

Inaccuracy in Traffic Forecasts

BENT FLYVBJERG, METTE K. SKAMRIS HOLM and SØREN L. BUHL

Department of Development and Planning, Aalborg University, Aalborg, Denmark

(Received 24 January 2005; revised 22 March 2005; accepted 5 April 2005)

ABSTRACT *This paper presents results from the first statistically significant study of traffic forecasts in transportation infrastructure projects. The sample used is the largest of its kind, covering 210 projects in 14 nations worth US\$58 billion. The study shows with very high statistical significance that forecasters generally do a poor job of estimating the demand for transportation infrastructure projects. The result is substantial downside financial and economic risk. Forecasts have not become more accurate over the 30-year period studied. If techniques and skills for arriving at accurate demand forecasts have improved over time, as often claimed by forecasters, this does not show in the data. For nine out of ten rail projects, passenger forecasts are overestimated; average overestimation is 106%. For 72% of rail projects, forecasts are overestimated by more than two-thirds. For 50% of road projects, the difference between actual and forecasted traffic is more than $\pm 20\%$; for 25% of road projects, the difference is larger than $\pm 40\%$. Forecasts for roads are more accurate and more balanced than for rail, with no significant difference between the frequency of inflated versus deflated forecasts. But for both rail and road projects, the risk is substantial that demand forecasts are incorrect by a large margin. The causes of inaccuracy in forecasts are different for rail and road projects, with political causes playing a larger role for rail than for road. The cure is more accountability and reference class forecasting. Highly inaccurate traffic forecasts combined with large standard deviations translate into large financial and economic risks. But such risks are typically ignored or downplayed by planners and decision-makers, to the detriment of social and economic welfare. The paper presents the data and approach with which planners may begin valid and reliable risk assessment.*

Introduction

Despite the enormous sums of money being spent on transportation infrastructure, surprisingly little systematic knowledge exists about the costs, benefits and risks involved. The literature lacks statistically valid answers to the central and self-evident question of whether transportation infrastructure projects perform as forecasted. When a project under performs, this is often explained away as an isolated instance of unfortunate circumstance; it is typically not seen as the

Correspondence Address: Bent Flyvbjerg, Department of Development and Planning, Aalborg University, Fibigerstraede 11, DK-9220 Aalborg, Denmark. E-mail: flyvbjerg@plan.aau.dk

0144-1647 print/1464-5327 online/06/010001-24 © 2006 Taylor & Francis
DOI: 10.1080/01441640500124779

2 B. Flyvbjerg et al.

particular expression of a general pattern of under performance in transportation infrastructure projects. Because knowledge is wanting in this area of research, until now it has been impossible to refute or confirm validly whether under performance is the exception or the rule.

In three previous papers, the present authors answered the question of project performance in detail as regards costs and cost-related risks. They found that projects do not perform as forecasted in terms of costs; almost nine out of ten projects fall victim to significant cost escalation. The authors also investigated the causes and cures of such under performance (Flyvbjerg *et al.*, 2002, 2003b, 2004; also Flyvbjerg *et al.*, 2003a). The present paper focuses on the benefit side of investments and answers the question about whether projects perform as forecasted in terms of demand and revenue risks. It compares forecasted performance in terms of demand with actual performance for a large number of transportation infrastructure projects. Knowledge about cost risk, benefit risk and compound risk is crucial to planners and decision-makers when developing projects and deciding which to build. For transportation infrastructure projects, the costs and benefits involved often run in the hundreds of millions of dollars, with risks being correspondingly high.

As pointed out by Pickrell (1990) and Richmond (1998), estimates of the financial viability of projects are heavily dependent on the accuracy of traffic demand forecasts. Such forecasts are also the basis for socio-economic and environmental appraisal of transportation infrastructure projects. According to the experiences gained with the accuracy of demand forecasting in the transportation sector, covering traffic volumes, spatial traffic distribution and distribution between transportation modes, there is evidence that demand forecasting—in a similar manner to cost forecasting and despite all scientific progress in modelling—is a major source of uncertainty and risk in the appraisal of transportation infrastructure projects.

Traffic forecasts are routinely used to dimension the construction of transportation infrastructure projects. Here, accuracy in forecasts is a point of considerable importance to the effective allocation of scarce funds. For example, Bangkok's US\$2 billion Skytrain was hugely over-dimensioned because the passenger forecast was 2.5 times higher than actual traffic. As a result, station platforms are too long for the shortened trains that now operate the system, a large number of trains and cars are idly parked in the train sheds because there is no need for them, terminals are too large, etc. The project company has ended up in financial trouble and even though urban rail is probably a good idea for a congested and air-polluted city such as Bangkok, over investing in idle capacity is hardly the best way to use resources, and especially not in a developing nation where capital for investment is scarce. Conversely, a UK National Audit Office study identified a number of road projects that were under dimensioned because traffic forecasts were too low. This, too, led to multi-million pound inefficiencies, because it is much more expensive to add capacity to existing fully used roads than it is to build the capacity up front (National Audit Office, 1988). For these and other reasons, accuracy in traffic forecasts matter.

Nevertheless, rigorous studies of accuracy are rare. Where such studies exist, they are characteristically small-N research, i.e. they are single-case studies or they cover only a sample of projects too small or too uneven to allow systematic, statistical analyses (Webber, 1976, Brooks and Trevelyan, 1979, Mackinder and Evans, 1981, Fullerton and Openshaw, 1985, National Audit Office, 1988, 1992, Fouracre *et al.*, 1990, Kain, 1990, Pickrell, 1990, Walmsley and Pickett, 1992, World

Bank, 1994, Richmond, 1998). Despite their value in other respects, with these and other studies, it has so far been impossible to give statistically satisfying answers to questions about how accurate traffic forecasts are for transportation infrastructure projects.

The objective of the present study has been to change this state of affairs by establishing a sample of transportation infrastructure projects that is sufficiently large to permit statistically valid answers to questions of accuracy. In addition to this intellectual objective, it has been a practical objective to give planners the tools to carry out realistic and valid risk assessment of projects as regards travel demand. Existing studies almost all conclude there is a strong tendency for traffic forecasts to be overestimated (Mackinder and Evans, 1981, p. 25; National Audit Office, 1985, appx 5.16; World Bank, 1986; Fouracre *et al.*, 1990, pp. 1, 10; Pickrell, 1990, p. x; Walmsley and Pickett, 1992, p. 2; Thompson, 1993, pp. 3–4). This paper will show below that this conclusion is a consequence of the small samples used in existing studies; it does not hold for the project population. When the sample of projects is enlarged by a factor 10–20 to a more representative one, a different picture is found where, for road projects, the forecasting problem is not simply one of overestimated traffic, whereas for rail, overestimation is very much the problem.

Measuring Inaccuracy in Traffic Forecasts

Traffic forecasts are routinely used to justify and dimension the construction of transportation infrastructure projects. To estimate the accuracy of such forecasts, it is necessary to compare forecasted with actual traffic. Common practice is followed here and the inaccuracy of a traffic forecast is defined as actual minus forecasted traffic in the percentage of forecasted traffic. Actual traffic is counted for the first year of operations (or the opening year). Forecasted traffic is the traffic estimate for the first year of operations (or the opening year) as estimated at the time of decision to build the project. Thus, the forecast is the estimate available to decision-makers based on which they made the decision to build the project in question. If no estimate was available at the time of decision to build, then the closest available estimate was used, typically a later estimate resulting in a conservative bias in our measure for inaccuracy. With this definition of inaccuracy, perfect accuracy is indicated by zero; an inaccuracy of –40%, for example, would indicate that actual traffic was 40% lower than forecasted traffic, whereas an inaccuracy of +40% would mean that actual traffic was 40% higher than forecasted traffic.

First Year as Basis for Comparison

Planners and promoters sometimes object to this way of measuring inaccuracy in traffic forecasts (Flyvbjerg *et al.*, 2003a). They say various forecasts are made at different stages of planning and implementation with forecasts typically becoming more accurate over time. Thus, the forecast at the time of making the decision to build is far from final. It is only to be expected, therefore, that such an early estimate would be highly inaccurate, and it would be unfair to use this estimate as the basis for assessing the accuracy of traffic forecasting, or so the objection goes. This method is defended here, however, because when the focus is on decision-making, and hence on the accuracy of the information available to

4 B. Flyvbjerg et al.

decision-makers, then it is exactly the traffic forecasted at the time of making the decision to build that is of primary interest. Otherwise, it would be impossible to evaluate whether or not decisions are informed. Forecasts made after the decision to build are by definition irrelevant to this decision. Whatever the reasons are for inaccurate forecasts, legislators and citizens—or private investors in the case of privately funded projects—are entitled to know the uncertainty of forecasted traffic and revenues. Otherwise, transparency and accountability suffer. It is also observed that if the inaccuracy of early traffic estimates was simply a matter of incomplete information and inherent difficulties in predicting a distant future, as project promoters and forecasters often say it is, then one would expect inaccuracies to be random or close to random. Inaccuracies, however, have a striking and highly interesting bias, as will be shown below.

Planners and promoters also sometimes object to using traffic in the first year of operations (or in the opening year) as the basis for measuring inaccuracy in forecasts. A manager at Eurotunnel, the owner of the Channel tunnel, which is one of the projects studied here, put it in the following manner in a comment on some of our previous work: “[I]t is misleading to make judgements about success or failure based on traffic revenues in the initial start-up years of the project” (letter from Eurotunnel to the authors, 1999). If projects experience start-up problems, which was very much the case for the Channel, then this may initially affect traffic negatively, but it would only be temporarily and it would be misleading to measure the inaccuracy of forecasts on this basis, according to this argument. When start-up problems are over, normal operations will ensue, traffic will increase, and this should be the basis on which inaccuracy is measured, the argument continues. Furthermore, it takes time before travellers effectively discover and make use of a new transportation facility and change their travel behaviour accordingly. Inertia is a factor. A project with lower-than-forecasted traffic during the first year of operations may well catch up with the forecast a few years down the line, and it would be more appropriate to measure inaccuracy on that basis. If the first year of operations is used as the basis for comparison, the result would be the identification of too many under performing projects, or so the opponents to using this basis argue.

At first sight the argument sounds convincing, and in principle (as opposed to in practice) there is nothing that prevents using another period than first year of operations as the basis for measuring inaccuracy. One might, for example, decide to use the fifth year of operations, because start-up problems might be expected to be ironed out by then; while important external changes in for instance land use will not have developed fully at this time either. Upon closer study, however, there are the following reasons for staying with first year of operations as the basis for measuring inaccuracy.

First, for projects for which there are data on actual and forecasted traffic covering more than one year after operations begin, it turns out that projects with lower-than-forecasted traffic during the first year of operations also tend to have lower-than-forecasted traffic in later years. Thus, using the first year of operations as the basis for measuring inaccuracy appears not often to result in the error of identifying projects as under performing that would not be identified as such if a different period were used as the basis for comparison. Actual traffic apparently does not quickly catch up with forecasted traffic for this type of project, and sometimes it never does. A follow-up study of seven of the ten urban rail projects analysed by Pickrell (1990) showed no significant gains in patronage over time;

for Baltimore, Buffalo and Pittsburgh patronage actually dropped over time (Richmond, 1998). For the Channel tunnel, more than 5 years after opening to the public, Eurostar train passengers numbered only 45% of that forecasted for the opening year; rail freight traffic was 40% of that forecasted; the result has been several near-bankruptcies. For the Humber bridge in the UK, 16 years after opening to the public actual traffic was still only about half of that forecasted. In Denmark, it took more than 20 years for actual traffic on the New Little Belt bridge to catch up with forecasted traffic, and for several years the difference between forecasted and actual traffic grew larger instead of smaller. Such findings fit well with Mierzejewski's (1995, pp. 31–32) observation that the conventional wisdom in forecasting, that in the long run forecasting errors tend to cancel each other out, is wrong; errors often reinforce each other with the result that inaccuracy becomes larger when measured against later years as compared with when measured against the first year of operations. Following this logic, using first year of operations as the basis for measuring inaccuracy would tend to underestimate overall inaccuracy of traffic forecasts.

Second, sightseeing traffic may be substantial during the first months of operations for the type of large-scale transportation infrastructure project of focus here, many of which are architectural and engineering marvels, in addition to being prosaic transportation machines designed to get people and goods from point A to B as efficiently as possible. Sightseeing traffic is traffic attracted by a project on the basis of people's desire to see and try the new transportation facility in question, for instance a new bridge or a new rail line. To illustrate, for the Øresund bridge between Sweden and Denmark, road traffic during the first month of operations was 19% higher than traffic for the same month a year later. The difference between the two months can mainly be ascribed to sightseeing traffic, which was somewhat lower than expected by the project company (Trafikministeriet, Finansministeriet and Sund & Bælt Holding, Ltd, 2002, appx 4:2). Sightseeing traffic may help offset the possible negative impacts on travel demand from start-up problems, etc. mentioned above, at least for projects that are sufficiently attractive in the public's eye. The existence of such countervailing influences on traffic during the start-up phase of projects help explain why first-year-of-operations tend to be a fairly precise basis for measuring inaccuracy in traffic forecasts.

Third, it may be observed as an empirical fact that forecasters and planners typically use first-year-of-operations as the principal basis for making their forecasts. For a given project, this is generally the main forecast presented to decision-makers and it forms part of the information decision-makers have at hand in making their decision of whether or not to build. If one wants to evaluate whether such decisions are informed, then it is the accuracy of this forecast that must be evaluated and one therefore needs to compare actual traffic in the first year of operations with forecasted traffic for that year.

Fourth, in practice only few projects can be found for which a traffic forecast exists for, say, the fifth year of operations *and* actual traffic was counted for this year so that inaccuracy may be systematically measured for this year. Many more projects can be found with information about forecasted and actual traffic for the first year of operations than for later years, because it appears to be common practice for both forecasters and those who evaluate the accuracy of forecasts to use first-year-of-operations as the basis for their work (Fouracre *et al.*, 1990; Pickrell, 1990; National Audit Office, 1992; Walmsley and Pickett, 1992; World Bank, 1994).

6 B. Flyvbjerg et al.

Finally, if newly opened transport infrastructure projects have a systematic adaptation period before traffic picks up, as claimed by many planners and promoters, this could and should be integrated in travel demand modelling. In this way, adaptation would be reflected in forecasts instead of being external to these.

Data Availability

Data that allow the calculation of inaccuracies in traffic forecasts unfortunately are relatively rare. For public sector projects, often the data are simply not produced. And even where the intention is to produce the data, projects may develop in ways that make it difficult or impossible to compare forecasted with actual traffic. For example, forecasted traffic for a given project may be estimated for the opening year, but due to delays, which are common, the actual opening date turns out to be several years later than that forecasted, and no forecast of traffic was made for the actual opening year. In more general terms, methodological differences in how and on what basis forecasted and actual traffic are estimated often make comparisons difficult or impossible. Finally, for large projects the elapse of time from forecasts are made, until the decision to build, until construction starts, until the project is completed, until operations begin, and until actual traffic can finally be counted may cover 5, 10 or more years. Over such long periods, the assumptions underlying forecasts may be dated and incommensurate when compared with the assumption underlying the way actual traffic is measured, or initial plans to compare actual with forecasted traffic may be given up or simply forgotten.

For private sector projects, traffic typically generates an income for the project owner. Budgeting and accounting is commercial, and therefore traffic forecasts and traffic counts tend to be more systematic and more conducive to comparative studies of forecasted and actual traffic than is the case for public sector projects. This typically does not help scholars much, however, because traffic data in private projects are often classified to keep them from competitors. And for both public and private projects, data that allow forecasted and actual traffic to be compared may be held back by project owners and managers because the size and direction of differences between forecasted and actual traffic may be of a kind that, if made public, would make the project look bad in the public eye, for instance where actual traffic is substantially lower than that forecasted.

Data Used in the Present Study

Despite the problems with a scarcity of data described above, after 4 years of data collection and refinement the authors developed a sample of 210 transportation infrastructure projects with comparable data for forecasted and actual traffic. The sample comprises a project portfolio worth approximately US\$58 billion in actual costs (2003 prices). The project types are urban rail, high-speed rail, conventional rail, bridges, tunnels, highways and freeways. The projects are located in 14 countries on five continents, including both developed and developing nations. The projects were completed during the 30 years between 1969 and 1998. The size of the projects range from construction costs of US\$22 million to US\$10 billion (2003 prices), with the smallest projects typically being stretches of roads in larger road schemes and the largest projects being rail links and fixed links (tunnels and

bridges). As far as is known, this is the largest sample of projects with comparable data on forecasted and actual traffic that has been established for this type of project.

The projects were selected for the sample based on data availability. All projects that are known for which comparable data on forecasted and actual traffic were obtainable were considered for inclusion in the sample. This was 485 projects. A total of 275 projects were then rejected because of unclear or insufficient data quality. More specifically, of the 275 projects rejected, 124 were rejected because inaccuracy had been estimated in ways different from and incomparable to the way it was decided to estimate inaccuracy (see above); 151 projects were rejected because inaccuracies for these projects had been estimated on the basis of adjusted data for actual traffic instead of using original, actual data as it was decided to do. All projects for which valid and reliable data were available were included in the sample. This covers both projects for which the authors collected the data and for projects for which other researchers in other studies collected data.

The authors' own data collection concentrated on large European projects, because too few data existed for this type of project to allow comparative studies. Primary data were collected on the accuracy of traffic forecasts for 31 projects in Denmark, France, Germany, Sweden and the UK and the number of large European projects with reliable data for both actual and estimated traffic were increased many times, allowing for the first time comparative studies for this type of project where statistical methods can be applied. Other projects were included in the sample from the following studies: Webber (1976), Hall (1980), National Audit Office (1985, 1988), Fouracre *et al.* (1990), Pickrell (1990), Walmsley and Pickett (1992), Skamris (1994) and Vejdirektoratet (1995). Statistical tests showed no differences between data collected through our own surveys and data collected from the studies carried out by other researchers.

As for any sample, a key question is whether the sample is representative of the population, here whether the projects included in the sample are representative of the population of transportation infrastructure projects. Since the criteria for sampling were data availability, validity and reliability, this question translates into one of whether projects with available, valid and reliable data are representative. There are four reasons why this is probably not the case. First, it has been argued that the very existence of data that make the evaluation of performance possible may contribute to improved performance when such data are used by project management to monitor projects (World Bank, 1994, p. 17). Such projects would have better than average, i.e. non-representative, performance. Second, it might be speculated that managers and promoters of projects with a particularly bad track record regarding traffic forecasts have an interest in not making traffic data available, which would then result in under representation of such projects in the sample. Conversely, managers and promoters of projects with a good track record for traffic forecasts might be interested in making this public, resulting in over representation of these projects. Third, even where managers have made traffic data available, they may have chosen to present data that place their projects in a favourable light. Often there are several forecasts of traffic to choose from and several compilations of actual traffic for a given project at a given time. If researchers collect data by means of survey questionnaires, as is often the case, there might be a temptation for managers to choose the combination of forecasted and actual traffic that suits them best, possibly a combination that makes their projects look good. An experienced researcher in a large European country who was presenting feedback on our research for that country commented on the data collection:

8 *B. Flyvbjerg et al.*

Most of the [research] is based on [national railway] replies to a questionnaire. This is likely to create a systematic bias. [The national railways] cannot be trusted to tell you the truth on these matters. As you know very well, the concept of 'truth' in these matters is particularly fragile. The temptation for [the national railways] to take, for the forecasts, the number that suits them best, this temptation must be great, and I don't think they could resist it. What you would need [in order to obtain better data] would be the original forecast documents, preferably from the archives of the Ministry of Transportation (not [from the national railways]), that were utilised to take the decision.

Other studies have documented the existence of such 'cooking' of data (Wachs, 1990). Unfortunately, in practice it proves difficult and often impossible to gain access to the original forecast documents. This is why we, and other researchers along with us, sometimes have to rely on the second-best methodology of survey questionnaires. This is also why data are likely to be biased. Finally, differences in the representativity of different subsamples may also result in non-representative data, for instance differences between rail and road. This point will be returned to below.

The available data do not allow an exact, empirical assessment of the magnitude of the problem of misrepresentation. But it is concluded, for the reasons given above, that most likely the sample is biased and the bias is conservative. In other words, accuracy in traffic forecasts estimated from the sample would be higher than accuracy in traffic forecasts in the project population. This should be kept in mind when interpreting the results from the statistical analyses below. The sample is not perfect by any means. Still, it is the best obtainable sample given the current state-of-the-art in this field of research.

In statistical analyses, the percentage difference in the sample between actual and forecasted traffic is considered to be distributed normally unless otherwise stated. Residual plots (not shown) indicate that a normal distribution might not be completely satisfied, the rail data having two outliers and the distribution for roads being somewhat skewed with larger upper tails. For the latter, a logarithmic transformation could improve normality, but this has not been considered worthwhile, partly because the tests are fairly robust to deviations from normality and partly because it complicates the interpretation.

The subdivisions of the sample implemented as part of analyses entail methodological problems of their own. Thus, the representation of observations in different combinations of subgroups is quite skewed for the data considered. Analysis would be improved considerably if the representation were more even. Partial and complete confounding occur, i.e. if a combination of two or more effects is significant, it is sometimes difficult to decide whether one or the other, or both, cause the difference. For interactions, often not all the combinations are represented, or the representations can be quite scarce. Actually, few useful results concerning subgroups could be found for these reasons, and our interpretations of the data have been adapted to these limitations. If better data could be gathered, sharper conclusions could be made.

The statistical models used are linear normal models, i.e. analysis of variance and regression analysis with the appropriate *F*- and *t*-tests have been made. The tests of hypotheses concerning means are known to be robust to deviations from normality. For each test, the *p* value has been reported.

Are Rail or Road Forecasts More Accurate?

Figures 1 and 2 show the distribution of the inaccuracy of traffic forecasts for the 210 projects in the sample split into rail and road projects. Inaccuracy, as said, is measured as actual minus forecasted traffic as a percentage of forecasted traffic. Thus, perfect accuracy is indicated by zero; a negative figure indicates that actual traffic is lower than forecasted traffic; a positive figure indicates that actual traffic is higher than forecasted traffic. The most noticeable attribute of Figures 1 and 2 is the striking difference between rail and road projects. Rail passenger forecasts are much more inaccurate and biased (inflated) than are road traffic forecasts.

Tests show that of the 27 rail projects included in the statistical analyses, two German projects should be considered as statistical outliers. These are the two projects represented by the two rightmost columns in the rail histogram in Figure 1 and the two uppermost plots in the rail box-plot diagram shown in Figure 2. Statistical tests with and without the two statistical outliers do not indicate any difference in terms of forecast inaccuracies between different types of rail projects, with the reservation that only urban rail has a reasonable representation (24 urban rail, two rail tunnels, one high-speed rail). Hence, the rail projects are considered as an aggregate. Excluding statistical outliers, the following results for the remaining 25 rail projects are found (results including the two statistical outliers, are given in brackets):

- The data document a massive problem with inflated rail passenger forecasts. For more than nine of ten rail projects, passenger forecasts are overestimated; for 72% of all rail projects, passenger forecasts are overestimated by more than two-thirds. [Including statistical outliers: for 67% of all rail projects, passenger forecasts are overestimated by more than two-thirds].
- Rail passenger forecasts were overestimated by an average of 105.6% (95% confidence interval of 66.0–169.9), resulting in actual traffic that was on average 51.4% lower than forecasted traffic (SD = 28.1, 95% confidence interval of –62.9 to –39.8). [Including statistical outliers: rail passenger forecasts were overestimated by an average of 65.2% (95% confidence interval of 23.1 to 151.3), resulting in actual traffic that was on average 39.5% lower than forecasted traffic (SD = 52.4, 95% confidence interval of –60.2 to –18.8)].
- A total of 84% of the rail projects have actual traffic more than 20% below forecasted traffic and none have actual traffic more than 20% above forecasted traffic. Even if one doubles the threshold to 40%, a solid 72% of all rail projects have actual traffic below that limit. [Including statistical outliers: the figures are 85 and 74%, respectively].

For road projects, with 95% confidence there is no significant difference ($p = 0.638$) in terms of forecast inaccuracies between vehicle traffic on highways, bridges and in tunnels (170 highways, 10 bridges, three tunnels). Hence, the 183 road projects are considered as an aggregate. The tests show the following (Table 1):

- A total of 50% of the road projects have a difference between actual and forecasted traffic of more than $\pm 20\%$. If one doubles the threshold to $\pm 40\%$, 25% of projects are above this level.
- There is no significant difference between the frequency of inflated versus deflated forecasts for road vehicle traffic ($p = 0.822$, two-sided binomial test). A

10 B. Flyvbjerg et al.

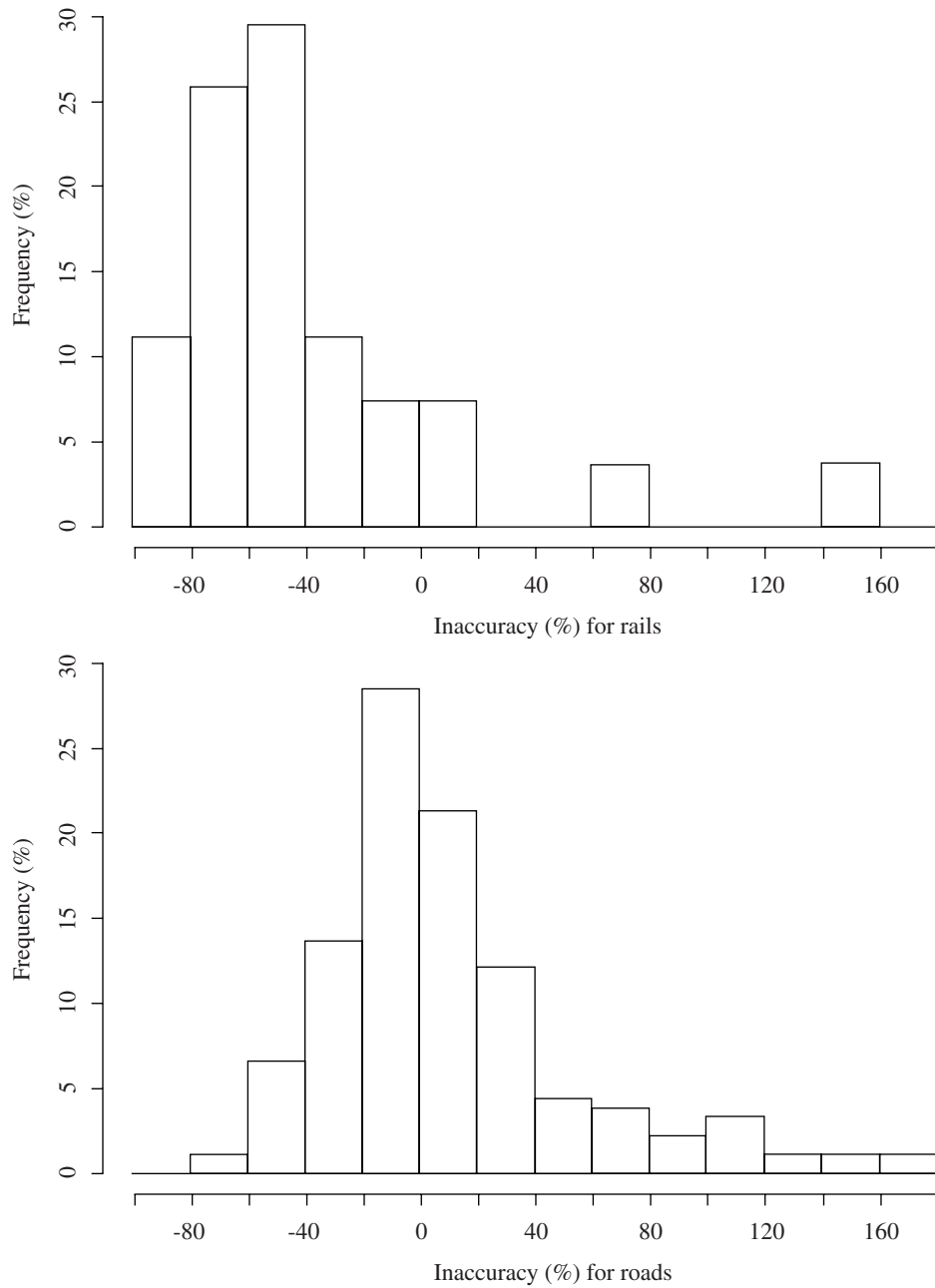


Figure 1. Inaccuracies of traffic forecasts in transportation infrastructure projects split into 27 rail and 183 road projects

total of 21.3% of projects have inaccuracies below -20% , whereas 28.4% of projects have inaccuracies above $+20\%$.

- Road traffic forecasts were underestimated by an average of 8.7% (95% confidence interval of 2.9–13.7), resulting in actual traffic that was on average 9.5% higher than forecasted traffic (SD = 44.3, 95% confidence interval of 3.0–15.9).

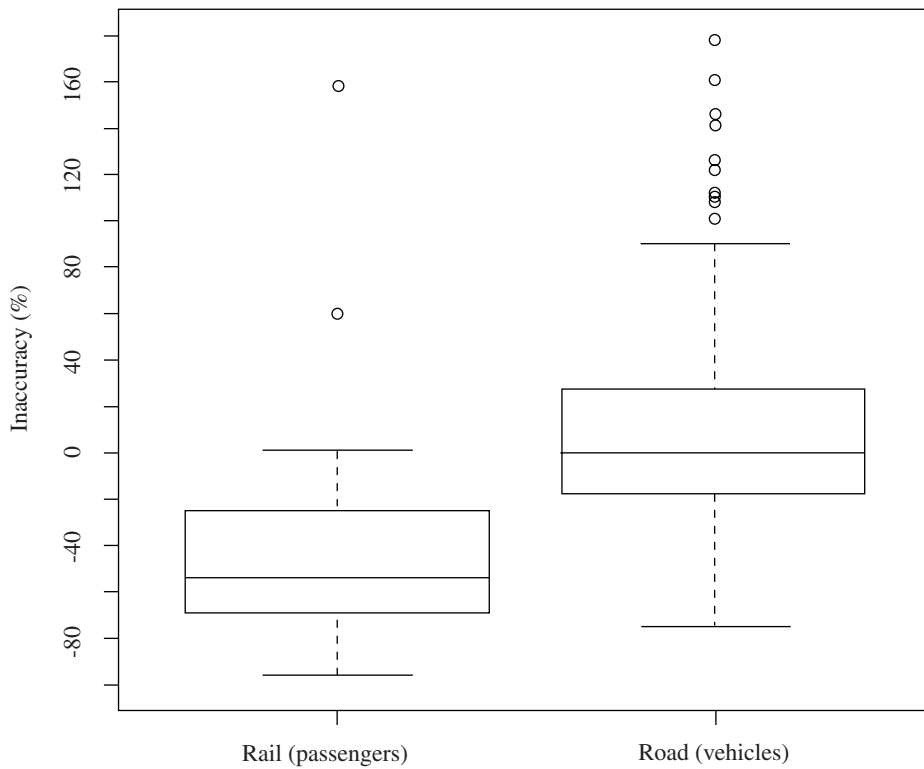


Figure 2. Inaccuracies of traffic forecasts in 210 transportation infrastructure projects

Thus, the risk is substantial that road traffic forecasts are incorrect by a large margin, but the risk is more balanced than for rail passenger forecasts. Testing the difference between rail and road, at a very high level of statistical significance, rail passenger forecasts are less accurate and more inflated than road vehicle forecasts ($p < 0.001$, Welch two-sample t -test). However, there is no indication of a significant difference between the standard deviations for rail and road forecasts: both are high, indicating a large element of uncertainty and risk for both types of forecasts ($p = 0.213$, two-sided F -test). Excluding the two statistical outliers for rail, the

Table 1. Inaccuracy in forecasts of rail passenger and road vehicle traffic

	Rail†	Road
Average inaccuracy (%)	-51.4 (SD = 28.1) [-39.5 (SD = 52.4)]	9.5 (SD = 44.3)
Projects (%) with inaccuracies larger than $\pm 20\%$	84 [85]	50
Projects (%) with inaccuracies larger than $\pm 40\%$	72 [74]	25
Projects (%) with inaccuracies larger than $\pm 60\%$	40 [41]	13

†Figures in brackets include two statistical outliers.

standard deviation for rail projects is significantly lower than for road projects, although still high ($p = 0.0105$).

It is concluded that the traffic estimates used in decision-making for rail infrastructure development are highly, systematically and significantly misleading. Rail passenger forecasts are consistently and significantly inflated. For road projects, the problem of misleading forecasts is less severe and less one-sided than for rail. But even for roads, for half the projects, the difference between actual and forecasted traffic is more than $\pm 20\%$. On this background, planners and decision-makers are well advised to take with a pinch of salt any traffic forecast that does not explicitly take into account the uncertainty of predicting future traffic. For rail passenger forecasts, this might not be enough. The data demonstrate to planners that risk assessment and management regarding travel demand must be an integral part of planning for both rail and road projects. The above data provide the empirical basis on which planners can found such risk assessment and management.

Have Forecasts Become More Accurate over Time?

Figures 3 and 4 show how forecast inaccuracy varies over time for the projects in the sample for which inaccuracy could be coupled with information about year of decision to build and/or year of completing the project. There is no indication that traffic forecasts have become more accurate over time. It is quite the opposite for road projects, where forecasts appear to become highly inaccurate toward the end of the period. Statistical analyses corroborate this impression.

For rail projects, forecast inaccuracy is independent of both the year of project commencement or the year of project conclusion. This is the case whether or not the two German projects (marked with 'K' in Figure 3) are treated as statistical outliers. It is concluded that forecasts of rail passenger traffic have not improved over time. Rail passenger traffic has been consistently overestimated during the 30-year period studied.

For road projects, inaccuracies are larger towards the end of the period with highly underestimated traffic. However, there is a difference between Danish and other road projects. For Danish road projects, it is found at a very high level of statistical significance that inaccuracy varies with time ($p < 0.001$). After 1980, Danish road traffic forecasts went completely wrong with gross underestimations of traffic, whereas this was not the case for Denmark before 1980, nor was it the case for other countries for which data exist. During a decade from the second half of the 1970s to the second half of the 1980s, the inaccuracy of Danish road traffic forecasts increased 18-fold, from 3 to 55% (Figure 5).

For Danish projects, the equation for the regression line for year of decision to build is:

$$I = 3.0 + 5.48(Y - 1970),$$

where I is the inaccuracy (%) of traffic forecast and Y is the year of the decision to build.

The Danish experience with increasing inaccuracy in road traffic forecasts is best explained by what Ascher (1979, pp. 52, 202–203) calls 'assumption drag', i.e. the continued use of assumptions after their validity has been contradicted by the

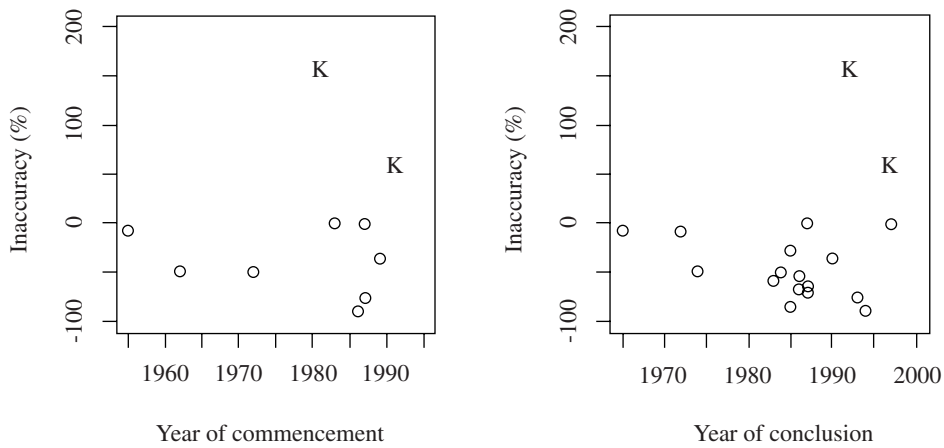


Figure 3. Inaccuracy in the number of rail passengers

data. More specifically, traffic forecasters typically calibrate forecasting models based on historical data. The so-called energy crises of 1973 and 1979 and the associated increases in petrol prices plus decreases in real wages had a profound, if short-lived, effect on road traffic in Denmark, with traffic declining for the first time in decades. Danish traffic forecasters adjusted and calibrated their models accordingly on the assumption that they were witnessing an enduring trend. The assumption was mistaken. When, during the 1980s, the effects of the two oil crises and related policy measures tapered off, traffic boomed again, rendering forecasts made on 1970s' assumptions highly inaccurate.

It is concluded that accuracy in traffic forecasting has not improved over time. Rail passenger forecasts are as inaccurate, i.e. inflated, today as they were 30 years ago. Road vehicle forecasts even appear to have become more inaccurate over time with large underestimations towards the end of the 30-year period studied. If techniques and skills for arriving at accurate traffic forecasts have improved over time, this does not show in the data. This suggests to planners that the most effective means for improving forecasting accuracy is probably not improved models but

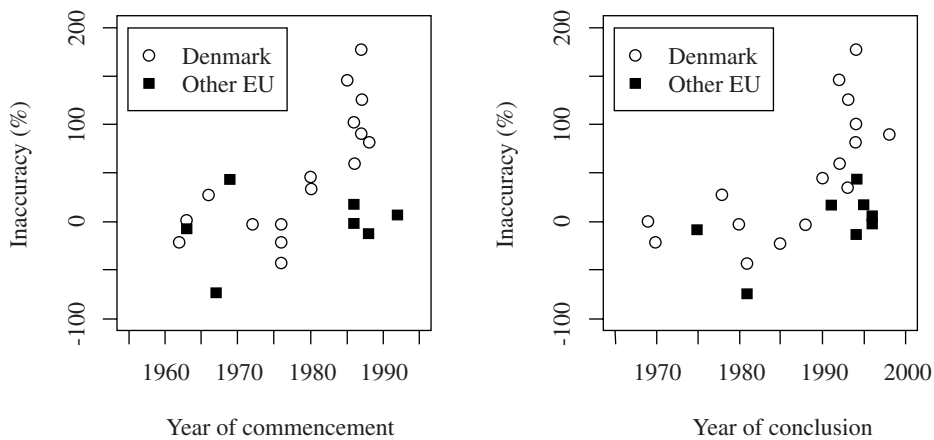


Figure 4. Inaccuracy in the number of road vehicles

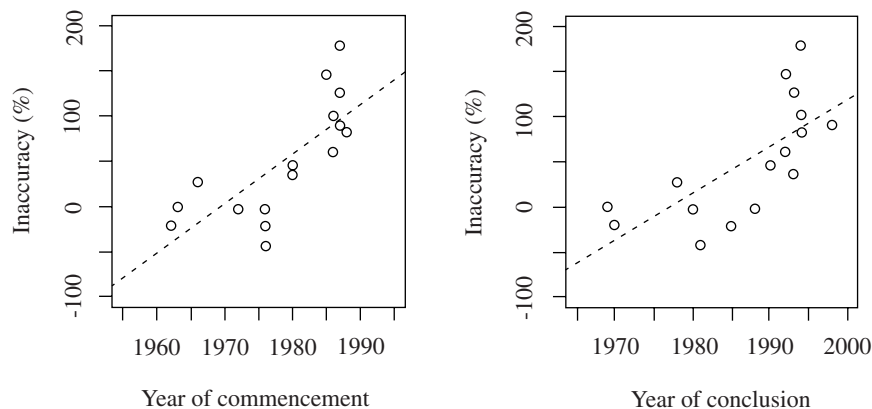


Figure 5. Inaccuracy in the number of road vehicles for Danish projects

instead more realistic assumptions and systematic use of empirically based assessment of uncertainty and risk. For rail, in particular, the persistent existence over time of highly inflated passenger forecasts invites speculation that an equilibrium has been reached where strong incentives and weak disincentives for overestimating passenger traffic may have taught project promoters what there is to learn, namely that overestimated passenger forecasts pay off: in combination with underestimated costs, such forecasts help misrepresent rail projects to decision-makers in ways that help get rail projects approved and built (Flyvbjerg *et al.*, 2003a). This suggests that improved accuracy for rail forecasts will require strong measures of accountability that would curb strategic misrepresentation in forecasts.

Effects of Project Size, Length of the Implementation Phase and Geography

Testing for the effect on forecasting inaccuracy as a dependent variable from the size of a project as independent variable, linear regression analyses were used to measure the size of the project by estimated costs, the estimated number of passengers and the estimated number of vehicles.¹ As the distributions of the estimated costs, the estimated number of passengers and the estimated number of vehicles are all skew, the logarithms of these have also been used as explanatory variables.

For rail projects, based on 17 cases, it was found that inaccuracies in passenger forecasts are not significantly dependent on costs ($p = 0.177$), but do have significance dependent on the logarithm of costs ($p = 0.018$), with higher costs leading to higher inaccuracies. Based on 27 cases, inaccuracies in passenger forecasts are not significantly dependent on the estimated size of number of passengers, neither directly ($p = 0.738$) nor taking logarithms ($p = 0.707$).

For road projects, based on 24 cases, inaccuracies in vehicle forecast are not significantly dependent on costs, neither directly ($p = 0.797$) nor logarithmically ($p = 0.114$). Based on 51 cases, inaccuracies in vehicle forecast are significantly dependent on the estimated number of vehicles, both directly ($p = 0.011$) and even stronger taking logarithms ($p < 0.001$), with smaller projects tending to have the most inaccurate, underestimated traffic forecasts.

The authors know of only one other study that relates inaccuracy in travel demand forecasting with the size of a project (Maldonado, 1990, quoted in Mierzejewski, 1995, p. 31). Based on data from 22 US airports, the present study



found that inaccuracy in aviation forecasting did not correlate with the size of the facility.

Additional tests indicate no effect on inaccuracy from the length of the project implementation phase, which is defined as the period from the decision to build a project to the start of operations. More data are needed to study the effect on inaccuracy from the geographic location of projects and the type of ownership. With the available data, there is no significant difference between geographical areas, which suggests that until such a time when more data are available, planners may pool data from different geographical areas when carrying out risk assessment.

Causes of Inaccuracies and Bias in Traffic Forecasts

The striking difference in forecasting inaccuracy between rail and road projects documented above may possibly be explained by the different procedures that apply to how each type of project is funded, where competition for funds is typically more pronounced for rail than for road, which creates an incentive for rail promoters to present their project in as favourable a light as possible, i.e. with overestimated benefits and underestimated costs (for more, see Flyvbjerg *et al.*, 2002). One might further speculate that rail patronage will be overestimated and road traffic underestimated in instances where there is a strong political or ideological desire to see passengers shifted from road to rail, for instance for reasons of congestion or protection of the environment. Forecasts here become part of the political rhetoric aimed at showing voters that something is being done—or will be done—about the problems at hand. In such cases, it might be difficult for forecasters and planners to argue for more realistic forecasts, because politicians here use forecasts to show political intent, not the most likely outcome.

To arrive at a more systematic analysis of causes of inaccuracies in traffic forecasts, such causes were identified for 234 transportation infrastructure projects. For a number of projects, we identified causes of inaccuracies but not the numerical size of inaccuracies. This explains why there are more projects (234) in this part of the analysis than in the previous part (210 projects).² Causes of inaccuracies are stated causes that explain differences between actual and forecasted traffic for the first year of operations or the opening year. For the projects for which we did the data collection, project managers were asked to account for the factors that would explain why actual traffic was different from forecasted traffic. For the other projects, the stated causes are a mixture of this type of statement by managers supplemented by statements by researchers about what caused such differences. For these projects, the data do not allow an exact distinction between manager statements and researcher statements, even though such a distinction would be desirable. It is a problem with using stated causes that what people say they do is often significantly different from what they actually do. Uncovering revealed causes for inaccuracy in traffic forecasting is therefore an important area for further research. For the time being, one has to make do with stated causes. Figure 6 shows the stated causes for inaccuracies in traffic forecasts for rail and road, respectively. For each transportation mode and stated cause, a column shows the percentage of projects for which this cause was stated as a reason for inaccuracy.

Again, the results are highly different for rail and road. For rail projects, the two most important stated causes are 'uncertainty about trip distribution' and 'deliberately slanted forecasts'. Trip distribution in rail passenger forecasts is often adapted to fit national or urban policies aimed at boosting rail traffic. But

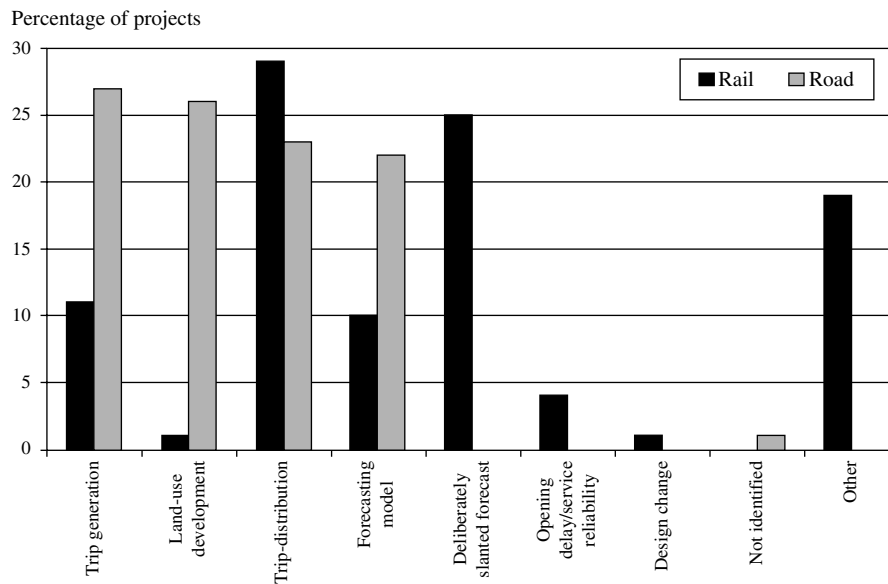


Figure 6. Stated causes of inaccuracies in traffic forecasts for 26 rail projects and 208 road projects

such policies frequently fail and the result is the type of overestimated passenger forecast which has been documented above as typical for rail passenger forecasting. As regards deliberately slanted forecasts, such forecasts are fabricated by rail promoters to increase the likelihood that rail projects are built (Wachs, 1990). Such forecasts exaggerate passenger traffic and thus revenues. The present authors have shown elsewhere that the massive overestimation of traffic and revenues documented above for rail goes hand in hand with an equally massive underestimation of costs (Flyvbjerg *et al.*, 2002, 2004). The result is cost-benefit analyses of rail projects that are highly inflated, with benefit-cost ratios that are contrived with a view to getting projects accepted and built.

For road projects, the two most often stated causes for inaccurate traffic forecasts are uncertainties about 'trip generation' and 'land-use development'. Trip generation is based on traffic counts and demographic and geographical data. Such data are often dated and incomplete and forecasters quote this as a main source of uncertainty in road traffic forecasting. Forecasts of land-use development are based on land-use plans. What is actually implemented is often quite different from what is planned, however. This, again, is a source of uncertainty in forecasting.

The different patterns in stated causes for rail and road, respectively, fit well with the figures for actual forecast inaccuracies documented above. Rail forecasts are systematically and significantly overestimated to a degree that indicates foul play on the part of rail forecasters and promoters. The stated causes, with 'deliberately slanted forecasts' as the second to largest category, corroborate this interpretation, which corresponds with findings by Wachs (1986) and Flyvbjerg *et al.* (2002). Road forecasts are also often inaccurate, but they are substantially more balanced than rail forecasts, which indicate a higher degree of fair play in road forecasting. This interpretation is corroborated by the fact that deliberately slanted forecasts are not quoted as a main cause of inaccuracy for road traffic forecasts, whereas more technical factors such as trip generation and land-use

development are quoted. This is not to say that road traffic forecasts are never politically manipulated. It is to say, however, that this appears to happen less often and less systematically for road than for rail projects. It is also not to say that road projects generally have a stronger justification than rail projects; just that they have less biased forecasts than rail projects.

What Forecasters Can Do To Reduce Inaccuracy, Bias and Risk

The above results show it is highly risky to rely on travel demand forecasts to plan and implement large transportation infrastructure investments. Rail passenger forecasts are overestimated in nine out of ten cases, with an average overestimate above 100%. Half of all road traffic forecasts are incorrect by more than $\pm 20\%$. Forecasts have not become more accurate over time. This state of affairs points directly to better risk assessment and management as something planners could and should do to improve planning and decision-making for transportation infrastructure projects. Today, the benefit risks generated by inaccurate travel demand forecasts are widely ignored or underestimated in planning, just as cost risks are neglected (Flyvbjerg *et al.*, 2003b).

When contemplating what planners can do to reduce inaccuracy, bias and risk in forecasting, one needs to distinguish between two fundamentally different situations: the situation where planners consider it important to get forecasts right and the situation where they do not. The first situation is considered in this section and the second in the following section.

If planners genuinely consider it important to get forecasts right, it is recommended they use a new forecasting method called 'reference class forecasting' to reduce inaccuracy and bias. This method was originally developed to compensate for the type of cognitive bias in human forecasting that psychologist Daniel Kahneman found in his Nobel Prize-winning work on bias in economic forecasting (Kahneman and Tversky, 1979; Kahneman, 1994). Reference class forecasting has proven more accurate than conventional forecasting. The present authors are currently developing this method in detail for practical demand and cost forecasting in transportation. For reasons of space, only an outline of the method, based mainly on Lovallo and Kahneman (2003), is presented.

Reference class forecasting consists in taking a so-called 'outside view' on the particular project being forecast. The outside view is established based on information from a class of similar projects. It 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 a group of reference projects. Reference class forecasting requires the following three steps for the individual project:

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

Kahneman relates the following story to illustrate reference class forecasting in practice (Lovallo and 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 note 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 years. 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 case 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.

Similarly—to take an example from city planning—planners in a city preparing to build a new subway would, first, establish a reference class of comparable projects. This could be the urban rail projects included in the sample for this paper. Through analyses the planners would establish that the projects included in the reference class were indeed comparable. Second, if the planners were concerned about getting patronage forecasts right, they would then establish the distribution of outcomes for the reference class regarding the accuracy of patronage forecasts. This distribution would look something like the rail part of Figure 1. Third, the planners would compare their subway project with the reference class distribution. This would make it clear to the planners that unless they had reason to believe they were substantially better forecasters and planners than their colleagues who did the forecasts and planning for projects in the reference class, they were likely to overestimate patronage grossly. Finally, planners may then use this knowledge to adjust their forecasts for more realism.

The contrast between inside and outside views has been confirmed by systematic research (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. However, most individuals and organizations are inclined to adopt the inside view in planning major initiatives. 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 cases seldom enters a planner's mind. This is the case in general, according to Lovallo and Kahneman (2003, pp. 61–62). And it is certainly the case for travel demand forecasting. Despite the many forecasts reviewed by the present authors, for instance for this paper, they have not come across a single genuine reference class forecast of travel demand.³ If readers have information about such forecasts, the authors would appreciate their feedback for on-going work on this issue.

While understandable, planners' preference for the inside 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. That is because it bypasses cognitive and organizational biases such as appraisal optimism and strategic misrepresentation, and cuts directly to outcomes. In the outside view, 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. Surely the outside view, being based on historical precedent, may fail to predict extreme outcomes, i.e. 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 is the road to inaccuracy.

The comparative advantage of the outside view is most pronounced for non-routine projects, understood as projects that planners and decision-makers in a certain locale have never attempted before—such as building an urban rail system in a city for the first time, or a new major bridge or tunnel where none existed before. It is in the planning of such new efforts that the biases toward optimism and strategic misrepresentation are likely to be large. To be sure, choosing the right reference class of comparative past projects becomes 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 large-scale transportation projects are both non-routine locally and use well-known technologies. Such projects are, therefore, particularly likely to benefit from the outside view and reference class forecasting.

When Forecasters Mislead with Numbers

This section considers the situation where planners and other influential actors do not find it important to get forecasts right and where planners, therefore, do not help to clarify and mitigate risk but, instead, generate and exacerbate it. Here planners are part of the problem, not the solution. This situation may need some explication, because it possibly sounds to many like an unlikely state of affairs. After all, it may be agreed that planners ought to be interested in being accurate and unbiased in forecasting. It is even stated as an explicit requirement in the AICP Code of Ethics and Professional Conduct that 'A planner must strive to

provide full, clear and accurate information on planning issues to citizens and governmental decision-makers' (American Planning Association, 1991, p. A.3), and the present authors certainly agree with the Code. The British Royal Town Planning Institute (2001) has laid down similar obligations for its members.

Then again, the literature is replete with things planners and planning 'must' strive to do, but which they do not. Planning must be open and communicative, but often it is closed. Planning must be participatory and democratic, but often it is an instrument to dominate and control. Planning must be about rationality, but often it is about power (Flyvbjerg, 1998, Watson, 2003). This is the 'dark side' of planning and planners identified by Flyvbjerg (1996) and Yiftachel (1998), which is remarkably under explored by planning researchers and theorists.

Forecasting, too, has its dark side. It is here we find Wachs' (1989) lying planners. They are busy, not with getting forecasts right and following the AICP Code of Ethics, but with getting projects funded and built. And accurate forecasts are often not an effective means for achieving this objective. Indeed, accurate forecasts may be counterproductive, whereas biased forecasts may be effective in competing for funds and securing the go-ahead for construction. 'The most effective planner', says Wachs (1989, p. 477), 'is sometimes the one who can cloak advocacy in the guise of scientific or technical rationality'. Such advocacy would stand in direct opposition to AICP's ruling that 'the planner's primary obligation [is] to the public interest' (American Planning Association, 1991, p. B.2). Nevertheless, seemingly rational forecasts that underestimate costs and overestimate benefits have long been an established formula for project approval (Flyvbjerg *et al.*, 2003a). Forecasting is here just another kind of rent-seeking behaviour. The consequence is a Machiavellian make-believe world of misrepresentation, which makes it extremely difficult to decide which projects deserve undertaking and which do not. The result is, as even one of the industry's own organs, the Oxford-based Major Projects Association, acknowledges, that too many projects proceed that should not. It should be added that many projects do not proceed that probably should had they not lost out to projects with 'better' misrepresentation (Flyvbjerg *et al.*, 2002).

In this situation, the question is not so much what planners can do to reduce inaccuracy and risk in forecasting, but what others can do to impose on planners the checks and balances that would give planners the incentive to stop producing biased forecasts and begin to work according to their Code of Ethics. The challenge is to change the rules of the power play that governs forecasting and project development. Here better forecasting techniques and appeals to ethics will not do; institutional change with a focus on accountability is necessary.

Two basic types of accountability define liberal democracies: (1) public sector accountability through transparency and public control, and (2) private sector accountability via competition and market control. Both types of accountability may be effective tools to curb planners' misrepresentation in forecasting and to promote a culture that acknowledges and deals effectively with risk. To achieve accountability through *transparency and public control*, the following would be required as practices embedded in the relevant institutions:

- National-level government should not offer discretionary grants to local infrastructure agencies for the sole purpose of building a specific type of infrastructure, for instance rail. Such grants create perverse incentives. Instead, national government should simply offer 'infrastructure grants' or 'transportation grants' to local governments, and let local political officials spend the

funds however they choose, but make sure that every dollar they spend on one type of infrastructure reduces their ability to fund another.

- Forecasts should be made subject to independent peer review. Where large amounts of taxpayers' money are at stake, such review may be carried out by national or state accounting and auditing offices, such as the General Accounting Office in the USA or the National Audit Office in the UK, who have the independence and expertise to produce such reviews. Other types of independent review bodies may be established, for instance within national departments of finance or with relevant professional bodies.
- Forecasts should be benchmarked against comparable forecasts, for instance using reference class forecasting as described above.
- Forecasts, peer reviews and benchmarkings should be made available to the public as they are produced, including all relevant documentation.
- Public hearings, citizen juries and the like should be organized to allow stakeholders and civil society to voice criticism and support of forecasts. Knowledge generated in this way should be integrated in planning and decision-making.
- Scientific and professional conferences should be organized where forecasters would present and defend their forecasts in the face of colleagues' scrutiny and criticism.
- Projects with inflated benefit–cost ratios should be reconsidered and stopped if recalculated costs and benefits do not warrant implementation. Projects with realistic estimates of benefits and costs should be rewarded.
- Professional and occasionally even criminal penalties should be enforced for planners and forecasters who consistently and foreseeably produce deceptive forecasts. An example of a professional penalty would be the exclusion from one's professional organization if its code of ethics is violated. An example of a criminal penalty would be punishment as the result of prosecution before a court or similar legal set-up, for instance where deceptive forecasts have led to substantial mismanagement of public funds (Garett and Wachs, 1996). Malpractice in planning should be taken as seriously as it is in other professions. Failing to do this amounts to not taking the profession of planning seriously.

To achieve accountability in forecasting via *competition and market control*, the following would be required, again as practices that are both embedded in and enforced by the relevant institutions:

- The decision to go ahead with a project should, where possible, be made contingent on the willingness of private financiers to participate without a sovereign guarantee for at least one-third of the total capital needs.⁴ This should be required whether or not projects pass the market test, i.e. whether or not projects are subsidised or provided for social justice reasons. Private lenders, shareholders and stock market analysts would produce their own forecasts or would critically monitor existing ones. If they were incorrect about the forecasts, they and their organizations would be hurt. The result would be more realistic forecasts and reduced risk.
- Full public financing or full financing with a sovereign guarantee should be avoided.
- Forecasters and their organizations must share financial responsibility for covering benefit shortfalls (and cost overruns) resulting from misrepresentation and bias in forecasting.



22 B. Flyvbjerg et al.

- Participation of risk capital should not mean that government gives up or reduces control of the project. On the contrary, it means that government can play the role it should be playing more effectively, namely as the ordinary citizen's guarantor for ensuring concerns about safety, environment, risk and a proper use of public funds.

If the institutions with responsibility for developing and building major transportation infrastructure projects would effectively implement, embed and enforce such measures of accountability, then the misrepresentation in transportation forecasting, which is widespread today, may be mitigated. If this is not done, misrepresentation is likely to continue, and the allocation of funds for transportation investments is likely to be wasteful.

Conclusions

This paper concludes that the patronage estimates used by planners of rail infrastructure development are highly, systematically and significantly misleading (inflated). This results in large benefit shortfalls for rail projects. For road projects, the problem of misleading forecasts is less severe and less one-sided than for rail. But even for roads, for half the projects, the difference between actual and forecasted traffic is more than $\pm 20\%$. On this background, planners and decision-makers are well advised to take 'with a pinch of salt' any traffic forecast that does not explicitly take into account the uncertainty of predicting future traffic.

The risks generated from misleading forecasts are typically ignored or downplayed in infrastructure planning, to the detriment of social and economic welfare. Risks, therefore, have a doubly negative effect in this particular type of planning since it is one thing to take on a risk that one has calculated and is prepared to take, much as done by insurance companies and professional investors, while it is quite another matter—that moves risk-taking to a different and more problematic level—to ignore risks. This is especially the case when risks are of the magnitude documented herein, with many demand forecasts being off by more than 50% on investments that measure in hundreds of millions of dollars. Such behaviour is bound to produce losers among those financing infrastructure, be they taxpayers or private investors. If the losers, or, for future projects, potential losers, want to protect themselves, then this study shows that the risk of faulty forecasts, and related risk assessment and management, must be placed at the core of planning and decision-making. The present goal has been to take a first step in this direction by developing the necessary data and approach.

The policy implications of the findings herein are clear. First, the findings show that a major planning and policy problem—namely misinformation—exists for this highly expensive field of public policy. Second, the size and perseverance over time of the problem of misinformation indicate that it will not go away by merely pointing out its existence and appealing to the good will of project promoters and planners to make more accurate forecasts. The problem of misinformation is an issue of power and profit and must be dealt with as such, using the mechanisms of transparency and accountability commonly used in liberal democracies to mitigate rent-seeking behaviour and the misuse of power. To the extent that planners partake in rent-seeking behaviour and misuse of power, this may be seen as a violation of their code of ethics, i.e. malpractice. Such malpractice should be taken seriously by the responsible institutions. Failing to do so amounts to not taking the profession of planning seriously.

Acknowledgements

The authors thank Don Pickrell, Martin Wachs and three anonymous referees for valuable help. Research was supported by the Danish Transportation Council and Aalborg University, Denmark.

Notes

1. The authors find that the estimated quantities are better than the actual quantities as a measure for project size in the evaluation of inaccuracy, because the estimates are what is known about size at the time of decision to build (and the time of making the forecasts) and using actual quantities would result in the mixing of cause and effect.
2. As in the other parts of the analyses, here too both projects are included for which the authors themselves collected primary data and projects for which other researchers did the data collection as part of other studies, which were then used as secondary sources. Again, the present authors' own data collection concentrated on large European projects because data were particularly wanting for this project type. By means of a survey questionnaire and meetings with project managers, primary data on causes of inaccurate traffic forecasts were collected for 16 projects, while secondary data for 218 projects were collected from the following studies: Webber (1976), Hall (1980), National Audit Office (1988), Fouracre *et al.* (1990), Pickrell (1990), Wachs (1990), Department of Transportation (1993), Leavitt *et al.* (1993), Skamris (1994) and Vejdirektoratet (1995).
3. The closest the authors have come to an outside view on travel demand forecasts 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.
4. The lower limit of a one-third share of private risk capital for such capital to influence accountability effectively is based on practical experience. For more, see Flyvbjerg *et al.* (2003a, pp. 120–123).

References

- American Planning Association (1991) *AICP Code of Ethics and Professional Conduct*. Adopted October 1978, as amended October 1991. Available at: <http://www.planning.org/ethics/conduct.html>
- Ascher, W. (1979) *Forecasting: An Appraisal for Policy-makers and Planners* (Baltimore, MD: Johns Hopkins University Press).
- Brooks, J. A. and Trevelyan, P. J. (1979) Before and after studies for inter-urban road schemes, in: *Highway Planning and Design: Proceedings of Seminar N Held at the PTRC Summer Annual Meeting, University of Warwick, England from 9–12 July 1979*, pp. 251–266 (London: Planning and Transport Research and Computation).
- Department of Transportation (1993) *Comparison of Forecast and Observed Traffic on Trunk Road Schemes* (London: DoT).
- Flyvbjerg, B. (1996) The dark side of planning: rationality and *Realrationalität*, in: S. Mandelbaum, L. Mazza and R. Burchell (Eds) *Explorations in Planning Theory*, pp. 383–394 (New Brunswick, NJ: Center for Urban Policy Research Press).
- Flyvbjerg, B. (1998) *Rationality and Power: Democracy in Practice* (Chicago, IL: University of Chicago Press).
- Flyvbjerg, B., Bruzelius, N. and Rothengatter, W. (2003a) *Megaprojects and Risk: An Anatomy of Ambition* (Cambridge: Cambridge University Press).
- Flyvbjerg, B., Holm, M. K. S. and Buhl, S. L. (2002) Cost underestimation in public works projects: error or lie?, *Journal of the American Planning Association*, 68(3), pp. 279–295.
- Flyvbjerg, B., Holm, M. K. S. and Buhl, S. L. (2003b) How common and how large are cost overruns in transport infrastructure projects?, *Transport Reviews*, 23(1), pp. 71–88.
- Flyvbjerg, B., Holm, M. K. S. and Buhl, S. L. (2004) What causes cost overrun in transport infrastructure projects?, *Transport Reviews*, 24(1), pp. 3–18.
- Fouracre, P. R., Allport, R. J. and Thomson, J. M. (1990) *The Performance and Impact of Rail Mass Transit in Developing Countries*. Research Report No. 278 (Crowthorne: Transportation and Road Research Laboratory).
- Fullerton, B. and Openshaw, S. (1985) An evaluation of the Tyneside Metro, in: K. J. Button and D. E. Pitfield (Eds) *International Railway Economics: Studies in Management and Efficiency*, pp. 177–208 (Aldershot: Gower).

24 B. Flyvbjerg et al.

- Garett, M. and Wachs, M. (1996) *Transportation Planning on Trial: The Clean Air Act and Travel Forecasting* (Thousand Oaks, CA: Sage).
- Gilovich, T., Griffin, D. and Kahneman, D. (Eds) (2002) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press).
- Gordon, P. and Wilson, R. (1984) The determinants of light-rail transit demand: an international cross-sectional comparison, *Transportation Research Part A: Policy and Practice*, 18A(2), pp. 135–140.
- Hall, P. (1980) *Great Planning Disasters* (Harmondsworth: Penguin).
- Kahneman, D. (1994) New challenges to the rationality assumption, *Journal of Institutional and Theoretical Economics*, 150, pp. 18–36.
- Kahneman, D. and Tversky, A. (1979) Prospect theory: an analysis of decisions under risk, *Econometrica*, 47, pp. 313–327.
- Kain, J. F. (1990) Deception in Dallas: strategic misrepresentation in rail transit promotion and evaluation, *Journal of the American Planning Association*, 56(2), pp. 184–196.
- Leavitt, D., Ennis, S. and McGovern, P. (1993) *The Cost Escalation of Rail Projects: Using Previous Experience to Re-evaluate the CalSpeed Estimates*. Working Paper No. 567 (Berkeley, CA: Institute of Urban and Regional Development, University of California).
- Lovullo, D. and Kahneman, D. (2003) Delusions of success: how optimism undermines executives' decisions, *Harvard Business Review*, July, pp. 56–63.
- Mackinder, I. H. and Evans, S. E. (1981) *The Predictive Accuracy of British Transportation Studies in Urban Areas*. Supplementary Report No. 699 (Crowthorne: Transportation and Road Research Laboratory).
- Maldonado, J. (1990) Strategic planning: an approach to improving airport planning under uncertainty. Unpublished MS thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Mierzejewski, E. A. (1995) *A New Strategic Urban Transportation Planning Process* (Tampa: University of South Florida, Center for Urban Transportation Research).
- National Audit Office (1985) *Department of Transportation Office: Expenditure on Motorways and Trunk Roads* (London: NAO).
- National Audit Office (1988) *Department of Transportation, Scottish Development Department and Welsh Office: Road Planning* (London: HMSO).
- National Audit Office (1992) *Department of Transportation: Contracting for Roads* (London: NAO).
- Pickrell, D. H. (1990) *Urban Rail Transit Projects: Forecast Versus Actual Ridership and Cost* (Washington, DC: US Department of Transportation).
- Richmond, J. E. D. (1998) *New Rail Transit Investments: A Review* (Cambridge, MA: John F. Kennedy School of Government, Harvard University).
- Royal Town Planning Institute (2001) *Code of Professional Conduct*, as last amended by the Council on 17 January 2001, available at: <http://www.rtpi.org.uk>
- Skamris, M. K. (1994) *Large Transportation Projects: Forecast Versus Actual Traffic and Costs*. Report No. 151 (Aalborg: Department of Development and Planning, Aalborg University).
- Thompson, L. S. (1993) Trapped in the forecasts: an economic field of dreams, *TR News*, no. 165.
- Trafikministeriet, Finansministeriet and Sund & Bælt Holding, Ltd (2002) *Udredning af økonomien i A/S Øresundsforbindelsen (de danske anlæg)* (Copenhagen: TFSBH).
- Vejdirektoratet (1995) *Notat om anlægsregnskaber* (Copenhagen: Danish Road Directorate).
- Wachs, M. (1986) Technique vs. advocacy in forecasting: a study of rail rapid transit, *Urban Resources*, 4(1), pp. 23–30.
- 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.
- Walmsley, D. A. and Pickett, M. W. (1992) *The Cost and Patronage of Rapid Transit Systems Compared with Forecasts*. Research Report No. 352 (Crowthorne: Transportation Research Laboratory).
- Watson, V. (2003) Conflicting rationalities: implications for planning theory and ethics, *Planning Theory and Practice*, 4, pp. 395–408.
- Webber, M. M. (1976) *The BART Experience: What Have We Learned?* Monograph No. 26 (Berkeley, CA: Institute of Transportation Studies, University of California).
- World Bank (1986) *Urban Transportation: A World Bank Policy Study* (Washington, DC: World Bank).
- World Bank (1994) *World Development Report 1994: Infrastructure for Development* (Oxford: Oxford University Press).
- Yiftachel, O. (1998) Planning and social control: exploring the dark side, *Journal of Planning Literature*, 12(4), pp. 395–406.