

Combining Reference Class Forecasting with Overconfidence Theory for Better Risk Assessment of Transport Infrastructure Investments

Steen Leleur¹

Department of Transport, Technical University of Denmark

Kim Bang Salling²

Department of Transport, Technical University of Denmark

Inga Pilkauskiene³

Rail Net Denmark

Morten Skou Nicolaisen⁴

Department of Planning, Aalborg University, Denmark

Assessing the risks of infrastructure investments has become a topic of growing importance. This is due to a sad record of implemented projects with cost overruns and demand shortfalls leading, in retrospect, to the finding that there is a need for better risk assessment of transport infrastructure investments. In the last decade progress has been made by dealing with this situation known as planners' optimism bias. Especially attention can be drawn to the use of reference class forecasting that has led to adjustment factors that, when used on the estimates of costs and demand, lead to cost-benefit analysis results that are modified by taking historical risk experience into account. This article seeks to add to this progress in risk assessment methodology in two ways: first it suggests to apply reference class forecasting (RCF) in a flexible way where the effort is focused on formulating the best possible reference pool of projects and second to apply overconfidence theory (OT) to interpret expert judgments (EJ) about costs and demand as relating to a specific project up for examination. By combining flexible use of RCF with EJ based on OT interpretation it is argued that the current adjustment factor methodology of RCF can be further developed. The latter is among other things made possible by the comprehensive project databases that have been developed in recent years. For this article the project database developed in the UNITE research project 2009-2013 has been employed. The presented simulation-based risk examination named SIMRISK is concluded to provide a new 'in-depth' possibility for dealing with uncertainties inherent to transport decision making based on socio-economic analysis. In addition a further research perspective is outlined.

Keywords: *Transport infrastructure investments, risk assessment, reference class forecasting, expert judgments, overconfidence theory, optimism bias.*

¹ Bygningstorvet 116 Vest, 2800 Kgs. Lyngby, Denmark, E-mail: sl@transport.dtu.dk

² Bygningstorvet 116 Vest, 2800 Kgs. Lyngby, Denmark, E-mail: kbs@transport.dtu.dk

³ Banedanmark, Amerika Plads 15, 2100 Copenhagen Ø, Denmark, E-mail: ipka@bane.dk

⁴ Aalborg Universitet, Skibbrogade 5, 9000 Aalborgm Denmark, E-mail: mortenn@plan.aau.dk

1. Introduction

Providing suitable decision support for strategic transport decision making is a topic of growing concern. This is due to a sad record of implemented projects with cost overruns and demand shortfalls leading, in retrospect, to the finding that there is a need for better risk assessment of transport infrastructure investments. In the last decade progress has been made dealing with this situation known as planners' optimism bias (Flyvbjerg et al., 2003). Especially attention can be drawn to the use of reference class forecasting that has led to adjustment factors that, when used on the estimates of costs and demand, lead to cost-benefit analysis results that are modified by taking historical risk experience into account (Flyvbjerg and COWI, 2004). Research on infrastructure investment decisions in relation to the uncertainty in construction costs and transport prognoses – referred to as the transport demand estimates – has revealed that these factors are especially critical in the cost-benefit analysis (Flyvbjerg, 2005). Thus Flyvbjerg et al. (2003) have disclosed a consistent tendency towards overestimating transport-related benefits and underestimating construction costs which shows the presence of estimation problems in the assessments. Many cases have been documented where a poorly executed estimation of both impacts has led to investments that later on turned out to be less than satisfactory (Priemus et al., 2008; Cantarelli et al., 2012a, 2012b; Nicolaisen, 2012). Therefore it becomes of great relevance to examine the construction costs and transport demand estimates found to be the major sources of uncertainty in order to minimize the risk of selecting inadequate projects. This is especially relevant for large infrastructure projects (Banister and Berechman, 2000).

In addition to current methodology based on reference class forecasting (e.g. Flyvbjerg and COWI, 2004) this article presents a new approach that makes flexible use of reference class forecasting (RCF) and expert judgments (EJ) in combination. After a presentation of RCF and EJ in Sections 2 and 3, respectively, and their combination in Section 4, a case study in Section 5 demonstrates the principles on a major transport investment. Section 6 is a discussion of the suggested new approach, among other things with a comparison with the currently recognised approach based on reference class forecasting which makes use of uplift factors. Section 7 presents a conclusion and perspective. The research behind the article was carried out in the UNITE project (2009-2013) about uncertainties in transport project evaluation funded by the Danish Strategic Research Council.

2. Reference Class Forecasting

2.1 Optimism Bias and Uplift factors

Reference class forecasting (RCF) is based on theories of judgment under uncertainty deriving from the Nobel Prize winning work of Daniel Kahneman and Amos Tversky (Kahneman, 2011). The main idea behind RCF is that a suitable set of projects similar in type to the project examined can serve as a reference in terms of their individually experienced uncertainty to inform the actual examination of uncertainty. Reference class forecasting has been used and operationalized by Flyvbjerg and others to set out various adjustment or 'uplift' principles (Flyvbjerg and COWI, 2004; Priemus et al., 2008). These principles are applied to modify given estimates by uplift factors based on a project database sample collected over the past decades depicting, in this case, costs and demands for transport infrastructure projects, respectively. The RCF project database contains information on each project as indicated in (1):

$$U = \frac{((X_a - X_f) \cdot 100)}{X_f} \quad (1)$$

where U is percent inaccuracy, X_a is the actual traffic demand/cost after the project is opened and X_f is the forecasted traffic demand/cost on which the decision to build has been taken.

To exemplify the mentioned uplift principles for costs, Table 1 shows some of the uplifts applicable within transport infrastructure projects that planners can use to take different levels of optimism bias into account in the appraisal. Altogether five different levels (the levels of certainty aimed at) are indicated ranging from 50 to 90% across three main categories of transport infrastructure projects: road, rail and fixed link. Each of these categories includes a huge variety of different project types, i.e. road comprises motorways, trunk roads, local roads, bus lane schemes etc., while rail comprises metro projects, light rail projects, high speed rail projects etc., and fixed link bridges and tunnels (Flyvbjerg and COWI, 2004).

Table 1. Applicable investment cost uplifts for selected percentiles applied to constant prices (adapted from Flyvbjerg and COWI, 2004)

Level of acceptable optimism bias	50%	60%	70%	80%	90%
Road	15%	24%	27%	32%	45%
Rail (and air)	40%	45%	51%	57%	68%
Fixed Link	23%	26%	34%	55%	83%

The certainty levels are classified according to the risk aversion of decision-makers in terms of cost overruns. In case the risk of a cost overrun must be for instance less than 20% (certainty level 80%) for a road type project, the construction cost estimate must be uplifted by 32%. Thereby an initial budget estimate equal to 100 million DKK is raised to 132 million DKK to ensure with 80% probability that the final investment cost will not surpass 132 million DKK. Flyvbjerg and COWI (2004) have suggested to apply the 50 percentile (low certainty) for decision-makers who are willing to take a high risk that cost overruns can occur, while 80% can be applied by investors demanding a higher degree of certainty that cost overruns will not occur. This will typically be the case when no additional funds are available.

Table 1 can guide practitioners and analysts within the field of transport project appraisal. However, as RCF assumes a known outcome distribution, it is a requirement that the examined project is likely to display similar properties to that of the projects in the reference class. Practitioners should thus consider how comparable the projects in the reference class are to the project being examined.

Similarly to cost overruns, adjustments can also be estimated for demand shortfalls. Instead of using overall RCF uplift factors that modify initial and optimistic values (adjusting too low construction costs and too high demand forecasts) it is in this context suggested to apply the historical information and experience in the reference class of projects as a base for derivation of probability distributions that can enter into e.g. a Monte Carlo simulation of costs and demand estimates, see Salling and Leleur (2009; 2011). The principles are given below with a description of the concept of certainty graphs.

2.2 The concept of certainty graphs

Basically the point estimate result benefit-cost ratio (BCR) from the cost-benefit analysis is replaced by a probability-based interval result referred to as the certainty graph. This graph is formulated through input probability distributions associated with the construction costs and transport demand, respectively – transport demand as expected benefits from future travelling time savings and revenue – to be entered into the Monte Carlo simulation (Vose, 2008; Salling, 2008). Specifically, the certainty graph (CG) is made up of the accumulated probability estimates for achieving at least the BCR indicated as argument; thus: $CG(x) = P(BCR \geq x)$. An example is shown in Figure 1 below.

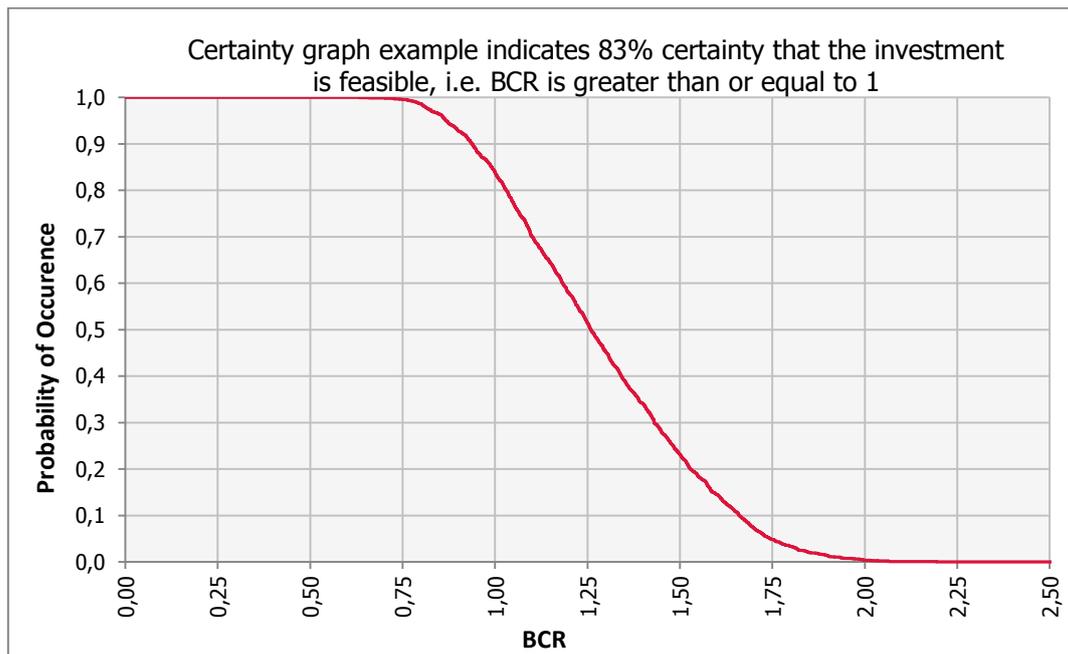


Figure 1. *The concept of a certainty graph*

The cost-benefit analysis applied should be based on a procedure or protocol (the actual national CBA-manual for example) that can generally be accepted in the study context. For a large transport infrastructure investment the impacts to be covered will generally consist of: construction and maintenance costs, time savings, operation costs, accidents, noise emissions, local air pollution and climate (Trafikministeriet, 2003; European Commission, 2008).

The validity of the CG depends on the probability distributions used for the construction costs and the transport demand. Salling (2008) has shown that the Erlang (Gamma) and Beta-PERT distributions are plausible probability distribution functions for construction costs and transport demand, the main sources of uncertainty in transport infrastructure appraisal. In the below demo-case in Section 5 the Erlang distribution is used for the construction costs, while a histogram distribution is used for the transport demand. The formulation of this histogram distribution making use of reference class forecasting and scenario-based expert judgments in combination is the main focus of the demo-case.

3. Expert judgments

Each feasibility study of a proposed large transport infrastructure project is carried out by a team comprising a number of experts with different areas of professional knowledge. As regards the experts to be consulted as part of the risk assessment, especially the design engineers and the planners doing the forecasts are of interest. These experts are usually willing to be involved in the interviews which are undertaken to lead to expert judgments about Min and Max estimates of the costs of the infrastructure and the expected demand with the latter represented by values such as number of vehicles and number of passengers in the target year of the prognosis.

These expert judgments can, as mentioned, be made by interviewing key persons that have been involved in the project planning. However, it may also be considered to use workshops where scenarios are made use of as described below.

3.1 Making use of scenarios

The way scenarios can be employed is described in relation to the Nuuk airport case, which is applied in Section 5 for a practical demonstration of the principles and methodology set out in this article. Below, focus has been limited to the scenarios that relate to the variability of demand, i.e. the Min and Max estimates to enter into the risk assessment.

In the Nuuk airport case, scenarios were formulated in a context of two major axes representing what was perceived as the two most influential drivers of uncertainty. Afterwards, each driver was subdivided into three categories, which thus expressed a suitable scenario-grid for the deliberations about how the demand for air travel to and from Nuuk would be influenced by the different future developments indicated by the scenarios. In this way the scenario-grid in Figure 2 was set up (Salling and Leleur, 2011).

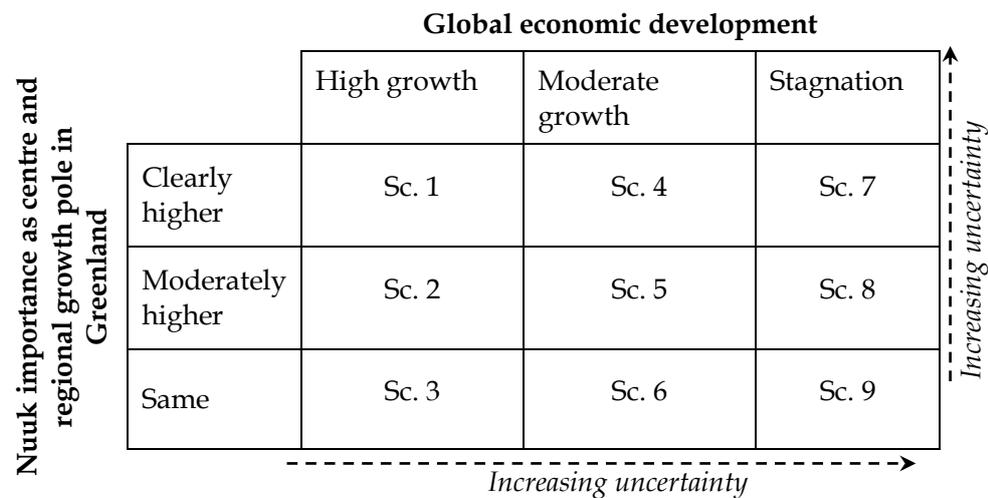


Figure 2. The concept of scenario-grid

The purpose of a scenario-grid is to assist the experts in making Min and Max estimates with each scenario as a plausible background. It is important to note that only Min and Max estimates are needed, and furthermore that the specific scenario producing the Min estimate can be expected to be different from the scenario producing the Max value. Therefore the estimates are global and not tied as pairs to one of the scenarios.

With several experts involved, the expert judgment to be part of the risk assessment is simply obtained by making use of average values from the experts or by setting these as agreed 'consensus' values. This provides what in the risk assessment can be seen as contextual information representing an 'Inside view' opposite to the 'Outside view' of reference class forecasting (RCF) (Kahneman, 2011, chapter 23). The next stage is to bring this together with the 'deviation-record' of similar projects implemented in the past. This is the RCF-component of the risk examination, which is dependent on which projects are used as references. For this purpose an established large project database must be made use of and adapted to the actual transport infrastructure project up for examination. Adapted here refers to the situation in which a number of previously implemented projects can be identified as being so similar to the actual one that they can be used to form a subgroup of projects in the project database providing historical experience that can be used to express possible future variability as regards the actual project under examination.

3.2 The adapted project database

As mentioned an established reference class of similar projects is part of the platform for the risk examination, where each project in the project database, established as part of the background for

working with RCF-information, is judged on the basis of its relevance as useful historical experience. This provides the subset of reference projects, which is referred to as the adapted project database, see the bottom of Figure 3 that gives an overview of the methodology applied in the model software (Salling and Leleur, 2015).

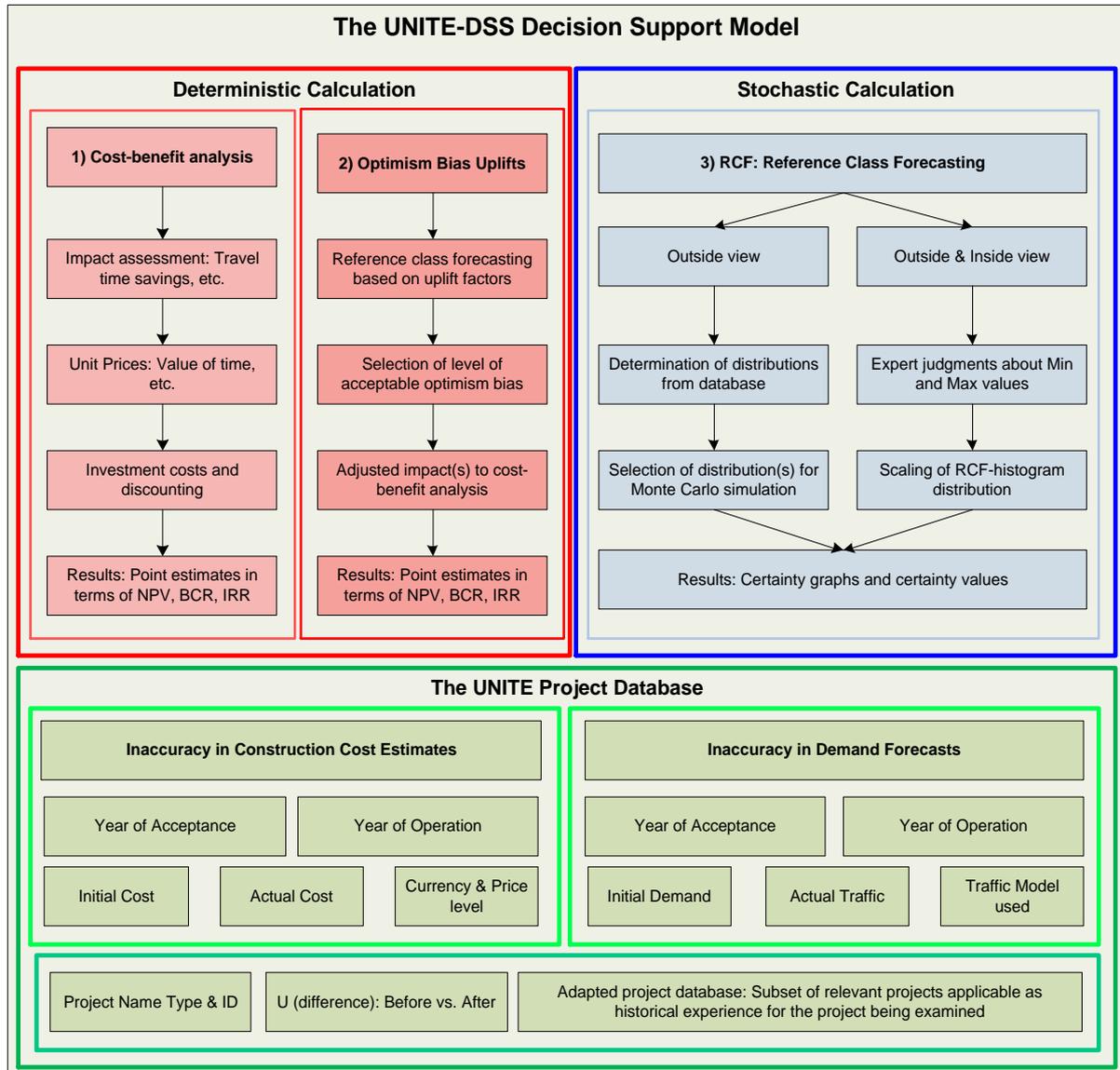


Figure 3. Overview of the UNITE-DSS software

As indicated in Figure 3 the UNITE-DSS software comprises both deterministic calculation (including cost-benefit analysis and optimism bias uplifts) and stochastic calculation comprising both a version taking an Outside view and a version taking an Outside & Inside view. As noted above the term 'outside' stems from Kahneman and indicates the main idea of RCF to make use of solely 'outside' reference projects for dealing with the project inaccuracy to be examined (Kahneman, 2011, chapter 23). The idea proposed in this article to make use of reference class forecasting in combination with expert judgments, with the latter representing the inside view of experts, has inspired the authors' use of the denomination Outside & Inside view.

The adapted project database at the bottom of Figure 3 is presented in more detail in Figure 4.

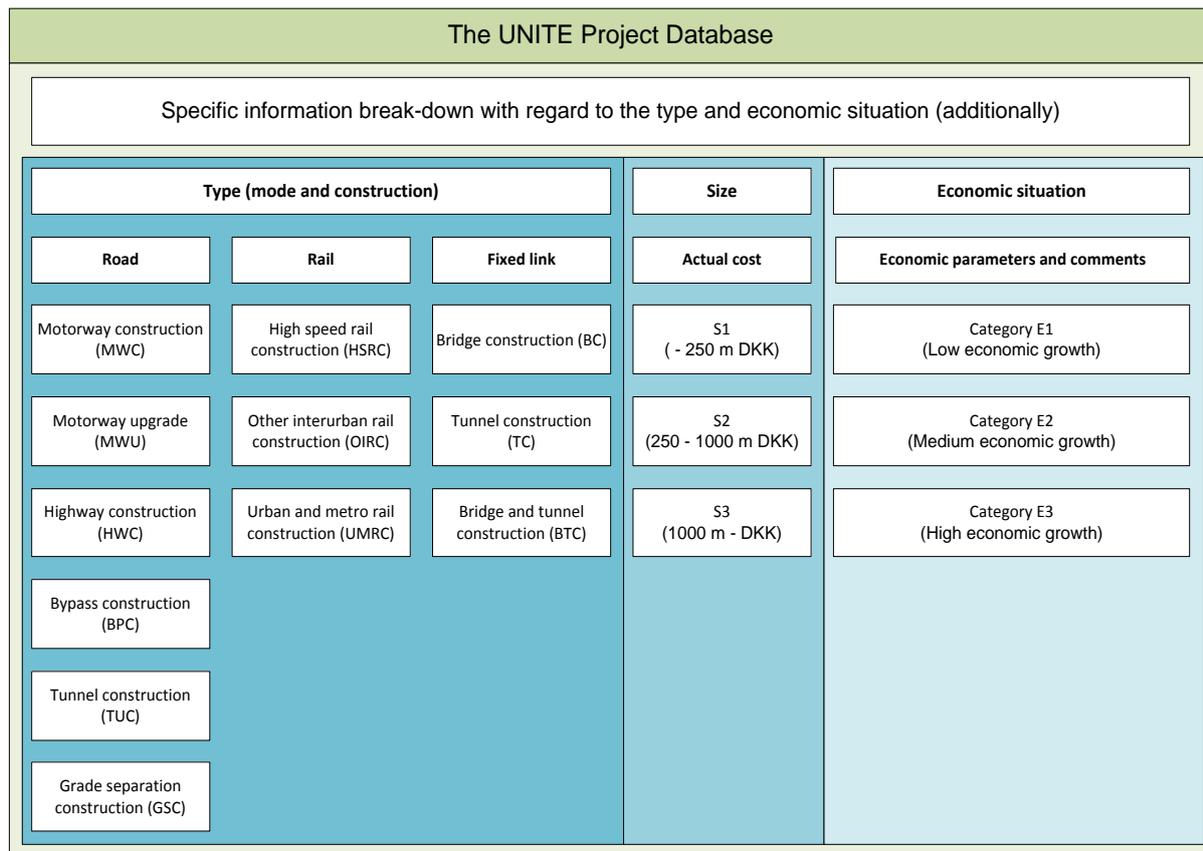


Figure 4. Information break-down within the UNITE Project Database with regard to type of infrastructure project

Figure 4 depicts the structure and content of the current UNITE Project Database containing information with regard to road, rail and fixed link projects, respectively. Previous research carried out in the field (from e.g. Flyvbjerg et al., 2003, Priemus et al., 2008) focussed on the difference between the initial cost and demand and the actual cost and demand, respectively, thereby providing for each project a value for U as presented in (1). In addition to such a focus, the UNITE research work has given priority also to broadening the information about the projects in the database so each project in the database has a direct link to an Access Database with detailed and comprehensive information about the project, including also relevant references and source details. This helps to investigate and ensure the verification and validity of data from the UNITE Project Database and will provide an important basis for further research. For example, it can be investigated what influence the actual economic growth regime has on the RCF-data.

As the adapted project database with its RCF-information plays a major role in the proposed methodology, so do the previously described experts judgments. How these respective inside and outside views can be combined is described below.

4. Combining RCF with Expert Judgments

The 'inside' uncertainty that is disclosed by the Min and Max values is used to 'calibrate' the 'outside' historical uncertainty revealed by the set of deviations from the adapted project database by using *overconfidence theory*, see (Van de Venter and Michayluk, 2008; Taleb, 2010). This theory states that people in general (including experts) are unaware of their lack of

capability to indicate a complete range of variation. This means that the ranges indicated by the experts' Min and Max estimates for costs and demand are too narrow. On the basis of examinations carried out in many different contexts, it has been found that the ranges provided by the people that have been questioned *typically only account for around 60% of the variation* (i.e. they are overconfident about their ability to indicate a full interval for the variability). This result is relatively stable (Van de Venter and Michayluk, 2008, p. 549; Goodwin & Wright, 2009, p. 251), which is what makes it relevant to be used in the proposed risk assessment approach referred to as SIMRISK (simulation-based risk assessment). In SIMRISK overconfidence theory is used to modify (extend or contract) the histogram distribution of deviations determined from the adapted project database.

By taking the Min and Max values as representative of the 20% and 80% percentiles (thereby defining the central 60% 'overconfidence interval') a histogram distribution of deviations based on the adapted project database can be transformed by rescaling it to fit this central interval. Thereby a case-specific histogram probability distribution is produced, which can be used as input for the Monte Carlo simulation by applying e.g. the @RISK software (Palisade Corporation, 2007) to produce a case-specific certainty graph, see Subsection 2.2 above. We refer to this distribution as the SIMRISK probability distribution function (pdf). The characteristic of this distribution is that it represents a mix of historical RCF-information and expert derived project-specific information combined by using overconfidence theory. A SIMRISK pdf can be worked out for both the construction costs and for the transport demand. In the case example below the SIMRISK methodology is demonstrated solely on transport demand uncertainty (prognosis for future traffic load), whereas the uncertainty of construction costs is modelled on the basis of an Erlang distribution, see (Salling and Leleur, 2009). However, with construction costs also being treated as scenario-sensitive the SIMRISK methodology can include also construction costs in the uncertainty examination similar to the way that transport demand is described to be included. In the simulation it is assumed that construction costs and transport demand are uncorrelated (Salling, 2008).

5. Nuuk Airport as demo-case

The case demonstration makes use of information described in (Leleur et al., 2007; Salling and Banister, 2009; Salling and Leleur, 2011), in which an examination of the international airport in Nuuk is presented by three case alternatives. These consist of two upgrade alternatives replacing the existing runway in Nuuk, i.e. increasing the current runway length to either 1799 metres (m) or 2200m, and as the third alternative the construction of a new, relocated airport to the south with a 3000m runway, consequently leading to the closing of the existing airport. Results from this study clearly pointed towards either of the two extension alternatives leaving the Nuuk 3000m alternative infeasible from a societal perspective (Leleur et al., 2007). In October 2007 information was released that the Home Rule Authorities in Greenland recommended the Nuuk 2200m alternative for implementation (Sermitsiaq.AG, 2007). However, construction work has not been started yet. In the light hereof this article examines the socio-economic robustness of this decision by applying reference class forecasting and combining an outside view (RCF-data) with an inside view based on expert judgments.

The UNITE Project Database currently consists of 262 projects of which altogether 204 projects were judged as not being relevant for the Nuuk airport case. The selection process as described above is simply a click on/click off procedure in the UNITE-DSS software developed in the UNITE project. In this way the *remaining 58 projects* were identified as a representative, relevant historical reference when considering the uncertainty that could be expected to influence the transport demand (representing the forecast of benefits from saved travel time and revenue). Together they define the histogram pdf below in Figure 5. The inaccuracy is based on the

percentage deviation U of the actual value X_a from the forecasted value as indicated in (1) above. For transport demand the green colour indicates a 'favourable' deviation (higher than forecasted) and the red colour a 'critical' deviation (lower than forecasted) with regard to the influence on the socio-economic feasibility of the examined alternative. The range between -20% and 20% is considered an acceptable practical deviation and is not coloured. Favourable and critical deviations thus highlight the influence on the BCR-value and have been indicated to ease the reading of the histograms.

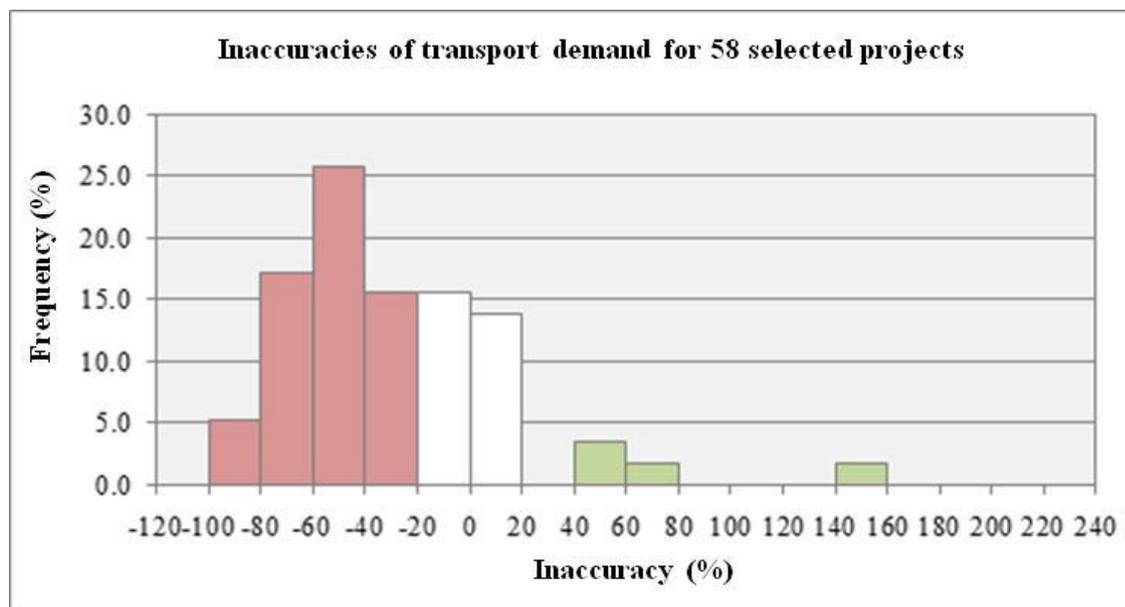


Figure 5. Histogram pdf based on RCF information for the Nuuk airport case

At an expert conference consisting of a small group of people with a special and detailed knowledge about the Nuuk project, the model estimate 170 (rounded overall benefit measure from monetary value) was presented, and the participants were informed about the underlying premises and the traffic demand modelling work behind this estimate. Furthermore, the major sources of uncertainty were explained to the participants. The background and the possible influence from the two drivers behind the rising uncertainty in Figure 2 were included in this presentation and discussion. Each participant was then asked to give a global Max and Min estimate and, thereby, an interval was determined that the expert was confident would include the value that was expected to be realised in the future by implementing the project.

By averaging the individual responses, the Max estimate 230 $((230-170)/170 = +35\%)$ and the Min estimate 90 $((90-170)/170 = -47\%)$ were obtained. Thus the participants seen as a group were confident that the value to be realized was somewhere in the interval from 90 to 230. Using the %-deviation from the model estimate 170 as a yardstick we find that the length of this interval is $35\% - (-47\%) = 82\%$. Using Figure 5 and surveying the probability mass between 35% and -47% and beyond we can note that the experts 'miss' some historically referenced variability. Their actual overconfidence can, however, be set against the expected overconfidence, which can be done in the following way. Using the adapted project database as shown in Figure 5, the intersections of the 20% and 80% percentiles (that distribute the probability mass in 20% LOW, 60% CENTRAL INTERVAL and 20% HIGH) can be identified and applied to determine the distance between these intersection points (again we use as the yardstick the %-deviation from the model estimate 170). This length of the central interval is found to be 70%. Thereby we can see that the full interval (in the experts' overconfident opinion) of variability is 'wider' (82%) than the central interval (70%) based on applying the RCF-data. The expert interval could also have been

'narrower' than the central interval with a distance of maybe 65%. In either situation, however, it is important to keep in mind that the group of experts is 'overconfident' about its ability to capture the overall variability of the value.

By applying the result from the overconfidence theory as presented in (Van de Venter and Michayluk, 2008) that the group's 'overconfidence' is 60%, the x-axis in Figure 5 is adjusted with a factor equal to: $82\%/70\% = 1.17$. By use of this factor, we transform the histogram in Figure 5 into the histogram in Figure 6 as shown below. Hereby, the probability mass in Figure 5 is 'stretched' by multiplying each project's percentage deviation U , see (1) above; the width of the histogram bins is kept equal to 20%.

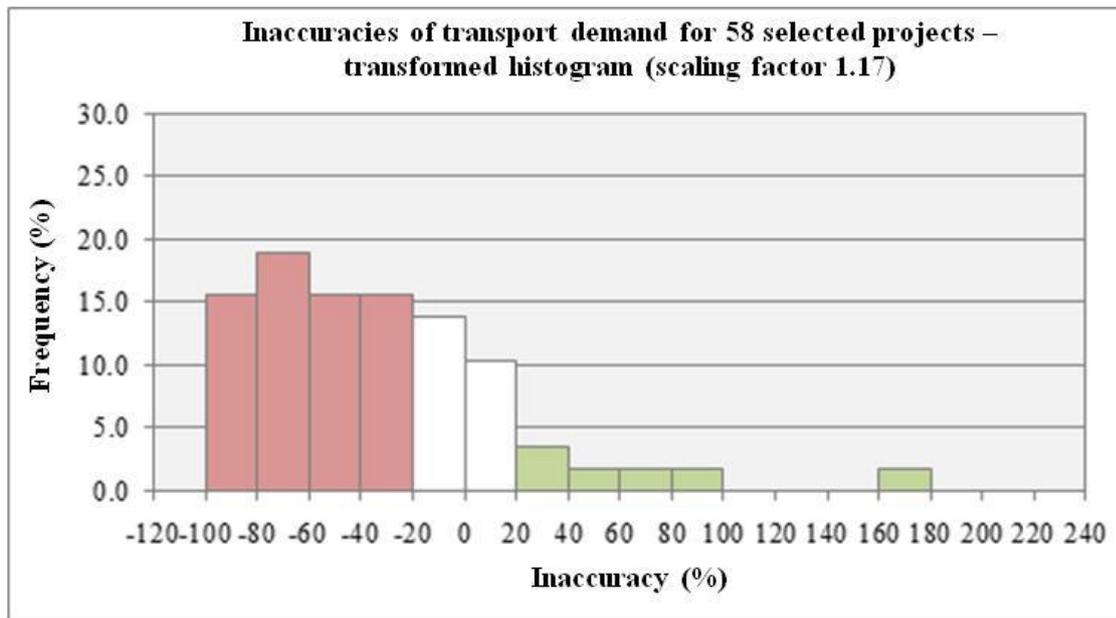


Figure 6. SIMRISK pdf based on scaling the RCF-histogram for the Nuuk airport case

Thus, Figure 6 shows the resulting SIMRISK histogram or pdf when the RCF-information is modified by using the expert judgments and making use of overconfidence theory to calibrate their interplay. As the scaling factor is higher than 1, the input from the experts 'adds' to the uncertainty provided by the historical information in the adapted project database. The opposite will be the case with a scaling factor less than 1, which will produce a narrower and steeper histogram distribution instead of a wider and flatter distribution.

By using the SIMRISK pdf from Figure 6 as input to a Monte Carlo simulation as regards traffic demand together with an Erlang pdf for construction costs, the certainty graph in Figure 7 is obtained.

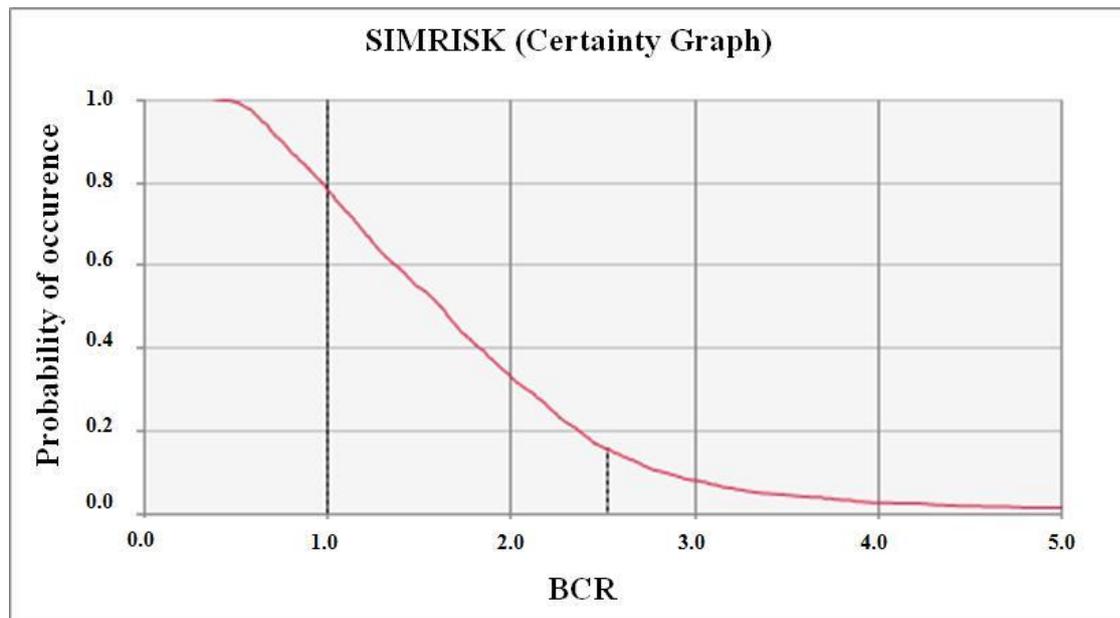


Figure 7. SIMRISK certainty graph result for the Nuuk airport case

It can be seen that compared to the BCR point estimate equal to 2.52 (Salling and Leleur, 2011), the certainty graph expresses that there is only 16% probability to achieve such a BCR. In general, there is a 78% probability that the investment is feasible from a socio-economic point of view, i.e. that the BCR is greater than or equal to 1.

6. Discussion

The robustness analysis of the Nuuk airport case shows that the result from a conventional cost-benefit analysis providing a BCR equal to 2.52 is to be seen as rather optimistic. With the BCR cut-off value equal to 1, the case result is that with rounded values the Nuuk 2200m alternative has around 80% probability of being a socio-economically feasible transport infrastructure investment. This finding is seen as highly relevant for the decision-makers. Furthermore, they will also note that the BCR equal to 2.52 (or higher) is only expected with around 15% certainty according to the determined certainty graph.

Evidently, the result obtained is dependent on the subset of projects selected from the project database. In the Nuuk case 58 projects were selected as relevant out of 262 projects. In this selection process it was necessary to scrutinise each project to decide whether it should be included or not. Based on the UNITE Project Database it became possible to obtain a reasonable number of projects in the subset while maintaining the focus on the relevance of each project selected for the actual examination. In the Nuuk-case especially similarity as regards size of investment and type of construction were seen as decisive selection parameters.

Generally the identification of the adapted database and the number of projects included are important activities that will influence the uncertainty examination. It is also evident that the expert judgments influence the outcome of the examination. Therefore, the appointment of experts and the conduct of their involvement are important issues. To establish a foundation as concerns the validity of the examination result, it is recommended that a log book is worked out as part of the examination. In this way the result can afterwards be reconsidered by inspecting the arguments behind the establishment of the adapted project database and the expert judgments behind the Min and Max estimates.

As concerns the validity of the results provided by the SIMRISK certainty graph it can be noted that in the case above the experts' input *added* to the uncertainty provided by the RCF-information. This is simply due to the fact that a scaling factor higher than 1 (in this case 1.17) was found based on applying overconfidence theory. In case a scaling factor below 1 had been found leading to a contraction of the RCF-histogram it could – instead of basing the SIMRISK certainty graph on this – have been considered to keep the RCF-histogram unscaled. With such a more cautious use of the methodology it produces a result that is in principle not different from a result based on the earlier mentioned RCF-uplift factors if these are updated based on the adapted project database. However, compared to the deterministic uplift factor approach the stochastic RCF-Outside & Inside view approach set out in this article may have the advantage of communicating the uncertainty of the investment in a graphical way that is easier for the decision makers to understand and make use of. The proposed cautious use of the SIMRISK methodology may have the advantage that the level of uncertainty due to RCF-information can be tested so that with scaling factors below 1 the RCF-information is kept unchanged while with scaling factors higher than 1 SIMRISK makes it possible to include and examine certain special circumstances from expert judgment indicating that *the uncertainty may well be higher than if solely RCF-information is made use of*. This possibility to pick up such special uncertainty is found to be a highly interesting feature as in some respects all large infrastructure projects are unique when being inspected closely. As concerns ease of use it should be noted that with the subset of projects determined in the UNITE Project Database and with the experts' Min and Max estimates laid down, the UNITE-DSS software promptly produces the RCF-histogram, the scaling factor, the SIMRISK pdf and the resulting SIMRISK certainty graph, see Figures 5-7. Furthermore, the software includes the possibility to change the overconfidence value from the default value 60% to a higher or lower value in case certain conditions indicate that this is relevant, see Lin and Bier (2008).

7. Conclusion and Perspective

A new approach based on the combination of reference class forecasting and expert judgments has been outlined for the purpose of undertaking simulation-based risk examination, SIMRISK, with regard to the socio-economic feasibility influenced by optimism bias. This is a topic that in recent years has been examined by several researchers based on the use of reference class forecasting, which, among other things, has led to a methodology applying uplift factors. No doubt this was an improvement compared to a conventional sensitivity analysis. This article proposes to combine historical RCF-experience based on a selected set of relevant projects in combination with expert judgments from a conducted expert conference.

A central assumption is that historical RCF-information can be combined with expert judgments and that overconfidence theory can be made use of in this respect. The principles have been demonstrated by using SIMRISK on the Nuuk airport case. As described the expert judgments can both add to the uncertainty revealed by the RCF-information but also deduct from this uncertainty which will depend on the scaling factor found. With a scaling factor equal to 1 no change will occur. A cautious way of using SIMRISK has been pointed out by suggesting that modification of the RCF-histogram is only carried out for scaling factors higher than 1, i.e. for situations where the inputs from the experts imply a possibly higher uncertainty than revealed by the RCF-information. Decision makers may perceive a risk assessment tool to detect such situations as very useful.

Naturally, the usefulness of the methodology outlined in this article to a great extent depends on the quality of the database in establishing suitable reference classes. This can be a major obstacle, as detailed data collection and archiving procedures for completed projects are rarely standard practice. However, the increasing focus on inaccurate forecasts has led to some improvements in

this area, and several countries now have mandatory auditing of certain transport infrastructure schemes that, in time, will allow for a greatly improved empirical foundation for the methodology outlined in this article.

Based on the work so far it can be concluded that the procedure followed in the developed uncertainty examination can be comprehended by the decision-makers. It can be noted that certainty graphs, when introduced in the right way, are found to communicate the uncertainty of the assessment in a proper and understandable way. Upcoming research will focus on establishing a larger empirical project database including a guidance on how to use it in the best and most flexible way. Guidelines for the interaction with experts will also be an important research task.

Overall we conclude that the presented SIMRISK approach based on combining reference class forecasting with overconfidence theory provides a new 'in-depth' possibility for dealing with the uncertainties inherent to transport infrastructure decision making based on socio-economic analysis. As especially large transport infrastructure projects are unique, the Outside & Inside view adapted in SIMRISK, see Figure 3, makes it possible to add to the way of risk examination based solely on reference class forecasting (the Outside view) with expert judgment (the Inside view). The potential of this novel approach, to our knowledge not dealt with so far in the reference class forecasting literature, will be further examined in new case studies supported by the developed UNITE-DSS software and the established UNITE Project Database, with the latter to be successively enlarged by adding new projects.

Acknowledgements

The authors would like to acknowledge the financial support by the Danish Council for Strategic Research for the project "Uncertainties in Transport Project Evaluation". In addition the authors are grateful for constructive comments from the three anonymous reviewers to a previous version of this manuscript.

References

- Banister, D. and Berechman, J. (2000). *Transport Investment and Economic Development*. UCL Press, London.
- Cantarelli, C.C, van Wee, B., Molin, E.J.E. and Flyvbjerg, B. (2012a). Different Cost Performance: Different Determinants? The Case of Cost Overruns in Dutch Transport Infrastructure Projects. *Transport Policy*, 22 (1), pp. 49-56.
- Cantarelli, C.C, Molin, E.J.E., van Wee, B. and Flyvbjerg, B. (2012b). Characteristics of Costs Overruns for Dutch Transport Infrastructure Projects and the Importance of the Decision to Build and Project Phases. *Transport Policy*, 22 (1), pp. 85-95.
- European Commission (2008). *Guide to Cost-Benefit Analysis of Investment Projects*, Evaluation Unit, DG Regional Policy, European Commission, Brussels.
- Flyvbjerg, B. (2005). Measuring Inaccuracy in Travel Demand Forecasting: Methodological Considerations regarding Ramp Up and Sampling. *Transportation Research Part A - Policy and Practice*, 39 (9), pp. 522-530.
- Flyvbjerg, B., Bruzelius, N. and Rothengatter, W. (2003). *Megaprojects and Risk: An Anatomy of Ambition*. Cambridge University Press, Cambridge, UK.

Leleur et al.

Combining Reference Class Forecasting with Overconfidence Theory for Better Risk Assessment of Transport Infrastructure Investments

Flyvbjerg, B. and COWI (2004). Procedures for Dealing with Optimism Bias in Transport Planning. Guidance Document prepared for the British Department for Transport, UK.

Goodwin, P. and Wright, G. (2009). *Decision Analysis for Management Judgment*. Fourth Edition, John Wiley and Sons, London.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, NY, USA.

Leleur, S., Salling, K.B. and Jensen, A.V. (2007). Documentation and Validation of the TGB Evaluation Model in Greenland (in Danish), report, Technical University of Denmark.

Lin, S.-W. and Bier, V.M. (2008). A study of expert overconfidence. *Reliability Engineering and System Safety*, 93, pp. 711-721.

Nicolaisen, M.S. (2012). Forecasts: Fact or Fiction? Uncertainty and Inaccuracy in Transport Project Evaluation, Ph.D. thesis, Aalborg University.

Palisade Corporation (2007). Manual for @RISK, Version 5, Palisade Corporation, NY, USA.

Priemus, H., Flyvbjerg, B. and van Wee, B. (2008). *Decision-Making on Mega-Projects: Cost-Benefit Analysis, Planning and Innovation*. Edward Elgar Publishing, Cheltenham, UK.

Salling, K.B. (2008). Assessment of Transport Projects: Risk Analysis and Decision Support, Ph.D. thesis, Technical University of Denmark.

Salling, K.B. and Banister, D. (2009). Assessment of Large Infrastructure Projects: The CBA-DK Model. *Transportation Research part A: Policy and Practice*, 43 (9-10), pp. 800-813.

Salling, K.B. and Leleur, S. (2009). Transport Appraisal and Monte Carlo Simulation by Use of the CBA-DK Model. *Transport Policy*, 18 (1), pp. 236-245.

Salling, K.B. and Leleur, S. (2011). Modelling of Transport Project Uncertainties: Feasibility Risk Assessment and Scenario Analysis. *European Journal of Transport and Infrastructure Research*, 12 (1), pp. 21-38.

Salling, K.B. and Leleur S. (2015). Accounting for the inaccuracies in demand forecasts and construction cost estimations in transport project evaluation. *Transport Policy*, Vol. 38 (2015), pp. 8-18.

Sermitsiaq.AG (2007). Atlantic Airport in Nuuk Saluted (in Danish), <http://sermitsiaq.gl/indland/article19110.ece>, accessed 17 December 2012.

Taleb, N.N. (2010). *The Black Swan – the Impact of the Highly Improbable*. Random House, NY, USA.

Trafikministeriet (2003). Manual for Socio-Economic Assessments (in Danish), Ministry of Transport, Copenhagen.

Van de Venter, G. and Michayluk, D. (2008). An Insight into Overconfidence in the Forecasting Abilities of Financial Advisors. *Australian Journal of Management*, 32 (3), pp. 545-557.

Vose, D. (2008). *Risk Analysis – A Quantitative Guide*. John Wiley and Sons, London.