EJTIR

Issue 15(3), 2015
pp. 286-303
ISSN: 1567-7141
tlo.tbm.tudelft.nl/ejtir

The Impact of External Quality Assurance of Costs Estimates on Cost Overruns: Empirical Evidence from the Norwegian Road Sector

James Odeck[1]
Department of Civil and Transport Engineering, Norwegian University of Science and Technology, Norway.

Morten Welde[2]
Department of Civil and Transport Engineering, Norwegian University of Science and Technology, Norway.

Gro Holst Volden[3]
Department for Applied Economics, SINTEF, Norway.

Cost overruns in transport infrastructure projects are prevalent and have been well documented in the literature; see, for instance, Flyvbjerg et al. (2003), Odeck (2004 and 2014) and Cantarelli et al. (2010). Governments may therefore exert efforts to reduce overruns by implementing strategies such as quality assurance of cost estimates, whereby external consultants are engaged to assure the accuracy of estimates. However, the literature has to a lesser extent provided evidence on what governments do to combat overruns and whether those efforts work in practice. This paper provides such evidence for the case of Norway, where the government implemented a quality assurance regime for cost estimates above 500 million NOK in the early 2000. Apart from explaining the Norwegian quality assurance regime, the paper uses statistical inferences to compare the magnitudes of cost overruns in the pre-and post- quality assurance periods. The statistically significant derived results are as follows: (i) quality assurance has led to a reduction in cost overruns; (ii) quality assurance has not, however, led to improved accuracy of the estimates provided by the authorities – rather, it has led to systematic overestimation by the authorities; and (iii) external consultants are more accurate than the authorities. We conclude that the quality assurance regime is achieving the objectives of reducing cost overruns and that similar regimes should be considered for smaller projects where overruns have been shown to be very large.

Keywords: Cost *overruns; government actions; road projects; quality assurance*.

1. Introduction

The planning of transport projects is a complex process involving legal, administrative, economic and technical issues. To provide decision makers with the best possible information for decision making, planners are required to assess and, if possible, quantify all the effects of implementing a

[1] A: 7491 Trondheim, Norway, T: +47 480 25 940 E: james.odeck@ntnu.no
[2] A: 7491 Trondheim, Norway, T: +47 977 300 31 E: morten.welde@ntnu.no
[3] A: PO Box 4760 Sluppen, 7465 Trondheim, Norway, T: +47 957 45 565 E: GroHolst.Volden@sintef.no

new facility or policy. The estimation of costs and benefits is thus crucial for rational transport planning and efficient resource allocation.

The investment cost is perhaps the most frequently used management parameter because it can be expressed quantitatively and with great precision. Costs can make participants accountable; they can be used to gauge progress and result attainment and to compare economic viability (Samset and Volden 2012). Transport planning and economics literature has attested that cost overruns are a prevalent problem that may jeopardise decision making and lead to inefficient resource allocation. Transport planners and economists will normally advise decision makers to prioritise according to the ratio of net benefit to funding through the government budget, i.e., the benefit-cost ratio, until the budget is used up. The eloquence of this decision-making rule is that it advises the decision makers to prioritise projects that generate the highest benefits per euro spent. In contrast, consider the case where the funding required for a given project, i.e., the investment cost, is underestimated but not known to the decision makers at the time of decision making. It then becomes clear that the underestimation of costs would lead to inefficient allocation of resources; had the decision makers more accurately known the costs at the time of decision making, they may have opted for other solutions, e.g., (i) not to implement the project at all; (ii) implementing the project in a different form or; (iii) implementing another project. The bottom line is that the use of more resources than necessary is tantamount to resource wastage whereby the excess resources wasted could have been used for other beneficial purposes. The problem of cost overruns is not only confined to inaccurate benefit-cost analyses as described in the example above. In the case of Norway, Sager and Ravlum (2005) indicated that cost-benefit analyses did not seem to play a particularly important role in Norwegian transport infrastructure policy-making. This may indicate that the importance of accurate cost estimates might be relatively minor. However, for those making cost estimates and for those making decisions, getting the estimates right leads to accountability and reliability of the estimates and the decision that are made especially with regards to the taxpayer. The magnitudes of cost overruns are thus a measure of performance for which decisions are judged. Notwithstanding, even if cost-benefit analyses does not matter, investments costs matter a great deal for decision-makers, see for instance Odeck (2014). Irrespective of this, it will be a violation of democratic practices if the parliament which makes decisions is misled with regards to investment cost of public projects.

The magnitude of transport project cost overruns has been addressed extensively in the literature in the recent years; see, for instance, van Wee (2007) and Siemiatycki (2009) for a review of the literature. In individual countries such as Norway, Sweden and England, auditor generals have addressed similar issues and have urged governments to consider improving the quality of their cost estimates; see, for instance, the UK National Audit Office (2007) in the case of England; Riksrevisjonen (2002) in the case of Norway and Riksrevisionsverket (1994) and Riksrevisionen (2011) in the case of Sweden. From the efforts that have been made in the literature to address and identify the magnitude of cost overruns of road projects across the globe, two important questions can be raised:

1. What efforts do governments actually make to reduce the magnitude of overruns and,
2. Do those efforts actually help to reduce the magnitude of overruns?

To our knowledge, there are only two papers in the literature that have attempted to address these two questions; Magnussen and Olsson (2006) and Odeck (2014). The former paper studied the magnitude of cost overruns across several public projects including transport while the latter dealt with the magnitude of overruns of road projects after reorganizations of the Norwegian Public Roads Administration charged with planning and building roads.

This paper contributes to the literature of cost overruns in the transport sector by addressing and examining efforts that governments make to reduce overruns and whether those efforts help to reduce overruns. We use the Norwegian road sector as a case study. Following several high profile public sector projects with cost overruns, the Norwegian government implemented a quality assurance regime in the year 2000 whereby all projects with estimated costs above 500 million NOK (app. 60,000 million EUR) must undergo scrutiny by external consultants. We first explain the Norwegian model of cost estimation and the quality assurance regime that aims to reduce cost overruns. Second, to ascertain whether the quality assurance regime has helped to reduce the magnitudes of overruns, we compare the magnitudes of overruns for projects estimated under the pre- and post-quality assurance regimes, using classical statistical tests for inference purposes.

The remainder of the paper is organised as follows. Section 2 provides a literature review of studies that have examined cost overruns. Section 3 outlines the Norwegian quality assurance regime. Section 4 describes the data, while section 5 presents the empirical results. Conclusions and recommendations are presented in Section 6.

2. Literature review

Several studies have demonstrated that underestimation of costs (also known as cost overrun) is a major source of risk in project appraisals. In the mid-1980s, Morris and Hough (1987) surveyed the literature on project successes and failures. Using data from over 3,500 projects in the public domain, they found that overruns of 40 to 200 percent were common. Furthermore, they found that the track records of public transport projects were particularly disappointing..

In the US, Pickrell (1992) found that seven of eight rail transit projects had capital costs well above forecasts. Capital spending overruns ranged from 17 percent to more than 150 percent. In the UK, the consultancy Mott MacDonald (2002) found that construction cost escalations in 50 UK projects varied from 24 percent to 36 percent. Based on data from over 250 transport projects worldwide, Flyvbjerg *et al.* (2003) demonstrated that estimates of construction costs in transport infrastructure projects were inaccurate and that substantial cost escalation was the rule rather than the exception for all categories of projects. Cost escalation appeared to be a global phenomenon, and average cost overruns ranged from 20 percent to 45 percent. The authors concluded that cost estimates and cost escalation have not improved over the last few decades and that cost estimates provided to decision makers are highly misleading. In Canada, Berechman and Wu (2006) studied 163 road projects from the Vancouver Island Highway Programme and found that 8 out of 10 projects had cost overruns, with an average overrun of 5.9 percent. Outturn cost varied from 52 percent below to 134 percent above estimates. Makovšek *et al.* (2012) studied highway projects in Slovenia and showed that a majority of the projects had cost overruns: 30 percent on average for projects opened in the years 1994-1999 and 19 percent for projects opened thereafter.

Odeck (2004 and 2014) has provided two studies of cost performance in the Norwegian road sector. The first study covered 620 projects from 1992 to 1995; the majority were small or very small projects (>10 mill. EUR). He found that approximately half the projects experienced cost overruns; the average among all projects was 7.9 percent, with variations from -58.5 percent to +183 percent. In the second study, which covered 1,045 projects from 1993 to 2007, he found that cost performance had improved over time and that the average cost overrun of projects carried out from 2004 to 2007 was just 4 percent. Furthermore, he suggested that the changes observed, especially with regard to the larger projects, may have been due to organisational changes that took place within the government authority. Another relevant study is Magnusson and Olsson (2006) who considered the magnitudes of cost overruns of several public projects including transport after the quality assurance regime was implemented. They found that the quality assurance had helped in reducing the magnitudes of overruns. They did, however, not consider transport projects separately.

Above all, these studies suggest that cost overrun is a global phenomenon and that overruns are more common than under-runs. Flyvbjerg (2005), Siemiatycki,(2009) and Cantarelli *et al.* (2010) have suggested three explanation types that can account for inaccuracy in forecasts of costs and benefits: technical, psychological, and political-economic.

Technical explanations stem from unforeseen events in the planning and/or construction phase, such as changes in input prices, inadequate data, lack of skills/experience and change of project scope and design. Technical explanations are considered to be variables that influence cost overruns rather than explain them. These explanations are circumstances that planners should ideally be able to manage and provide contingencies for in their estimates. Failing to do so could clearly influence cost overruns without necessarily being the main cause.

Psychological explanations account for cost overruns and benefit shortfalls in terms of what psychologists call planning fallacy and optimism bias, i.e., the tendency to ignore risk, even when faced with projects that may involve high risk. As a result, planners and promoters pursue initiatives that are unlikely to come in on budget or on time or to ever deliver the expected returns.

Political-economic explanations see planners and promoters as deliberately and strategically overestimating benefits and underestimating costs when forecasting the outcomes of projects. Projects that promise to deliver high benefits at low costs may appear more attractive to decision makers, and project promoters can thus be tempted to deliberately underestimate costs or overestimate benefits so that their projects are selected for implementation. Forecasts are adjusted to derive the most politically or organisationally attractive outcomes. There is a range of studies (see, e.g., Wachs, 1987, 1989; Mackie and Preston, 1998; Flyvbjerg, 2007) supporting claims of dishonesty in the transport community. Decisive proofs of these claims have, however, been difficult to find.

A general conclusion that can be drawn from the review above is that the literature has extensively and adequately addressed the magnitudes of cost overruns and also suggested some probable causes. Furthermore, the magnitudes of overruns in Norway seem to be much lower as compared to those of other countries; se for instance Siemiatycki (2009) who reports average cost overruns in the range of 50-100% for some studies, while the comparable figure for Norway is about 10%. There is no apparent reason why Norway should show such a low level of overruns as compared to other countries. A probable reason may be as pointed out by Odeck (2004 and 2014); the road sector has in the last decades been under constant pressure from the Auditor General to reduce cost overruns.

However, very few studies have addressed what governments do to overcome the observed overruns and how successful governments have been in those endeavours. Thus, the current paper is a useful contribution to the literature on the topic of cost overruns in transport projects, especially with respect to what governments do to combat overruns and whether the endeavours of governments have been successful.

3. The Norwegian Quality Assurance regime

In Norway, the 1980s and 1990s were littered with scandals surrounding large-scale public investment projects. The government thus commissioned a study on the planning, implementation and follow-up of large public investment projects. The working group found that 8 out of 11 projects had cost overruns and concluded that, even if there were signs of improvement, there was a need to standardise planning procedures and cost estimation methodologies (Berg et al., 1999). The suggested solution was a Quality Assurance (QA) regime for public projects whose main objective was to combat overruns[4]. This recommendation was accepted and implemented by the Ministry of Finance, which oversees funds allocated to and spent on public projects. The QA regime was effective beginning in 2000. This regime has come to be known as the QA2 regime, and it is conducted at the second and final stage of planning process. The QA1 is discussed later.

The concept of QA is not new and has been applied for decades in the private sector. QA is a system for ensuring a desired level of quality in the development, production or delivery of products and services. With regards to the cost estimates of transport projects, QA implies that the estimates made by the authorities are subjected to external scrutiny to ensure that quality standards of cost estimates are being met and to detect and correct any deviations. Unfortunately, the good intentions of QA are seldom practiced with respect to cost estimation in the public sector. Norway is an exception, as external QA has been mandated for almost all large public sector projects.

The Norwegian QA2 regime as implemented in 2000 required all cost estimates for individual projects equal to and above 500 million NOK[5] conducted by all public organisations (such as the Norwegian Public Road Authorities, which provides cost estimates for roads projects) to be scrutinised by external consultants chosen by the Ministry of Finance following very strict competition. Any consultants who have previously been involved with the project cannot be awarded tender for that particular project. Investments in the oil sector and investments made by state-owned limited companies are exempt from the QA2 regime. The limit for costs estimates that must be subjected to QA was revised in 2011 and is now 750 million NOK.

From 2005 onwards, the regime was expanded to early QA of the chosen concept (QA1). This process takes place at the end of the pre-study phase, before a decision is made by the Cabinet to start a pre-project. This requirement implies that both benefits and costs are subjected to a QA1 regime at the earliest stage of project development. QA1 is seen as a vital step to ensure that the most economically viable concept is planned further. At this stage, only one or no alternatives for further planning are recommended.

Next, and important for our study, the QA2 is performed before the project is submitted to Parliament for approval and funding. Normally, this is done at the end of the pre-project phase. The consultant reviews the documentation behind the proposition presented to Parliament, with an emphasis on cost estimates. In addition, the QA2 charts the management challenges in the remaining phases of the project. As a basis for the QA2 exercise, the following documents are prepared by the responsible ministry/agency: (i) an overall project management document (steering document); (ii) a complete base estimate; and (iii) an assessment of at least two alternative contract strategies. The external consultant reviews these documents, undertakes an assessment of success factors and pitfalls, and quantifies the uncertainty of the cost estimate. On that basis, the consultant gives recommendations regarding the cost frame of the project, including necessary contingency reserves to account for uncertainty. The consultant also gives recommendations as to how the project should be managed to maximise the probability that the cost frame will hold. Pertinent to the analysis that follows, the consultant gives an alternative cost estimate that can be compared to those of the ministry or agency responsible for the original estimates.

Major projects are always affected by uncertainty during their implementation. Cost estimates calculated as part of the QA2 scheme are therefore based on stochastic cost estimation. Two key figures that are estimated by both the road authorities and the external consultants are termed P50 and P85. P50 is the median, which means that there is a 50 percent probability that the cost

---

[4] For a comprehensive description of the Norwegian QA scheme, see "Concept Research Programme" (2015) at http://www.concept.ntnu.no/english.
[5] 1 NOK ≈ 0,12 € as of February 5, 2015.

will be within this numerical value. The P85 value is higher, as it represents an 85 percent likelihood that the cost will be within this numerical value.

The cost frame approved by the Parliament is normally slightly lower than the P85 value. The implementing party will, however, have to manage the project within a lower steering frame, which generally corresponds to the P50 value. Cost increases above this figure require consent at the ministry level.

The proposed cost frame is normally P85 with deductions for possible simplifications and reductions (reduction list) that can be handled during the project if the cost frame is in danger of being exceeded. The agency's steering frame is lower to avoid incentives for the use of contingency reserves (Samset and Volden, 2013).

The Parliament and the responsible ministry are of course not required to follow the recommendations from the QA consultants which could be higher or lower than the agency's estimates. The final cost frame for the project is decided by Parliament. The result of the cost estimation process can be illustrated by means of a cumulative graph that shows the range of possible outcomes and their relative likelihood of occurrence. This is illustrated in Figure 1.

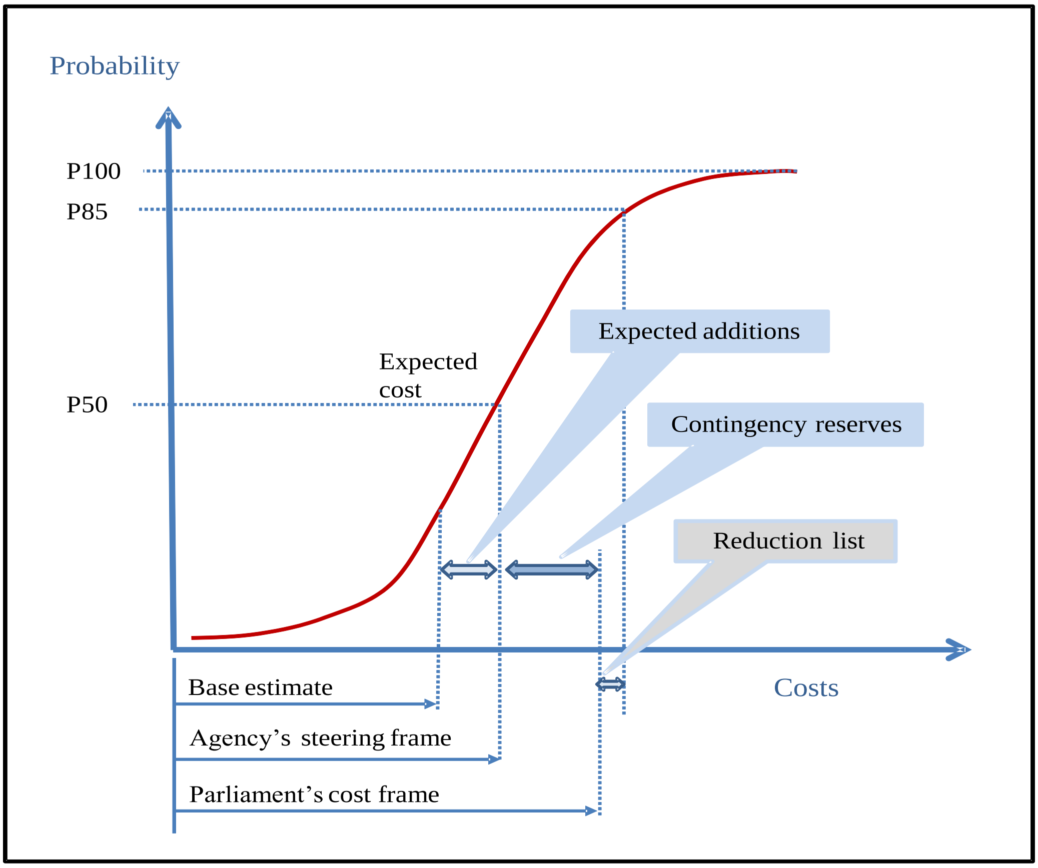

Figure 1. Stochastic cost estimation and definition of key terms

Figure 2 illustrates the different stages of the Norwegian QA regimes and their roles in project approval.

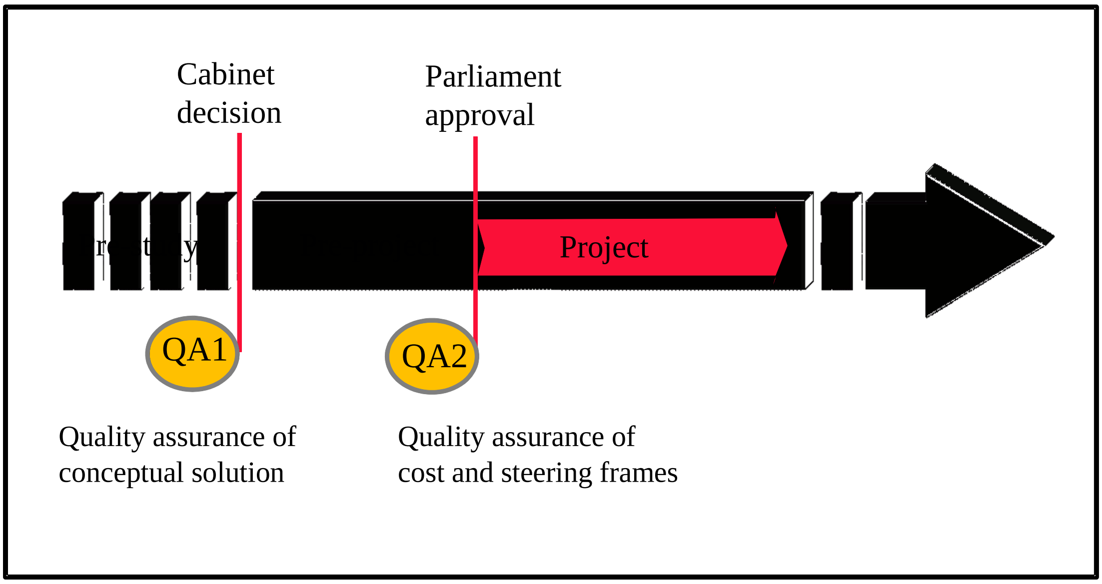

Figure 2. The Norwegian Quality Assurance regime for major public investment projects

From Figure 2, it is clear that QA2 is the most critical regime for decision making, for the simple reason that the decision maker relies on the cost estimates made at that stage, where the figures from the external consultants estimates are assumed to be the correct ones. As of 2015, QA2 has been operational for 15 years and QA1 for 10 years. In total, some 70 QA1 reports and some 150 QA2 reports have been compiled across the Norwegian public sector.

The first results from the regime, based on 40 road, rail, defence, ICT and construction projects subjected to QA2 and now in their operational phase, were published in Samset and Volden (2013). Of the 40 projects, 32 (80 percent) were completed within the cost frame set by the Parliament. The final costs of the projects were distributed symmetrically around the expected value. This result suggests that the QA2 scheme and the methodology used for cost estimation may have contributed to more accurate cost estimation.

The above-mentioned study did not include pre-QA data, however, and did not study the road sector specifically. The main hypothesis in this paper is that QA of cost estimates has contributed to reducing cost overruns by increasing the overall quality of estimates and by creating a disciplining effect that may have reduced promoters' and planners' propensity to strategically underestimate.

4. Data and Methodology

We use data from the Norwegian road sector for the simple reason that it has a larger pool of public projects that must undergo QA scrutiny and furthermore, data are adequately monitored and are retrievable. The data used in this study were retrieved from official documents available at the road authorities' archives that contain planning data and documents on completed projects, including estimated costs provided by the authorities. The data also include the reports compiled by the external QA2 consultants and, finally, the actual costs of the projects. The data set is divided into two periods: pre-QA2 and post-QA2. In the pre-QA2 period, there was no quality assurance, and the estimates provided to decision makers were solely those made by the road authorities. The data set includes projects that were estimated in the period 1993 – 2012. The cut-off year marking the inauguration of the QA2 regime is the year 2000, when QA2 was implemented. Thus, projects with cost estimates beginning in 2000 were classified as post-QA2 projects and those before 2000 as pre-QA2 projects.

As QA2 is conducted for very large projects with estimated cost above 500 million NOK, the projects from the pre-QA2 period had to be of the same cost estimate magnitude to make appropriate comparisons with those of the post-QA2 period. All project cost estimates and actual costs were adjusted to the 2012 cost levels.

A closer examination of the data set revealed that the project size as measured by investment cost (and also number of road kilometres) has changed a great deal since the quality assurance regime was implemented in 2000. For instance, had quality assurance been implemented in 1993, only three projects would have been scrutinised that year, and in 1995, only one would have been scrutinised. For the period after 2000, an average of approximately four projects each year has fulfilled the requirement of quality assurance.

In selecting the projects to be included in this study two important criteria were used: (1) that the project had an estimated cost above 500 mill NOK and thus satisfied the conditions for quality assurance irrespective of the estimation period and (2) that the actual costs were validated and submitted to the authorities after inspection to confirm that the project was delivered according to the terms in the contract. It was revealed that some projects did not satisfy these requirements; hence, such projects are not included in this study. Following these considerations, 18 and 22 projects were identified for the pre- and post-quality assurance periods, respectively.

Next, the derivation of cost overruns must be defined. We adopted the definition that is commonly used in the forecasting literature that measures overruns (forecasting bias) in percentages as follows:

$$PCO_i = \left(\frac{Y_i - F_i}{F_i}\right) \times 100. \tag{1}$$

where PCO_i is percent cost overrun for project *i*; Y_i is the actual cost for project *i* and; F_i is the estimated (forecasted) costs for project *i*. Note that F_i represents the cost frame, i.e., the P85 value, cf. Figure 2. Other authors in the literature on forecasting, e.g., Makridakis *et al.* (1998), have advocated for a measure of forecasting bias as a percentage of actual cost, where Y_i should be the denominator. For comparison purposes with studies that have examined overruns among road projects, such as Flyvbjerg *et al.* (2003), we resolved to measure overruns using Equation (1). An important point here is that we are interested in measuring overruns (bias) relative to estimates, on the understanding that it is the estimates that can be improved to reduce the magnitude of overruns through, for example, quality assurance. Furthermore, PCO_i can also be interpreted as a measure of estimate accuracy. If it is equal to 0, accuracy prevails; if less than 0, under-run prevails, and if greater than 0, overruns prevail.

When comparing two cost estimation regimes with regards to pre- and post-quality assurance, some averaging of the observed PCOs across individual projects is required. The appropriate averaging is the so-called Mean Percentage Cost Overrun (MPCO), which is defined based on (1), as follows:

$$MPCO_R = \frac{1}{n_R}\sum_{i=1}^{n_R} PCO_i \tag{2}$$

Where $MPCO_R$ is the mean percentage cost overrun in regime *R*. $MPCO_R$ is likely to have a small value because the negative and positive values tend to offset each other. It does, however, show which way the overruns tend to go, namely, positive or negative. If under-overruns are equally a problem, the Mean Absolute Percentage Overrun (MAPCO) can be used:

$$MAPCO_R = \frac{1}{n_R}\sum_{i=1}^{n_R}\left|PCO_i\right| \tag{3}$$

Where $MAPCO_R$ is the mean absolute percentage overrun in regime R; it is defined by taking the absolute value of the $MAPCO_R$ to gain knowledge regarding the absolute size of estimate inaccuracy. The summary statistics of the projects included in the analysis are show in Table 1 below.

Table 1. Summary Statistics

	Pre-quality assurance			Post-quality assurance			Overall		
	Estimated	Actual	PCO	Estimated	Actual	PCO	Estimated	Actual	PCO
Mean	778.76	912.03	13.0 %	1314.08	1159.95	-10.8 %	1073.18	1048.39	0 %
Median	673.13	703.12	6.5 %	1199.00	1178.00	-10.8 %	921.00	892.36	0 %
Min	442.71	446.36	-22.7 %	599.00	432.00	-47.0 %	442.71	432.00	-47 %
Max	1499.28	1997.43	70.5 %	2967.00	2636.00	19.6 %	2967.00	2636.00	71 %
St.dev	314.97	492.94	21.5 %	568.17	505.68	17.2 %	541.41	514.98	22 %
No. obs.	18			22			40		
% accurate estimates	5.56			4.55			5.00		
% overruns	72.22			27.27			47.50		
% under-runs	22.22			68.18			47.50		

Several initial observations can be made with regards to Table 1. First the number of observations between the pre- and post-quality assurance periods towards the bottom of Table 1 shows that they are relatively close to each other, 18 versus 22. It is therefore evident that they can be compared to each other statistically with respect to mean differences in magnitudes of overruns. Second, projects in the post-quality assurance period are larger than in the pre-quality assurance period in terms of estimated and actual costs, as can be judged by observing the mean, min and max. This result verifies the earlier remarks that Norwegian road projects have increased in size since quality assurance was implemented in early 2000s. Third, the summary statistics reported in the table show that the mean, median and standard deviation of the overruns are higher in the pre-quality assurance period compared to the post-quality assurance period. Fourth, values for overall measures across all periods show that the cost estimates have been accurate with a mean overrun of zero because of the under-runs in the post-quality assurance period. Finally, consider the percentage number of projects with accurate estimates, overruns and under-runs as shown on the lowermost part of Table 1. While the percentage of projects with accurate estimates is almost the same between the two regimes, the pre-quality assurance has 72.22 percent overruns and only 22.22 percent under-runs. For the post-quality assurance period, the results are almost the opposite of the pre-quality assurance: 68.18 percent under-runs and only 27.27 percent overruns.

These tentative results so far indicate that the external quality assurance scheme has led to improvements in the magnitudes of overruns, but they tell us little about the comparative accuracy of cost estimates between the two regimes. Consider next that cost under-runs are as problematic as overruns because, if under-runs are prevalent, excess funds could have been used to implement more projects and thereby incur benefits to users at an earlier stage[6]. This particular

[6] This proposition that cost overestimates are an equally serious problem can be disputed; in this case, a positive decision to build the project would still have been made, unlike the case of cost underestimates.

issue has been considered less frequently in the literature on the accuracy of cost estimation for transport projects. To infer this aspect of cost estimation from our data set, consider the absolute values of mean overruns in Table 1. This table shows that estimate inaccuracy is prevalent in both cases, with mean values of 11 and 13 percent for post- and pre-quality assurance, respectively. These results indicate that while quality assurance has reduced overruns, it cannot be concluded that it has led to improvements in the accuracy of forecasts; accuracy of forecasts is not to be understood as a reduction of overruns alone but as forecasts meeting outturns.

The observations concerning the summary statistics presented in Table 1 are merely tentative. To draw firm conclusions regarding the impact of quality assurance on overruns, a succinct statistical analysis is called for. This is the subject matter of the next section.

5. Empirical results

In the preceding section, we established that there are some general differences in the magnitudes of mean percentage cost overruns (MPCOs) between the pre- and post-quality assurance periods. In this section, we conduct statistical assessments to ascertain the statistical significance of those observations. We proceed as follows: (i) we conduct a host of statistical tests to confirm whether there are any significant differences between PCOs in the pre- and post-quality assurance periods; (ii) we check for selection bias to ensure that results do not depend on the selection of projects analysed; and (iii) we test for the accuracy of quality assurers' estimates versus the authorities' estimates after the quality assurance regime was implemented. The latter test will reveal whether authorities or consultants more accurately estimate project costs.

5.1 Are there significant differences in PCOs between pre- and post- quality assurance regimes?

A more visual initial assessment, similar to those presented in the previous section, can be provided by comparing PCO distribution plots between the pre-and post- quality assurance regimes. If the distribution of PCOs in a given regime has a mean closer to zero and a smaller variation compared to the other, it can be considered to be more accurate than the other regime. A comparison of the distributions is provided in Figure 3.

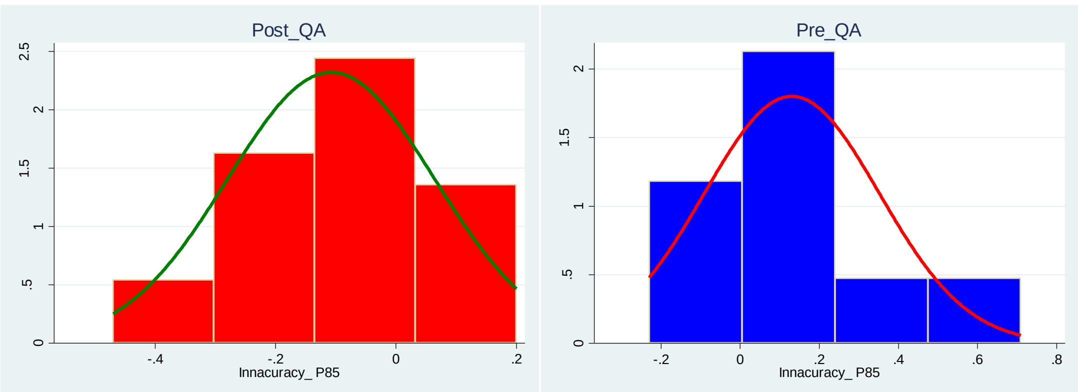

Figure 3. A comparison of the distributions of PCOs between post- and pre-quality assurance periods

It is clear from Figure 3 that the post-quality assurance distribution of PCOs is skewed to the left, while that of the pre-quality assurance is skewed to the right. The mean percentage cost overruns (MPCOs) are negative and positive for the post- and pre-quality assurance periods, respectively, indicating that quality assurance led to under-runs while pre-quality assurance had overruns.

The next question to address is whether the means for the two periods observed above are significantly different from zero and whether they are significantly positive or negative; the former would indicate significant overruns and the latter significant under-runs. To address these issues, we performed one-sample t-tests to compare the mean of each sample to a given number, in our case, whether they are different from zero, greater than zero or less than zero. The t-statistic and its p-value are calculated under the assumption that the sample comes from an approximately normal distribution. If the p-value associated with the t-test is small (0.05 is often used as the threshold), there is evidence that the mean is different from the hypothesised value. If the p-value associated with the t-test is not small (p > 0.05), the null hypothesis is not rejected, and one can conclude that the mean is not different from the hypothesised value. The results of the one-sample test for each of the regimes are presented in Table 2.

From Table 2, the t-statistic for post-QA is -2.934, with 21 degrees of freedom. The corresponding two-tailed p-value is .0078, which is less than 0.05. We conclude that the mean post-QA is different from 0. The same results apply in the case of pre-QA, where the t-statistic is 2.499 and the corresponding p-value is 0.0023. Regarding whether the means are significantly positive or negative, the table shows that post-QA, the mean overruns are significantly negative, whereas they are significantly positive pre-QA. The opposite test, i.e., of the hypotheses that post-QA means are positive and that the pre-QA means are negative, rejects these hypotheses at the 5 percent level.

Table 2. Single sample t-test for mean different from zero

Post-QA						
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.Interval]	
Innacuracy	22	-0.108	0.037	0.172	-0.184	-0.031
				t = -2.9388 Df. = 21		
Null Hypothesis		Ho: mean = 0				
Alternative hypotheses		Ha: mean < 0	Ha: mean=0	Ha: mean>0		
Significance level		Pr(T < t) = 0.0039	Pr(|T| > |t|) =0.0078	Pr(T > t)=0.9961		

Pre-QA						
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.Interval]	
Inaccuracy	18	0.131	0.052	0.222	0.020	0.241
				t = 2.4994 Df. = 17		
Null Hypothesis		Ho: mean = 0				
Alternative hypotheses		Ha: mean < 0	Ha: mean=0	Ha: mean>0		
P-values		Pr(T < t) = 0.9885	(|T| > |t|) =0.023	Pr(T > t)=0.0115		

The next question that should be addressed before considering whether there is statistical evidence as to the difference between the two regimes is the normality test. The test tells whether the distribution of PCOs in the two regimes is normally distributed around their mean values.

We used the Kolmogorov-Smirnov test (K-S) and the Shapiro-Wilk (S-W) test, which test the null hypothesis that a data sample comes from a normally distributed population. Given that the null hypothesis is that the PCOs are normally distributed, if the p-value is less than the chosen significance level, then the null hypothesis is rejected (i.e., one concludes the data are not from a normally distributed population). Table 3 reports the test results.

Table 3. The Kolmogorov-Smirnov test (K-S) and Shapiro-Wilk (S-W) test for normality in PCOs[1]

	Kolmogorov-Smirnov test (K-S)[1]			Shapiro-Wilk (S-W) test		
	Statistics	df	Significance	Statistics	df	Significance
Post-QA	0.835	22	0.965	0.978	22	0.891
Pre-QA	0.903	18	0.037	0.029	18	0.065

[1]the test statistics reported are for skeweness only

The results in Table 3 show that while we cannot rule out normality with respect to the distribution of PCOs in the post-quality assurance period, the hypotheses of normality of PCOs in the pre-quality assurance period is rejected at the 5 and 10 percent significance levels by the Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests, respectively. One interpretation of these results is that during the pre-quality assurance period, PCOs were positively skewed with a positive mean at 13 percent, while in the post-quality assurance period, PCOs are normally distributed with a negative MPCO at approximately -11 percent, i.e., under-runs rather than overruns became prevalent after the quality assurance programme was implemented.

The results in Table 3 are tentative in the sense that the tests are conducted individually for each group of observation, i.e., sequentially for each estimation regime. There may be overlaps between these observations such that a firm conclusion as to the differences between them cannot be drawn. Statistical tests that compare overruns between the two regimes are therefore necessary before any firm conclusions can be drawn as to the differences observed. To test for significant differences, several non-parametric tests can be performed. We used three alternative tests: the Kruskal-Wallis test, the Median independent sample test and the two sample mean-comparison t-test. The difference between these tests is that the Kruskal-Wallis method tests the null hypothesis that the distributions of the samples (i.e., PCOs) are the same across the two periods being compared, whereas the Median method tests the null hypothesis that the samples are taken from populations with the same median. The two sample mean-comparison t-test tests the null hypothesis that the means are equal under two different assumptions: (a) the variances are equal, and (b) the variances are not equal. The test statistics with a very high p-value in both of the methods indicate that there are no significant differences in mean and median values across the two regimes. Table 4 reports the results of the four statistical tests.

Table 4. Testing for significance of differences in PCO between the pre- and post-quality assurance periods

	Null Hypothesis	*Kruskal-Wallis*		**Decision**
		$c^2(1)$	p-value	
1	The distribution of the samples are the same across the estimations regimes	11.273	0.001	Reject the null hypothesis
		Median Test		
		Person $c^2(1)$	p-value	
2	The samples are from a population with the same median	8.0211	0.005	Reject the null hypthesis
		Two sample t-test		
		t-value	p-value	
3(a)	The two groups of overruns have equal means and equal variances	3.8306	0.0005	Reject the null hypthesis
3(b)	The two groups of overruns have equal means and unequal variances	3.7341	0.0007	Reject the null hypthesis

The tests in Table 4 show that the null hypotheses are rejected, which implies that the distributions of the PCOs between the pre-and post-quality assurance periods are significantly different and that the MPCOs in the pre- and post-quality assurance periods are also different. These tests clearly confirm that there is enough statistical evidence to conclude the following: (a) quality assurance has reduced the magnitudes of percentage cost overruns of large projects, but (b) there is no evidence that the accuracy per se has improved, as quality assurance has led to significant under-runs.

5.2 Selection bias

Because we only compare large projects between two distinct periods, the pre- and post-quality assurance periods, it may be that the magnitudes of overruns changed not only between the large projects but overall across all projects. In such a situation, it may be difficult to conclude that the implementation of quality assurance for large projects alone was the major cause of the differences in the magnitudes of cost overruns among large projects. Other mechanisms may have been at play, leading to the improvement of estimates across all sizes of projects. If this is the case, our results above would be biased because they only account for the developments in one section of investments. This problem is termed selection bias, which implies that a statistical bias exists in the sense that there is an error in choosing the sample or groups for inclusion in a scientific study.

To check whether this problem is present in our study, we provide a comparison of the magnitudes of overruns of the larger projects discussed above versus smaller projects for the same period of time. If the trends are different to the extent that the post-quality assurance period shows a larger improvement for larger projects compared to the others, there is no reason to suspect selection bias. If the magnitudes of overruns improved for all projects irrespective of project size, we suspect some selection bias. However, it may still be argued that the message of the quality assurance regime disseminated downwards such that those charged with cost estimates improved the quality of their estimates even for smaller projects.

Table 5 reports the results of a comparison of overruns across projects of different sizes divided by pre- and post-QA periods. Note that large projects that we have so far considered are now included as large projects with cost estimates greater than 100 million NOK.

Table 5. Testing for the significance of differences in PCO between pre- and post-quality assurance for all project sizes

	Tw-sample t-test of mean of differences			
Mean percentage error(MPE)	Pre-QA	Post-QA	P-value	Decision
Small (less 50 mill. NOK)	10.89 %	11.40 %	0.989	Do not reject null hypothesis
Medium (50-100 mill. NOK)	6.61 %	7.39 %	0.768	Do not reject the null hypothesis
Large (greater than 100 Mill. NOK)	9.37 %	-3.82 %	0.000	Reject the null hypothesis
No. Of projects				
Small (less 50 mill. NOK)	578	304		
Medium(50-100 mill. NOK)	35	31		
Large (greater than 100 mill. NOK)	47	51		
Max MPE	182.67 %	131.91 %		
Small (less 50 mill. NOK)	56.52 %	42.86 %		
Medium (50-100 mill. NOK)	44.88 %	28.70 %		
Large (greater than 100 mill. NOK)				
Min MPE				
Small (less 50 mill. NOK)	-58.50 %	-40.00 %		
Medium (50-100 mill. NOK)	-27.45 %	-20.77 %		
Large (greater than 100 mill. NOK)	-22.70 %	-38.80 %		
Stdev MPE				
Small (less 50 mill. NOK)	36.52 %	22.30 %		
Medium (50-100 mill. NOK)	19.45 %	11.52 %		
Large (greater than 100 mill. NOK)	13.42 %	10.31 %		

As Table 5 shows, there is no evidence that the mean percentage cost overrun of smaller projects are different between the two periods under consideration, while for larger projects with cost estimates above 100 million NOK, there is a difference with significance. These results therefore prove that selection bias is not present in our study. However, the larger projects not under the QA regime also seem to have experienced significant reduction in the magnitudes of overruns. This result may be an indication that the QA regime led to an overall improvement in the cost estimation of larger projects in general. A question may be raised as to why one would expect the QA spill-over effects only with regards to larger projects and not the smaller ones. The larger projects undergo the same type of cost estimation process as those under the QA regime and are also tendered. The smaller projects do not undergo the same type of cost estimation procedures and are not tendered but are directly acquired. Thus, the implementation of QA was most likely to have a spill-over effect on to larger projects as compared to smaller ones.

5.3 Who is more accurate: the authorities or the quality assurers?

The impact of quality assurance on cost overruns has been shown above to have reduced the PCO of road projects beyond any statistical doubt. Because the quality assurance regime ceteris-paribus encourages authorities to provide improved estimates that are then scrutinised by external quality assurers (external consultants), it is of interest to investigate which of the two estimates provided is the most accurate. Our hypothesis is that the estimates provided by the authorities are better than those provided by the external consultants. This hypothesis builds on the argument that the authorities are much more experienced in cost estimations given the large number of projects that they have conducted and that quality assurance is only an incentive to be more accurate. The external consultants, on the other hand, are not as experienced as the authorities and have diverse experiences on different types of projects and hence, are not as specialised as the authorities are with respect to road projects.

Our data set contains estimates provided by the authorities and external consultants, allowing for comparisons similar to those provided in sub-section (i) above. We performed the Median independent sample test and the two sample mean-comparison t-test; the results are reported in Table 6.

Table 6. Comparing the PCOs of road authorities and external consultant estimates

	Null hypothesis	*Median Test*		**Decision**
		Person c^2 (1)	p-value	
1	The median of the populations are the same	3.272	0.07	Reject the null hypothesis
		Two sample t-test		
		t-value	p-value	
2(a)	The two groups of overruns have equal means and equal variances	3.272	0.07	Reject the null hypothesis
2(b)	The two groups of overruns have equal means and unequal variances	1.81	0.07	Reject the null hypothesis
		Authorities	***Consultants***	
	Mean percentage overrun	-0.11	-0.03	
	Median percentage overrun	-0.11	-0.03	

Table 6 shows that all the null hypotheses were rejected at the 10 percent significance level. Because both the medians and the means are larger for the authorities' PCOs in absolute terms, these results reveal that the external consultants are more accurate with their estimates than the road authorities are. The authorities seem to inflate their estimates in order to avoid overruns.

6. Concluding remarks

Our objective in this paper has been to investigate an issue that is less addressed in the transport literature on cost overruns, namely, whether efforts made by governments to reduce overruns are actually effective. Our case study is Norway, where the government has instituted a QA regime that requires that the cost estimates for large projects made by authorities must be scrutinised by external consultants who then make alternative estimates.

Although Norway is a small country, the framework for quality assurance and its results should be of relevance to other countries too. In fact, the principle of ensuring quality at entry should be even more important in countries with larger and more complex projects as the consequences of failure then would be larger.

Quality assurance of cost estimates in Norwegian road projects is part of a comprehensive governance framework which above all seeks to ensure quality at entry, i.e. to ensure that the grounds for decision-making is of a sufficient quality so that projects not only meet their budgets but also meet their purpose. The project management literature has traditionally been concerned with "doing the project right" and often ignores the front-end of projects from when the project concept is conceived until the final decision to build is made. However, governments and private investors have increasingly recognised the importance of high quality front-end management for good project delivery. The Norwegian QA regime fits well with these recent developments in the project management literature (see Flyvbjerg (2013) and the references therein for a review of the literature). The idea is that many of the planning failures and cost overruns witnessed over the last decades could have been mitigated by taking the "outside view" of projects i.e. using experiences from similar projects elsewhere and where cost estimates are scrutinized by independent consultants, rather than focusing on the specific planned actions in the project. Since the introduction of the Norwegian QA regime other countries such as Denmark and the UK have introduced similar regimes where front-end estimates are subjected to external scrutiny before funds are tied up to projects. Norway provides an interesting case from where a system of quality assurance has been in operation for over a decade. The encouraging results presented in this paper could thus provide an inspiration for other countries to follow.

By comparing the magnitudes of overruns among large road projects in the pre- and post-QA periods, we used statistical tests to infer whether the QA regime impacted the magnitudes of overruns. With statistical significance, we reached the following three conclusions:

1. Quality assurance of cost estimates has led to a reduction in cost overruns.
2. Quality assurance has not led to improved accuracy of the estimates provided by the authorities; however, it has led to systematic overestimation by the authorities.
3. The estimates provided by external consultants are more accurate than those made by the authorities. External consultants also seem to overestimate costs but to a much lesser degree than the estimates provided by the authorities.

Relative to the question that we initially posed, our answer given the results above is an affirmative yes: the efforts made by governments to reduce overruns in the case of Norway's QA regime have helped to reduce the magnitudes of cost overruns. The estimates provided by consultants prove that the estimates provided to the decision makers have become more accurate. Our results lead to two inferences. The first is that cost overruns among smaller projects have been observed to be very large. The inference, then, is that the implementation of a QA regime for smaller projects should be considered. The second is that costs are only a part of cost-benefit analysis. Projects may be viable in the long term even if construction costs escalate. It is therefore important to focus on both costs and benefits. How projects perform in the long term – their impact, their relevance and their sustainability – should therefore be the focus of future research. Finally, the Norwegian experience given its impacts on the magnitudes of cost overruns is an invaluable experience that other countries that constantly experience cost overruns of infrastructure projects can learn from.

Acknowledgements

Thanks are due to the organisers of UNITE conference 2013 at DTU-Denmark. Without their initiative, this paper would not have been realised. Thanks are also due to the Concept Research Programme at the Norwegian University of Science and Technology, which financed the research reported in this paper.

References

Berg, P., Andersen, K., Østby, L.E., Lilleby, S., Styrvold, S., Holand, K., Korsnes, U., Rønning, K. and Johansen, F. (1999). *Styring av statlige investeringer. Sluttrapport fra styringsgruppen*. Oslo: Finansdepartementet.

Berechman, J. and Wu, Q. (2006). *Cost overruns Risk Analysis in Transport Infrastructure Investments*. Working Paper 2006-05. Phelps Centre for the Study of Government and Business. University of British Columbia.

Cantarelli, C.C., Flyvbjerg, B., Molin, E.J.E. and van Wee, B. (2010). Cost Overruns in Large-scale Transport Infrastructure Projects: Explanations and Their Theoretical Embeddedness. *European Journal of Transport Infrastructure Research*, 10(1), 5-18.

Concept Research Programme (2015). Ministry of Finance scheme for quality assurance of major public investments (the QA scheme). [online] Available at: http://www.ntnu.no/concept [Accessed: February 5, 2015].

Flyvbjerg, B. (2005). *Policy and Planning for Large Infrastructure Projects: Problems, Causes, Cures*. World Bank Policy Research Working Paper 3781, December 2005.

Flyvbjerg, B. (2007). Cost Overrun and Demand Shortfalls in Urban Rail and Other Infrastructure. *Transport Planning and Technology*, 30(1), 9-30.

Flyvbjerg, B. (2013). Quality control and due diligence in project management: Getting decisions right by taking the outside view. *International Journal of Project Management*, 31(5), 760-774.

Flyvbjerg, B., Skamris, M.K. and Buhl, S.L. (2003). How common and how large are cost overruns in transport infrastructure projects? *Transport Reviews*, 23(1), 71-88.

Mackie, P. and Preston, J. (1998). *Twenty-one sources of error and bias in transport project appraisal*. Transport Policy, 5(1), 1-7.

Magnusson, O.M. and Olsson, N.O.E (2006). Comparative analysis of cost estimates of major public investment projects. *International Journal of Project Management*, 24(4), 281-288.

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

Makridakis, S.G., Wheelwright, S.C. and Hyndman, R.J. (1998). *Forecasting: Methods and Applications*. New York: John Wiley & Sons, Inc.

Morris, P. W. G. and Hough, G. H. (1987). *The Anatomy of Major Projects*. Chichester: Wiley and Sons.

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

Odeck, J. (2004). Cost overrun in road construction – what are their sizes and determinants? *Transport Policy*, 11(1), 43-53.

Odeck, J. (2014). Do reforms reduce the magnitudes of cost overruns in road projects? Statistical evidence from Norway. *Transportation Research Part A: Policy and Practice*, 65, 68-79.

Pickrell, D.H. (1992). A Desire Named Streetcar. Fantasy and Fact in Rail Transit Planning. *Journal of the American Planning Association*, 58(2), 158-176.

Riksrevisionsverket (1994). *Infrastrukturinvesteringar- en kostnadsjämförelse mellan plan och utfall i 15 större projekt inom Vägverket och Banverket*. Swedish National Audit Bureau. RRV 1994: 23

Riksrevisionen (2011). *Kostnadskontroll i stora järnvägsinvesteringar?* Swedish National Audit Bureau Audit. Report RiR 2011:6.

Riksrevisjonen (2002). *Riksrevisjonens undersøkelse av planlegging og oppfølging av store veganlegg i Statens vegvesen*. Oslo: Riksrevisjonen, Dokument nr. 3:3 (2002-2003).

Samset, K. and Volden, G. (2012). The proposal. In Williams, T.M. and Samset, K. (eds.) *Project Governance. Getting Investments Right*. Basingstoke: Palgrave Macmillan. Ch. 2.

Samset, K. and Volden, G.H. (2013). *Investing for Impact. Lessons with the Norwegian State Project Model and the first investment projects that have been subjected to external quality assurance*. Concept report no. 36. Trondheim: Ex Ante Academic Publisher.

Sager, T. and Ravlum, I-A. (2005). The political relevance of planner's analysis: the case of a parliamentary standing committee. *Planning Theory*, 4(1), 33-65.

Siemiatycki, M. (2009). Academics and Auditors. Comparing Perspectives on Transport Project Cost Overruns. *Journal of Planning Education and Research*, 29(2), 142–156.

UK National Audit Office (2007). *Estimating and monitoring the costs of building roads in England*. London: The Stationary Office.

Wachs, M. (1987). Forecasts in urban transport planning: Uses, methods, and dilemmas. *Climatic Change*, 11(1-2), 61-80.

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

Van Wee, B. (2007). Large infrastructure projects: a review of the quality of demand forecasts and cost estimations. *Environment and Planning B: Planning and Design*, 34(4), 611-625.