Transportation Research Part A 120 (2019) 277–294

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

The accuracy of benefit-cost analyses (BCAs) in transportation: An ex-post evaluation of road projects

James Odeck[a,b,c,*], Anne Kjerkreit[c]

[a] *Molde University College, P.O. Box 6405, Molde, Norway*
[b] *The Norwegian University of Science and Technology, 7491 Trondheim, Norway*
[c] *Norwegian Public Roads Administration, P.O. Box 8142 Dep, 0033 Oslo, Norway*

ARTICLE INFO

Keywords:
Ex-post evaluation
Benefit-cost analyses (BCA)
Road projects
Norway

ABSTRACT

Ex-post evaluations of benefit-cost analyses (BCAs) of transportation projects are scarce in the literature. If conducted frequently, they could reveal the extent to which objectives are achieved and may give inputs that can improve the quality of ex-ante BCAs. We first explain the usefulness of ex-post BCA evaluations of transportation projects, depending on which planning phase a project is in. We then perform ex-post BCAs on 27 Norwegian road projects that have been in service for at least 5 years and compare the results with the ex-ante BCAs that were presented to the decision-makers. We use two different measures of aggregating the magnitudes of accuracy; the Mean Percentage Error (MPE) and Mean Absolute Percentage (MAPE). Using MPE, we find that: (1) ex-ante BCAs underestimate the actual net present value (NPV) by 50% on average and, the NPV per dollar invested by a mere 0.14% points on average; (2) the traffic level and traffic growth rates are often underestimated ex-ante, leading to an underestimation of the benefits and; (3) construction costs are underestimated by 5% on average, which is too small to offset the observed underestimation of benefits. By default, MAPE shows higher values of inaccuracies implying that one should not be indifferent as to the measure used. Overall, the Norwegian ex-ante BCAs perform fairly well. We urge that the authorities concerned must improve their traffic forecasts and construction cost estimates to assure that the ex-ante objectives are met. The upkeep of ex-ante data is essential to enable ex-post evaluation which, ceteris paribus, will enhance the transparency and credibility of BCA as an appropriate decision-making tool.

1. Introduction

Decision-makers in the transportation sector rely on the expected benefits and costs that a given project should generate throughout its lifetime when making their decisions. To help aggregate such benefits and costs into a single measure of project worthiness, transportation economists/planners regularly conduct ex-ante Benefit-Cost Analyses (BCAs). Ex-ante means that the analyses are an integrated part of the planning process and that the analyses are based on forecasts, which may or may not match the real outcomes. An ex-ante BCA proceeds by first evaluating the expected change in the benefits and costs of an undertaking compared to a “do-nothing or do-minimum” situation, and all the benefits and costs are measured in monetary terms. For road projects in particular, the monetary benefits and costs include travel-time savings, reductions in accident costs, vehicle operating costs, environmental impacts such as increased/reduced noise from vehicles, and the investment and maintenance/operational costs of the

* Corresponding author at: The Norwegian University of Science and Technology, 7491 Trondheim, Norway.
E-mail address: james.odeck@vegvesen.no (J. Odeck).

https://doi.org/10.1016/j.tra.2018.12.023
Received 23 August 2017; Received in revised form 20 December 2018; Accepted 20 December 2018
Available online 09 January 2019
0965-8564/

road network. An ex-ante BCA further proceeds by comparing the discounted monetized benefits to the discounted costs. The result of such a comparison is the Net Present Value (NPV). If the NPV is positive, then the project is considered to be profitable from a socioeconomic perspective because its benefits exceed its costs; otherwise, the project is deemed unprofitable. A further advantage of using a BCA as a tool for judging the socioeconomic merit of projects concerns the selection of the most profitable projects from a pool of projects when government investment funds are limited. In such a case, the BCA rule states that projects should be ranked according to the value of their NPV–Cost ratio until the available funds are used completely. NPV–Cost ratios are calculated as the NPV divided by the financial costs of the project provided through government funds/budgets. If the ratio is 0.20, for example, the interpretation is that the return from government/societal investment in the project is 20%.

Although there has been criticism of using BCAs as appropriate tools for decision-making because a BCA does not include all factors worth considering in decision-making and some important impacts are not valued in monetary terms, there is one additional criticism that the transportation literature has addressed to a lesser extent, and that is the focus of this paper. This criticism is that BCAs are rarely conducted ex-post, i.e., at some point after a project has been implemented. There are several reasons that legitimize this criticism. Ex-post assessments, if conducted regularly, can have the following benefits: (1) provide input for the development of BCA techniques over time, (2) provide important lessons to BCA designers looking for practical evidence regarding delivering transportation projects, (3) show results to policymakers who want to know whether schemes deliver planned benefits and whether road policy is as effective as intended, and (4) show communities whether their concerns are being addressed effectively.

In this paper, we investigate the accuracy of ex-ante BCAs in transportation by using Norwegian road projects as a case study. We first define the different classes of BCAs according to the phases of planning to which they are most applicable because an ex-post BCA may not be equally useful in all planning phases/stages. We then argue that ex-post BCAs of road or transportation projects, where useful, should be conducted regularly to investigate whether an ex-ante BCA has delivered what it promises and that an ex-post BCA is an appropriate way to gauge areas for potential improvements to an ex-ante BCA. We then conduct ex-post BCAs on 27 Norwegian road projects that have been implemented and then compare the results with the ex-ante BCAs available to decision-makers at the time of decision-making to gauge the accuracy of the ex-ante analyses. It should be noted, however, that our form of ex-post evaluation is not a comparison between the outcome of a BCA and a counterfactual situation that describes what would have occurred in the absence of the projects, as counterfactual situations are almost impossible to replicate once a project is already built; see, for instance, Bråthen (2001). Instead, our form of ex-post evaluation in this paper is to test the extent to which what was promised is being delivered; we essentially measure the inaccuracy of the ex-ante BCA results.

The contribution of this paper to the transportation literature is clear; it is a valuable addition to a topic that has scarcely been addressed in the transportation literature and may help improve the legitimacy of ex-ante BCA analyses as a useful tool for decision-making. Furthermore, it schematically emphasizes when ex-post BCA analyses are most useful, as emphasized by Boardman et al. (1994), but does so in the context of road planning stages.

The paper proceeds as follows. Section 2 is a literature review of past studies that have examined the ex-post evaluation of BCAs of transportation projects. Section 3 discusses the need for ex-post BCA studies. Section 4 describes the methodology used. Section 5 describes the data and calculation strategy used. Section 6 presents the results of the ex-post BCA evaluation of the 27 Norwegian projects studied. Concluding remarks and discussions are included in Section 7.

2. Literature review

Ex-post evaluations of BCAs in the transportation literature are scarce, at least in scientific journals such as Transportation Research Parts A, B, C and D; Transportation Policy; Transportation Reviews; Transportation Economics and Policy; and Transportation. To our knowledge, Boardman et al. (1994, 2011) made the earliest and most thorough attempt to address the importance of ex-post BCA evaluations of transportation projects. They concluded that the comparisons of ex-ante and ex-post BCAs of transportation projects were lacking in the literature. They then proposed that ex-ante and ex-post evaluations be compared to other evaluations of the same project; without such studies, it would be impossible to evaluate the practical value of BCA as a decision-making tool. This proposition is in line with the purpose of the present paper. Boardman et al. (1994) applied their proposition to the Coquihalla Highway. They found that contrary to what might have been expected, the largest source of differences between ex-ante and ex-post BCA evaluations was not errors in forecasts or differences in the evaluation of intangible benefits but rather the major differences between the declared and actual construction costs of the project. That is, the largest errors arose from what most analysts would have thought were the most reliable figures entered into the BCA. They concluded that comparison studies are potentially the most useful studies for learning about the accuracy and efficacy of cost-benefit analysis for decision-makers and evaluators. An oft-cited paper in the literature is that of Anguera (2005), who conducted an ex-post economic evaluation of the Channel Tunnel. He concluded that overall, the British economy would have been better off had the Tunnel never been constructed, as the total resource cost outweighed the benefits generated. Ceteris paribus, this means that the ex-post results exhibited negative NPV values, whereas the ex-ante results exhibited positive NPV values. The single biggest component of users' gain was not, as originally expected, travel-time savings but rather the transfer from producers; producers were the greatest losers. The longer-term evaluation of the project confirmed the poor viability of the investment in both financial and cost-benefit terms.

The recent decade has seen renewed interest in the ex-post evaluation of transportation projects. The governments of Norway, UK, France, Australia and New Zealand all have frameworks for the ex-post BCA evaluation of road projects. For instance, in the UK, Highways England produces Post Opening Project Evaluation (POPE) reports '1 year after' and '5 years after' the opening of a road scheme; see POPE (2016). The NZ Transport Agency conducts post-implementation reviews every year on a small sample of completed projects or packages in which it has invested; see NZ Transport Agency (2016). In the case of France, Taroux et al. (2005) and

Meunier (2010), in their conference papers, explained the French framework and then examined the results of reports that used the framework. Their major findings were that ex-post BCAs are difficult because of the problem of replicating the reference situation, particularly regarding traffic, and that there are divergences between the forecasted benefits and costs and their actual outcomes. The conference papers of Kjerkreit et al. (2008) and Kjerkreit and Odeck (2015) reported ex-post studies of Norwegian road projects and concluded that the BCA outcomes are generally higher than forecasted.

The above mentioned frameworks for ex-post evaluations across the globe have, however, not been extensively published in scientific journals such as Transportation, Transport Policy, Transportation Research Part A, Socio-economic Planning Sciences and Transport Geography. Thus, large groups of readers cannot access the results of how ex-ante BCAs perform in reality; many ex-post evaluations of BCAs and the reasons behind their inaccuracies/accuracies have not reached the wider audience of transportation journals.

Thus far, to our knowledge, only two recent studies have involved ex-post BCA evaluation and have reached the wider audience of transportation journals. Kelly et al. (2015) reported ex-post BCA studies of projects across European countries. They studied the project-level outcomes with respect to the BCAs of 10 large transport projects spread over eight countries. The projects considered had benefited from EU Cohesion and Instrument for Structural Policies for Pre-Accession (ISPA) funding. They compared the ex-ante and ex-post cost-benefit analyses and found that although much attention in the literature has been paid to the issue of optimism bias over the last decade, optimism bias remained prevalent. The ex-ante BCAs yielded significantly higher NPV results compared to the ex-post results. They also found a clear need to improve the quality and consistency of ex-ante analyses, particularly in the areas of capital cost estimation, travel demand modeling and risk analyses. Börjesson et al. (2014) conducted an ex-post BCA for the Stockholm Metro built in the 1950s. In their study, they did not compare their results with those presented to the decision-makers in the 1950s because BCAs were not conducted at that time. Instead, they inserted values for the 1950s into the current BCA framework to derive relevant BCA results for that time. They found that the Stockholm Metro was socially beneficial and that its greatest benefit is its capacity, which makes it possible for many people to travel to and from the city center. They also examined the wider economic impacts due to labor market distortions and the land-use effects of the metro. The results showed that the wider economic impacts increased consumer surplus by 48% and that the yearly income increased by 1.5%.

For the sake of completeness, the wider body of literature regarding the ex-post evaluation of elements of BCA must be referenced here. This part of the literature focuses primarily on the overruns of construction costs and/or traffic forecasts of road transportation projects, including tolled roads, but does not account for their overall effects on the accuracy of BCA results, as we do here. Such studies include but are not limited to those by Nicolaisen and Driscoll (2014), Flyvbjerg et al. (2004), Odeck (2004), Baeza and Vassallo (2012), Love et al. (2012), Gomez et al. (2015), Odeck et al. (2015), Welde and Odeck (2011) and, most recently, Welde and Odeck (2017). Among these studies, the Norwegian studies have shown that the accuracy of both traffic and cost forecasts presented to the decision-makers at the time of decision-making are reasonably accurate and are within ± 5 to ± 10% of the original estimates; see, for instance, Welde and Odeck (2011), Odeck et al (2015) and, most recently, Odeck and Welde (2017). This is in contrast to the international literature, in which both traffic forecasts (including tolled roads) and cost estimates exhibit a higher level of inaccuracy; traffic forecasts are overestimated, and costs are underestimated, with both being inaccurate by more than 20% on average; see, for instance, Nicolaisen and Driscoll (2014), Gomez et al. (2015), and Flyvbjerg et al. (2004).

3. The need for ex-post BCAs

As BCAs are used for decision-making and may differ by the planning phase of a project, the best way of describing/demonstrating the usefulness of ex-post BCAs is to develop a matrix that compares the usefulness of the different types/classes of ex-post BCAs according to each planning and decision-making stage. In such an endeavor, the usefulness of the ex-post BCA, according to when in

Table 1
Usefulness of ex-post BCA by planning and decision-making stages.

		Types of BCA			
		Ex ante BCA		*Ex post BCA*	
No.	Type of decision	(1) Early project phase *Ex-ante BCA*	(2) Detailed project phase *Ex-ante BCA*	(3) Early *Ex-post* (*In medias res*) BCA	(4) Full *Ex-post BCA*
1	Feasibility of a project; decision to go ahead planning for different options for the same project.	Very useful to demonstrate the potential of a project in the early project phase	Of little use as more information has accrued and improved BCA is required	Of no use; the project is already built using even more detailed information	Not useful -the project is over
2	Selection of alignments/alternatives to compete for funds with others of a different project	Of little help; contains too little or no information of potential alignments	Very useful to select the best alignment	If low sunk costs, may recommend a change of alignment. If high sunk costs, not very useful as the recommendation is proceed with the alignment	Not useful; it is too late and the alignment is already built
3	Resource allocation between competing projects	Of little help; contains too little detailed information	Very useful for allocation of resources to the most socioeconomically profitable projects	Not useful as the resources are already allocated -recommendation is proceed with the project	Not useful; it is too late and the project is already built
4	Learning about the actual outturn ex ante BCA estimates used at the final decision making	Of no help - contains poor estimates hence high uncertainty	Of little help -high uncertainty about future benefits and costs. Can be used as baseline for comparison	Of much help - uncertainty is reduced	Excellent -but some difficulties in tracing changes that occurred may be a problem.
5	Contributing to the improvement of *Ex ante* BCA framework	Not helpful; takes BCA framework as given	Not helpful; takes BCA framework as given	Of good help but the contribution increases if performed later	Very useful but some errors may still go undetected
6	Learning about accuracy of unit prices of impacts that enter BCA	Not helpful; does not consider the accuracy of unit prices	Not helpful; does not consider the accuracy of unit prices	Not helpful; does not consider the accuracy of unit prices	Not helpful; does not consider the accuracy of unit prices

the planning/decision-making process it is most relevant, will emerge. Such a matrix was first developed by Boardman et al. (1994). Table 1 presents such a revised matrix within the context of the most common road planning stages in the case of Norway and many other countries in the Western world.

We first consider the different types/classes of BCAs listed in the second row of the table. The BCAs are divided into the following two basic classes, as defined previously: ex-ante and ex-post BCAs. Next, we consider the third row. Ex-ante and ex-post BCAs can be further divided into two basic classes depending on when they are conducted. For the ex-ante case, the two classes are the early project phase, when the decision to go ahead with planning is made, and the detailed project phase, when the decision to select the appropriate alignment to follow and/or the decision to build is made. For the ex-post case, the two classes involve an evaluation at some point after the project has been implemented, also known as in medias res evaluation, and after the project has been implemented and terminated, which is the full ex-post evaluation.

Next, we consider the decision-making stages listed in the second column of the table. The different stages/phases of road planning involve the following: (1) a feasibility study of the project where the decision to continue planning for a future final decision is made, (2) the selection of an appropriate project plan (alignment/alternative) to compete for funds with other projects, (3) resource allocation between competing projects from a pool of projects where resources are limited, (4) learning about the actual outcomes of the ex-ante BCA estimates used in the decision-making stage, and (5) learning about the accuracy of the unit prices of impacts.

Next, we consider the usefulness of the different classes of BCAs according to the different planning/decision-making stages, as indicated in Table 1. To aid readers in understanding where and when ex-post BCAs are useful, the grey shaded cells in Table 1 indicate where and when ex-post analyses are most useful.

It is clear that the early project phase ex-ante BCA is only useful for demonstrating the potential that a project has at that early stage. This is because very little is known and because there is no detailed information at the feasibility study stage. The decision where it is helpful is therefore the go-ahead decision to continue planning for different options for the same project. Then, we consider the detailed project ex-ante BCA. This class of BCA is most useful for (1) selecting the most appropriate alignment/option of the same project and (2) selecting the appropriate projects from a pool of projects for funding/resource allocation when funds/resources are limited. This class of BCA is the most common in the transportation literature, in which the major issue seems to be the allocation of funds. As the table shows, this class of BCA is not useful for the early decision to proceed and plan or for evaluating performance. The next two BCA classes to consider are the early ex-post BCA (in medias res) at some point after the project has been implemented and is still in operation and the full ex-post BCA after the project has been implemented and completed, i.e., when the project is no longer in operation. The table indicates that both of these classes of BCAs are most useful for learning about the actual outcome of the ex-ante BCAs that were used in the final decision to select the appropriate alignment and/or allocate resources to the most appropriate projects. It should be noted here that although the full ex-post BCA is the most useful for this purpose, the early ex-post BCA is probably the most practical to conduct. This is because several conditions in the transportation system may have changed by the time that the project lifetime has expired, thus making it difficult to disentangle the real effects of a project from other changes that have occurred during its long lifetime. Furthermore, many large projects may last for a hundred years, and it would be unreasonable to wait and conduct an ex-post evaluation after such a long time. From the discussion above, several reasons for conducting ex-post BCAs from time to time are evident: (i) they provide valuable information about how a project is meeting its prestated objectives, (ii) they are useful for efficiency monitoring purposes, and (iii) they provide a measure of ex-ante BCA precision and are therefore a measure of how much confidence the decision-makers should put in BCA as a decision-making tool, especially given that ex-ante BCA has been heavily criticized for not being accurate. Thus, in a manner similar to an efficacy measurement, ex-post BCA may help improve the credibility of ex-ante BCA as an appropriate decision-making tool if the latter is proven to meet expected outcomes. Finally, an ex-post BCA provides information for potential areas in which an ex-ante BCA can be improved.

4. Methodology

There are several ways to investigate the magnitudes of ex-ante BCA inaccuracies. Using jargon from the literature about the forecasting/measurement of inaccuracy, ex-ante BCAs are the forecasts, and ex-post BCAs are the actual outcomes. Thus, a measurement of the difference (deviation) between them on the same project is a measurement of inaccuracy. The most commonly used approach for measuring inaccuracy of forecasts is the percentage error (PE); see, for example, Makridakis et al. (1998)); for applications in the field of ex-post evaluation in transportation, see, for instance, Flyvbjerg et al. (2004), Odeck (2014), Nicolaisen and Driscoll (2014) and Gomez et al. (2015). Using NPV as the measure of the BCA results, the PE for an individual project is expressed as follows:

$$PE_{NPV} = \frac{(NPV_{ex-post} - NPV_{ex-ante})}{NPV_{ex-ante}} \times 100 \tag{1}$$

where the subscripts indicate forecasts (ex-ante) and actual outcomes (ex-post).

It becomes clear that if $PE_{NPV} > 0$, then the ex-ante BCA was an underestimate, whereas if $PE_{NPV} < 0$, then the ex-ante BCA was an overestimate. If $PE_{NPV} = 0$, the ex-ante BCA exactly matches the ex-post BCA outcomes, and the ex-ante BCA was accurate.

To exploit formula (1) above, it is possible to investigate the inaccuracy of important factors that enter the NPV, such as time savings, accident costs and investment costs, to infer their inaccuracies; some factors that enter the NPV may have been estimated more accurately than others. Such an investigation may help BCA practitioners identify factors that need more attention to ensure the accuracy of their forecasts; e.g., in the case of construction costs, formula (1) above becomes

$$PE_{Cost} = \frac{(Cost_{ex-post} - Cost_{ex-ante})}{Cost_{ex-ante}} \times 100 \quad (2)$$

Note that when investigating the PE of factors that are reported in percentages, the appropriate measure is the Percentage Point Error (PPE). In our case, this will involve the traffic growth rate forecasts and the NPV-Cost ratio[1], both of which, in the BCA analyses, are often reported in percentages. In contrast to the PE, the PPE measures the arithmetic error of two percentages. For instance, if the traffic forecast was estimated to be 1.5% and it turned out to be 4.2%, then the PPE (4.2–1.5%) is 2.7%; i.e., it indicates the difference (deviation) between the ex-post and the ex-ante percentages. However, should one still desire a PE measure for it, then it means that the PE increase is (4.2–1.5%)/1.5%) = 180%. PPE is much easier to interpret, as long as results are presented in percentages; the estimated forecast was 1.5% but it turned out to be 4.2%, for an increase of 2.6% in relation to the original estimate. Finally, when summarizing the means across all projects, Mean PE (MPE) and Mean PPE (MPPE) are commonly used and are written as follows; for instance, see Makridakis et al. (1998):

$$MPE = 1/n \sum_{i=1}^{n} PE_i \;\; and \;\; MPPE = 1/n \sum_{i=1}^{n} PPE_i \quad (3)$$

Further, when percentage errors (PEs/PPEs) change from negative to positive or vice versa, the MPE, minima and maxima across observations become difficult to interpret. To illustrate, consider two observations with the following ex-ante and ex-post NPV-Cost-ratios, respectively: Project 1: −0.90 and 0.9 and Project 2: −0.2 and −0.4. Using Eq. (2) to calculate the PE, Project 1 obtains −200%, and Project 2 obtains 100%. If now one was to judge the minimum change across these two projects, Project 1 would be considered the one with the minimum change, but it has the greatest change in PE. Consequently, calculating the mean changes according to Eq. (3) is not informative enough. A remedy for this problem is, of course, to use the absolute value and/or the square of PE and PPE when calculating the mean changes across observations. Essentially and has been suggested by e.g., Makridakis et al.,1998, this means using Mean Absolute Percentage Error (MAPE) and/or Mean Percentage Squared Error (MPSE) to summarize changes across units of observation as follows:

$$MAPE = 1/n \sum_{i=1}^{n} |PE_i| \;\; and \;\; MPPE = 1/n \sum_{i=1}^{n} |PPE_i| \quad (4)$$

$$MPSE = 1/n \sum_{i=1}^{n} (PE_i)^2 \;\; and \;\; MSPPE = 1/n \sum_{i=1}^{n} (PPE_i)^2 \quad (5)$$

The advantages/disadvantages between these alternative measures to infer mean change in percentage errors should be explained. MPE in Eq. (3) measure the mean percentage of error, MAPE in Eq. (4) measure the mean absolute value of percentage of error, and MSPE measures the mean percentage of errors squared. These measures are therefore useful to express mean changes in different ways depending on what one intends to reveal. MPE tends to be small because the negative and positive values will offset one another. It is only useful as an indication of whether there is systematic under- or overestimation in forecasting. In our context, it is useful in gauging the biases of BCA forecast made ex-ante. On the other hand, because MAPE is based on absolute values, it indicates the absolute magnitude of all the errors regardless of direction of bias. It thus gives an indication of the magnitude of the overall biases of forecasts, irrespective of direction. MAPE assumes that all biases, irrespective of direction, are equally important to consider. Finally and because of squaring the PE's (and PPE's), MPSE gives more weight to larger values than smaller values, but in other respects it is similar to MAPE. Thus, MPSE reflect the fact that large errors represent a much more serious problem than small errors; meaning that it is appropriate to focus primarily on larger errors to effectively reduce total errors. In the results section we report all the three measures.

Given the differences between MPE, MAPE and MPSE discussed above, a question that still remains is which one among them is most useful to calculate/address from a policy perspective.[2] Since our objective is to infer the magnitudes of inaccuracies across and within individual BCA estimates, MPSE is of less relevance as we are not specifically interested in magnifying the differences in over-/underruns between large and small projects. MPE and MAPE are the most relevant to compare. MPE is most useful to consider if the policy is to infer the performance of a pool projects as a group (or a portfolio of projects). MAPE on the other hand is most useful to consider if the objective is to determine the variation in performances across individual projects where overruns and underruns are assumed to be equally bad; it is an absolute measure.

In what follows, we derive the PE for NPV, investment costs, traffic volumes and accident rates and, PPE for NPV-cost ratio and traffic forecast, which are normally presented in percentages in BCA analyses. These are the factors that are considered the major determinants of BCA sizes. When summarizing across projects however, MPE, MAPE and MPSE indices are provided whereas, MPE and MAPE are the major measures that are addressed in conclusions. The project data used and how the estimations were performed follow in the next section.

5. Data and calculation strategy

The data used are from 27 large road projects in Norway and were retrieved from the Norwegian Public Roads Administration's (NPRA) databases and official documents used at the time of the budgetary decision-making in addition to observed data from after

[1] Note that NPV-Cost ratio may not be the best project ranking criterion. See for instance Minken (2016), De Rus (2010) and Boardman et al. (2011).

[2] Thanks to an observant reviewer who proposed addressing this issue.

the projects were built, such as project completion reports and traffic counts. In particular, this involved obtaining the data input that was used in the ex-ante BCA calculations and, preferably, the software used for the calculations. Equally important was the retrieval of the transport model data used in the ex-post calculations. In this process, some lessons about the availability of ex-ante BCA data that should be available for ex-post evaluation were learned: (1) there is no systematic upkeep of ex-ante data, and their retrieval is a struggle, in which an enormous effort must be made to trace and replicate the data from the archives; and (2) even in cases in which BCA data could be retrieved, the planning stage for which they were intended was not evident. Therefore, some assumptions had to be made. For instance, if BCA figures were not explicitly shown in the decision documents, we assumed that the last BCAs performed by the NPRA before decisions were made, were the ones most likely used. The same applied to the traffic estimates and the version of the standard software used.

Thus, our first conclusion regarding the availability of data for conducting ex-post BCA evaluations of road projects is that road authorities need a system for data upkeep to allow for ex-post evaluations, which will, in turn, enhance the transparency and credibility of the ex-ante BCAs. Kelly et al. (2015) made a similar observation regarding the European projects that they examined. Another important observation made while preparing the data for analysis was that the NPRA charged with maintaining the frameworks for conducting BCA routinely updates its frameworks. For instance, for the 27 projects examined, three different versions of the software for performing BCA were used, depending on the time of study. The differences in these versions were mainly the unit values of factors used, such as the value of time, accidents, vehicle operating costs that change over time, and the interest rate. Thus, another conclusion is that the NPRA seems to be up-to-date with its frameworks, and there is no apparent potential for software improvement. Note here that we are referring to the software and not the methodology, which may need improvements, such as monetizing more impacts and/or updating the existing monetary values, which is not the focus of this paper.

The basic requirement for a project's inclusion in the study was that the project be large and situated outside the complex infrastructure systems that are commonly found in cities. This requirement was necessary because complex systems are prone to other changes in the system, making it difficult to trace or disentangle the impacts caused by the project. The size of projects chosen was according to the NPRA's definition of large projects: the total investment costs should not be less than approximately 200 million NOK.[3] A second requirement was that the project must have been in operation for at least 5 years and still be in operation. Consequently and in relation to the sketch in Table 1, the ex-post BCA type that we conducted in this study is that of the early phase after projects have been built (in medias res) and the ex-ante CBAs were at detailed projects phase, i.e., just before the decision were made.

Furthermore, for comparability purposes, the ex-post calculations were performed for only the first 5 years after project opening for traffic because many projects had been in operation for only 5 years. It is important to note that the limitation to 27 projects was because only a maximum of five projects have been ex-post evaluated per year, starting in the year 2006, with the Norwegian Ministry of Transport and Communication's (MTC) authorization; in many of the years, there were only 2–3 projects that could be subjected to ex-post evaluation.

An important step after collecting the data and relevant material was obtaining project overviews and histories for each individual project. This step included acquiring all the necessary details of the project, such as objectives, design and changes over time (e.g., changes in design, environmental requirements added after construction had started). This type of information adds value in explaining the possible reasons for eventual inaccuracies.

Table 2 presents an overview of the projects studied, including project type and their main objectives.

We first consider the geographical distribution of the projects studied, listed in the third column of Table 2. A majority of projects are in the Eastern region (11 projects) and in the western region (seven projects) of the country. This is not surprising because these are the regions with the highest population densities and thus have a high demand for road infrastructure. The western region also has a relatively poor road infrastructure, in which there are still many ferry crossings. All regions are, however, represented in the dataset. Next, we consider the third column, which specifies the projects' objectives. It is clear that Norwegian road projects are mostly aimed at reducing travel time/travel costs for road users and reducing accident costs/improving traffic safety. A third prevalent objective is to improve the local environment, such as by reducing noise, local air pollution and barriers. Table 2 reveals that this objective is also relatively common across projects.

The calculation strategy for comparing the ex-ante and the ex-post BCAs across the 27 projects using the data described above is as follows:

(1) The same software package was used to re-estimate both the ex-ante and ex-post BCAs using the data known in the two periods. The underlying argument is that we wanted to know the extent to which the ex-post evaluation delivered to decision-makers was accurate, given what was known at the time.
(2) The results derived in (1) for the ex-ante and ex-post results were then compared to infer the magnitudes of inaccuracies.
(3) Individual projects' aspects were examined to infer causes of inaccuracies that may or may not be accredited to the BCA methodology.

6. Results

The accuracy of ex-ante BCA estimates will depend on the accuracy of estimates of several important components. In this results section, in addition to presenting the overall results, i.e., the accuracy of the BCA in terms of NPV-Cost ratio, we examine the accuracy

[3] 1 US dollar is equal to 9 NOK (2016).

Table 2
Overview of the projects studied.

Project no.	Project	Region	Project type	Project objective*
1	Rv.23 Oslofjordforbindelsen	East	Strait crossing; sub sea tunnel	[1,2,3]
2	E18 Rannekleiv - Temse	Central	Trunk Road; bypass	[1,2,3]
3	Rv.714 Hitra - Frøya	Central	Strait crossing; sub sea tunnel	[1,2,4]
4	E18 Teigeland - Håland	West	Trunk Road, mainly tunnel	[1,5]
5	Rv.62 Øksendalstunnelen	West	Trunk Road, mainly tunnel	[1,5]
6	E8 Nordkjosbotn - Laksvatnbukt	North	Trunk Road	[1,2,3]
7	E18 Gutu - Helland - Kopstad	South	Trunk road; bypass Sande and Holmestrand	[1,2,3]
8	E39 Kleivedammen - Andenes	West	Trunk Road	[1,2]
9	E134 Hegstad - Damåsen	South	Trunk Road; bypass Darbu	[1,2,3]
10	Rv.616 Kolset - Klubben	West	Strait crossing; sub sea tunnel	[1]
11	Rv.580 Hop- Midttun	West	Trunk Road within city; bypass Nestun	[1,3]
12	E6: Akershus grense - Patterød	East	Trunk road	[1,2]
13	E18 Ørje - Eidsberg	East	Trunk Road	[1,2]
14	Rv35 Grualia - Kneppe	East	Trunk Road, “new link”	[6]
15	E6 Halmstad - Patterød	East	Trunk Road	[1,2]
16	E18 Brokelandsheia - Vinterkjær	South	Trunk Road; Bypass Søndeled	[1,2,3]
17	E39 Svegatjørn - Moberg	West	Trunk Road; Bypass Osøyro	[1,2,3]
18	E18 Sekkelsten- Krosby	East	Trunk Road; Bypass Askim	[1,2,3]
19	E6 Ny Svinesundforbindelse	East	Trunk Road	[1,2]
20	E6 Skjerdingstad - Jaktøyen	Central	Trunk Road; Bypass Melhus?	[1,2,3]
21	E16 Kløfta- Nybakk	East	Trunk Road	[1,2,3]
22	Rv.4 Reinsvoll - Hunndalen	East	Trunk Road; Bypass Raufoss	[1,2,3]
23	Fv.43 Aunevik - Bukkesteinen	South	Trunk Road	[1,2,4]
24	E6 Svingenskogen - Åsgård	East	Trunk Road	[1,2]
25	E39 Gammelsæter - Nipetjørn	West	Trunk Road	[1,5]
26	E18 Langåker - Bommestad	East	Trunk Road	[1,2,3]
27	E6 Vist -Jevik -Selli	Central	Trunk Road; Expansion through Steinkjer	[2,3]

* 1 = Reduce road user costs; 2 = Reduce accident costs/improve traffic safety; 3 = Improve local environment; 4 = Improve regional development; 5 = Reduce risk of avalanche/landslide; 6 = Improve accessibility to airports/ports and other modes of transport.

of construction costs, traffic forecasts and accident costs. Thus, the following results are presented: (i) overall inaccuracy of BCA, (ii) inaccuracy of investment cost estimates, (iii) inaccuracy of traffic forecasts, (iv) inaccuracy of accident costs (iv) individual project characteristics that explain the observed deviations, (v) whether the ex-post results would have changed the project rankings had they been known at the time of decision-making and, (vii) the developments in the accuracy of BCA estimates over time.

(i) Overall inaccuracy of BCA

Table 3 presents the overall inaccuracy measurement of ex-post BCA, i.e., the total net discounted benefits, total net discounted costs, net present values (NPV) and NPV-Cost ratios for individual projects. The inaccuracy of the NPV-Cost ratio, as measured by PPE, which summarizes the overall inaccuracy of BCA results for individual projects, is provided in the last column of the table.

Consider first the summary statistics provided in the first part of the bottom section of Table 3. These are the MPE measures according to Eq. (3). The ex-post results are always higher than the ex-ante results, meaning that both benefits and investment costs are higher ex-post than in ex-ante calculations. This phenomenon implies that the mean benefits are underrun and the investment costs are overrun. The latter confirms previous studies regarding the overrun of cost estimates, as in, e.g., Flyvbjerg et al. (2004) and Odeck (2004). However, the overall BCA results, as summarized in the NPV and NPV-Cost-ratio columns, show that the overall mean ex-post NPV results are higher than those estimated ex-ante. The MPE of NPVs is 50% higher, with a median of 0, and the MPPE of NPV-Cost-ratio is 0.14% points higher. These results imply that the mean underruns in benefits are large and offset the cost overruns. These results concur with those of Kelly et al. (2015). However, it should be noted that there are great variations across projects studied, as observed in the minima, maxima and standard deviations of the results.

As discussed in the methodology section, one should not be indifferent towards whether the changes were positive or negative; by default the negative and positive values cancels out in a MPE measurement of changes. To account for the fact the negative and positive changes are equally important and that larger values should have more weight, the MAPE and MPSE should be considered. The two lowest sections of Table 3 reports the summary statistics of mean change results according to MAPE and MPSE.

Firstly, consider the MPE results shown on the second uppermost part of the summary statistics section. They are simply an averaging of the PE for benefits, investment costs, NPV and, an averaging of PPE for NPV-Cost- ratio. As explained earlier, MPE values tends to be small because the negative and positive values will often offset one another. Another problem with MPE is that if some values change from negative to positive, the minimum and maximum values of changes may be misleading. Secondly, consider the case of PE for benefits. The MPE shows that the Minimum value of change is −760 while the maximum is 231. This is illogical as a percentage change of −760 is more than threefold a percentage of 231. Despite that MPE may give some reasonable mean values when considering the magnitudes of maximum and minimum mean change values, a different measures is required. Such a measure is provided in the second last section of the summary statistics i.e., the MAPE which measure the mean changes in absolute terms.

Table 3
Overall inaccuracy measurement of ex-post BCA.

Project no.	Project name	Benefits (mill.NOK)			Investment costs (mill.NOK)			NPV (mill.NOK)			NPV/cost- ratio (in % return from investments)		
		Ex-ante	Ex-post	PE	Ex-ante estimates	Ex-post estimates	PE	Ex-ante estimates	Ex-post estimates	PE	Ex-ante estimates	Ex-post estimates	PPE*
1	Rv 23 Oslofjordforbindelsen	6041	6256	4	1479	1692	14	4563	4565	0	4.97	5.36	0.39
2	Ev 18 Rannekleiv - Temse	745	912	22	337	414	23	409	498	22	1.15	1.13	−0.02
3	Rv 714 Hitra - Frøya	228	446	96	470	290	−38	−242	156	−165	−0.85	0.87	1.72
4	Ev 134 Teigeland - Håland	81	130	60	526	548	4	−445	−418	−6	−0.80	−0.72	0.08
5	Rv 62 Øksendalstunnellen	313	330	5	256	262	3	57	67	18	0.21	0.24	0.03
6	E8 Norkjosbotn-Laksvatnbukt	323	282	−13	429	530	24	−81	−219	168	−0.19	−0.41	−0.22
7	E18 Gutu-Helland-Kopstad	830	2749	231	4896	4771	−3	−4066	−2022	−50	−1.04	−0.53	0.51
8	E39 Kleivedammen-Andenes	112	173	54	256	249	−3	−144	−76	−47	−0.53	−0.29	0.24
9	E134 Hegstad - Damåsen	631	794	26	328	322	−2	303	473	56	0.87	1.39	0.52
10	Rv.616 Kolset - Klubben	−5	33	−760	394	517	31	−355	−428	21	−0.90	−0.80	0.10
11	Rv.580 Hop- Midttun	741	663	−11	401	459	14	340	203	−40	1.01	0.53	−0.48
12	E18 Ørje- Eidsberg grense	404	609	51	248	246	−1	60	311	419	0.37	2.04	1.67
13	E6 Akershus grense- Patterød	383	453	18	291	302	4	27	72	163	0.16	0.41	0.25
14	Rv. 35 Lunner - Gardermoen	1248	1286	3	670	780	16	395	285	−28	0.54	0.33	−0.21
15	E6 Halmstad - Patterød	479	1147	139	384	379	−1	95	768	708	0.40	3.25	2.85
16	E18 Brokelandsheia - Vinterkjær	818	1018	24	818	1010	23	184	450	145	0.28	0.76	0.48
17	E39 Svegatjørn - Moberg	214	327	53	323	280	−13	−90	62	−169	−0.28	−0.22	0.06
18	E18 Sekkelsten- Krosby	791	1206	52	627	704	12	165	502	204	0.25	0.67	0.42
19	E6 Ny Svinesundforbindelse	1833	2087	14	1083	1141	5	811	1011	25	0.75	0.89	0.14
20	E6 Skjerdingstad - Jaktøyen	435	597	37	904	830	−8	−418	−186	−56	−0.46	−0.22	0.24
21	Rv4. Reinsvoll - Hunndalen	504	837	66	355	522	47	149	315	111	0.74	1.07	0.33
22	Fv.43 Aunevik - Bukkesteinen	142	180	26	183	297	62	−41	−118	189	−0.21	−0.37	−0.16
23	E16 Kløfta - Nybakk	438	599	37	970	996	3	−532	−397	−25	−0.52	−0.38	0.14
24	E6 Svingenskogen - Åsgård	2208	3295	49	2851	3083	8	−642	209	−133	−0.38	0.19	0.57
25	E39 Gammelsæter - Nipetjørn	65	94	45	318	305	−4	−252	−211	−16	−0.75	−0.65	0.10
26	E18 Langåker - Bommestad	2926	1961	−33	1478	1253	−15	1202	449	−63	8.10	2.10	−6.00
27	E6 Vist -Jevik -Selli	1334	1785	34	1207	1744	44	94	−12	−113	0.07	−0.01	−0.08
Summary statistics													
	MPE	899	1120	12	833	886	9	57	234	50	0.48	0.62	0.14
	Median	479	663	34	429	522	4	57	156	0	0.16	0.33	0.14
	Min	−5	33	−760	183	246	−38	−4066	−2022	−169	−1.04	−0.80	−6.00
	Max	6041	6256	231	4896	4771	62	4563	4565	708	8.10	5.36	2.85
	StDev	1215	1291	160	975	982	20	1238	1005	181	1.87	1.33	1.38
	MAPE	–		73	–	–	16	–	–	117	–	–	0.67
	Median	–		37	–	–	12	–	–	56	–	–	0.24
	Min	–		3	–	–	1	–	–	0	–	–	0.02
	Max	–		760	–	–	62	–	–	708	–	–	6.00
	StDev	–		143	–	–	16	–	–	146	–	–	1.22
	MPSE	–		25603	–	–	505	–	–	35133	–	–	1.93
	Median	–		1351	–	–	151			3148	–	–	0.06
	Min	–		9	–	–	1			0	–	–	0.00

(continued on next page)

Table 3 (*continued*)

		Benefits (mill.NOK)			Investment costs (mill.NOK)			NPV (mill.NOK)			NPV/cost- ratio (in % return from investments)		
Project no.	Project name	Ex-ante	Ex-post	PE	Ex-ante estimates	Ex-post estimates	PE	Ex-ante estimates	Ex-post estimates	PE	Ex-ante estimates	Ex-post estimates	PPE*
	Max	–		577600	–	–	3883			501860	–	–	36.00
	StDev	–		108750	–	–	888			97456	–	–	6.88

* PPE stands for Percentage Points Error and is the most relevant to consider when measuring change in NPV/cost-ratio which, initially is expressed as percentages.

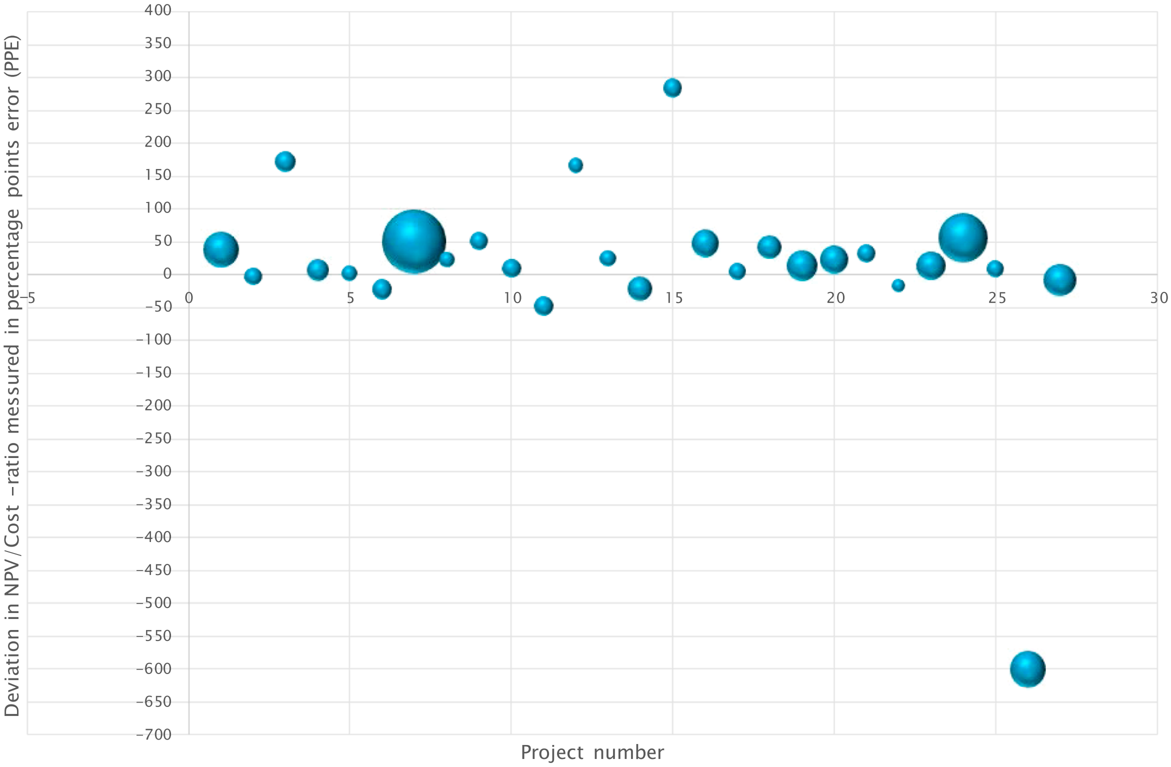


Fig. 1. Distribution of deviations in NPV–Cost ratio by size of project measured by ex-ante construction cost estimates.

There it can be seen that the minimum and maximum changes are quite different from those of the MPE; they are much larger in values indicating that there is a large variation in PEs across individual projects when it is assumed that positive and negative values are equally a problem. If MPSE reported in the last section is considered, the minimum and maximum changes are squares of the MAPE values. Our conclusions with respect to the mean discrepancies between ex-ante and ex-post estimates of BCAs must therefore be interpreted with care. The mean change as measured by MPE for the NPV-Cost ratio, is relatively small at only 0.14% points. The standard deviation is however large at 1.38% points. If however, one wants to be indifferent with respect to higher or lower values i.e., that the over-/underestimations are equally serious, then the MAPE is the right measure to use. It shows that the mean change is relatively large at 0.67% point, but the standard deviation at 1.22% points and is smaller as compared to the case of MPE.

An interesting issue to study is how these overall results may vary by project size, as measured by estimated project costs. A question that is readily asked is whether larger projects tend to have larger overall BCA inaccuracies, as measured by the NPV-Cost ratios. Fig. 1 is a so-called bubble plot of the PPE (deviation) of NPV-Cost ratio versus project number. The size of the bubbles represents the size of the projects as measured by ex-ante estimated costs. Fig. 1 appears to show that the PPE of NPV-Cost ratios of larger projects are lower than those of smaller projects; their deviations as measured by PPE are closer to zero than are those of smaller projects. These observations are merely visual and can statistically be deceitful. To ensure that this is the case, we performed a Kruskal–Wallis equality of populations rank test. We tested for the equality of the median NPV-Cost ratio between large and small projects, as measured by the ex-ante estimated using three different criteria: (1) large projects are those that have estimated cost larger than the mean of all projects in the data set, (2) large projects are those that have had estimated cost at approximately 200 million NOK which has been the definition of large projects in Norway and, (3) large projects are those that have had estimated costs more than one standard deviation above the mean. The tests yielded the following results: ($\rho = 0.77$, $\rho = 0.75$ and $\rho = 0.54$). From these results, one cannot reject the hypothesis that large and small projects exhibit the same level of inaccuracy as far as the NPV-Cost ratio is concerned; i.e., the hypothesis that the level of inaccuracy between them is different is rejected at any reasonable significance level.

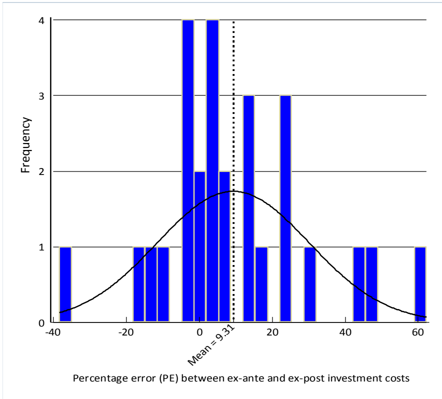
(ii) Inaccuracy of investment cost estimates

The conclusions from the previous section notwithstanding, investment costs are an important component of BCA; hence, its level of inaccuracy in terms of the PE measure merits closer examination. The differences between the estimated investment costs in the ex-ante situation and the actual costs ex-post as measured by PE are summarized in Table 4. The table provides both the summary statistics and a distribution plot of the deviation between ex-ante and ex-post investment costs.

The summary statistics in Table 4 indicate that the mean change between the estimated and the actual costs as measured by MPE

Table 4
Deviation and distribution of construction costs.

	Investment costs				
	Ex-ante	Ex-post	MPE	MAPE	MPSE
Mean	833	886	9.31	15.81	504.67
Median	429	522	4.12	12.28	150.82
min	183	246	−38.25	0.78	0.60
Max	4896	4771	62.32	62.32	3883.24
StDev	975	982	20.44	15.96	887.60
No. of projects with overrun	–	–	17	–	–
No. of projects with underrun	–	–	10	–	–
Total no. of projects	27	27	27	27	27



reported in the third last column is 9% but varies between −38% and 62%. The mean change at 9% is small and would meet the accuracy requirement for the cost estimate at the zoning plan level, which is ± 10%. The MAPE and MPSE are reported in the last two columns respectively. By default, both these values are higher than the MPE values. MAPE is the MPE in absolute terms and MPSE is MAPE squared. Thus, the value of MAPE at 15.8% indicate that the deviation in cost estimates, irrespective of the direction of change, is large. These results indicate that the distribution of the deviations is slightly skewed to the right, as shown in the distribution plot, which means that overruns are slightly prevalent compared to underruns. Overall, approximately 63% (17/27) of the projects had overruns, whereas 37% (10/27) had cost underruns. These results concur with Odeck (2004), Odeck (2014), and Odeck et al. (2015), who used MPE as the appropriate measure and concluded that there were cost overruns among large road projects in Norway but that those overruns were small and at a tolerable level. Further, these results confirm that cost overruns of Norwegian road projects are less than those of other countries, as has been observed by, e.g., Flyvbjerg et al. (2004). However, we warn that as far as inaccuracy of estimates are concerned where one should be indifferent with regards to the direction of inaccuracy, MAPE is the preferred measure. Our results using it shows an inaccuracy of about 16% which, all else equal and in terms of cost control purposes, should be regarded as borderline acceptable.

(iii) Inaccuracy of traffic forecasts

Both traffic volumes and traffic growth rates are important inputs for calculating the benefits of road projects. Inaccurate estimates of the expected traffic volumes and traffic growth rates will most likely lead to inaccurate BCA estimates. Table 5 provides summary statistics regarding the inaccuracy of traffic volumes in the first full year after project opening, average traffic growth rate in the first 5 years after opening, and traffic volume 5 years after project opening. Note that for traffic growth rate, MPPE is used instead of MPE.

The table demonstrates that both traffic volumes and traffic growth rates are higher ex-post compared to ex-ante, attesting that traffic forecasts are generally higher than forecasted. For the first full year after project opening, the MPE between ex-ante estimates and ex-post outcomes is 14.4%, but it varies widely in the interval [−32%, 85%]. For the mean traffic growth rate for the first 5 years after project opening, which is normally expressed in percentages, the MPPE is 1.9%, and it varies in the interval [−1.2%, 14%]. For traffic volume 5 years after project opening, the MPE is 29%, and it varies in the interval [−20%, 173%].

A comparison of the deviation in traffic volume in the first full year after opening and deviation at 5 years after opening is shown in Fig. 2, which is a box and whiskers plot. The two plots shown in the figure are similar, with the small difference that the deviation is greater after 5 years than it is in the first full year after opening. This difference is a result of the cumulative effect of the underestimate of traffic growth rate discussed above.

What these results tell overall is that whereas the MPE (deviation) depicts higher traffic than estimated ex-ante, there are projects at both ends of the scale; some exhibit higher underestimates, and others exhibit higher overestimates. However, a majority of projects (59%) exhibit higher traffic volumes and higher traffic growth than forecasted. These results confirm those of Odeck (2013), who concluded that the regional traffic forecasts for road projects in Norway are inaccurate, but the magnitudes of the inaccuracies are relatively small. Given that inaccuracy is two-sided, i.e., either over- or underestimated, these results indicate that inaccuracy is common and sometimes large with respect to the traffic forecasts of road projects. A recommendation therefore is that the NPRA should constantly strive to improve the accuracy of its forecasts. We discovered that the BCA results presented to decision-makers are made at an early stage, sometimes up to 4 years before decision-making. Thus, a second recommendation is that the BCAs that include traffic forecasts should be updated to the time of decision-making. Third and concerning traffic forecasts, we discovered that forecasts

Table 5
Deviation in traffic forecasts.

	Traffic volume - first full year (AADT)					Mean Traffic growth rate- 5 first years					Traffic volume 5 years after opening(AADT)				
	Ex-ante	Ex-post	MPE	MAPE	MPSE	Ex-ante	Ex-post	MPPE	MAPE	MPSE	Ex-ante	Ex-post	MPE	MAPE	MPSE
Mean	8738	9140	14.38	25.20	1091.74	0.02	0.04	1.93	1.97	11.30	8937	10,120	29.03	34.48	2652.98
Median	6814	8000	9.51	21.25	451.73	0.01	0.03	1.45	1.43	2.05	7150	7412	16.67	20.06	402.47
min	88	155	−32.03	1.27	1.62	0.01	0.00	−1.23	0.00	0.00	90	173	−19.82	0.59	0.35
Max	26,300	26,954	85.19	85.19	7256.52	0.02	0.18	14.25	14.25	203.06	28,550	32,000	173.30	173.30	30031.85
StDev	7649	7627	29.14	21.37	1733.25	0.00	0.04	2.78	2.72	38.81	8333	9001	41.42	38.26	6017.29
No. of projects with traffic underestimation	–	–	16	–	–	–	–	24	–	–	–	–	20	–	–
No. of projects with traffic overestimation	–	–	11	–	–	–	–	3	–	–	–	–	7	–	–
Total no. of projects	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27

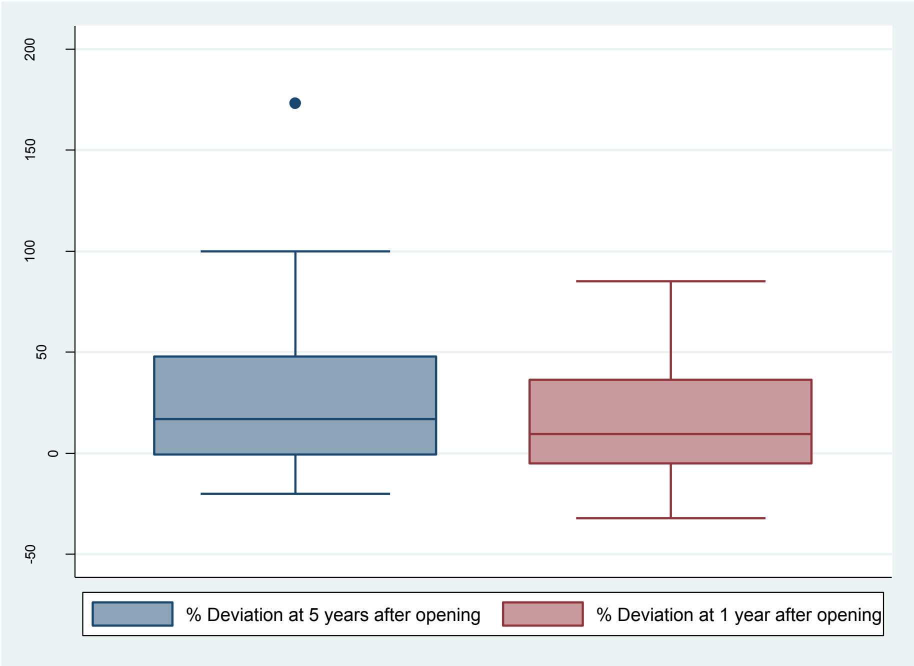


Fig. 2. Box and whiskers chart of deviation in traffic forecasts.

often used are those for the regions (see, e.g., Odeck (2013) and are not project-specific. We therefore recommend that project-specific traffic forecasts be used and presented to decision-makers.

(iv) Inaccuracy of accident costs

Many road projects are built as a way to reduce the rate of accidents across corridors, as evidenced in Table 1. The extent to which this objective is met across the projects studied can be examined in Fig. 3, which plots the stipulated reduction in the number of accidents with personal injury ex-ante versus actual outcomes ex-post, 5 years after the projects have been implemented.

In the figure, if the points are concentrated along the 45° line, then the ex-post outcomes exactly match the ex-ante estimates; if they lie above the line, then the ex-post outcomes produce higher reductions, and vice versa if they lie below the line.

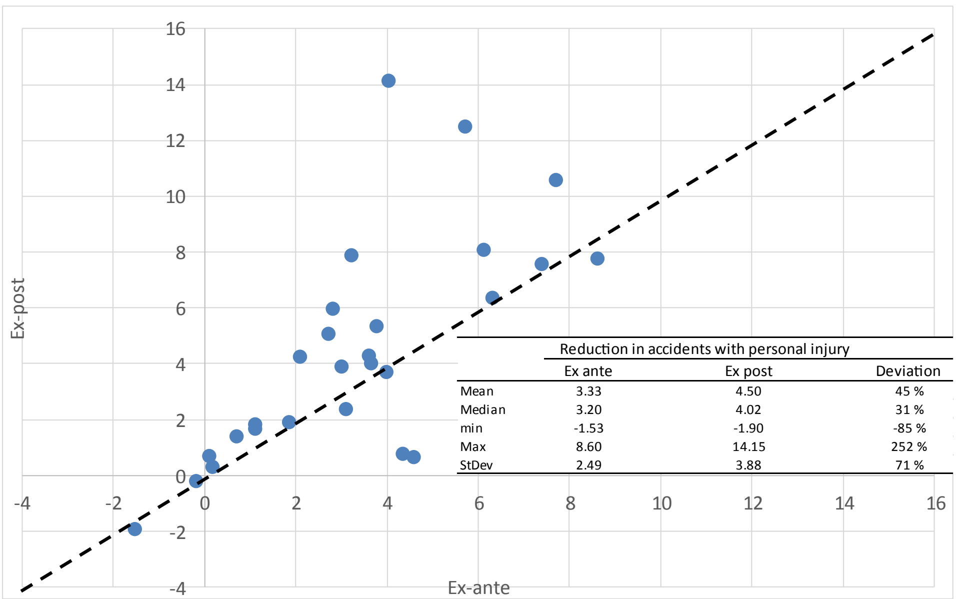


Fig. 3. Ex-ante versus ex-post plot of number of accidents, first 5 years after opening.

As is evident from the figure, most of the points lie above the line, verifying that most projects produce higher reductions in accidents than estimated ex-ante; thus, projects outperform in terms of the goal of reducing accidents with personal injury. This result is verified by the summary statistics in the figure, showing a mean deviation of 45%, although there are some negative deviations, and the standard deviation is large at 71%. These results imply that new road investments most likely lead to savings in accidents even though they lead to increased traffic; new roads are designed to reduce road accidents despite increasing traffic. In fact, a major objective of the Norwegian government's investments in roads, as profiled in the Norwegian National Transport (NTP) for the period of 2018–2029, is to reduce the rate of road accidents.

Note that we are here comparing the number of accidents before and 5 year after road projects were built. The results of such a comparison may not appropriately account for all the confounding factors that explained the change in the number of accidents. This means that road implementation alone may not be the only contributor to the reductions/increases observed. There may have been other contributors to accident reduction/increases in general such as increased police control, education of new drivers and/or general trends. Failing to account for all other confounding factors may lead to false conclusions about the effectiveness of road construction as a measure to reduce traffic safety. This problem has been widely addressed in the literature of evaluation of traffic safety measures and goes under the name regression-to the mean; see for instance, Elvik (2002) and Hauer (1997). We hence warn that our results must be interpreted with care since we have not accounted for all possible confounding factors.

(v) Projects' individual characteristics that explain the observed deviations

It has been observed that deviations between ex-ante BCA estimates and ex-post outcomes are present but relatively small on average, whereas the deviation intervals are large. A question that is readily asked therefore is why there are such variations across projects. An answer can be provided by examining the project-specific factors that caused deviations from ex-post estimates.

In general, the documents that we studied revealed that one of the reasons for inaccuracy was the long planning horizon, which is the time between the planning and implementation of projects. Projects may develop differently than assumed at the time of decision making. The reasons may include a change of plans and cost considerations, among others. The history and development of projects after the decision to implement them are important factors that may help explain some of the discrepancies observed. Other explanations include the length of the construction period, changes in construction such as the number of protective environmental measures, and changes in safety standards and geo-technical difficulties. All such changes will naturally impact the benefits and cost of projects, such that BCA results will appear inaccurate. Flyvbjerg (2005), Siemiatycki (2009) and Cantarelli et al. (2010) suggested similar explanations that can account for the inaccuracy of forecasts in the transport sector but mainly in the context of cost overruns.

The costs of projects can be grouped as follows: technical, psychological, and political-economic explanations.

To illustrate how such changes impact the inaccuracy of the BCA results of the projects in our study, we consider three projects for which such changes occurred and may have impacted the inaccuracy of the BCA results obtained.

a) Project no. 1: Rv. 23 Oslofjordforbindelsen

The ex-ante BCA analysis of this project assumed two construction phases. The first stage was a single tube tunnel serving traffic from both directions. In the second stage, a twin-tube tunnel was to be constructed such that each tube served traffic from only one direction. The second stage's completion was set for 2013. At the time of the ex-post evaluation, however, the plans had changed, and only one tube tunnel serving traffic in both directions was built. The second-stage plans were abandoned, and/or there was still discussion about when or why the second tunnel should be built. Therefore, the benefits and costs of the second stage tube tunnel could not be included in the ex-post evaluation, whereas they were included in the ex-ante evaluation. How this change of plans impacted the deviation between the ex-ante and ex-post BCA analysis is difficult to analyze. Not including the construction costs of the second stage in the ex-post analysis would lead to favorable results, whereas not including benefits would have the opposite result. However, the exclusion of the construction costs of the second stage should more than outweigh the exclusion of the possible additional benefits of a second tunnel such that the results derived here may exhibit an ex-post BCA result that is too large and positive compared to the ex-ante estimates. This is because the project was the first fast link in that location; hence, an additional tunnel cannot be expected to generate benefits similar to the first tunnel, but the costs would be of the same magnitude.

b) Project no. 3: Rv. 714 Hitra - Frøya

The planned project consisted of both a sub-sea tunnel connecting the two islands of Hitra and Frøya and an improvement of the connecting roads from the mainland. At the time of the ex-post evaluation, the project had still not been completed because the design of the connecting road had not been determined. Thus, the ex-post evaluation included only the construction costs of what had been built at that time, which was only the tunnel. At the time of ex-post analysis, the traffic followed the old road. Thus, both the benefits and costs of what was not built could not be accounted for ex-post. This delay in construction may explain the inaccuracy observed for this project.

c) Project no. 4: E134 Teigeland – Håland

A part of this project was not built because the project management decided not to build the full project to avoid possible cost overruns. The part of the project that was not built could not be subtracted from the ex-post analysis because of insufficient data. Note from Table 3 that this project exhibited high benefits and costs. Again, as in the two projects above, the cost side would have been higher than the benefit side had the whole project been built by the time of the ex-post analysis.

(vi) Would ex-post results change the project rankings?

Thus far, our results indicate that MPPE of NPV-Cost ratio is 0.14% points higher, with great variations across the projects studied. An issue that comes to mind is whether the ex-post results would have changed the decision-makers rankings had the ex-post results been known at the decision-making time.[4]

Given that NPV-Cost ratio is the measure that should be used for project ranking, we first ranked projects in descending order according to the ex-ante and ex-post NPV-Cost ratios. Thereafter, we conducted both Spearman's rho (ρ) and Kendall's tau (τ) correlation tests to infer whether the two rankings were significantly correlated. The results are reported in Table 6.

Both the Spearman's ρ and Kendall's τ rank correlations provide evidence that the correlation between rankings in terms of ex-ante and ex-post NPV-Cost-ratios are moderately high and positive. This implies that the higher a project is ranked ex-post, the more likely that it will also rank higher ex-post. The $|t|$- value for the H_o hypothesis that NPV-Cost-ratio ex-ante and ex-post are independent is rejected at the 1% significance level. The similarities/differences between project ranking ex-ante versus ex-post can further be visualized using Fig. 4.

The interpretation of Fig. 4 is that had the ex-ante and ex-post given identical rankings, then all points which, indicate individual projects rankings, would lie along the 45° stippled line that equates ex-ante and ex-post rankings. As the figure shows, only three projects have the same ranking according to both ex-ante and ex-post ranking, but it is evident that most of the points lie close to the dotted line. Second, points above the 45° line indicate that the rankings are higher ex-ante, and vice versa for ex-post assessment. Projects 4, 7 and 25, as marked in the figure, are examples of how the rankings differ; project 4 has a higher ex-ante ranking, project 7 has the same ranking both ex-post and ex-ante and project 25 has a higher ex-post ranking. The overall picture, however, is that the project rankings ex-ante versus ex-post are fairly similar despite few individual cases where rankings may change.

With respect to Norway where the data used are derived, there are however, two reservations regarding the applicability of the results that must be mentioned. The first is that the Norwegian decision-makers have been shown not to fully use the results of BCAs when ranking projects; see, for instance, Eliasson et al. (2015). However, the results derived here are still highly useful to them since they use BCA results partially in addition to other non-monetized impacts when ranking projects; see, for instance, Odeck (2010). Thus, in totality, the accuracy of BCA estimates is important to ensure accurate and informed decision-making. The second issue involves the precisions with which NPV-Cost ratios are presented to the decision-makers. If, for instance, the confidence intervals of the ex-ante and ex-post results overlap, then there is no reason to expect that the rankings would change because both the results are equally likely to occur. Unfortunately, the confidence intervals of the NPV-Cost ratio estimates are rarely provided to decision-makers. For the latter, we call upon providers of BCA's to also provide confidence intervals of their estimates.

(vii) Are there improvements in accuracies over time?

An interesting aspect of ex –post evaluations if conducted regularly, is that theyenable an examination of the extent to which improvements in inaccuracies are made over time. Despite the fact that the data set used in this study is not large, i.e., does not contain a large number of projects and the period in which the projects were implemented stretches only from 2000 to 2009, we investigated the extent to which there have been improvements in the inaccuracies of BCA estimates over time. We thus warn that the results derived should be interpreted with care. Our motivation for examining improvements over time is motivated by Post Opening Project Evaluation –POPE (2015).[5] They found that the accuracy of traffic forecasts improved over time leading to improvements in BCA over time in the case of England.

Fig. 5 shows the developments in accuracies of some of the most important components/elements of BCA across projects and by year of project implementation in the case of Norway. Ceteris paribus, projects are implemented successively according to their year of analyses.

Consider all the panels shown in Fig. 5 which are plots of PE (or PPE in the case of NPV-Cost ratio) versus the opening year of projects. In each of the panels, plots are provided and a linear regression line shown including the coefficients, R-squared and the p-values of the estimated equation. In Panel A which correlates the PE of cost estimates to the opening year of project, the Adjusted-R^2 is 0.0284 while the ρ-value is 0.4. Opening year therefore explains too small a variation in PE of cost estimates and the relationship is not significant. Similar conclusions can be made with regards to Panel C which correlates PE of NPV to project opening; the Adjusted-R^2 is 0.0054 while the ρ-value is 0.72. The correlation between PE of traffic forecasts 5 years on and the correlation between PPE of NPV-Cost ratio and year of project opening are however significant at the 10% significance level. The ρ-values are 0.070 and 0.072 for PE of traffic forecasts 5 years on and PPE of NPV-Cost ratio respectively. A probable reason for these latter observations is that the transport models from which traffic forecasts are made have been under scrutiny since the early 2000 and are continuously revised; what we observe here may be the results of the continuous revisions. As matter of fact, Odeck (2013) and Odeck and Welde (2017) observed that the PE of Norwegian traffic forecasts were much lower than observed elsewhere, and argued that this may be the result of continuous revision of the transport model in use.

[4] We thank one anonymous referee for asking us to consider this issue.
[5] Thanks are due to an observant reviewer who advised on addressing this issue.

Table 6
Spearman's ρ and Kendall's τ tests for correlations between ex-ante and ex-post rankings.

Test	Correlation coeff.	|t|- value for the test of Ho: NPV/cost-ratio ex-ante and ex-post are independent
Spearman's p	0.827**	0.000
Kendalls's τ	0.675**	0.000
Number of obs = 27		

** Correlation is significant at the 0.01 level (2-tailed).

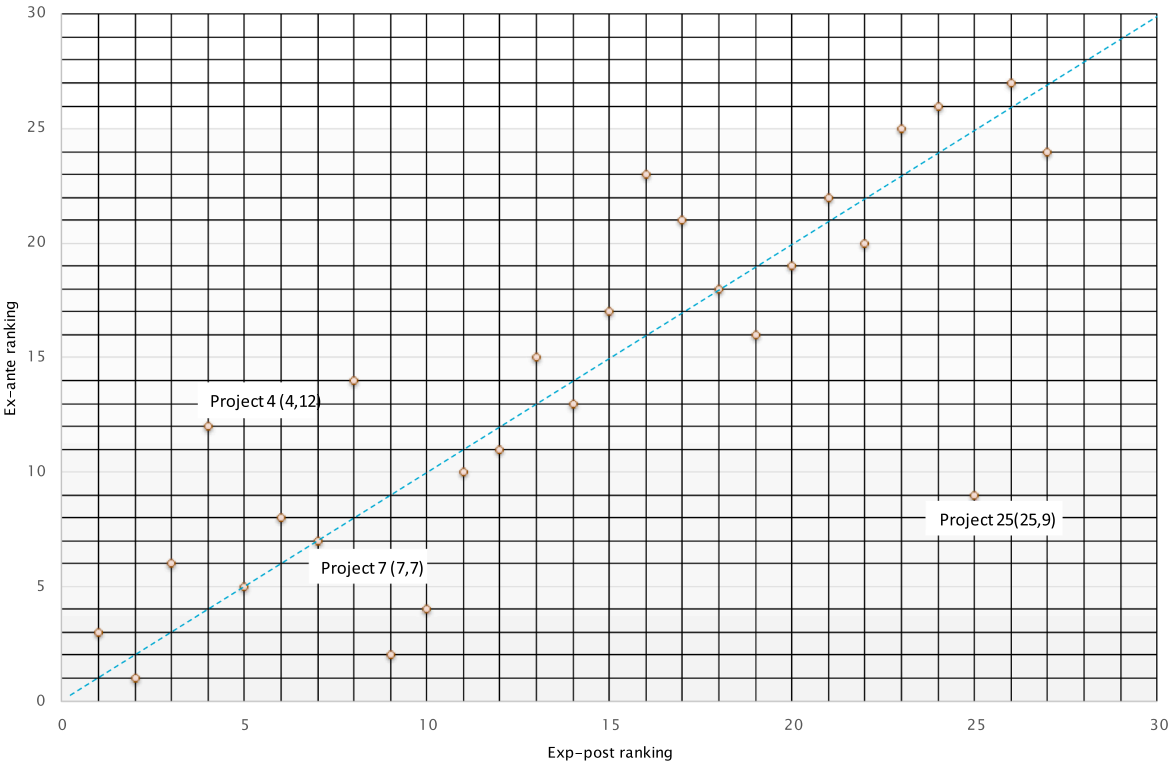


Fig. 4. A plot of the rankings ex-post versus rankings ex-post.

7. Discussion and concluding remarks

This paper has argued and suggested that ex-post BCAs should be conducted at some point after projects have been implemented to gauge the extent to which the projects deliver and that they may help identify potential areas for improving ex-ante BCAs. It then empirically examines the accuracy of BCAs in the context of Norwegian road projects. Ex-post BCA analyses were performed 5 years after projects were realized. In total, 27 large projects were analyzed.

We find that the BCA results such as NPV and NPV-Cost ratio are, on average, higher ex-post compared to the ex-ante estimates. A major explanation of benefits that are higher than forecasted is that traffic levels tend to be under-estimated in the ex-ante BCAs. Thus, an immediate recommendation to the Norwegian authorities charged with providing traffic forecast/transport models is that they should constantly strive to improve the accuracy of their traffic forecasts. There is also a tendency for cost overruns, but these are relatively small in magnitude, averaging 9%. We warn however, that one should not be indifferent as to the aggregation method used to summarize the results i.e., whether MPE or MAPE. If the objective is to examine the performance of a portfolio of projects, MPE is the most appropriate measure. However, if the objective is to measure performances across individual projects and where underruns and overruns are equally undesirable, then MAPE is the most appropriate measure. By default, the former will show lower values as compared to the latter.

The results above do not indicate any signs of optimism bias to the same extent, as has been observed elsewhere in the literature. A possible explanation for less optimism bias in Norway compared to other European countries is that Norway has a so-called Quality Assurance (QA) regime initiated by the ministry of Finance, under which the cost estimates provided by the road authorities for all large projects with cost estimates above 500 mill. NOK are scrutinized by external consultants appointed by the Ministry of Finance. Note that this scrutiny is for projects larger than those defined as large by the NPRA and which we analyze in this paper. This form of scrutiny has led to a tremendous reduction in cost overruns of large road projects above/equal to 500 mill. NOK; see, for instance,

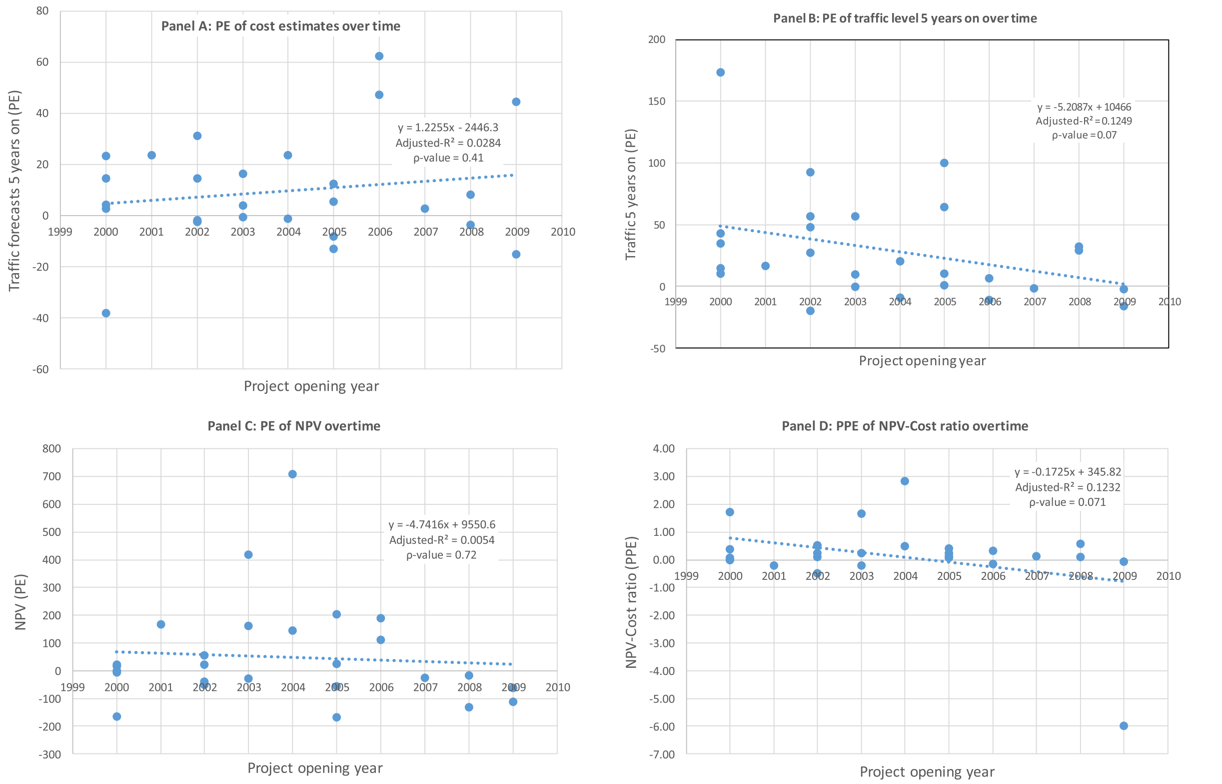


Fig. 5. Development in accuracies over time.

Odeck et al. (2015), who also addressed the workings of the Norwegian QA regime. As a result, it is therefore quite possible that the NPRA has adjusted internally by scrutinizing cost estimates of projects lower than 500 mill. NOK. Further, on the benefit side of BCAs, where transport modeling/traffic forecasts are quite important, the Norwegian Public Roads Authority (NPRA) has centralized its development of transport modeling. In that framework, a standardized model is used across all projects and is quality-assured by the NPRA. Thus, the NPRA has no incentives for biases, as can be observed in other countries where transport modeling/forecasts are not centralized but rather are made by concessionaires who may be biased to see their projects go through; see, for instance, Odeck and Welde (2017), who evaluated the accuracy of toll road traffic forecasts in Norway compared to other countries.

Whereas optimism bias does not exist in Norway for the reasons explained above, the results reveal clearly that there are some large variations between ex-ante and ex-post estimates; especially when aggregating across individual projects as measured by MAPE. A question that is readily asked is what the NPRA is doing to narrow the variations observed. Until now, ex-post evaluations have not been forthcoming; thus, the NPRA has not been aware of the problems regarding these variations. However, the QA regime and the continual revision of transport models/forecast imply that the NPRA is constantly alert to ensure that variations in actual outcomes are reduced; for the latter, see, for instance, Odeck (2013), who addressed the accuracy of the national road traffic growth-rate forecasts in Norway.

We also find that ex-post NPV-Cost-ratios produce a relatively similar ranking of projects compared to rankings based on their ex-ante counterparts. Thus, we have no reasons to believe that the ex-post results would have changed the decision-makers' priorities had they been known at the time of decision-making

A few issues encountered while working on this paper merit some discussion. The first is that an ex-post BCA of complex projects is a challenge in the sense that simplifications of the transport network are needed in this process because it is impossible to replicate a transport model ex-post. The uncertainty and consequences of these simplifications should be addressed carefully. The results should only be interpreted as indications of whether the impacts are greater than, equal to, or smaller than forecasted. A second issue involves the upkeep/storage of the ex-ante data and other material that is necessary for ex-ante analyses. We found that there was no standard for keeping ex-ante data; hence, a significant amount of time was needed to retrieve the data. We therefore urge the involved authorities to develop a data upkeep system that will enable ex-ante evaluations; the use of this system will enhance the transparency of ex-ante BCAs and the institutions that conduct them.

References

Anguera, R., 2005. The channel tunnel—an ex post economic evaluation. Transp. Res. Part A 40 (4), 291–315.
Baeza, M.A., Vassallo, J.M., 2012. Traffic uncertainty in toll motorway concessions in Spain. Proc. Inst. Civ. Eng. Transp. 165 (TR2), 97–105.

Boardman, A.E., Mallery, W.L., Vining, A.R., 1994. Learning from Ex -Ante/Ex-post cost-benefit comparisons: the Coquihalla highway example. Socio-Econ. Plann. Sci. 28 (2), 69–84.

Boardman, A.E., Mallery, W.L., Vining, A.R., 2011. Cost-Benefit Analysis: Concepts and Practice, fourth ed. Prentice Hall.

Bråthen, S., 2001. Do fixed links affect local industry? A Norwegian case study. J. Transp. Geogr. 9 (1), 25–38.

Börjesson, M., Jonsson, R.D., Lundberg, M., 2014. An ex-post CBA for the Stockholm Metro. Transport. Res. Part A: Policy Pract. 70, 135–148.

Cantarelli, C.C., Flyvbjerg, B., Molin, E.J.E., van Wee, B., 2010. Cost overruns in large-scale transport infrastructure projects: explanations and their theoretical embeddedness. Eur. J. Trans. Infrastruct. Res. 10 (1), 5–11.

De Rus, G., 2010. Introduction to Cost-Benefit Analysis: Looking for Reasonable Shortcuts. Edward Elgar Publishing.

Eliasson, J., Börjesson, M., Odeck, J., Welde, M., 2015. Does Benefit-Cost Efficiency Influence Transport Investment Decisions? J. Trans. Econ. Policy (JTEP) 49 (3), 377–396.

Elvik, R., 2002. The importance of confounding in observational before-and-after studies of road safety measures. Accid. Anal. Prev. 34 (5), 631–635.

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

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

Gomez, J., Vassallo, J.M., Herraiz, I., 2015. Explaining light vehicle demand evolution in interurban toll roads: a dynamic panel data analysis in Spain. Transportation 42, 1–27.

Hauer, E., 1997. Observational before/after studies in road safety. Estimating the effect of highway and traffic engineering measures on road safety.

Kelly, C., Laird, J., Costantini, S., Richards, P., Carbajo, J., Nellthorp, J., 2015. Ex post appraisal: What lessons can be learnt from EU cohesion funded transport projects? Transp. Policy 37, 83–91.

Kjerkreit, A., Odeck, J., Sandvik, K.O., 2008. Post opening evaluation of road investment projects in Norway: How correct are the estimated future benefits? In: European Transport Conference 2008, pp. 1–16.

Kjerkreit, A., Odeck, J., 2015. The accuracy of ex-ante cost benefit analysis – evidence from the Norwegian road sector. In: XXVth World Road Congress, Seoul, November 3–5 2015.

Makridakis, S., Wheelwright, S.C., Hyndman, R.J., 1998. Forecasting: Methods and Applications, 3rd ed. John Wiley & Sons, New York.

Love, P.E.D., Sin, C.-P., Wang, X., Irani, Z., Thwala, D.W., 2012. Overruns in transportation infrastructure projects. Struct. Infrastruct. Eng. Maintenance, Manage., Life-Cycle Des. Perform. 1–19.

Meunier, D., 2010. Ex-post evaluation of transport infrastructure projects in France: old and new concerns about assessment quality. In: Paper for the 12thWorld Conference on Transportation Research 2010, Lisbon.

Minken, H., 2016. Project selection with sets of mutually exclusive alternatives. Econ. Transport. 6, 11–17.

Nicolaisen, M.S., Driscoll, P.A., 2014. Ex-post evaluations of demand forecast accuracy: a literature review. Trans. Rev. 34 (4), 540–557.

NZ Transport Agency, 2016. https://www.nzta.govt.nz/planning-and-investment/planning/transport-monitoring/audits/post-implementation-reviews/.

Odeck, J., 2004. Cost overruns in road construction – what are their sizes and determinants? Transp. Policy 11, 43–53.

Odeck, J., 2010. What determines decision-makers' preferences for road investments? Evidence from the Norwegian road sector. Trans. Rev. 30 (4), 473–494.

Odeck, J., 2013. How accurate are national road traffic growth-rate forecasts?—The case of Norway. Transp. Policy 27, 102–111.

Odeck, J., 2014. Do reforms reduce the magnitudes of cost overruns in road projects? Statistical evidence from Norway. Transport. Res. Part A: Policy Pract. 65, 68–79.

Odeck, J., Welde, M., Volden, G.H., 2015. The impact of external quality assurance of costs estimates on cost overruns: Empirical evidence from the Norwegian road sector. Eur. J. Trans. Infrastruct. Res. 15 (3), 286–303.

Odeck, J., Welde, M., 2017. The accuracy of toll road traffic forecasts: An econometric evaluation. Transport. Res. Part A: Policy Pract. 101, 73–85.

POPE, 2016. https://www.gov.uk/government/collections/post-opening-project-evaluation-pope-of-major-schemes.

Post Opening Project Evaluation –POPE, 2015. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/497241/POPE__Meta_2015_Final_210116_-_FINAL.pdf.

Siemiatycki, M., 2009. Academics and auditors – comparing perspectives on transportation project cost overruns. J. Plan. Educ. Res. 29, 142–156.

Taroux, J.P., Chapulut, J.N., Mange, E., 2005. The New Ex-post Evaluation Methods for Large Projects in France, ETC 2005 proceedings.

Welde, M., Odeck, J., 2011. Do planners get it right? The accuracy of travel demand forecasting in Norway. Eur. J. Trans. Infrastruct. Res. 11 (1), 80–95.

Welde, M., Odeck, J., 2017. Cost escalations in the front-end of projects – empirical evidence from Norwegian road projects. Transport. Rev. 1–19. https://doi.org/10.1080/01441647.2016.1278285.