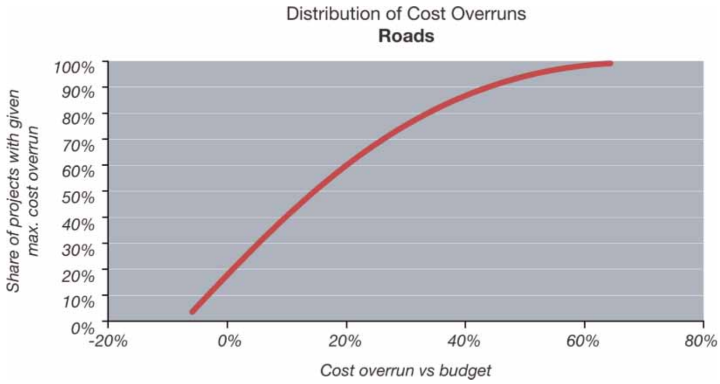

Figure 2. Probability distribution of cost overrun for roads, constant prices ($N = 172$). *Source:* Flyvbjerg database on large-scale infrastructure projects

accept only a 10% risk for cost overrun, then the required uplift would be 68% for rail. Figures 5–7 all share the same basic S-shape, but at different levels, demonstrating that the required uplifts are significantly different for different project categories for a given level of risk of cost overrun. Figures 5–7 also show that the cost for additional reductions in the risk of cost overrun is different for the three types of projects, with risk reduction becoming increasingly expensive (rising marginal costs) for roads and fixed links below 20% risk, whereas for rail the cost of increased risk reduction rises more slowly, albeit from a high level.

Table 4 presents an overview of applicable optimism bias uplifts for the 50% and 80% percentiles for all the project categories listed in Table 3. The median or average uplift would be pertinent to the investor with a large project portfolio, where cost overruns on one project may be offset by cost savings on another. The 80% percentile—corresponding to a risk of cost overrun of 20%—is the level of risk that the UK Department for Transport is typically willing to accept for large investments in local transportation infrastructure.

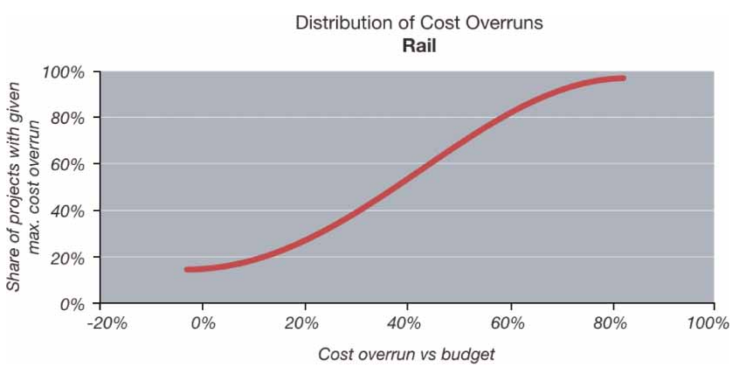

Figure 3. Probability distribution of cost overrun for rail, constant prices ($N = 46$). *Source:* Flyvbjerg database on large-scale infrastructure projects

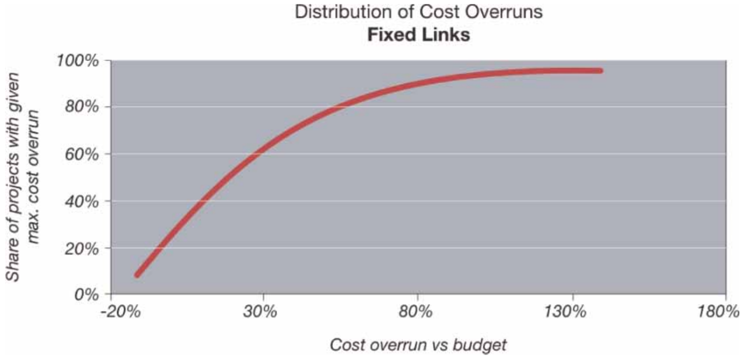

Figure 4. Probability distribution of cost overrun for fixed links, constant prices ($N = 34$). *Source:* Flyvbjerg database on large-scale infrastructure projects

The established uplifts for optimism bias should be applied to estimated budgets at the time of decision to build a project. In the UK, the approval stage for a large transportation project is equivalent to the time of presenting the business case for the project to the Department for Transport with a view to obtaining the go or no-go for that project.

If, for instance, a group of planners were preparing the business case for a new motorway, and if they or their client had decided that the risk of cost overrun must be less than 20%, then they would use an uplift of 32% on their estimated capital expenditure budget. Thus, if the initially estimated budget were £100 million, then the final budget—taking into account the bias at the 80%-level—would be £132 million (£1 = \$1.7). If the planners or their client decided instead that a 50% risk of cost overrun was acceptable, then the uplift would be 15% and the final budget £115 million.

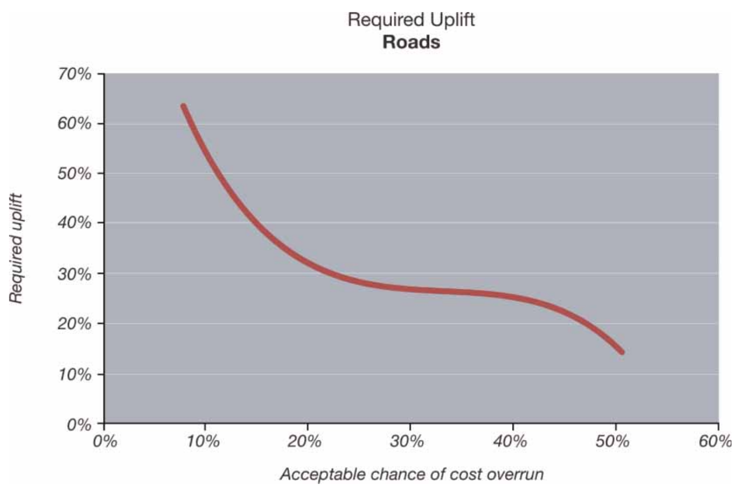

Figure 5. Required uplift for roads as a function of the maximum acceptable level of risk for cost overrun, constant prices ($N = 172$).
Source: Flyvbjerg database on large-scale infrastructure projects

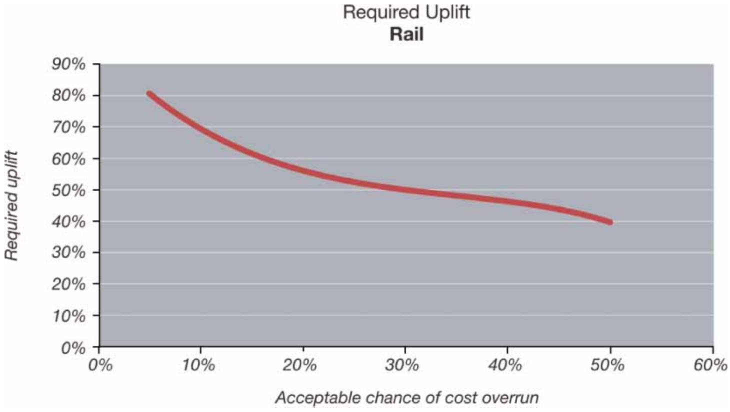

Figure 6. Required uplift for rail as a function of the maximum acceptable level of risk for cost overrun, constant prices ($N = 46$).
Source: Flyvbjerg database on large-scale infrastructure projects

Similarly, if a group of planners were preparing the business case for a metro rail project, and if they or their client had decided that with 80% certainty they wanted to stay within budget, then they would use an uplift on capital costs of 57%. An initial capital expenditure budget of £300 million would then become a final budget of £504 million. If the planners or their client required only 50% certainty they would stay within budget, then the final budget would be £420 million.

It follows that median or average uplifts should be used only in instances where investors are willing to take a high degree of risk that cost overrun will occur and/or in situations where investors are funding a large number of projects and where cost savings (underruns)

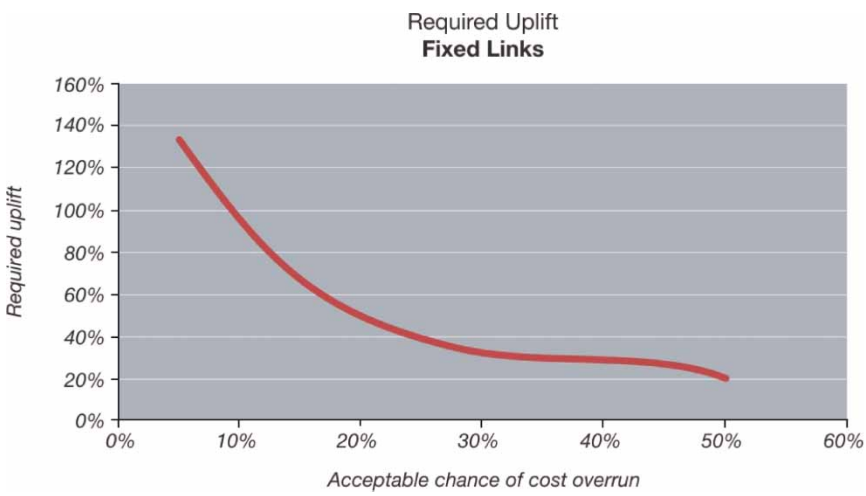

Figure 7. Required uplift for fixed links as a function of the maximum acceptable level of risk for cost overrun, constant prices ($N = 34$).
Source: Flyvbjerg database on large-scale infrastructure projects

Table 4. Applicable capital expenditure optimism bias uplifts for 50% and 80% percentiles, constant prices

Category	Types of projects	Applicable optimism bias uplifts	
		50% percentile	80% percentile
Roads	Motorway Trunk roads Local roads Bicycle facilities Pedestrian facilities Park and ride Bus lane schemes Guided buses on wheels	15%	32%
Rail	Metro Light rail Guided buses on tracks Conventional rail High speed rail	40%	57%
Fixed links	Bridges Tunnels	23%	55%
Building projects	Stations Terminal buildings	4–51%[a]	
IT projects	IT system development	10–200%[a]	
Standard civil engineering	Included for reference purposes only	3–44%[a]	
Non-standard civil engineering	Included for reference purposes only	6–66%[a]	

[a]Based on Mott MacDonald (2002, p. 32); no probability distribution available.

on one project may be used to cover the costs of overruns on other projects. The upper percentiles (80–90%) should be used when investors want a high degree of certainty that cost overrun will not occur, for instance in stand-alone projects with no access to additional funds beyond the approved budget. Other percentiles may be employed to reflect other degrees of willingness to accept risk and the associated uplifts as shown in Figures 5–7.

Only if planners have historical evidence to demonstrate that they would be significantly better at estimating costs for the project at hand than their colleagues were for the projects in the reference class would the planners be justified in using lower uplifts than those described earlier. Conversely, if there is evidence that the planners are worse at estimating costs than their colleagues, then higher uplifts should be used.

The methodology described earlier for systematic, practical reference class forecasting for transportation projects was developed 2003–2004 with publication by the Department of Transport in August 2004. From this date local authorities applying for funding for transportation projects with the Department for Transport or with HM Treasury were required to take into account optimism bias by using uplifts as described earlier and as laid out in more detail in guidelines from the two ministries.

Forecasting Costs for the Edinburgh Tram

In October 2004, the first instance of practical use of the optimism bias uplifts was recorded, in the planning of the Edinburgh Tram Line 2. Ove Arup and Partners Scotland (2004) had been appointed by the Scottish Parliament's Edinburgh Tram Bill Committee to provide a review of the Edinburgh Tram Line 2 business case developed on behalf of Transport Initiatives Edinburgh. Transport Initiatives Edinburgh is project promoter and is a private limited company owned by the City of Edinburgh Council established to deliver major transport projects for the Council. The Scottish Executive is a main funder of the Edinburgh Tram, having made an Executive Grant of £375 million ($670 million) towards lines 1 and 2 of which Transport Initiatives Edinburgh proposed spending £165 million towards Line 2.

As part of their review, Ove Arup assessed whether the business case for Tram Line 2 had adequately taken into account optimism bias as regards capital costs. The business case had estimated a base cost of £255 million and an additional allowance for contingency and optimism bias of £64 million—or 25%—resulting in total capital costs of approximately £320 million. Ove Arup concluded about this overall estimate of capital costs that it seemed to have been rigorously prepared using a database of costs, comparison to other UK light rail schemes, and reconciliations with earlier project estimates. Ove Arup found, however, that the following potential additional costs needed to be considered in determining the overall capital costs: £26 million for future expenditure on replacement and renewals and £20 million as a notional allowance for a capital sum to cover risks of future revenue shortfalls, amounting to an increase in total capital costs of 14.4% (Ove Arup and Partners Scotland, 2004, pp. 15–16).

Using the UK Department for Transport uplifts for optimism bias presented above on the base costs, Ove Arup then calculated the 80th percentile value for total capital costs—the value at which the likelihood of staying within budget is 80%—to be £400 million (i.e. £255 million × 1.57, see Table 4). The 50th percentile for total capital costs—the value at which the likelihood of staying within budget is 50%—was £357 million (i.e. £255 × 1.4). Ove Arup remarked that these estimates of total capital costs were likely to be conservative, that is, low, because the UK Department for Transport recommends that its optimism bias uplifts be applied to the budget at the time of decision to build, which typically equates to business case submission, and Tram Line 2 had not yet even reached the outline business case stage, indicating that risks would be substantially higher at this early stage as would corresponding uplifts. On that basis Arup concluded that "it is considered that current optimism bias uplifts [for Tram Line 2] may have been underestimated" (Ove Arup and Partners Scotland, 2004, p. 27).

Finally, Ove Arup mentioned that the Department for Transport guidance does allow for optimism bias to be adjusted downward if strong evidence of improved risk mitigation can be demonstrated. According to Ove Arup, this may be the case if advanced risk analysis has been applied, but this was not the case for Tram Line 2. Ove Arup therefore concluded that "the justification for reduced Department for Transport optimism bias uplifts would appear to be weak" (Ove Arup and Partners Scotland 2004: 27–28). Thus the overall conclusion of Ove Arup was that the promoter's capital cost estimate of app. £320 million was optimistic. Most likely Tram Line 2 would cost significantly more.

By framing the forecasting problem to allow the use of relevant empirical distributional information, Ove Arup was able to take an outside view on the Edinburgh Tram Line 2 capital cost forecast and thus debias what appeared to be a biased forecast. As a result Ove Arup's client, The Scottish Parliament, was provided with a more reliable estimate of what the true costs of Line 2 was likely to be.

After the initial Edinburgh Tram reference class forecast, several other UK large-scale transportation infrastructure projects have been submitted to reference class forecasting, including other light rail schemes, London's £15 billion Crossrail project, and the Taunton Third Way in Somerset County. The implementation of reference class forecasting in the UK has not been without debate, because the new method identifies optimism bias and strategic misrepresentation as the major sources of error in forecasts, whereas previous methods focused on poor data and models. Thus the focus has been shifted from the behaviour of those forecasted to the behaviour of forecasters. Outside the UK, reference class forecasting has been applied to cost and patronage forecasting for the planned Dutch Zuiderzee high-speed rail line and the A15 Maasvlakte-Vaanplein motorway and to patronage forecasting in a major private-finance rail scheme in an emerging economy. Finally, the APA has officially endorsed the method, as already said.

With the first reference class forecasts carried out 2004–2005, it is too early to draw definitive conclusions about the impacts of adopting the practice of reference class forecasting. This is an area for further research. However, it may be observed already that in the UK reference class forecasting has lead to increased awareness of optimism bias and strategic misrepresentation in the planning of major local transport schemes. This has in turn lead the Department for Transport to consult on changes to the way such schemes are promoted and approved, including requirements for local authorities to find 10% of costs locally and to fund 50% of an additional risk allowance and 100% of any cost overruns above that. These changes would help shift the responsibility of inaccurate forecasts to where it should be, i.e. with promoters and their forecasters. In this way reference class forecasting has become part and parcel of risk mitigation, as the theory behind the method advocates.

Potentials and Barriers for Reference Class Forecasting

As mentioned earlier, two types of explanation best account for forecasting inaccuracy, optimism bias and strategic misrepresentation. Reference class forecasting was originally developed to mitigate optimism bias, but reference class forecasting may help mitigate any type of human bias, including strategic bias, because the method bypasses such bias by cutting directly to empirical outcomes and building forecasts on these. Even so, the potentials for and barriers to reference class forecasting will be different in situations where (1) optimism bias is the main cause of inaccuracy as compared to situations where (2) strategic misrepresentation is the chief reason for inaccuracy. We therefore need to distinguish between these two types of situation when endeavouring to apply reference class forecasting in practice.

In the first type of situation—where optimism bias is the main cause of inaccuracy—we may assume that planners and forecasters are making honest mistakes and have an interest in improving accuracy. Consider, for example, the students mentioned earlier, who were asked to estimate their future academic performance relative to their peers. We may reasonably believe that the students did not deliberately misrepresent their estimates, because they had no interest in doing so and were not exposed to pressures that would push them in that direction. The students made honest mistakes, which produced honest, if biased, numbers regarding performance. And, indeed, when students were asked to take into account outside-view information, we saw that the accuracy of their estimates improved significantly. In this type of situation—when forecasters are honestly trying to gauge the future—the potential for using the outside view and reference class forecasting will be good. Forecasters will be welcoming the method and barriers to using it will be low, because no one has reason to be against a methodology that will improve their forecasts.

In the second type of situation—where strategic misrepresentation is the main cause of inaccuracy—differences between estimated and actual costs and benefits are best explained by political and organizational pressures. Here planners and forecasters would still need reference class forecasting if accuracy were to be improved, but planners and forecasters may not be interested in this because inaccuracy is deliberate. Biased forecasts serve strategic purposes that dominate the commitment to accuracy and truth. Consider, for example, planners with responsibility for estimating costs and benefits of urban rail projects. Here, the assumption of innocence regarding outcomes typically cannot be upheld. Cities compete fiercely for approval and for scarce national funds for such projects, and pressures are strong to present projects as favourably as possible, that is, with low costs and high benefits, in order to beat the competition. There is no incentive for the individual city to debias its forecasts, but quite the opposite. Unless all other cities also debias, the individual city would lose out in the competition for funds. Planners are on record confirming that this is a common situation (Flyvbjerg & Cowi, 2004, pp. 36–58). The result is the same as in the case of optimism: planners promote ventures that are unlikely to perform as promised. But the causes are different as are possible cures.

Under these circumstances the potential for reference class forecasting is low—the demand for accuracy is simply not there—and barriers are high. In order to lower barriers, and thus create room for reference class forecasting, incentives must be aligned to reward accurate forecasts and punish inaccurate ones. A simple measure would be for national government to cap grants for local projects at the estimated cost at the time of decision to build (the final business case) and pass the full risk of cost overrun and benefit shortfall to local taxpayers. An even better measure may be block grants to local authorities, which would ensure that one dollar spent on one type of local infrastructure would not be available for other types. Governments and banks, furthermore, may have to make reference class forecasting mandatory and make funding contingent on the application of this particular method, as was done in the UK. Finally, promoters' forecasts should be made subject to due diligence by other parties, including banks and independent bodies such as national auditors or independent analysts. Such bodies would need reference class forecasting or similar methodology to do their work. Projects that were shown to have inflated benefit-cost ratios should be stopped or placed on hold. The higher the stakes, and the higher the level of political and organizational pressures, the more pronounced will be the need for such measures. The wider issue of the politics and social practices of evaluation are detailed in Flyvbjerg *et al.* (2003) and Flyvbjerg *et al.* (2005), including the design of accountability measures and incentive alignment and how they may be implemented in practical policy and planning.

The existence of strategic misrepresentation does not exclude the simultaneous existence of optimism bias, and vice versa. In fact, it is realistic to expect such co-existence in forecasting in large and complex ventures and organizations. This again underscores the point that improved forecasting methods—here reference class forecasting—and measures of accountability must go hand-in-hand if the attempt to arrive at more accurate forecasts is to be effective.

Finally, it could be argued that in some cases the use of reference class forecasting may result in such large reserves set aside for a project that this would in itself lead to risks of inefficiencies and overspending. Reserves will be spent simply because they are there, as many believe in the construction business. For instance, it is important to recognize that for the earlier mentioned examples the introduction of reference class forecasting and optimism-bias uplifts would establish total budget reservations (including uplifts) which for some projects would be more than adequate. This may in itself create an incentive which works against firm cost control if the total budget reservation is perceived as being available to the project and its contractors. It is therefore important to combine the introduction of reference class forecasting and optimism bias uplifts with tight contracts, maintained incentives for promoters to undertake good risk assessment, and prudent control during procurement and project implementation. How this may be done is described in Flyvbjerg and Cowi (2004).

Acknowledgments

Bent Flyvbjerg would like to acknowledge the contributions of Daniel Kahneman and Dan Lovallo to the underlying ideas in this article. The author would also like to thank David Banister, Christian Brockmann, Antonio Estache, Phil Goodwin, Peter Hall, Christoph Lieb, Dan Lovallo, Hugo Priemus, Emile Quinet, Mateo Turro, José Manuel Vassallo, and Roger Vickerman.

Notes

1. Inaccuracy is measured in percentages as (actual outcome/forecast outcome − 1) × 100. The base year of a forecast for a project is the time of decision to build that project. An inaccuracy of zero indicates perfect accuracy. Cost is measured as construction costs in constant prices. Demand is measured as number of vehicles for roads and number of passengers for rail.
2. The closest thing to an outside view in large infrastructure forecasting is Gordon and Wilson's (1984) use of regression analysis on an international cross section of light-rail projects to forecast patronage in a number of light-rail schemes in North America.
3. The fact that this is, indeed, the first instance of practical reference class forecasting has been confirmed with Daniel Kahneman and Dan Lovallo, who also know of no other instances of practical reference class forecasting. Personal communications with Daniel Kahneman and Dan Lovallo, author's archives.

References

Altshuler, A. & D. Luberoff, D. (2003) *Mega-Projects: The Changing Politics of Urban Public Investment* (Washington, DC: Brookings Institution).

American Planning Association (APA) (2005) JAPA article calls on planners to help end inaccuracies in public project revenue forecasting. Available at http://www.planning.org/newsreleases/2005/ftp040705.htm (accessed 15 November 2007).

Ascher, W. (1979) *Forecasting: An Appraisal for Policy-Makers and Planners* (Baltimore, MD: The Johns Hopkins University Press)

Buehler, R., Griffin, D. & Ross, M. (1994) Exploring the 'planning fallacy': Why people underestimate their task completion times, *Journal of Personality and Social Psychology*, 67, pp. 366–381.

Flyvbjerg, B. (2005a) Design by deception: The politics of megaproject approval, *Harvard Design Magazine*, 22, Spring/Summer, pp. 50–59.

Flyvbjerg, B. (2005b) Measuring inaccuracy in travel demand forecasting: Methodological considerations regarding ramp up and sampling, *Transportation Research A*, 39(6), July, pp. 522–530.

Flyvbjerg, B. & Cowi, (2004) *Procedures for Dealing with Optimism Bias in Transport Planning: Guidance Document* (London: UK Department for Transport).

Flyvbjerg, B., Bruzelius, N. & Rothengatter, W. (2003) *Megaprojects and Risk: An Anatomy of Ambition* (Cambridge: Cambridge University Press).

Flyvbjerg, B., Skamris Holm, M. K. & Buhl, S. L. (2002) Underestimating costs in public works projects: Error or lie?, *Journal of the American Planning Association*, 68(3), Summer, pp. 279–295.

Flyvbjerg, B., Skamris Holm, M. K. & Buhl, S. L. (2004) What causes cost overrun in transport infrastructure projects?, *Transport Reviews*, 24(1), pp. 3–18.

Flyvbjerg, B., Skamris Holm, M. K. & Buhl, S. L. (2005) How (in)accurate are demand forecasts in public works projects? The case of transportation, *Journal of the American Planning Association*, 71(2), Spring, pp. 131–146.

Gilovich, T., Griffin, D. & Kahneman, D. (Eds) (2002) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press).

Gordon, P. & Wilson, R. (1984) The determinants of light-rail transit demand: An international cross-sectional comparison, *Transportation Research A*, 18A(2), pp. 135–140.

HM Treasury (2003a) *The Green Book: Appraisal and Evaluation in Central Government, Treasury Guidance* (London: TSO).

HM Treasury (2003b) *Supplementary Green Book Guidance: Optimism Bias* (London: HM Treasury).

Kahneman, D. (1994) New challenges to the rationality assumption, *Journal of Institutional and Theoretical Economics*, 150, pp. 18–36.

Kahneman, D. & Lovallo, D. (1993) Timid choices and bold forecasts: A cognitive perspective on risk taking, *Management Science*, 39, pp. 17–31.

Kahneman, D. & Tversky, A. (1979a) Prospect theory: An analysis of decisions under risk, *Econometrica*, 47, pp. 313–327.

Kahneman, D. & Tversky, A. (1979b) Intuitive prediction: Biases and corrective procedures, in: S. Makridakis & S. C. Wheelwright (Eds) *Studies in the Management Sciences: Forecasting*, p. 12 (Amsterdam: North Holland).

Lovallo, D. & Kahneman, D. (2003) Delusions of success: How optimism undermines executives' decisions, *Harvard Business Review*, July, pp. 56–63.

Mott MacDonald (2002) *Review of Large Public Procurement in the UK*, Study for HM Treasury (London: HM Treasury).

Ove Arup and Partners Scotland (2004) *Scottish Parliament, Edinburgh Tram Line 2 Review of Business Case* (South Queensferry: Ove Arup and Partners Scotland).

Vanston, J. H. & Vanston, L. K. (2004) Testing the tea leaves: Evaluating the validity of forecasts, *Research-Technology Management*, 47(5), pp. 33–39.

Wachs, M. (1989) When planners lie with numbers, *Journal of the American Planning Association*, 55(4), pp. 476–479.

Wachs, M. (1990) Ethics and advocacy in forecasting for public policy, *Business and Professional Ethics Journal*, 9(1–2), pp. 141–157.