TRANSPORT REVIEWS, 2003, VOL. 23, NO. 1, 71–88

How common and how large are cost overruns in transport infrastructure projects?

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

Department of Development and Planning, Aalborg University, Fibigerstraede 11, DK-9220 Aalborg, Denmark

Despite the hundreds of billions of dollars being spent on infrastructure development — from roads, rail and airports to energy extraction and power networks to the Internet — surprisingly little reliable knowledge exists about the performance of these investments in terms of actual costs, benefits and risks. This paper presents results from the first statistically significant study of cost performance in transport infrastructure projects. The sample used is the largest of its kind, covering 258 projects in 20 nations worth approximately US$90 billion (constant 1995 prices). The paper shows with overwhelming statistical significance that in terms of costs transport infrastructure projects do not perform as promised. The conclusion is tested for different project types, different geographical regions and different historical periods. Substantial cost escalation is the rule rather than the exception. For rail, average cost escalation is 45% (SD = 38), for fixed links (tunnels and bridges) it is 34% (62) and for roads 20% (30). Cost escalation appears a global phenomenon, existing across 20 nations on five continents. Cost estimates have not improved and cost escalation not decreased over the past 70 years. Cost estimates used in decision-making for transport infrastructure development are highly, systematically and significantly misleading. Large cost escalations combined with large standard deviations translate into large financial risks. However, such risks are typically ignored or underplayed in decision-making, to the detriment of social and economic welfare.

1. Age of infrastructure

A war is on. The Great War of Independence from Space (Cairncross 1997, Bauman 1998). The key weapon in this war is infrastructure development. Transport and telecommunication infrastructure plays a central role in nothing less than the creation of what many see as a new world order where people, information, goods, energy and money move about with unprecedented ease. Thus, the past decade has seen a virtual explosion in infrastructure building. Hundreds if not thousands of billions of dollars — public and private — are currently tied up in the provision of new infrastructure around the world (Flyvbjerg *et al*. 2003, chapter 1).

Despite the high level of activity and the enormous sums of money being spent on infrastructure, surprisingly little systematic knowledge exists about the costs, benefits and risks involved. Existing studies of costs, benefits and uncertainties in transport infrastructure development are few. Where such studies exist they are typically small-N research, i.e. they are single-case studies or they cover a sample of infrastructure

*Author for correspondence; e-mail: flyvbjerg@i4.auc.dk

Transport Reviews ISSN 0144-1647 print/ISSN 1464-5327 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0144164022000016667

projects too small or too uneven to allow systematic, statistical analyses. Examples of such studies are Hall (1980), Fouracre*et al.* (1990), Kain (1990), Pickrell (1990), Walmsley and Pickett (1992), Szyliowicz and Goetz (1995), Skamris and Flyvbjerg (1997), Bruzelius*et al.* (1998), Nijkamp and Ubbels (1999) and Richmond (2001). To our knowledge, only one study exists that, with a sample of 66 transport projects, approaches a large-N study and takes a first step toward statistical analysis (Merewitz 1973a, b). [Merewitz's study was aimed at comparing cost overrun in urban rapid transit projects, and especially overrun in the San Francisco Bay Area Rapid Transit (BART) system, with overrun in other types of public works projects. Merewitz's aims were thus different from ours and his sample of transport projects was substantially smaller: 17 rapid transit projects and 49 road projects compared with our 58 rail projects, 167 highway projects, and 33 bridges and tunnels. In addition to issues of small-N sampling, in our attempt to replicate Merewitz's analysis we found that his handling of data raises a number of other issues. First, Merewitz did not correct his cost data for inflation, i.e. current prices were used instead of constant ones. This is known to be a major source of error due to varying inflation rates between projects and varying duration of construction periods. Second, in statistical tests Merewitz compared the mean cost overrun of subgroups of projects, e.g. rapid transit, with the grand mean of overrun for all projects, thus making the error of comparing projects with themselves. Subgroups should be tested directly against other subgroups in deciding whether they differ at all and, if so, which ones differ. Third, Merewitz (1973a, b) are inconsistent. Merewitz (1973a) calculates the grand mean of cost overrun as the average of means for subgroups, i.e. the grand mean is unweighted where common practice is to use the weighted mean, as appears to be the approach taken in Merewitz (1973b). Fourth, due to insufficient information the p calculated by Merewitz are difficult to verify; most likely they are flawed, however, and Merewitz's one-sided p are misleading. Finally, Merewitz used a debatable assumption about symmetry, which has more impact for the non-parametric test used than non-normality has for parametric methods. Despite these shortcomings, the approach taken in Merewitz's study was innovative for its time and in principle pointed in the right direction about how to analyse cost escalation in public works projects. The study cannot be said to be a true large-N study for transport infrastructure, however, and its statistical significance is unclear.]

Notwithstanding their value in other respects, these and other studies have not produced statistically valid answers to the central and self-evident question of whether transport infrastructure projects perform as promised. Moreover, because of the small and uneven samples used, different studies reach very different conclusions. For costs, for instance, Nijkamp and Ubbels (1999: 23) claimed that 'in general cost estimates tend to be rather reliable', whereas Pickrell (1990, 1992: 158) concluded that costs were 'grossly' underestimated.

Among project promoters and decision-makers, when a project under performs, this is often explained away as an isolated instance of unfortunate circumstance: it is only the Channel Tunnel, or the Washington Metro or the Humber Bridge that is in dire straits, and this is not necessarily part of a general pattern of under performance, or so the argument goes. Because samples are small, until now it has been impossible to validly refute or confirm such claims.

The objective of the study reported here, therefore, is to answer in a statistically valid and reliable manner the following questions: Do forecast costs and benefits of transport infrastructure projects compare well with actual costs and benefits? Or are costs and benefits highly uncertain phenomena? How large and how common are differences between forecast and actual development? Are they significant? What are the consequences for policy-making and planning? In this paper we examine how common and how large cost escalations are in transport infrastructure development. In later publications, we will examine what causes cost escalation and we will assess the reliability of traffic forecasts.

2. Sampling, data collection and methodology

Given the objective stated above, our first task was to establish a sample of transport infrastructure projects substantially larger than what is common in this area of research, a sample large enough to allow statistical analyses of costs and benefits. Here a first problem is that data on cost development in such projects are relatively difficult to come by. One reason is that it is quite time-consuming to produce data of this kind. For public sector projects, funding and accounting procedures are typically unfit for keeping track of the multiple and complex changes that occur in total project costs over time. For large projects, the relevant period may cover 5, 10 or more fiscal years, from decision to build, until construction starts, until the project is completed and operations begin. Reconstructing the total costs of a public project therefore typically entails long and difficult archival work and complex accounting. For private projects, even if funding and accounting practices may be more conducive to producing data on total costs, such data are often classified to keep them from the hands of competitors. Unfortunately, this also tends to keep data from the hands of scholars. And for both public and private projects, data on cost development may be held back by project owners, because cost development more often than not equals cost overrun, and cost overrun is normally considered somewhat of an embarrassment to promoters and owners. In sum, establishing reliable data on cost development for even a single transport infrastructure project is often highly time-consuming or even impossible.

This state of affairs explains why large-N studies have so far been absent in this field of scholarship. However, despite the problems mentioned, after 4 years of data collection and refinement we were able to establish a sample of projects with data on construction cost development for 258 transport infrastructure projects. We chose to focus on land-based transport. Consequently, the project types are rail (high-speed rail, urban rail, conventional rail), fixed links (bridges, tunnels) and roads (highways, freeways). The distribution of the 258 projects on the three types of project was as follows:

- Rail: 58.
- Fixed link (bridges and tunnels): 33.
- Road: 167.

The projects were located in 20 nations on five continents, including both developed and developing nations. The geographical distribution of projects was:

- Europe: 181.
- North America: 61.
- Other: 16.

The project portfolio is worth approximately US$90 billion (constant 1995 prices). All costs are construction costs. To correct for inflation, all costs were converted to the 1995 level using the appropriate historical, sectoral (construction industry) and geographical indices for discounting and the appropriate exchange rates for conversion between currencies. Thus, all costs in the sample are given in constant 1995 prices, i.e. prices corrected for inflation. The size of projects range from US$1.5 million to $8.5 billion, with the smallest projects typically being stretches of roads in larger road schemes and the largest projects being rail links, tunnels and bridges. The distribution of projects regarding size was the following (constant 1995 prices):

- > US$500 million: 32.
- US$100 million – $500 million: 35.
- < US$100 million: 191.

The projects were completed between 1927 and 1998. Older projects were included in the sample in order to test whether the accuracy of estimated costs improve over time. As far as we know, this is the largest sample of projects with data on cost development that has been established in this field of research.

In statistical analysis, data should be a sample from a larger population, and the sample should represent the population properly. These requirements are ideally satisfied by drawing the sample by randomized lot. Randomization ensures with high probability that non-controllable factors are equalized. A sample should also be designed such that the representation of subgroups corresponds to their occurrence and importance in the population. In studies of human affairs, however, where controlled laboratory experiments often cannot be conducted, it is frequently impossible to meet these ideal conditions. This is also the case for the current study and we therefore had to take a different approach to sampling and statistical analysis.

We selected the projects for the sample on the basis of data availability. All projects that we knew of for which data on construction cost development were obtainable were considered for inclusion in the sample. We follow international convention and define cost development as the difference between actual and estimated costs in percentage of estimated costs. Actual costs are defined as real, accounted costs determined at the time of completing a project. Estimated costs are defined as budgeted, or forecast, costs at the time of decision to build a project. Even if the project planning process varies with project type, country and time, it is typically possible to locate for a given project a specific point in the process that can be identified as the time where the formal decision was made to build the project. Usually a cost estimate was available for this point in time as information for decision-makers. If not, the closest available estimate was used, typically a later estimate resulting in a conservative bias in our measurement of cost development. Often the real decision to build a project has been made well before the formal decision and based on informally developed forecasts that are substantially more optimistic than those developed during the subsequent formal planning and decision-making process. Ideally, we would calculate cost development on the basis of the cost estimate at the time of the real decision to build. However, in most cases, it is virtually impossibly to identify the specific, real decision date and to obtain information about the informal cost estimate used by decision-makers at this point in time. Using the cost estimate at the time of the formal decision to build again results in a conservative bias in our measurement of cost development.

Cost data were collected from a variety of sources, i.e. annual project accounts, questionnaires, interviews and other studies. Data on cost development were available for 343 projects. We then rejected 85 projects because of insufficient data quality. For instance, for some projects we could not obtain a clear answer regarding what was included in costs, or whether cost data were given in current or constant prices, or which price level (year) had been used in estimating and discounting costs. More specifically, of the 85 projects we rejected 27 because we could not establish whether or not cost data were valid and reliable. We rejected 12 projects because they had been completed before 1915 and no reliable indices were available for discounting costs to the present. Finally, we excluded 46 projects because cost development for these turned out to have been calculated before construction was completed and operations begun; therefore, for these projects the true final costs may be different from the cost estimates used to calculate cost development, and no information was available on true final costs. In addition to the 85 rejects mentioned here, we also rejected a number of projects to avoid double counting. This was typically projects from other studies that appeared in more than one study or where we had a strong suspicion that this might be the case. Some such projects would appear with the same data in different studies, but other projects had different data in different studies, for instance because one study used current prices while another used constant prices. In the latter type of instance we would choose the data and the study of the highest quality.

In sum, all projects for which data were considered valid and reliable were included in the sample. This covers both projects for which we ourselves collected the data, and projects for which other researchers in other studies did the data collection (Merewitz 1973a, Hall 1980, National Audit Office and Department of Transport 1985, Lewis 1986, National Audit Office, Department of Transport, Scottish Development Department and Welsh Office 1988, Fouracre*et al.* 1990, Pickrell 1990, National Audit Office and Department of Transport 1992, Walmsley and Pickett 1992, Leavitt*et al.* 1993, Riksrevisionsverket 1994, Vejdirektoratet 1995). Cost data were made comparable between projects by discounting prices to the 1995 level and calculating them in one currency (euros or dollars), using the appropriate geographical, sectoral and historical indices for discounting and the appropriate exchange rates for conversion between currencies (€1.00 = US$1.29, 1995 level).

Our own data collection concentrated on large European projects, because too few data existed for this type of project to allow comparative studies. For instance, for projects with actual construction costs larger than €500 million (constant 1995 prices), we were initially able to identify from other studies only two (2) European projects for which data were available on both forecast and actual costs. If we lowered the project size and looked at projects larger than €100 million, we were able to identify such data for eight European projects. We saw the lack of reliable cost data for European projects as particularly problematic since the Commission of the European Union had just launched its policy for establishing the so-called trans-European transport networks (TTEN), which would involve the construction of a large number of major transport infrastructure projects across Europe at an initial cost of €220 billion (Commission of the European Union 1993: 75). As regards costs, we concluded that the knowledge base for the Commission's policy was less than well developed and we hoped to help remedy this situation through our data collection. Our efforts proved successful. We collected primary data on cost for 37 projects in Denmark, France, Germany, Sweden and the UK and were thus able to increase many times the number of large European projects with reliable data for both actual and estimated costs, allowing for the first time comparative studies for this type of project where statistical methods can be applied.

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 transport infrastructure projects. Since the criteria for sampling were data availability, this question translates into one of whether projects with available data are representative. There are five reasons why this is probably not the case. First, it may be speculated that projects that are managed well with respect to data availability may also be managed well in other respects, resulting in better-than-average, i.e. non-representative, performance for such projects. Second, 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: 17). Again, such projects would not be representative of the project population. Third, we might speculate that managers of projects with a particularly bad track record regarding cost escalation have an interest in not making cost data available, which would then result in underrepresentation of such projects in the sample. Conversely, managers of projects with a good track record for costs might be interested in making this public, resulting in overrepresentation of these projects. Fourth, even where managers have made cost data available they may have chosen to give out data that present their projects in as favourable a light as possible. Often there are several forecasts of costs to choose from and several calculations of actual costs 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 forecast and actual costs that suits them best, possibly a combination that makes their projects look good. An experienced researcher in a large European country, who was giving us feedback on our research for that country, commented on the data collection (the quote has been made anonymous for obvious reasons):

> 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 utilized 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 find or gain access to the original forecast documents. This is why we and other researchers sometimes have to rely on the second-best methodology of survey questionnaires. It is also a reason why data are likely to be biased. Fifth, and finally, differences in the representativity of different subsamples may also result in non-representative data.

The available data do not allow an exact, empirical assessment of the magnitude of the problem of misrepresentation. However, the little data that exist that shed light on this problem support the thesis that data are biased. When we compared data from the Swedish Auditor General for a subsample of road projects, for which the problems of misrepresentation did not seem to be an issue, with data for all road projects in our sample, we found that cost escalation in the Swedish subsample is significantly higher than for all projects (Holm 1999: 11 – 15). We conclude, for the reasons given above, that most likely the sample is biased and the bias is conservative. In other words, the difference between actual and estimated costs estimated from the sample is likely to be lower than the difference in the project population. This should be kept in mind when interpreting the results from statistical analyses of the sample. 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 the statistical analyses, percentage cost development in the sample is considered normally distributed. Residual plots, not shown here, indicate that normal distribution might not be completely satisfied, the distributions being somewhat skewed with larger upper tails. However, transformations, e.g. the logarithmic one, does not improve this significantly. For simplicity, therefore, no transformation is made here. In one case, namely comparing the negative and the positive part of the distribution of percentage cost development, the resulting truncated distributions can obviously not be considered normal. Hence, a non-parametric test is used.

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 skew for the data considered. The 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. We have adapted our interpretations of the data to these limitations, needless to say. 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 mean values are known to be robust to deviations from normality. Also, χ^2-tests for independence have been used for count data. For each test the p has been reported. This value is a measure for rareness if identity of groups is assumed. Traditionally, $p < 0.01$ is considered highly significant, < 0.05 significant, whereas a larger p means that the deviation could be due to chance.

3. Cost performance in 258 projects

Figure 1 shows a histogram with the distribution of construction cost escalation for all 258 projects in the sample. Cost development is calculated, as mentioned, as actual cost minus forecast cost as percentage of forecast cost. A cost development of zero for a project means that the forecast costs for the project were correct and thus equalled actual costs. If errors in forecasting costs were small, the histogram would be narrowly concentrated around zero. If errors in overestimating costs were of the same size and frequency as errors in underestimating costs, the histogram would be symmetrically distributed around zero. Neither is the case. We make the following

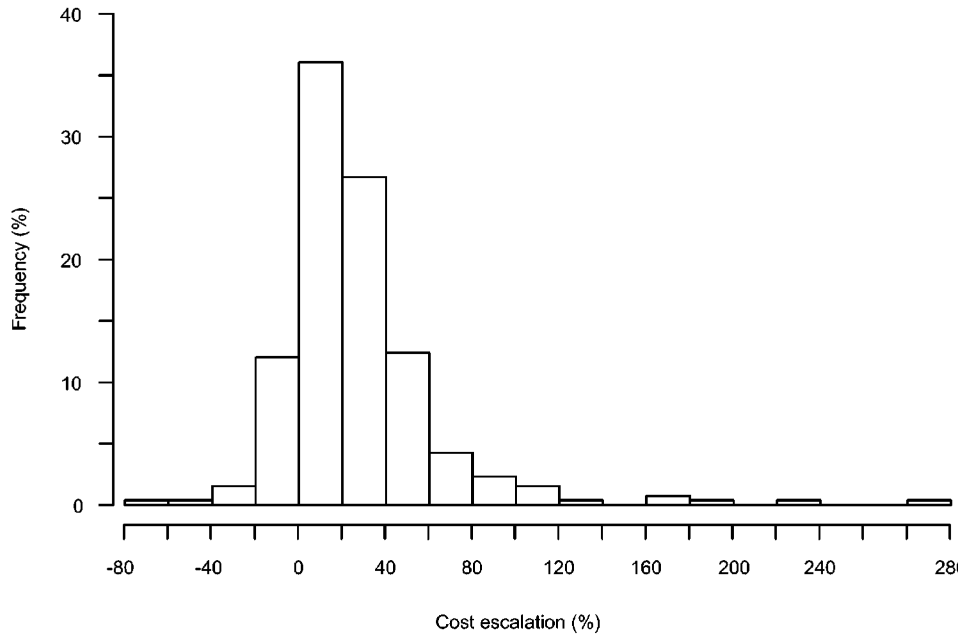

Figure 1. Cost escalation in 258 transport infrastructure projects (constant prices).

observations regarding the distribution of cost development (figures rounded off to nearest integer value):

- Cost escalation happens in almost nine out of 10 projects. For a randomly selected project, the likelihood of actual costs being larger than forecast costs is 86%. The likelihood of actual costs being lower than or equal to forecast costs is 14%.
- Actual costs are on average 28% higher than forecast costs (SD = 39).
- We reject with overwhelming significance the thesis that the error of overestimating costs is as common as the error of underestimating costs ($p < 0.001$; two-sided test, using the binomial distribution). Forecast costs are biased and the bias is caused by systematic underestimation.
- We reject with overwhelming significance the thesis that the numerical size of the error of underestimating costs is the same as the numerical size of the error of overestimating costs ($p < 0.001$; non-parametric Mann–Whitney U-test). Costs are not only underestimated much more often than they are overestimated or correct, costs that have been underestimated are also wrong by a substantially larger margin than costs that have been overestimated.

We conclude that the error of underestimating costs is significantly much more common and much larger than the error of overestimating costs. Underestimation of costs at the time of decision to build is the rule rather than the exception for transport infrastructure projects. Frequent and substantial cost escalation is the result.

In what follows we study cost escalation for different types of projects and for different geographical regions.

4. Which projects perform best: rail, road or fixed link?

After we have established beyond statistical doubt that large cost escalations are the rule rather than the exception in transport infrastructure projects, in this section we test whether different project types perform differently as regards cost escalation. For this purpose we subdivide the sample into the following three types of project: (1) rail projects (high-speed; urban; and conventional, inter-city rail), (2) fixed links (bridges and tunnels) and (3) road projects (highways and freeways). Figure 2 shows histograms with cost development for each project type. Table 1 shows the expected (average) value of cost development and standard deviation for each type of project.

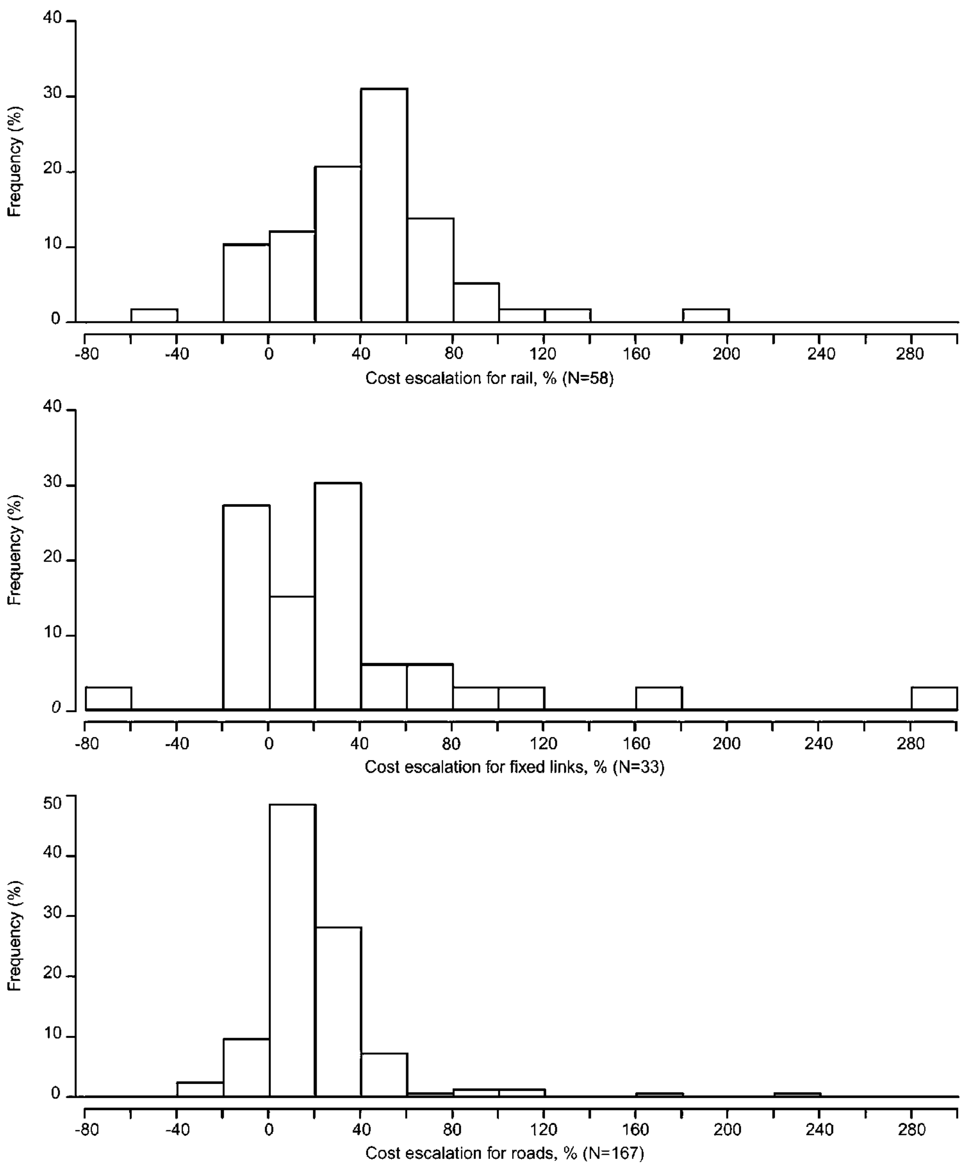

Figure 2. Cost escalation for rail, fixed links and roads (constant prices).

Table 1. Average cost escalation for rail, fixed links and roads, respectively (constant prices). For all project types, the average cost escalation is different from zero with extremely high significance.

Type of project	Number of cases (n)	Average cost escalation (%)	SD	Level of significance, p
Rail	58	44.7	38.4	<0.001
Fixed links	33	33.8	62.4	0.004
Road	167	20.4	29.9	<0.001
All projects	258	27.6	38.7	<0.001

Statistical analyses of the data in table 1 show both means and standard deviations to be different with a high level of confidence. Rail projects incur the highest difference between actual and estimated costs with an average of no less than 44.7%, followed by fixed links averaging 33.8% and roads with 20.4%. An F-test falsifies at a very high level of statistical confidence the null hypothesis that type of project has no effect on percentage cost escalation ($p < 0.001$). Project type matters. The substantial and significant differences between project types indicate that pooling the three types of projects in statistical analyses, as we did in the previous section, is not appropriate strictly speaking. Therefore, in the analyses that follow, each type of project will be considered separately.

Based on the available evidence we conclude that rail projects appear to be particularly prone to cost escalation, followed by fixed links. Road projects appear to be relatively less predisposed for cost escalation, although actual costs are higher than forecast costs much more often than not also for roads.

If we subdivide the sample a second time and split fixed links into tunnels and bridges we find an average cost escalation of 48% for tunnels (SD = 44) and 30% for bridges (SD = 67). However, by subdividing the sample this second time we reach the limits of its usefulness as a basis for statistical analysis. The number of observations in each category now becomes too small to attain significant results. The difference between tunnels and bridges is statistically non-significant. Only by further data collection for more tunnels and bridges would we be able to change this state of affairs and again arrive at statistically significant results.

Similarly, if we subdivide rail projects into high-speed rail, urban rail and conventional rail, we find that high-speed rail tops the list of cost escalation with an average of 52% (SD = 48), followed by urban rail with 45% (SD = 37) and conventional rail with 30% (SD = 34). Again the differences are statistically non-significant, and again the reason is that the subsamples are too small. Furthermore, for high-speed rail the average conceals what might be important geographical differences (see below).

We conclude that the question of whether there are significant differences in cost escalation for rail, fixed links and roads, respectively, must be answered in the affirmative. Average cost escalation for rail projects is substantially and significantly higher than that of roads, with fixed links in a statistically non-significant middle position between rail and road. Cost escalation for rail is more than twice that of roads. For all three project types, the evidence shows that it is sound advice for policy and decision-makers as well as investors, bankers, media and the public to take any estimate of construction costs with a grain of salt, and especially for rail projects and fixed links.

5. Geographical variations in cost escalation

In addition to testing whether cost escalation differs for different kinds of projects, we have also tested whether escalation varies with geographical location, between Europe, North America, and 'other geographical areas' (a group of 10 developing nations plus Japan). Table 2 shows cost development for these geographical areas for fixed links, rail and road, respectively. There is no indication of statistical interaction between geographical area and type of project. We therefore consider the effects from these variables on cost escalation separately.

For all projects, we find that the difference between geographical areas in terms of cost development is highly significant ($p < 0.001$). Geography matters for cost escalation.

If Europe and North America are compared separately, which is compulsory for fixed links and roads because no observations exist for other geographical areas here, then comparisons can be made by t-tests (as the standard deviations are rather different, the Welch version is used). For fixed links average cost escalation is 43.4% in Europe versus 25.7% in North America, but the difference is non-significant ($p = 0.414$). Given the limited number of observations and the large standard deviations for fixed links, we would need to enlarge the sample with more fixed links in Europe and North America in order to test whether the differences might be significant for a larger sample.

For rail, cost escalation is 34.2% in Europe versus 40.8% in North America. For roads, the similar numbers are 22.4 versus 8.4%. Again these differences are non-significant ($p = 0.510$ and 0.184, respectively).

We conclude that the highly significant differences in cost escalation we found above for geographical location are due to 'other geographical areas', with their poor track record of cost escalation for rail, averaging 64.6%. Figure 3 shows the box plot of cost escalation for rail according to geographical area. In addition to more data on projects in Europe and North America, a particularly interesting question for further research is whether data on fixed links and roads in 'other geographical areas' would show the same tendency at poor cost performance and high risk as does rail.

6. Are we learning yet? Cost escalation over time

In Sections 4 and 5 we saw how cost performance in transport infrastructure projects varies with project type and geography. We conclude our analysis here by studying how cost performance varies over time. We ask and answer the question of whether project performance, as regards cost escalation, has improved over time. If promoters, forecasters and decision-makers learn from past experience, one might expect such improvement.

Table 2. Cost escalation in Europe, North America and other geographical areas (constant prices).

	Europe			North America			Other geographical areas		
Type of project	Number of projects	Average cost escalation (%)	SD	Number of projects	Average cost escalation (%)	SD	Number of projects	Average cost escalation (%)	SD
Rail	23	34.2	25.1	19	40.8	36.8	16	64.6	49.5
Fixed links	15	43.4	52.0	18	25.7	70.5	0	–	–
Roads	143	22.4	24.9	24	8.4	49.4	0	–	–
Total	181	25.7	28.7	61	23.6	54.2	16	64.6	49.5

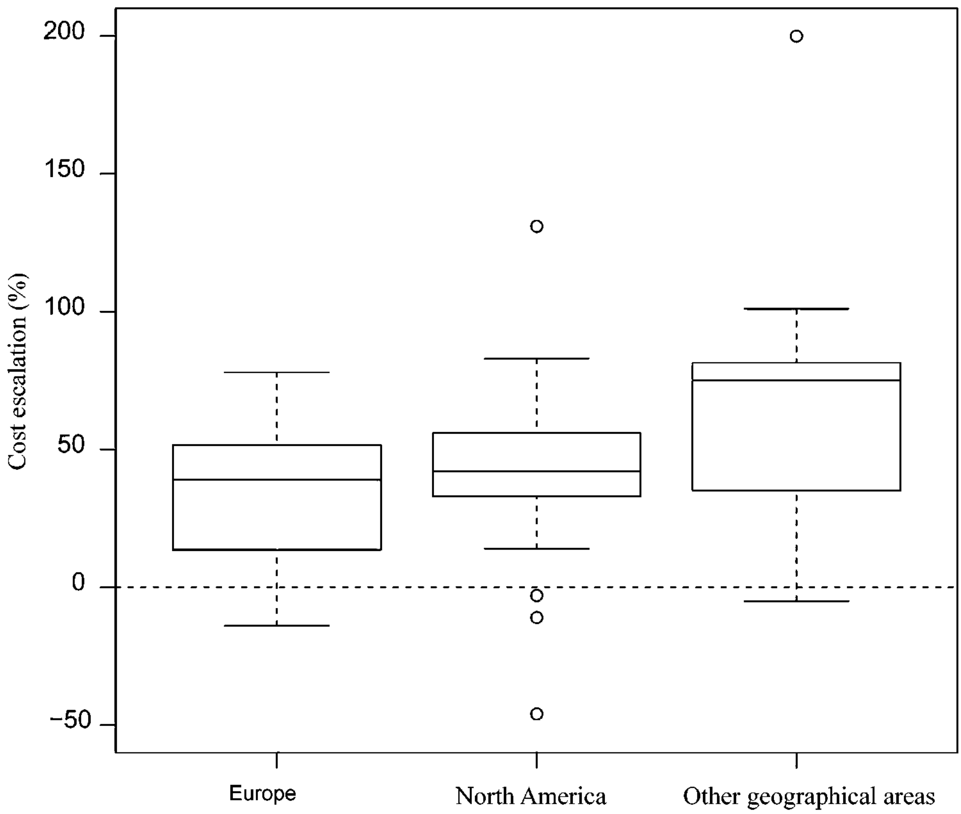

Figure 3. Box plots of cost escalation for rail split into geographical areas (constant prices).

Time may be measured by year of decision to build a project or by year of completion (the year operations begin). The year of completing a project, with inauguration and start of operations, is historically substantially more manifest than the year of decision to build. Consequently, it has been a great deal easier to obtain data on year of completion than on year of decision to build. Data were available on the year of decision to build for only 111 of the 258 projects in the sample, whereas data on the year of completion were available for 246 projects. We have tested development in cost escalation over time for both sets of data, although when evaluating the dependence of cost escalation on year, it is better to use year of decision to build rather than year of completion; the latter includes length of implementation phase, which has influence on cost escalation, causing confounding.

Figure 4 shows a plot of cost escalation against year of decision to build for the 111 projects in the sample for which these data are available. The diagram does not seem to indicate an effect from time on cost escalation. Statistical analyses corroborate this impression. The null hypothesis that year of decision has an effect on cost escalation cannot be supported ($p = 0.22$, F-test). A test using year of completion instead of year of decision (with data for 246 projects) gives a similar result ($p = 0.28$, F-test). Similar analyses have been carried out with year of decision combined with the logarithm of estimated cost as a measure of the size of projects, also split into rail, fixed links and roads. Year of completion and logarithm of actual

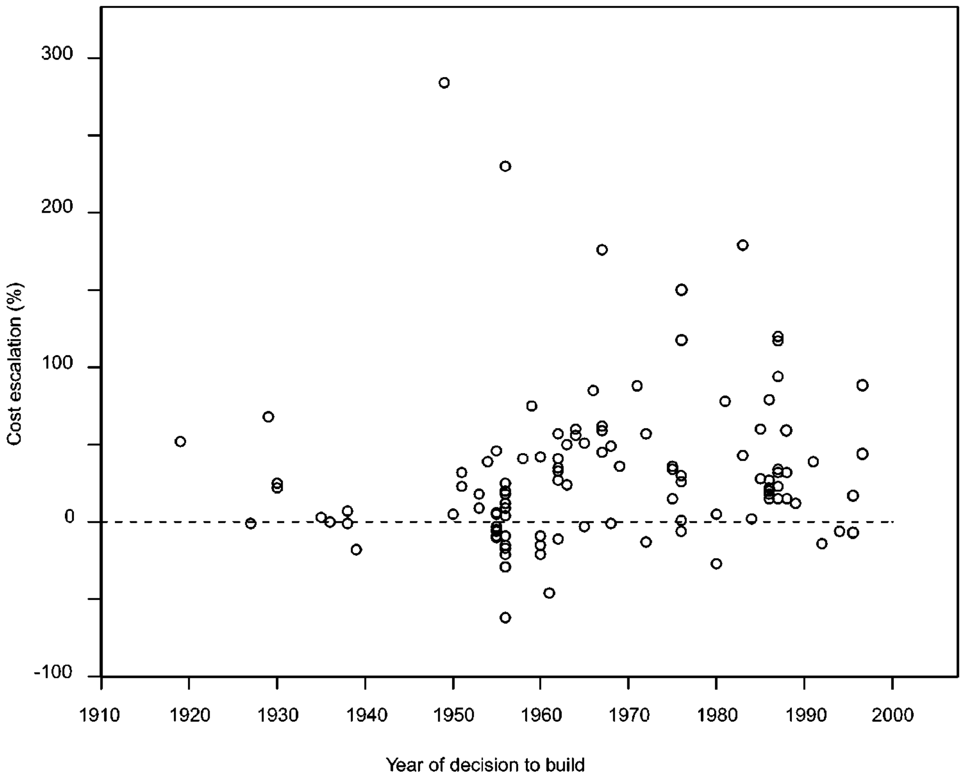

Figure 4. A century of cost escalation (constant prices).

cost was also tried. In no case could any statistical significant result be established, neither with main effects nor with interactions (in no case was a $p<0.10$ found). However, if two bridge projects with extreme data were considered statistical outliers and removed from the tests, cost escalation rose with the logarithm of estimated cost ($p=0.022$, F-test), but there was no interaction with year of decision.

We therefore conclude that cost performance has not improved over time. Cost escalation today is in the same order of magnitude as it was 10, 30 or 70 years ago. If techniques and skills for estimating and forecasting costs of transport infrastructure projects have improved over time, this does not show in the data. No learning seems to take place.

One might argue that in recent times the public and lobby groups have to a greater extent the possibility to intervene in the decision-making process and to effect project changes that are likely to cause cost escalation, for instance environmentalists who in order to safeguard nature successfully call for a tunnel to bury transport infrastructure that was planned to be constructed above ground. Our data cover a historical period long enough to include the situation before as well as after the public gained such increased possibilities to influence decision-making and costs. And again we see that cost estimates are as inaccurate and cost escalation as large before as after the new role of the public in transport infrastructure decision-making.

Where the pattern of cost underestimation and escalation is strikingly similar between different historical periods and different projects, the causes of escalation typically differ. To illustrate, for the Channel tunnel between England and France, changed safety requirements were a main cause of cost escalation. For the Great Belt link in Denmark, environmental protests and lobbying combined with accidents with flooding and a devastating fire made the budget balloon. For the Øresund link between Sweden and Denmark, it proved more costly than estimated to carve major new transport infrastructure into densely populated Copenhagen. And so on. *What*, exactly, causes cost escalation in projects is substantially more difficult to predict than the fact *that* cost escalation is likely to haunt projects. However, knowledge of the latter fact — the 'that' of cost escalation — is the appropriate, necessary and sufficient point of departure for financial risk analysis and management. Our data show this to be a lesson that has not been learned by project promoters and forecasters. Cost underestimation and escalation persist unabated and financial risk analyses based on real risks are sorely lacking in the planning of most major infrastructure projects, today and historically.

At first sight, it may seem strange that no learning appears to be taking place in this important and highly costly sector of public and private decision-making and that cost underestimation and escalation are allowed to continue unchecked decade after decade. After all, project promoters and forecasters, as we know them, seem to be as smart and as capable of learning as are other people. However, perhaps they have already learned what there is to learn? The behaviour of promoters and forecasters invite speculation that the persistent existence over time and space and project type of significant and widespread cost escalation is a sign that an equilibrium has been reached where strong incentives and weak disincentives for cost underestimation and related escalation may have taught project promoters that cost underestimation pays off. If this is the case, cost underestimation and escalation must be expected and it must be expected to be intentional. We have tested this and other explanations of cost underestimation and escalation elsewhere (Flyvbjerg*et al.* 2002). We found that cost underestimation and escalation indeed appear to be intentional and appear to be part of power games played by project promoters and forecasters aimed at getting projects started. Cost underestimation is used strategically to make projects appear less expensive than they really are in order to gain approval from decision-makers to build the projects. Such behaviour best explains why cost escalations are so consistent over time, space and project type (Wachs 1986, 1989, 1990, Flyvbjerg 1996, 1998).

7. Areas for further research

This paper presents a start only of the analysis of the sample of 258 transport infrastructure projects on which the paper is based. Immediate topics for further research would be, first, to analyse how length of project implementation correlates with cost escalation; very large projects, that take a long time to construct, might perform worse due to higher possibilities of changes over time. Second, the dependence of cost escalation on project size is also a topic for further research; do large projects perform worse than small ones, as is sometimes assumed in the literature? Answers to questions of this type may help explain the differences in cost escalation for different project types documented above. Third, the relationship between cost escalation and ownership of projects, for instance private versus public ownership, is an area of high interest for further research; do private projects perform better than public ones? Fourth, and finally, it would be interesting to know whether traffic forecasts are as systematically biased as are forecasts of costs; and if there is a bias in traffic forecasts does it exacerbate or compensate the bias in cost forecasts when calculating benefit–cost ratios of projects? These are areas of further research, which we will explore as we continue to mine the sample for additional data and results.

As mentioned, in our knowledge the sample used to arrive at the conclusions in this paper is the largest and the best of its kind. Nevertheless, there is scope for improvement, and this is a separate area for further research. Even if the sample is relatively large, it is too small to allow more than a few subdivisions if comparative statistical analyses are still to be possible. A central task for further research is therefore to enlarge the sample to represent better both different types of projects and different geographical locations of projects. Regarding types of projects, data for more fixed link and rail projects would be particularly useful. Such data would allow a better, i.e. a statistically corroborated, understanding of cost development and risk for subtypes of projects like tunnels, bridges, high-speed rail, urban rail and conventional rail. Such an understanding is non-existent today. Regarding different geographical locations of projects, immediate reward could be gained from data for projects outside Europe and North America, and especially for fixed links and roads. However, even for Europe and North America data on more projects are needed to allow better comparative analysis. A final area for further research is additional assessment of the conservative bias in the sample to get a better idea of its extent.

8. Summary and conclusions

Despite the enormous sums of money being spent on infrastructure development around the world, surprisingly little systematic and reliable knowledge exists about the costs, benefits and risks involved. The objective of the study reported here is to produce such knowledge. More specifically, the objective is to provide answers to the question of whether transport infrastructure projects perform as promised in terms of costs and benefits, or whether costs and benefits are highly uncertain phenomena involving significant elements of risk? The present paper covers the cost side of transport infrastructure development, based on a sample of 258 projects worth approximately US$90 billion (constant 1995 prices).

The answer to this question is, with overwhelming statistical significance, No, transport infrastructure projects do not perform as promised, and, Yes, costs are highly uncertain involving substantial elements of downside risk. The main findings from the study are (all highly significant, and most likely conservative) the following:

- Nine out of 10 transport infrastructure projects fall victim to cost escalation.
- For rail average cost escalation is 45% (SD = 38).
- For fixed links (tunnels and bridges) average cost escalation is 34% (SD = 62).
- For roads average cost escalation is 20% (SD = 30).
- For all project types average cost escalation is 28% (SD = 39).
- Cost escalation exists across 20 nations and five continents; it appears to be a global phenomenon.
- Cost escalation appears to be more pronounced in developing nations than in North America and Europe (data for rail only).
- Cost escalation has not decreased over the past 70 years. No learning seems to take place. Or, alternatively, project promoters and forecasters have learned what there is to learn, namely that cost escalation pays off; cost escalation is a

simple consequence of cost underestimation and underestimation is used tactically to get projects approved and built.

We conclude that cost estimates used in public debates, media coverage and decision-making for transport infrastructure development are highly, systematically and significantly deceptive. Cost–benefit analyses are typically centrally placed in infrastructure decision-making to calculate viability and to rank projects. However, cost–benefit analyses will be as misleading as the estimates of the costs and benefits that enter into such analyses, which in turn will result in the misallocation of scarce resources.

Moreover, the risks generated from misleading cost estimates are typically ignored or underplayed in infrastructure decision-making, to the detriment of social and economic welfare. Risks, therefore, have a doubly negative effect in this particular policy area, since it is one thing to take on a risk that one has calculated and is prepared to take, much as insurance companies and professional investors do, while it is quite another matter — that moves risk-taking to a different level — to ignore risks, especially when they are of the magnitude we have documented here. 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 our study shows that the risk of cost escalation, and related risk assessment and management, must be placed at the core of decision-making. Our goal with this paper has been to take a first step in this direction by producing the type of knowledge that is necessary to initiate such risk assessment and management.

The policy implications of our findings are clear. First, the findings show that a major policy problem exists for this highly expensive field of public policy. The problem is the pervasiveness of misinformation in the planning of transport infrastructure projects, and the systematic bias of such misinformation toward justifying project implementation. 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 their forecasters to make less deceptive forecasts. The problem of misinformation is an issue of power and profit and must be dealt with as such, using the mechanisms of accountability we commonly use in liberal democracies to control power and rent-seeking behaviour that have got out of hand. Institutional checks and balances must be put in place to curb misinformation, including financial, professional or even criminal penalties for ignoring or giving misleading information about risk and for consistent or foreseeable estimation 'errors'. The work of developing such checks and balances has been begun in Bruzelius*et al.* (1998) and Flyvbjerg*et al.* (2003), with a focus on four basic instruments of accountability in transport infrastructure planning and policy-making: (1) increased transparency, (2) the use of performance specifications, (3) explicit formulation of the regulatory regimes that apply to project development and implementation and (4) the involvement of private risk capital, even in public projects.

Acknowledgements

Research was supported by the Danish Transport Council and Aalborg University, Denmark. The comments of two anonymous referees were highly useful in preparing the final version of the paper.

References

BAUMAN, Z., 1998, *Globalization: The Human Consequences* (Cambridge: Polity).

BRUZELIUS, N., FLYVBJERG, B. and ROTHENGATTER, W., 1998, Big decisions, big risks: improving accountability in mega projects. *International Review of Administrative Sciences*, **64,** 423–440.

CAIRNCROSS, F., 1997, *The Death of Distance: How the Communications Revolution Will Change Our Lives* (Boston: Harvard Business School Press).

COMMISSION OF THE EUROPEAN UNION, 1993, *Growth, Competitiveness, Employment: The Challenges and Ways Forward into the 21st Century*. White Paper (Brussels: Commission of the EU).

FLYVBJERG, B., 1996, The dark side of planning: rationality and *Realrationalität*. In S. Mandelbaum, L. Mazza and R. Burchell (eds), *Explorations in Planning Theory* (New Brunswick: Center for Urban Policy Research Press), pp. 383–394.

FLYVBJERG, B., 1998, *Rationality and Power: Democracy in Practice* (Chicago: University of Chicago Press).

FLYVBJERG, B., BRUZELIUS, N. and ROTHENGATTER, W., 2003, *Megaprojects and Risk: An Anatomy of Ambition* (Cambridge: Cambridge University Press) (forthcoming).

FLYVBJERG, B., HOLM, M. K. SKAMRIS and BUHL, S. L., 2002, Cost underestimation in public works projects: error or lie? *Journal of the American Planning Association*, **68,** 279–295.

FOURACRE, P. R., ALLPORT, R. J. and THOMSON, J. M., 1990, *The Performance and Impact of Rail Mass Transit in Developing Countries*. TRRL Research Report 278 (Crowthorne: Transport and Road Research Laboratory).

HALL, P., 1980, *Great Planning Disasters* (Harmondsworth: Penguin).

HOLM, M. K. S., 1999, Inaccuracy of traffic forecasts and cost estimates in Swedish road and rail projects. Unpublished paper, Aalborg University, Department of Development and Planning.

KAIN, J. F., 1990, Deception in Dallas: strategic misrepresentation in rail transit promotion and evaluation. *Journal of the American Planning Association*, **56,** 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, University of California, Berkeley, Institute of Urban and Regional Development.

LEWIS, H., 1986, *The Metro Report: The Impact of Metro and Public Transport Integration in Tyne and Wear* (Newcastle: Tyne and Wear Passenger, Transport Executive).

MEREWITZ, L., 1973a, *How do Urban Rapid Transit Projects Compare in Cost Estimate Experience?* Reprint no. 104, University of California Berkeley, Institute of Urban and Regional Development.

MEREWITZ, L., 1973b, Cost overruns in public works. In W. A. Niskanen, A. C. Harberger, R. H. Haveman, R. Turvey and R. Zeckhauser (eds), *Benefit–Cost and Policy Analysis*. Reprint no. 114, University of California Berkeley, Institute of Urban and Regional Development (Chicago: Aldine), pp. 277–295.

NATIONAL AUDIT OFFICE, DEPARTMENT OF TRANSPORT, 1985, *Expenditure on Motorways and Trunk Roads* (London: National Audit Office).

NATIONAL AUDIT OFFICE, DEPARTMENT OF TRANSPORT, 1992, *Contracting for Roads* (London: National Audit Office).

NATIONAL AUDIT OFFICE, DEPARTMENT OF TRANSPORT, SCOTTISH DEVELOPMENT DEPARTMENT, AND WELSH OFFICE, 1988, *Road Planning* (London: HMSO).

NIJKAMP, P. and UBBELS, B., 1999, How reliable are estimates of infrastructure costs? A comparative analysis. *International Journal of Transport Economics*, **26,** 23–53.

PICKRELL, D. H., 1990, *Urban Rail Transit Projects: Forecast versus Actual Ridership and Cost* (Washington, DC: US Department of Transportation).

PICKRELL, D. H., 1992, A desire named streetcar: fantasy and fact in rail transit planning. *Journal of the American Planning Association*, **58,** 158–176.

RICHMOND, J. E. D., 2001, A whole-system approach to evaluating urban transit investments. *Transport Reviews*, **21,** 141–179.

RIKSREVISIONSVERKET, 1994, *Infrastrukturinvesteringar: En kostnadsjämförelse mellan plan och utfall i 15 större projekt inom Vägverket och Banverket* (Stockholm: Riksrevisionsverket).

SKAMRIS, M. K. and FLYVBJERG, B., 1997, Inaccuracy of traffic forecasts and cost estimates on large transport projects. *Transport Policy*, **4,** 141–146.

Szyliowicz, J. S. and Goetz, A. R., 1995, Getting realistic about megaproject planning: the case of the new Denver international airport. *Policy Sciences*, **28,** 347–367.

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,** 23–30.

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

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

Walmsley, D. A. and Pickett, M. W., 1992, *The Cost and Patronage of Rapid Transit Systems Compared with Forecasts*. Research Report 352 (Crowthorne: Transport Research Laboratory).

World Bank, 1994, *World Development Report 1994: Infrastructure for Development* (Oxford: Oxford University Press).