

Morten Welde, Magne Jørgensen,
Per Fridtjof Larssen og Torleif Halkjelsvik

Estimering av kostnader i store statlige prosjekter: Hvor gode er estimatene og usikkerhetsanalysene i KS2-rapportene?

Concept-rapport nr. 59





Morten Welde, Magne Jørgensen,
Per Fridtjof Larssen og Torleif Halkjelsvik

Estimering av kostnader i store statlige prosjekter: Hvor gode er estimatene og usikkerhetsanalysene i KS2-rapportene?

Concept-rapport nr. 59

Concept-rapport nr. 59

Estimering av kostnader i store statlige prosjekter: Hvor gode er estimatene og usikkerhetsanalysene i KS2-rapportene?

Morten Welde

Norges teknisk-naturvitenskapelige universitet

Magne Jørgensen

Simula Metropolitan og OsloMet - storbyuniversitetet

Per Fridtjof Larssen

Norges teknisk-naturvitenskapelige universitet

Torleif Halkjelsvik

Simula Metropolitan

ISSN: 0803-9763 (papirversjon)

ISSN: 0804-5588 (nettversjon)

ISBN: 978-82-93253-81-5 (papirversjon)

ISBN: 978-82-93253-82-2 (nettversjon)

RETTIGHETSHAVER

© Forskningsprogrammet Concept

Publikasjonen kan siteres fritt med kildeangivelse.

SAMMENDRAG: Rapporten presenterer en oppdatert oversikt over kostnadskontroll i store statlige prosjekter som har vært gjennom KS2. Den gjør i tillegg en del statistiske analyser av kostnadsestimatene i KS2-rapportene.

DATO: Desember 2019

UTGIVER: Ex ante akademisk forlag

Concept-programmet

Norges teknisk- naturvitenskapelige universitet

7491 NTNU – Trondheim

www.ntnu.no/concept

Ansvar for informasjonen i rapportene som produseres på oppdrag fra Concept-programmet ligger hos oppdragstaker. Synspunkter og konklusjoner står for forfatternes regning og er ikke nødvendigvis sammenfallende med Concept-programmets syn. Concept-rapportserie er godkjent som vitenskapelig publiseringskanal på Nivå 1. Alle bidrag kvalitetssikres av uavhengige fagfeller.

Concept-rapportserien

Forskningsprogrammet Concept er forankret ved NTNU og arbeider med forskning knyttet til utviklingen og kvalitetssikringen av store investeringsprosjekter i Norge. Dette er tverrfaglig forskning innenfor fagområdene prosjektledelse, offentlig finansiering, statsvitenskap, samfunnsøkonomisk analyse og evaluering. Rapportserien presenterer forskningsresultater på programmets fagområder og er godkjent som vitenskapelig publiseringskanal på nivå 1. Målgruppen omfatter primært forskere på respektive fagområder og fagpersoner i offentlig forvaltning og utredningsmiljøer.

Redaksjon

Knut Samset, professor, NTNU, redaktør

Gro Holst Volden, forskningssjef Concept

Morten Welde, seniorforsker, NTNU

Redaksjonsråd

Tom Christensen, professor, Universitetet i Oslo

Petter Næss, professor, Norges miljø- og biovitenskapelige universitet

Nils Olsson, professor, NTNU

Ingeborg Rasmussen, styreleder, Vista Analyse

Jørn Rattsø, professor, NTNU

Tore Sager, professor emeritus, NTNU

Heidi Ulstein, partner, Menon Economics

Vibeke Binz Vallevik, gruppeleder, DNV GL

Bjørn Otto Elvenes, førsteamanuensis, NTNU

Forord

Forskningsprogrammet Concept driver forskning på store statlige prosjekter som har vært gjennom ordningen med ekstern kvalitetssikring av konseptvalg (KS1) og av kostnadsoverslag og styringsunderlag (KS2). Et av våre forskningsområder er studier av kostnadskontroll, herunder mulige årsaker til kostnadsoverskridelser.

Den siste større studien av dette temaet ble utgitt for to og et halvt år siden (Concept rapport nr. 51). Selv om utvalget av prosjekter som har vært gjennom KS2 og hvor en endelig sluttkostnad er tilgjengelig ikke har økt med flere enn sju siden da, har vi likevel valgt å publisere en ny Concept rapport basert på samme overordnede tema. I tillegg til at det er behov for jevnlige og oppdaterte studier av kostnadskontroll i disse prosjektene så er hensikten med denne studien å få økt kunnskap om reell usikkerhet i prosjektene gjennom dypere studier av estimatene/usikkerhetsanalysene som ligger til grunn for Stortingets fastsettelse av styrings- og kostnadsrammer. Dette er med andre ord både en studie av kostnadskontroll og av estimatene i KS2-rapportene.

For innhenting av sluttkostnader i form av prosjektrekskap, sluttrapporter, med mer er vi avhengige av bistand og velvillighet fra departementene og de statlige etatene som er ansvarlige for prosjektene. Vi takker som vanlig våre kontaktpersoner der. En referansegruppe bestående av Olav Torp fra institutt for bygg- og miljøteknikk ved NTNU og Kjetil Hove fra Forsvarets forskningsinstitutt ga råd ved oppstart av studien. Til sist vil vi også takke Stein Berntsen fra Dovre Group AS og Vibeke Binz Vallevik fra DNV GL for verdifulle innspill.

Trondheim, desember 2019

Knut Samset

Programansvarlig, Concept programmet, NTNU Trondheim

Innhold

SAMMENDRAG	4
SUMMARY	9
1 BAKGRUNN	13
2 TIDLIGERE STUDIER	15
2.1 TIDLIGERE STUDIER AV PROSJEKTER SOM HAR VÆRT GJENNOM KS2	15
2.2 ANDRE STUDIER AV KOSTNADSESTIMERING AV OFFENTLIGE PROSJEKTER.....	17
2.3 STUDIER AV USIKKERHETSANALYSER.....	19
3 DATA	21
3.1 PROSJEKTENE SOM INNGÅR I ANALYSEN.....	21
3.2 PROSJEKTDATA.....	22
4 FORSKNINGSPØRSMÅL OG METODE.....	24
4.1 FORSKNINGSPØRSMÅL.....	24
4.2 SANNSYNLIGHETSBASERTE KOSTNADSESTIMATER.....	25
4.3 EVALUERING AV SANNSYNLIGHETSBASERTE ESTIMATER.....	27
<i>Terminologi</i>	<i>27</i>
<i>Evaluering av sannsynlighetsbaserte punktestimater.....</i>	<i>28</i>
<i>Evaluering av prediksjonsintervaller og estimerte utfallsfordelinger.....</i>	<i>32</i>
5 STYRINGS- OG KOSTNADSRAMMER.....	36
5.1 OVERHOLDELSE AV STYRINGS- OG KOSTNADSRAMMER	36
5.2 UTVIKLING AV ESTIMATSKJEVHET OG AVVIK OVER TID	40
6 ESTIMATENE I KS2-RAPPORTENE	45
6.1 P50 SOM PUNKTESTIMAT: AVVIK OG SKJEVHET.....	46
6.2 PREDIKSJONSINTERVALL OG ESTIMATFORDELINGER: KALIBRERING OG INFORMASJONSVERDI	49

6.3	FORSKJELLER MELLOM ETATER, KS-RÅDGIVERE, SAMMENHENG MED PROSJEKTSTØRRELSE, OG UTVIKLING OVER TID	57
6.4	HVOR GODE ER ESTIMATENE MOT EN ENKLERE REFERANSEMODEL?.....	67
7	AVSLUTTENDE OPPSUMMERING	70
	REFERANSER	74
	VEDLEGG: LISTE OVER PROSJEKTER BENYTTET I STUDIEN	79

Sammendrag

Ordningen med ekstern kvalitetssikring av store statlige investeringsprosjekter (KS-ordningen / statens prosjektmodell) har blant annet som formål å sikre at budsjettrammer er realistiske og at usikkerhetsvurderingene av kostnadsestimatene gjenspeiler faktisk kostnadsusikkerhet. I hvilken grad rammer, estimater og usikkerhetsvurderinger er realistiske, og hvor det ligger forbedringsmuligheter er hovedtema for denne studien.

Kapittel 1 beskriver bakgrunnen og motivasjonen for denne studien. Utgangspunktet er at forskningsprogrammet Concept fortløpende samler inn sluttkostnader i prosjekter som har vært gjennom KS2. Det gir grunnlag for studier av utvikling i kostnadskontroll. Etter hvert som utvalget av prosjekter øker, kan det også gi grunnlag for dypere undersøkelser av realismen i estimatene som lå til grunn for Stortingets fastsettelse av styrings- og kostnadsramme.

Studien har tre hovedtema. Vi ser på:

1. Realismen i styrings- og kostnadsrammer
2. Realismen til punktestimater i KS2, og
3. Realisme og informasjonsverdi til prediksjonsintervallene og estimatfordelingene

Kapittel 2 gir en gjennomgang av tidligere studier av kostnadskontroll i prosjekter som har vært gjennom KS2. De viser alle relativt gode resultater både med hensyn til budsjettavvik og usikkerhetsvurderinger. Mens gjennomsnittlig kostnadsoverskridelse rapportert i internasjonale studier typisk ligger rundt 30 prosent, så rapporterer de tidligere norske studiene gjennomsnittlige overskridelser på mellom to og seks prosent målt mot styringsramme. Tilsvarende rapporterer andre undersøkelser gjerne sterk undervurdering av usikkerhet, mens P50 og P85-estimatene fra KS2 (det vil si estimater som forventes å ikke overskrides i henholdsvis 50 og 85 prosent av tilfellene) har vært rimelig godt kalibrerte, selv om flere forfattere har pekt på at spredningen er noe høyere enn det man la til grunn på beslutningstidspunktet.

Datagrunnlaget, som beskrives i Kapittel 3, omfatter et større antall prosjekter enn tidligere undersøkelser, samt at analysene i større grad fokuserer på estimatene som brukes som grunnlag for budsjetttrammene. Analysen av styringsramme og kostnadsramme omfatter henholdsvis 83 og 85 prosjekter. Tilstrekkelig data for vår analyse av kostnadsestimatene fantes for 70 av disse prosjektene.

I Kapittel 4 beskriver vi detaljerte forskningsspørsmål og metode for analysene. I dette motiverer og angir vi, basert på nyeste forskning på området, hvordan sannsynlighetsbaserte kostnadsestimater bør evalueres.

Vi innfører en analyse av estimatavvik og estimatskjevhet basert på hva som er en rimelig «tapsfunksjon», der tapsfunksjonen er det som forsøkes minimert når estimatet gis. En slik analyse som tar hensyn til at feilen man forsøker å minimere ved estimering samsvarer med feilen man blir evaluert mot er viktig for å få en riktig vurdering av hvor gode estimatene er. Tilsvarende beskrives målestørrelser for usikkerhetsvurderinger basert på kalibrering og hvor informative prediksjonsintervall og estimatfordelinger er. Vi argumenterer for at godt kalibrerte sannsynlighetsbaserte estimater (f.eks. at ca. 50 prosent av P50-estimatene ikke overskrides) ikke er et tilstrekkelig evalueringskriterium. I tillegg trenger vi målestørrelser for hvor informative de sannsynlighetsbaserte estimatene er.

I analysene av styrings- og kostnadsramme i Kapittel 5 finner vi at medianen av avvikene mellom faktiske kostnader og styringsramme, målt som absolutt prosentvis avvik (uavhengig av fortegn), er på 10 prosent (gjennomsnitt = 12,5), og at median overskridelse av styringsrammene er på 1,5 prosent (gjennomsnitt = 3,4 prosent). Det er med andre ord, for alle prosjektene over ett, kun en svak tendens til overskridelser, og mye lavere enn rapportert i internasjonale undersøkelser. Over tid har det imidlertid vært en noe bekymringsfull utvikling. Mens det tidligere var en tendens til kostnadsunderskridelser (i gjennomsnitt 6 prosent underskridelse av styringsramme for prosjekter med beslutningsår i perioden 2001-2003) har det i de senere vært en tendens til kostnadsoverskridelser (i gjennomsnitt 12 prosent overskridelser i perioden 2010-2012).

Styringsrammene baserer seg på P50-estimer, mens kostnadsrammene baserer seg normalt på P85-estimer. Dette vil si, gitt godt kalibrerte estimater, at faktisk kostnad bør være under styringsrammen i omtrent 50 prosent og under kostnadsrammen i omtrent 85 prosent av tilfellene. Vi finner

imidlertid at dette kun gjelder i 40 prosent av tilfellene for styringsrammen og 73 prosent for kostnadsrammene. Andelene har vært synkende over tid. Mens det i 2001-2003 var 62 og 100 prosent innenfor henholdsvis styrings- og kostnadsramme, så var det i 2010-2012 kun 21 og 43 prosent innenfor, riktignok basert på et lavere antall prosjekter enn i tidsperiodene før. Grunnen til at treffrater for styrings- og kostnadsramme for alle prosjektene under ett ikke er så langt fra de intenderte er altså i stor grad at man har gått fra for høye styrings- og kostnadsrammer, til for lave. Tendensen til for lave styrings- og kostnadsrammer bør forsøkes snus gjennom bedre estimeringsarbeid i fremtidige KS-prosjekter, eventuelt bedre styring av prosjektene i gjennomføringsfasen.

Analysen av estimatene i Kapittel 6 finner omtrent samme grad av overskridelser og estimatavvik for P50 og P85-estimatene som de rapportert for styrings- og kostnadsrammene. P50 estimatene viste en median estimatskjevhet på -1 prosent (gjennomsnitt = 3 prosent). Median prosentvis avvik (uavhengig av fortegn) var på 12 prosent (gjennomsnitt = 14 prosent). Vi beregnet at det forventede avviket ikke kan bli mindre enn 8-10 prosent, gitt en del antagelser som at styringsrammene ikke påvirker prosjektene til å tilpasse leveranser for å redusere avvik. Selv om sistnevnte antagelse neppe er oppfylt, antyder denne beregningen at avvikene ikke er spesielt høye.

Vi observerer at det typisk skjer en reduksjon fra estimat til ramme. Styringsrammen var i gjennomsnitt sju prosent lavere enn P50-estimatet og kostnadsrammen sju prosent lavere enn P85-estimatet. Til tross for at det var flere prosjekter som burde ha beholdt de opprinnelige P50 og P85-estimatene som henholdsvis styrings- og kostnadsramme, så fant vi ikke at justeringene totalt sett reduserte realismen. Mange av justeringene synes å være godt begrunnet.

Estimatene gitt i KS2 omfatter både punktestimater, prediksjonsintervall og estimatfordelinger (S-kurver). Analysene våre omfatter alle disse og har som hovedfunn at:

- Estimatifordelingene og prediksjonsintervallene er typisk for smale for å reflektere faktisk usikkerhet. Det er for eksempel slik at hele 19 prosent av prosjektene har lavere kostnad enn P10-estimatet og 20 prosent mer enn P90-estimatet. Fremtidig estimering bør ta inn over seg at utfallsrommet for prosjektkostnader er bredere enn det som tidligere typisk har vært antatt.

- Estimert kostnadsusikkerhet, anslått gjennom bredden på prediksjonsintervall og estimatfordeling, korrelerer ikke med faktisk kostnadsusikkerhet, målt gjennom kostnadsavvik og overskridelser. Dette tyder på lav evne til å skille mellom prosjekter med høy og lav kostnadsusikkerhet. En bedre evne vil blant annet kunne redusere behovet for påslag fra styrings til kostnadsramme, uten at det går ut over kostnadskontroll og prosjektgjennomføring. Vi viser, gitt noen antagelser, at kostnadsrammene ville kunne være 17 prosent lavere dersom evnen til å skille mellom lav- og høyrisikoprosjekter hadde vært god. Målinger av og tiltak for å forbedre denne evnen bør prioriteres i estimeringsarbeidet.
- Det er forskjeller i estimeringsprestasjoner både mellom etater og konsulentfirma ansvarlig for kvalitetssikring av estimatene (KS-rådgivere). Forsvarsprosjekter skiller seg ut ved å ha en sterk tendens til å overestimere kostnader (gjennomsnittlig underskridelse av P50-estimatet med 19 prosent) og alt for smale prediksjonsintervaller (29 prosent av prosjektene innenfor 80 prosent prediksjonsintervallet). Statens vegvesen har også en tendens til for smale prediksjonsintervaller (57 prosent av prosjektene innenfor 80 prosent prediksjonsintervallet). Blant KS-rådgiverne er det ingen store forskjeller i estimatavvik, men større forskjeller i hvor realistisk usikkerheten anslås. Det kan være ulikheter i prosjektkompleksitet eller andre forhold som forklarer disse forskjellene.
- Gitt den manglende evnen til å skille mellom lav- og høyrisikoprosjekter i estimeringsarbeidet, vil en enkel, mekanisk påslagsmodell (klimatologisk modell) i teorien kunne gjøre det like godt som det mer krevende KS2-estimeringsarbeidet. Vi undersøkte dette, der påslagene var basert på historisk estimatavvik, men fant at KS2-estimatene gjorde det noe bedre på de fleste målestørrelsene. Dette indikerer at arbeidet som gjøres i KS2-estimeringen gir merverdi, målt mot enkle påslagsmodeller.

Vi oppsummerer og diskuterer funnene i Kapittel 7. Hovedkonklusjonene er at totalt sett så synes KS2-rammer og kostnadsestimater å være realistiske og være rimelig godt kalibrerte. Utviklingen over tid er imidlertid bekymringsfull og bør føre til forbedringer i estimeringsarbeidet. To vesentlige forbedringsområder er å angi bredere estimatfordelinger, det vil si å anerkjenne at kostnadsusikkerheten typisk er større enn den som tidligere har

fremkommet i estimeringsarbeidet, samt å bli bedre på å skille mellom prosjekter med lav og med høy kostnadsusikkerhet.

Summary

The external quality assurance scheme for large government investment projects (the QA scheme / the state project model) aims, among other things, to ensure that budgets are realistic and that the risk analyses of the cost estimates reflect real cost uncertainty. The extent to which budgets, estimates and risk analyses are realistic, and where there may be potentials for improvements, are the main themes of this study.

Chapter 1 describes the background and motivation for the study. The starting point is that the Concept research programme collects final costs in projects that have been through QA2 (quality assurance of cost estimate before the parliament's investment decision). That provides a basis for studies of cost performance. As the sample of projects increase, more detailed studies of the estimates that formed the basis for the parliament's investment decision becomes possible.

The study has three main topics. We look at:

1. The realism in the projects' budgets
2. The realism in the point estimates in the QA2 reports, and
3. The realism and information value in the prediction intervals and estimation distributions.

Chapter 2 provides a review of previous studies of cost performance in projects that have been through QA2. They all show relatively good results both in terms of deviation from budgets and risk assessments. While average cost overruns reported in international studies typically have been around 30 per cent, Norwegian studies report average overruns of between two and six per cent. Other studies also typically report a strong underestimation of uncertainty. The P50 and P85 estimates from the QA2 reports on the other hand (that is, estimates that are not expected to be exceeded in 50 and 85 per cent of cases, respectively) seem to have been reasonably well calibrated. However, several authors have pointed out that the distribution of final costs to the budgets have been somewhat higher than assumed at the time of the investment decision.

The data used in the study, which is described in Chapter 3, is based on a larger sample of projects than previous studies. The analyses focus more on the estimates than previous studies have done. The analysis of the P50 and P85 estimates is based on samples of 83 and 85 projects respectively. Sufficient data for our analysis of the cost estimates were found for 70 of these projects.

In Chapter 4, we outline detailed research questions and the methodology for the analyses. In this, we motivate and indicate, based on the latest research on the area, how probability-based cost estimates should be evaluated.

We introduce an analysis of estimate deviations and estimation bias based on what is a reasonable "loss function", where the loss function is what we attempt to minimise in the estimates. We evaluate the extent to which we have been successful in estimating the real uncertainty of projects *ex ante*. We also assess how informative prediction intervals and estimate distributions have been. We argue that well-calibrated probability-based estimates (e.g., that 50 per cent of P50 estimates should not be exceeded) are not a sufficient evaluation criterion. In addition, we need indicators for how informative the probability-based estimates have been.

In Chapter 5, we find that the median deviations between actual costs and the P50, measured as absolute percentage deviation, is 10 per cent (mean = 12.5), and that the median deviation from the P85 is 1.5 per cent (average = 3.4). In other words, for all the projects, there is only a slight tendency for overruns, and much lower than what has been reported in international studies. Over time, however, there has been a somewhat worrying development. While there was a tendency for cost underruns in the past (an average of 6 per cent underruns of the P50 for projects with an investment decision between 2001 and 2003), there has been a tendency for cost overruns in the later years (an average of 12 per cent overruns in the period 2010-2012).

Given well-calibrated estimates, the actual cost should be below the P50 in about 50 per cent of the cases and below the P85 in about 85 per cent of the cases. However, we find that this only applies in 40 per cent of the cases for the P50 and 73 per cent for the P85. The shares have been declining over time. While in 2001-2003, 62 and 100 per cent were within the P50 and P85 respectively, in 2010-2012 there were only 21 and 43 per cent within, albeit based on a smaller sample than in the time-periods before. The reason why hit rates for the P50 and the P85 for all projects together are not so far from the

intended targets is because we have gone from overestimation to underestimation. The tendency for underestimation should be reversed through better estimation and governance in future projects.

The analyses of the estimates in Chapter 6 find about the same degree of overruns and estimate deviations for the P50 and P85 estimates as those reported in Chapter 5. The P50 estimates showed a median estimate bias of -1 per cent (mean = 3 per cent). The median percentage deviation (regardless of sign) was 12 per cent (mean = 14 per cent). We calculated that the expected deviation from the P50 budget could not be less than 8-10 per cent, given some assumptions that the projects do not adapt deliveries to reduce deviations. Although the latter assumption hardly is met, this calculation suggests that the deviations are not particularly high.

We observe that there is typically a reduction from estimate to budget. The P50 budget was on average seven per cent lower than the P50 estimate and the P85 budget seven per cent lower than the P85 estimate. Although there were several projects that should have retained the original P50 and P85 estimates as P50 and P85 budget, respectively, we did not find that the adjustments overall reduced the realism. Many of the adjustments seem to be well justified.

The estimates in the QA2 reports include both point estimates, prediction intervals and estimate distributions (S-curves). Our analyses include all of these and have as their main findings are as follows:

- The estimation distributions and prediction intervals are typically too narrow to reflect actual uncertainty. For example, as many as 19 per cent of the projects have a lower cost than the P10 estimate and 20 per cent more than the P90 estimate. Future estimation should take into account that the scope for project costs is broader than previously typically assumed.
- Estimated cost uncertainty, estimated through the width of the prediction interval and estimate distribution, does not correlate with actual cost uncertainty, measured by cost deviations and overruns. This indicates a low ability to distinguish between projects with high and low cost uncertainty. If we become better at identifying the high-risk projects, we could potentially reduce the need for risk contingency without compromising cost performance and project execution. We show, given some assumptions, that the P85 could be

17 per cent lower if the ability to distinguish between low and high risk projects had been better. Measures to improve this capability should be given priority in the estimation work.

- There are differences in estimation performance between agencies and between consultancies carrying out the external QA. Defence projects stand out by having a strong tendency to overestimate costs (their average underrun of the P50 estimate is 19 per cent) and overly narrow prediction intervals (29 per cent of projects within the 80 percent prediction range). The Norwegian Public Roads Administration also tends to estimate too narrow prediction intervals (57 per cent of projects within the 80 per cent prediction interval). Among the QA consultancies, there are no major differences in estimate deviations, but larger differences in how realistic the uncertainty is estimated. There may be differences in project complexity or other issues that explain these differences.
- Given the inability to distinguish between low- and high-risk projects in the estimation work, a simple mechanical mark-up model could in theory do just as well as the more demanding QA2 estimation work. We investigated this, where the uplifts were based on historical estimate deviations, but found that the QA2 estimates did better. This indicates that the work done in the QA2 estimation provides added value, measured against simple mark-up models.

In Chapter 7, we summarize and discuss the findings. Overall, the main conclusions are that the QA2 framework is useful and that cost estimates appear to be realistic and reasonably well calibrated. However, developments over time are worrying and should lead to improvements in the estimation work. Two major areas of improvement are to specify broader estimate distributions, that is, to recognize that cost uncertainty is typically greater than that which has previously been identified in the estimation work, as well as to better distinguish between projects with low and high cost uncertainty.

1 Bakgrunn

Ekstern kvalitetssikring av styringsunderlag og kostnadsoverslag, KS2, ble innført som forbedringstiltak etter en rekke store kostnadsoverskridelser i statlige investeringsprosjekter på 1980- og -90-tallet. Hensikten med innføring av ordningen var å bedre kontrollen og styringen med kostnadene i statlige prosjekter. KS2 innebærer en kvalitetssikring av prosjektenes kostnadsestimater før investeringsbeslutning, samt av styringsunderlaget for prosjektgjennomføringen. Senere har ordningen blitt utvidet med ytterligere ett kvalitetssikringspunkt i tidligfasen, KS1, hvor ekstern kvalitetssikring skal bistå med å gi et faglig og uavhengig grunnlag for den politiske beslutningen om et konseptvalg.

Forskningsprogrammet Concept driver følgeforskning på prosjektene som har vært gjenstand for ekstern kvalitetssikring. For å vurdere hvor vellykket et prosjekt har vært, gjennomfører Concept både brede etterevalueringer og nærmere studier av konkrete suksesskriterier i prosjekter. Det viktigste er at prosjektene leverer de tiltenkte effekter for brukere og samfunn og om de svarer på et langsiktig strategisk behov også i fremtiden, men det er selvsagt også viktig at budsjetter og kostnadsestimater er realistiske. Blir utbyggingskostnadene høyere enn det man har budsjettet med, kan det gå ut over fremdrift ved at man blir avhengig av ekstrabevilgninger undervegs eller det kan gå ut over finansieringen av andre prosjekter. Kostnadsoverskridelser kan også true lønnsomhet, og ikke minst kan det representere et omdømmeproblem som setter den ansvarlige organisasjonen i et dårlig lys. Også for høye estimater og budsjetter kan ha uheldige effekter. Det bindes opp mer budsjettmidler enn strengt tatt nødvendig, samt at for høye estimater og budsjetter kan føre til lavere produktivitet (Nan og Harter, 2009).

Etter innføringen av ordningen med ekstern kvalitetssikring av kostnadsanslag og styringsunderlag, KS2, i 2000 og deretter av konseptvalg, KS1, i 2005 er store statlige investeringsprosjekter gjennom omfattende utredninger før Stortinget kan fastsette et endelig budsjett og faktisk prosjektgjennomføring kan starte. Utbyggings- eller anskaffelseskostnad fastsettes gjennom sannsynlighetsbasert, eller stokastisk kostnadsestimering. Det medfører at kostnaden kan angis som en sannsynlighetsfordeling som ideelt sett representerer den reelle usikkerheten i prosjektet (for en nærmere beskrivelse

av ordningen med KS2 og kostnadsestimering under usikkerhet, se Forskningsprogrammet Concept (2019) og Drevland (2013)). Antall prosjekter som har vært gjennom KS2 og som er ferdigstilt øker hvert år. Med jevne mellomrom gjør Concept derfor studier av kostnadskontrollen i prosjektene og utviklingen over tid. Etter hvert som utvalget av prosjekter øker, gir det også rom for grundigere studier av både kostnadskontroll og av estimatene som kostnadsrammene har vært basert på.

Denne studien ser på to nært tilgrensende tema. Vi gjennomfører først en oppdatert studie av i hvilken grad prosjekter som har vært gjennom KS2 overholder Stortingets vedtatte styrings- og kostnadsrammer med noen flere prosjekter og et bredere sett av analyser enn da dette sist ble studert (Welde, 2017). Vi gjør også en evaluering av selve kostnadsestimaterne og usikkerhetsanalysene i KS2-rapportene som lå til grunn for Stortingets endelige budsjettvedtak. Vi forsøker med det å gå bak resultatene i prosjektene ved å se hvor gode estimatene er ved endelig investeringsbeslutning. Blant annet ser vi på evnen til å skille mellom prosjekter med høy og lav usikkerhet. Hensikten med studien er både å få økt kunnskap om reell usikkerhet i prosjektene, og å identifisere mulige forbedringer ved dagens praksis.

Rapporten er organisert som følger: I Kapittel 2 gir vi en oversikt over tidligere studier relevante for innholdet i denne rapporten. Vi viser både til studier av kostnadskontroll i prosjekter som har vært gjennom KS2, og til noen internasjonale studier av relevans for denne rapporten. Kapittel 3 beskriver datagrunnlaget for studien og i Kapittel 4 formulerer vi forskningsspørsmålene som skal besvares i de kommende kapitlene. Kapittel 4 inneholder også en gjennomgang av metode for evaluering av sannsynlighetsbaserte estimater. I Kapittel 5 inneholder en analyse av styrings- og kostnadsrammer. Vi gjør med andre ord en oppdatert analyse av kostnadskontroll i prosjekter som har vært gjennom KS2, og supplerer med noen nye analyser. I Kapittel 6 analyserer vi estimatene i KS2-rapportene med bruk av metoden beskrevet i Kapittel 4. Til slutt, i Kapittel 7, oppsummerer vi resultatene og kommer med noen anbefalinger til forbedret praksis.

2 Tidligere studier

2.1 Tidligere studier av prosjekter som har vært gjennom KS2

Concept har over årene gjennomført flere studier av i hvilken grad prosjektene som har vært gjennom KS2 har blitt gjennomført innenfor henholdsvis kostnadsramme og styringsramme. Concept rapport nr. 35 var basert på en studie av Aass (2013) og presenterte en gjennomgang av kostnadsestimater og kostnadstall i de første 40 KS-prosjektene som var gjennomført og hvor det forelå tall for sluttkostnad (Aass, 2013). Det gjaldt prosjekter gjennomført i perioden 2000 til 2012. Stortingets vedtatte kostnadsramme for prosjekter tilsvarer som regel (men se analyser senere i rapporten av justeringer fra estimat til ramme) et budsjett som skal overholdes med 85 prosent (P85) sannsynlighet, og styringsrammen tilsvarer som regel budsjett som skal overholdes med 50 prosent (P50) sannsynlighet. Resultatene viste at 80 prosent av de første 40 prosjektene var blitt gjennomført til en kostnad lavere enn kostnadsrammen og 45 prosent innenfor styringsrammen. Studien konkluderte dermed at forutsigbarheten og kontrollen med kostnadene i de 40 første ferdigstilte kvalitetssikrede prosjektene hadde vært god.

Welde (2014b) gjorde en oppfølgende studie året etter. Utvalget var da økt til 51 prosjekter. Studien viste at 75 prosent av prosjektene hadde en sluttkostnad på eller innenfor kostnadsrammen og 45 prosent av prosjektene hadde sluttkostnader innenfor styringsrammen. Året etter, med et utvalg på 67 prosjekter, fant Welde (2015) at andelen prosjekter under kostnadsrammen var 79 prosent og andelen under styringsrammen var 48 prosent. Den siste studien av kostnadskontroll i prosjekter som har vært gjennom KS2 er den av Welde (2017). Med et utvalg på 78 prosjekter fant han at om lag 77 prosent av prosjektene hadde sluttkostnader under kostnadsrammen og 48 prosent under styringsrammen. Tabell 2-1 oppsummerer resultatene fra de ulike studiene. Avvikene er målt som differansen mellom kostnader og budsjett, delt på budsjett, slik at positive verdier tilsvarer gjennomsnittlige kostnadsoverskridelser.

Tabell 2-1: Ulike studier av overskridelser i prosjekter som har vært gjennom KS2

Studie	Utvalg	Styringsramme	Kostnadsramme	Andel under kostnadsramme	Andel under styringsrammen
Aass (2013)	40	+2,1 %	-7,6 %	80 %	45 %
Welde (2014)	51	+6,0 %	-4,4 %	75 %	45 %
Welde (2015)	67	+2,8 %	-6,9 %	79 %	48 %
Welde (2017)	78	+2,3 %	-6,9 %	77 %	48 %

Gjennomsnittlig har faktiske kostnader havnet ca. 7 prosent under kostnadsrammen og 2-3 prosent over styringsrammen. Unntaket kan se ut til å være studien fra 2014 som ga noe høyere overskridelser. Det kan skyldes at denne studien inneholdt en relativt høyere andel prosjekter som ble igangsatt i årene før finanskrisen i 2008. Welde (2017) viste at det var en noe større andel overskridelser blant prosjekter som ble igangsatt i denne perioden enn prosjekter igangsatt før og etter.

Kostnadsrammen settes, som tidligere beskrevet, normalt til det som er estimert som P85, ofte med fratrukk for mulige mindre forenklinger og reduksjoner (kuttliste) som kan gjennomføres underveis i prosjektet dersom rammen skulle stå i fare for å overskrides. Det innebærer at ideelt sett så burde omtrent 85 prosent av prosjektene gjennomføres med en sluttkostnad under kostnadsrammen. Som Tabell 2-1 viser så er ikke den målsetningen fullt ut oppnådd, med 77 prosent (60 av 78 prosjekter) under kostnadsrammen. Med andre ord ser det ut som at kostnadsrammene typisk settes noe lavt, i forhold til hvordan prosjektene har vært gjennomført.

Styringsrammen, som oftest disponeres av utførende etat, settes altså normalt til P50-estimatet eller noe i nærheten, hvilket innebærer at i en portefølje av prosjekter burde om lag halvparten gjennomføres under og resten over styringsrammen. Tabell 2-1 indikerer at prosjektene sett under ett var svært nær målsetningen, med for eksempel 48 prosent (37 av 78 prosjekter) under styringsrammen i den siste undersøkelsen.

Ser man på estimerte kostnader i tidligfasen før KS, hvor prosjektene ikke er fullt utviklet, viser disse et mer negativt bilde i form av flere store kostnadsøkninger (Welde, 2014a; Welde, 2017, Aspenberg Jordal, 2019). Det er også verdt å merke seg at dersom man ikke hadde justert styringsrammen eller estimatene med de riktige kostnadsindeksene (justerer for redusert kjøpekraft per estimert krone for de påfølgende årene) vil flere prosjekter vise kostnadsoverskridelser (Welde, 2014c).

I tillegg til å dokumentere forholdet mellom budsjetter og faktiske kostnader, har Concept gjort en rekke studier av metodegrunnlaget som benyttes i usikkerhetsanalyser, se Austeng m.fl. (2005a), Drevland m.fl. (2005), Austeng m.fl. (2005b), Austeng m.fl. (2005c), Drevland (2013) og Torp m.fl. (2015).

2.2 Andre studier av kostnadsestimering av offentlige prosjekter

Concept rapport nr. 51 (Welde, 2017) inkluderte en omfattende oversikt over tidligere studier på kostnadsoverskridelser i offentlige prosjekter i Norge og andre land. Nesten alle studiene viste gjennomsnittlig overskridelser, og overskridelser i størrelsesordenen 30 prosent og høyere var ikke uvanlig (se Welde, 2017, Tabell 2-1). Det kan dermed se ut som om situasjonen for KS-prosjektene i Norge er bedre enn for tilsvarende prosjekter i andre land. Å sammenligne på tvers av land kan imidlertid være problematisk med hensyn til beslutningstidspunkt, estimeringsmetodikk og praksis for prisomregning. For eksempel blir kostnadsrammene i statlige prosjekter i Norge vedtatt på et tidspunkt da prosjektenes modenhet er høy – i veg- og jernbaneprosjekter etter godkjent kommunedelplan og normalt også reguleringsplan. Mange interessenter vil imidlertid anse at investeringsbeslutningen er tatt lenge før kostnadsrammen blir vedtatt. En sen fastsettelse av endelig budsjett kan være fornuftig hvis hensikten er å unngå kostnadsoverskridelser, men hvis den reelle investeringsbeslutningen har blitt tatt på et tidligere tidspunkt så burde man kanskje heller ha sammenliknet sluttkostnaden med et kostnadsanslag utarbeidet på et tidligere tidspunkt. Frem til nå har sannsynligvis verken ansvarlige etater eller departement oppfattet tidlige kostnadsestimat som bindende slik at kostnadsøkninger i tidligfasen har blitt sett på som relativt normal praksis.

Årsaker til overskridelser

Den overveldende dokumentasjonen av store overskridelser i offentlige prosjekt har ført til flere spekulasjoner angående mulige årsaker. Welde (2017) gjennomgikk litteraturen og diskusjonene rundt de forklaringene som har fått mest oppmerksomhet. Bent Flyvbjerg og kollegaer (se for eksempel Flyvbjerg m.fl., 2018) har påstått at overskridelser i hovedsak skyldes taktisk underestimering, et begrep som egentlig betyr å lyve for å få vedtatt prosjekter man ønsker å få gjennomført. Denne påstanden har blitt kritisert fra flere hold (se for eksempel Love m.fl., 2018, og Halkjelsvik og Jørgensen, 2018).

En annen forklaring på tendensen til overskridelser er at prosjektet som er estimert ikke er det samme som blir gjennomført, altså at omfangsendringer og mangler i kontraktsgrunnlaget fører til et mer omfattende prosjekt enn det som lå til grunn for budsjettet. Flere studier internasjonalt (Love m.fl., 2015; Welde, 2017) viser at dette kan være en viktig faktor ved større overskridelser.

En tredje potensiell grunn til overskridelser er relatert dynamikken mellom estimater/budsjetter og prosjektgjennomføringen. Prosjektleder kan ha få insentiver til å ende opp med lavere kostnader enn budsjettet, så uten aktiv involvering av prosjekteier vil man bruke de midlene man har til rådighet (Parkinson's prinsipp). I en portefølje av prosjekter vil dermed prosjektene som i utgangspunktet ville ha endt opp med underskridelser ende opp med kostnader nært budsjett og derfor ikke kunne kompensere for prosjekter med store overskridelser (Austeng, 2005; Welde, 2017).

En fjerde potensiell grunn kan knyttes til evalueringsmetodene. Siden P50-estimatene ikke reflekterer forventningsverdiene (gjennomsnittsverdiene) i skjevfordelte estimater, vil man derfor kunne observere overskridelser dersom man ser på gjennomsnittet eller summen av flere prosjekter. I tildelingsbrevet til Statsbygg (Kommunal og Moderniseringsdepartementet, 2018) står det for eksempel: «Samlet sluttkostnad for porteføljen av ferdigstilte prosjekter siste fem år skal ikke overstige samlet styringsramme». Men slike antagelser passer bedre til estimater som representerer forventet gjennomsnitt (forventningsverdi) enn P50-estimatene som styringsrammene typiske er basert på.

Andre problemer kan ligge i selve estimeringsmetodene, som feil ved innhenting av materialkostnader (Kaming m.fl., 1997), feil i aggregering av estimat (Halkjelsvik og Jørgensen, upublisert) eller mangel på kunnskap om kompleksiteten i prosjektet og eksterne forhold som tilgjengelig arbeidskraft (Jergeas og Ruwanpura, 2010). En faktor som har blitt trukket frem i litteraturen er at man ikke i tilstrekkelig grad tar hensyn til prosjekters usikkerhet (Flyvbjerg m.fl., 2003). Dette vil kunne påvirke nøyaktigheten til punktestimater som blir brukt som grunnlag for budsjetter, men kanskje i enda større grad nøyaktigheten til prediksjoner av ytterpunktene (halene) av de estimerte utfallsfordelingene, slik som P90 og P10, noe vi kort gjennomgår litteraturen på i de neste avsnittene.

2.3 Studier av usikkerhetsanalyser

Dersom vi ser bort fra studier av hvor stor andel av prosjektene som går over og under (treffrate) hva som kan betraktes som *ensidige prediksjonsintervaller* (for eksempel P50 og P85-estimer), så er det få studier som har evaluert usikkerhetsanalyser av offentlige prosjekter. Concept rapportene nr. 10-13 gjennomgikk en rekke aspekter ved usikkerhetsanalyser i KS-ordningen, men det ser ikke ut til at man har gjort en kvantitativ vurdering av kvaliteten til usikkerhetsestimaten utover treffrate.

Det finnes derimot en litteratur på estimering av usikkerhet i eksperimentelle beslutnings- og bedømmelsessituasjoner. I denne litteraturen er det typisk mindre oppgaver og ikke store prosjekter som blir estimert. For eksempel fikk deltakere i en studie av Connolly og Dean (1997) i oppgave å vurdere tiden som de var 99 prosent sikre på å ikke overskride for en programmeringsoppgave og minimumstiden de var 99 prosent sikre på at de kom til å bruke. Hele 43 prosent av deltakerne havnet utenfor rammene av dette som skal tilsvare et 98 prosent prediksjonsintervall. Denne andelen skulle ha vært 2 prosent dersom usikkerhetsvurderingene hadde vært velkalibrerte.

Tendensen til å oppgi alt for smale prediksjonsintervaller blir ofte kalt overkonfidens, men det er ikke sikkert at det er en overdreven tro på egne og andres prestasjoner som ligger bak smale prediksjonsintervall/konfidensintervall. Det kan heller se ut som om folk ikke helt klarer å skille mellom ulike nivå av sannsynlighet når man ber om usikkerhetsanalyser. Undersøkelser av Jørgensen m.fl. (2004) og Jørgensen og Teigen (2005) viste at det var liten forskjell mellom estimatene når man ba om for eksempel 99 prosent og 75 prosent prediksjonsintervall. Det så ut til at deltakerne valgte seg et rimelig lavt estimat og et rimelig høyt estimat, nesten uavhengig av sannsynlighetene disse estimatene skulle representere. Det er ikke urimelig å anta at man også i estimeringen av større prosjekter har problemer med usikkerhetsanalyser av elementer ved bruk av trepunktsestimering. Dette vil kunne føre til underestimering av usikkerhet også på aggregert nivå. Teigen m.fl. (2018) finner at det å forstå og bruke sannsynlighetsbaserte kostnadsestimater byr på utfordringer, selv for personer med akademisk utdanning og relevant praksis.

Grunnen til at de fleste oppgir for smale prediksjonsintervall kan være at det føles mer riktig å angi maksimum og minimum som er informative, og ikke så ekstreme at de så og si aldri vil skje. Vi vil ofte ønske å unngå at vi oppfattes som å helgardere oss med det som andre, og vi selv, kan oppfattes som ekstremt vide, usannsynlige og ikke-informative intervaller (se Jørgensen m.fl., 2004).

3 Data

3.1 Prosjektene som inngår i analysen

Utvalget i studien omfatter prosjekter som har vært underlagt krav om ekstern kvalitetssikring i form av KS2, hvor prosjektresultatet er ferdigstilt og sluttoppgjør for kontrakter er avsluttet eller kommet så langt at den endelige investeringskostnaden er kjent. Per desember 2019 inneholder Concepts database Trailbase 200 prosjekter som har vært gjennom KS2. Av disse er nærmere 130 registrert som ferdige og av dem har vi tilgang til troverdige sluttkostnader for 85 prosjekter.

Årsaken til at det er en såpass stor differanse mellom prosjekter som er ferdige og prosjekter hvor det foreligger en sluttkostnad er at det kan ta tid før endelig prosjektrengskap er avsluttet. Selv etter at en veg er åpnet eller et bygg er tatt i bruk, kan det ta tid før sluttoppgjør med entreprenør er ferdig, og det kan være behov for restarbeider eller reklamasjoner. I mange prosjekter vil kostnadspådraget kunne pågå i flere år etter at prosjektet er ferdigstilt og idriftsatt. Dertil kommer at prosjekteier i mange tilfeller ikke stiller krav til når økonomisk sluttrapport skal være ferdig. Eksempelvis pågikk utbyggingen av E6 gjennom Østfold i årene 2001 til 2008, men endelig økonomisk sluttrapport forelå ikke før i 2014 (Ulstein m.fl., 2017). Rv. 13 Hardangerbrua åpnet for trafikk i 2013, men teknisk sluttrapport forelå ikke før høsten 2018 (Wifstad m.fl., 2018). Ole Johan Dahls hus ved Universitetet i Oslo, IFI2, åpnet formelt i 2011. Da prosjektet ble ettervaluert i 2018 forelå det ennå ikke en sluttrapport som anga formell sluttkostnad (Welde m.fl., 2018). Det finnes, etter det vi er kjent med, ingen systematisk oversikt over hvor lang tid det tar for store statlige prosjekter er endelig avsluttet. Det kan derfor være grunn til å behandle sluttkostnader oppgitt umiddelbart etter prosjektåpning med varsomhet.

Det er grunn til å tro at en del av prosjektene som er ferdige, men hvor sluttkostnad enda ikke foreligger, har opplevd utfordringer som medfører at endelig kostnad vil overstige kostnadsrammen i betydelig grad. Et eksempel på det er Statens vegvesens prosjekt for nytt kjøretøyregister, Autosys, som startet i 2002 (den gang under navnet Au2sys). Prosjektets budsjett har vært økt en rekke ganger og pågår fortsatt. Det første formelle budsjettet var 285 millioner

kroner, noe som var under terskelverdien for KS2. I Prop. 1 S (2015–2016) ble det gitt en omtale av det videre arbeidet i prosjektet med en kostnadsramme for hele prosjektet på 1.948 millioner kroner. En mulig svakhet med vårt datasett er derfor at prosjekter som har opplevd store utfordringer og som kan ende opp med store overskridelser ikke er inkludert.

3.2 Prosjektdata

Studien er i hovedsak basert på sammenlikning av faktiske sluttkostnader med styringsrammer, kostnadsrammer, og kostnadsestimater, inkludert estimater fra usikkerhetsanalyser.

I store statlige prosjekter omfattet av ordningen med ekstern kvalitetssikring benytter de ansvarlige etatene stokastisk kostnadsestimering hvor usikkerheten i de ulike elementene i estimatet estimeres gjennom gruppeprosesser. Metodene som benyttes er i all hovedsak i tråd med internasjonal beste praksis (Welde og Torp, 2016). I tillegg gjennomgås etatenes estimater av eksterne eksperter gjennom KS2-prosessen. Kostnadsestimatene gir grunnlaget for fastsetting av styrings- og kostnadsrammene (budsjettene).

For analyser av kostnadsestimatene som ligger bak styrings- og kostnadsrammene har vi benyttet data fra KS2-rapportene og valgt å fokusere på estimatene for P10, P50, P85, og P90. Data er hentet ut av KS2-rapportene slik de foreligger i Concepts database, Trailbase. Kostnadsestimatene i KS2-rapportene er som regel å finne under kapitlet «Usikkerhetsanalyse», KS2-rapporter har og en innledende *superside*, en oppsummering av sentrale verdier fra rapporten. Data er hentet fra kapittelteksten og ikke supersiden, dette for at data skal være så nært det presenterte underlaget som mulig.

KS2-rapportene er i hovedsak like i utforming, men siden det er forskjellige konsulentkonstellasjoner og etater involvert, så forekommer det mindre variasjoner i hvilke verdier som er oppgitt. Samtlige rapporter illustrerer S-kurven (den kumulative sannsynlighetsfordelingen) for kostnadsestimatet; de oppgir P50 og nesten alle oppgir P85-estimer. Noen rapporter stopper der, mens andre oppgir samtlige PX-verdier i intervaller på ti. Der hvor P10, P90 og P85-estimer ikke er oppgitt numerisk, så har vi brukt S-kurven til avlesning. Av 280 verdier (70 prosjekter à 4 verdier) er 81 S-kurver avlest av forfatterne. For noen prosjekter har vi kun data for styrings- og

kostnadsrammene, men mangler kostnadsestimatene fra KS2. Denne forskjellen i datapunkter reflekteres i analysene.

Vi har indeksjustert estimatene og budsjettene med de etatsspesifikke indeksene slik at det blir meningsfylt å sammenligne de med de faktiske sluttkostnadene. Når Stortinget vedtar rammene for et prosjekt, forutsetter man også at rammene blir justert etter en definert indeks. Gjennom en regulering av rammene basert på prisutviklingen innen den relevante sektoren skjermes prosjektene for en del av den systematiske kostnadsusikkerheten knyttet til generell og sektorspesifikk prisvekst. Prisregulering av rammene er en praktisk ordning da indeksbasert kompensasjon for lønns- og prisstigning som oftest er kontraktsfestet med leverandørene, se Welde (2014c, 2017) for en ytterligere beskrivelse av prisjustering i statlige prosjekter.

Grunnlaget for estimatene i en KS2-rapport er en såkalt *grunnkalkyle* for prosjektet. En grunnkalkyle kan defineres som den deterministiske summen av sannsynlig kostnad for alle spesifiserte kostnadsposter på analysetidspunktet. Flere av KS2-rapportene oppgir ikke grunnkalkyle men heller *basisestimat*. Basiskostnad blir gjerne definert som sum av grunnkalkyle og uspesifisert kostnad for alle konkrete poster. *Uspesifisert* er kostnader som man av erfaring vet vil komme, men som ikke er kartlagt på grunn av manglende detaljeringsgrad (Finansdepartementet, 2008). Vi antar at estimatene, og dermed også rammene, blir utledet fra grunnkalkylen eller basiskostnaden gjennom en prosess som tar høyde for usikkerheten i elementene eller prosjektet som helhet. Med dette som utgangspunkt kan vi undersøke om prosessen fra grunnkalkyle/basiskostnad til estimat gir bedre kostnadsprediksjoner enn enkle påslagsmodeller.

4 Forskningsspørsmål og metode

4.1 Forskningsspørsmål

Det overordnede formålet med studien er å vurdere realismen i styrings- og kostnadsrammer vedtatt av Stortinget, og realismen i estimatene fra KS2 som typisk danner grunnlaget for budsjettene, og å se på muligheter til forbedringer. Forskningsspørsmålene brytes ned som angitt under:

Tema 1: Realismen i styrings- og kostnadsrammene

- Hva er avviket mellom styrings- og kostnadsrammer og faktisk kostnad?
- Hvor mye overskrider/underskrider faktisk kostnad rammene?
- Hvor godt er rammene kalibrert, det vil si holder styringsrammen i ca. 50 prosent kostnadsrammen i ca. 85 prosent av prosjektene?
- Er styrings- og kostnadsrammene mer eller mindre realistiske enn deres respektive estimater (P50 og P85-estimer fra KS2)?

Tema 2: Realisme til punkttestimatene i KS2

- Hva er avviket mellom P50-estimat og faktisk kostnad?
- Hvor mye overskrider/underskrider faktisk kostnad P50-estimatet?
- Hvilke faktorer kan forklare overskridelser og avvik i estimatene?

Tema 3: Realisme og informasjonsverdi til prediksjonsintervallene og estimatfordelingene

- Hvor godt er P10, P50, P85 og P90-estimatene kalibrert?
- Hvor godt er de avledede estimatfordelingene som helhet kalibrert?
- Hvor informative er prediksjonsintervall og estimatfordelinger? Hvor godt skiller de mellom høy og lav usikkerhet?
- Hvor gode er punkttestimatene og prediksjonsintervallene sammenlignet med enkle påslagsmodeller?

For de fleste spørsmålene gjør vi sektorvise analyser og analyser av utvikling over tid. I tillegg har vi valgt å legge inn noen analyser på utvalgte problemstillinger, inkludert hvilke faktorer som henger sammen med

estimeringsavvik og usikkerhet, og om det er forskjeller mellom konsulentfirmaenes kvalitet på estimater.

4.2 Sannsynlighetsbaserte kostnadsestimater

Grunnlaget for probabilistiske kostnadsestimater er tanken om at det finnes et *utfallsrom*, det vil si at det er en rekke mulige kostnadsutfall. Dersom det samme prosjektet hadde blitt gjennomført mange ganger uavhengig av hverandre, så ville ikke prosjektet ha kostet det samme hver gang. Kostnadene for det samme prosjektet vil i stedet ofte kunne variere mye, avhengig av hvem som gjennomfører det, hvordan arbeidsprosessene er, kundens bidrag, flaks/uflaks og mange andre faktorer. Et utfallsrom kan beskrives som en *fordeling* av mulige utfall med det vi tror er de tilknyttede *sannsynligheter*. En slik hypotetisk fordeling av ulike utfall for enkeltprosjekt vil vi benevne G i analysene våre.

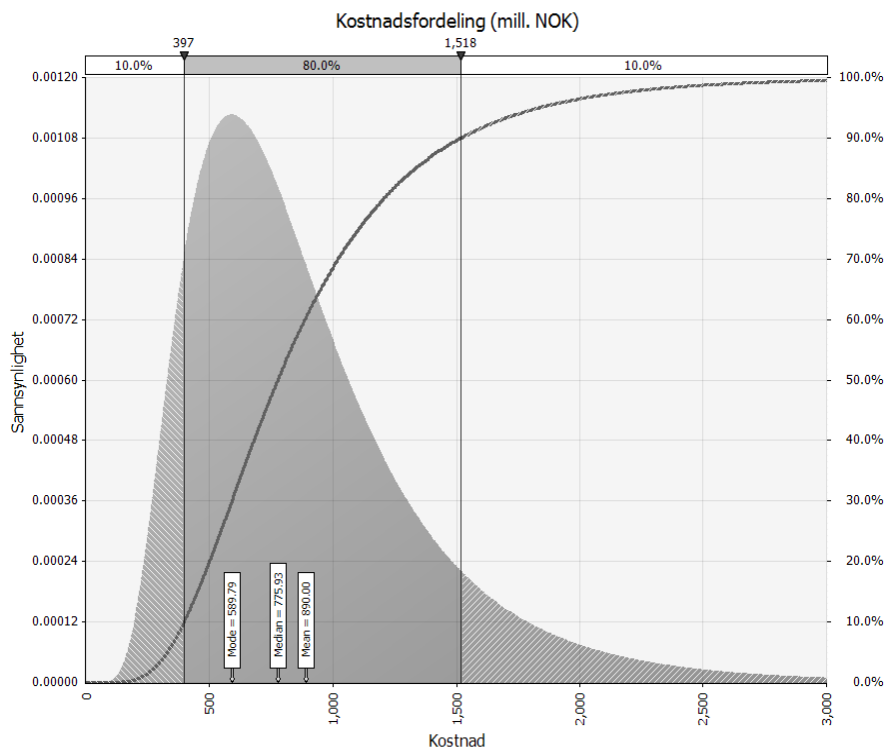
Når vi estimerer kostnader probabilistisk, så forsøker vi å estimere hele eller deler av den faktiske fordelingen G. Den estimerte fordelingen av kostnader blir benevnt F i analysene. F vil ofte ha en S-form og kalles derfor ofte for prosjekts estimerte S-kurve for kostnader. Det er denne estimerte fordelingen (S-kurven), eller deler av den som P50 og P85-estimerer, som vil bli analysert i denne studien. Når vi evaluerer punkttestimaters realisme, så er det altså punkter på fordelingen F. Et P50-estimat (median) er for eksempel den verdien vi tror det er 50 prosent sannsynlig at faktisk kostnad ikke vil overskride.

For å gi et eksempel kan vi anta at vi har estimert utfallsrommet for et prosjekt og kommet frem til estimatfordelingene vist i Figur 4-1. Den viser det estimerte utfallsrommet både som en tetthetsfordeling (den grå fordelingen) og en kumulativ fordeling (S-kurven fra 0 til 100 prosent og det vi benevner F i denne studien). Slike fordelinger utarbeides som en del av KS2-arbeidet og bygger typisk på at man ber fagpersoner om sannsynlighetsvurderinger og gjør antagelser om hvordan den bakenforliggende fordelingen ser ut, for eksempel at kostnadsfordelinger er log-normalt eller gamma-fordelte.

Ser vi nærmere på den grå tetthetsfordelingen i Figur 4-1 så ser vi at det mest sannsynlige utfallet (topp-punktet, eller modalverdi) er på ca. 590 millioner kroner. Vi ser også at fordelingen er høyreskjev (lang hale mot høyre). Når vi har høyreskjeve fordelinger, som er svært vanlig innen kostnadsestimering, så vil ofte topp-punktet ligge til venstre for median (som tilsvarer P50-estimatet)

og forventningsverdi (gjennomsnittet av fordelingen). I eksemplet i figuren er P50-estimatet på ca. 775 millioner, altså vesentlig høyere enn det mest sannsynlige utfallet på 590 millioner. Bruker vi S-kurven, så kan vi avlese hvor høyt vi må gå for at vi skal være 85 prosent sikre på å ikke overskride estimatet. Dette vil være P85-estimatet. Starter vi med 85 prosent på y-aksen til høyre og finner korresponderende kostnadsverdi på S-kurven, så finner vi en verdi på ca. 1300 millioner. Forventningsverdien (gjennomsnittet), som i dette eksemplet er på 890 millioner, er også av interesse, særlig for porteføljestyring. Dette fordi den forventede totalkostnaden for en portefølje av prosjekter tilsvarer summen av alle forventningsverdiene for prosjektene som inngår i porteføljen.

Figuren illustrerer en typisk høyreskjev fordeling som viser at det er langt mer sannsynlig at kostnaden blir vesentlig høyere enn forventet enn vesentlig lavere. I en høyreskjev fordeling vil forventningsverdien kunne være en god del høyere enn P50, i dette tilfellet 15 prosent høyere.



Figur 4-1 Tetthetsfordeling og kumulativ fordeling

Den faktiske kostnaden til et prosjekt kan ses på som en tilfeldig trekning fra utfallsrommet beskrevet ved kostnadsfordelingen G. Det at faktisk kostnad kan ligge hvor som helst i utfallsrommet (men hvor ulike utfall har ulike sannsynlighet) sier oss at vi ikke uten videre kan si at et punkt estimat fra F (for eksempel P50-estimatet) er feil dersom faktisk kostnad ble mye høyere eller lavere enn dette. Dersom vi i eksemplet vårt brukte P50-estimatet (775 millioner) som punkt estimat (eller som styringsramme) og faktisk kostnad ble på 850 millioner, så kan fortsatt estimatfordelingen F ha vært riktig. Selv med «perfekte» estimater så vil man forvente estimeringsavvik.

For å finne ut hvor mye estimeringsavvik man minst kan forvente for et sett av prosjekter kan man bruke estimerte kostnadsfordelinger og trekke tilfeldige verdier fra fordelingene gjentatte ganger. Dette vil simulere variasjonen i faktiske kostnader under perfekt estimering. Vi gjorde denne øvelsen for datasettet vi skal bruke i denne studien og fant at det gjennomsnittlige estimatavviket man kan forvente med perfekte P50-estimater er mellom 8 og 10 prosent. Her er estimatavvik målt som avvik (uavhengig av fortegn) mellom estimert og simulert kostnad delt på estimert kostnad.

4.3 Evaluering av sannsynlighetsbaserte estimater

Fokuset i dette kapitlet er hovedsakelig på metoder og målestørrelser for evaluering av de sannsynlighetsbaserte estimatene, og i mindre grad studier av kostnadskontroll (der vi i stor grad følger praksis for hvordan dette har blitt behandlet i tidligere Concept-studier, se kap. 2.1).

I evalueringen av kostnadsestimatene, som utgjør grunnlaget for fastsettelse av rammene, innfører vi i dette kapitlet til dels nye evalueringspraksiser og målestørrelser både for punkt estimater, prediksjonsintervaller og estimatfordelinger.

Kapitlet innfører først nødvendig terminologi, og gir deretter en motivasjon for og en beskrivelse av gode praksiser for evaluering av sannsynlighetsbaserte punkt estimater, prediksjonsintervaller og estimatfordelinger. Til slutt definerer vi målestørrelsene brukt i selve evalueringen.

Terminologi

Tabellen nedenfor angir terminologi for evaluering av sannsynlighetsbaserte kostnadsestimater.

Tabell 4-1: Terminologi

Terminologi	
G	Faktisk utfallsfordeling (kumulativ) for kostnader forbundet med gjennomføringen av et prosjekt. Representerer den iboende usikkerheten i hvor mye et prosjekt vil koste.
Faktisk kostnad (kostnad)	Faktisk kostnad for et prosjekt. Kan anses som et punkt trukket fra faktisk utfallsfordeling G.
F	Estimert utfallsfordeling (kumulativ) av kostnader for et prosjekt. En bredere (mindre konsentrert) estimatfordeling angir en større estimert usikkerhet i hva et prosjekt vil koste.
Estimert kostnad (estimat)	Punktestimert for kostnad. Kan være hvilket som helst punkt på estimatfordelingen F.
PX-estimat	Estimat anslått til å være X prosent sannsynlig at faktisk kostnad ikke vil overskride. Kan i noen tilfeller evalueres som et punktestimert (f.eks. avvik fra P50), og kan alltid evalueres som et prediksjonsintervall (intervall fra 0 til og med PX).
X % prediksjonsintervall	Prediksjonsintervall for kostnader. Estimert intervall [minimum, maksimum] som forsøker å inkludere faktisk kostnad med X prosent sannsynlighet.
Tapsfunksjon	En funksjon som tar punktestimert og faktisk kostnad som inndata og produserer et tall som representerer tapet («cost», «loss») eller feilen forbundet med avviket mellom dem. En lineær tapsfunksjon som $ kostnad - estimat $ antar at det er like problematisk med underskridelser som overskridelser og at det faktiske avviket i kroner er viktig (ikke prosentvis avvik). Utfallet av en evaluering er avhengig av hvilken tapsfunksjon som ligger bak målestørrelsen for estimeringsfeil.

Evaluering av sannsynlighetsbaserte punktestimater

Et punktestimert for kostnader til et prosjekt kan i teorien være et hvilket som helst punkt fra estimatfordelingen. Vi kan for eksempel velge å angi vårt estimert som et P50-estimat, et P85-estimat eller forventningsverdien til estimatfordelingen. Hvilket punkt vi velger bør bestemmes av hva vi skal bruke estimatet til, som for eksempel planlegging, til budsjettering eller porteføljestyling. Sammen med et formål til estimatet vil det være en, ofte ikke eksplisitt formulert, *tapsfunksjon*. Denne tapsfunksjonen vil si noe om hvor problematisk (hvor stort tap) det er med ulike utfall av faktisk kostnad. Det kan for eksempel være slik at store overskridelser av kostnader er mer problematisk enn store underskridelser. Oppgaven til et estimert er å minimere den forventede verdien til denne tapsfunksjonen, det vil si å minimere det forventede tapet forbundet med at faktisk kostnad avviker fra estimert kostnad.

Eksempler på mulige tapsfunksjoner som er relevante i en kostnadsestimerings-sammenheng, og som vi ønsker å minimere omfatter:

1. Differansen mellom faktisk og estimert kostnad
2. Sannsynligheten for under- eller overskridelser større enn en gitt verdi
3. Differansen mellom sum av faktisk og sum av estimert kostnad for en portefølje av prosjekter
4. Avvik mellom faktisk og estimert kostnad, der kostnadsoverskridelser vektet sterkere enn kostnadsunderskridelser, for eksempel at overskridelser vektet med 0,85 mens underskridelser vektet med 0,15.
5. Prosentvis avvik mellom faktisk og estimert kostnad

Viktigheten av å være bevisst tapsfunksjonen når man gir et punkt estimat illustreres ved at man i de fem ovennevnte eksemplene burde bruke ulike punkter fra estimatfordelingen F som vårt estimat. Medianverdien (P50-estimatet) vil minimere den første tapsfunksjonen, modalverdien (mest sannsynlig utfall) vil minimere den andre, forventningsverdien vil minimere den tredje, og et P85-estimat vil minimere den fjerde tapsfunksjonen. Tapsfunksjonen til den femte er noe mer komplisert og vi vil komme tilbake til den senere i kapitlet.

En konsekvens av at ulike typer estimat kan ha ulike tapsfunksjoner er at målestørrelser for hvordan vi evaluerer punkt estimatet bør samsvare med hva vi mener er en fornuftig tapsfunksjon. Et eksempel på manglende samsvar er at vi har angitt et P85-estimat (som impliserer en tapsfunksjon der overskridelser vektet 0,85 og underskridelser med 0,15) men evaluerer estimatet ved å måle prosentvis avvik (der det ikke er noen slik vektning) mellom faktisk kostnad og P85-estimatene. Dersom man ønsker å evaluere P85-estimer som punkt estimater (og ikke som prediksjonsintervall) så er det mulig å la målestørrelsene gjenspeile vektningen i tapsfunksjonen. I dette tilfelle vil det kunne være målestørrelsen:

$$L_{P85} = \begin{cases} 0.85 \cdot |kostnad - estimat|, & \text{hvis } estimat \leq kostnad \\ 0.15 \cdot |kostnad - estimat|, & \text{ellers} \end{cases}$$

Hovedpoenget er at det er mulig å oppnå et samsvar mellom hva vi ønsker å oppnå med kostnadsestimatene (representert med tapsfunksjonen som vi ønsker å minimere) og hvordan vi evaluerer dem. Denne sammenheng kan formuleres slik (se Zellner, 1986; Gneiting, 2011):

Det optimale punkttestimatet ($\widehat{estimat}$) fra en estimatfordeling F er det som minimerer det forventede (\mathbb{E}) tapet (L) for fordelingen av faktiske kostnader G , dvs. $\widehat{estimat} = \arg \min_{estimat} \mathbb{E}_F L(estimat, G)$.

Vi vet vanligvis ikke hvordan fordelingen av faktiske kostnader for et prosjekt (G) ser ut, og vi kan ikke anta at estimatfordeling vår F er identisk med G . Vi har altså ingen garanti for at valgt estimat faktisk minimerer tapsfunksjonen. Det vi derimot kan tilstrebe er *samsvar* mellom hva vi angir som estimert kostnad (hvilket punkt på F vi velger som vårt estimat), tapsfunksjon og målestørrelser for hvordan vi evaluerer estimatene. En viktig motivasjon for å etterstrebe et slikt samsvar er at de som estimerer vil bli belønnet med lav estimeringsfeil for å gi «ærlige» estimater som minimerer tapsfunksjonen, og ikke strategisk angi for lave eller for høye estimater for å forbedre målt estimeringsavvik. Et samsvar mellom type estimat og målestørrelsen for estimeringsavvik vil bidra til en mer meningsfull evaluering (Granger og Pesaran, 2000; Gneiting, 2011; Carvalho, 2016).

Et eksempel på manglende samsvar mellom tapsfunksjon og evalueringsfunksjon, med tilhørende insitament for strategiske estimater, er den hyppige brukte målestørrelser basert på prosentvis estimeringsfeil, for eksempel målestørrelsen prosentvis absolutt estimatavvik:

$$|kostnad - estimat|/estimat.$$

Denne målestørrelsen blir blant annet brukt for å evaluere P50-estimater som punkttestimat. P50-estimater minimerer, som før nevnt, verdien til tapsfunksjonen:

$$|kostnad - estimat|.$$

Det vil si at det å velge P50-estimatet som estimat minimere de absolutte avvikene i kroner mellom kostnader og estimat (gitt at estimert og faktisk fordeling av kostnader samsvarer), men *ikke* absolutt prosentvis estimatavvik. Når avvik blir evaluert ved å dele avviket på estimatet, slik prosentvis estimatavvik er definert over, vil det lønne seg å gi estimat som er noe høyere enn P50. Man får altså bedre evalueringer av estimatene ved å angi et P50-estimat som ikke er P50. Mer om problemene med å bruke målestørrelser basert relativ estimeringsfeil finnes i blant annet i Gneiting (2011) og Foss Stensrud m.fl. (2003). Når vi likevel velger å bruke målestørrelser basert på prosentvis estimeringsavvik i denne studien, er det for å kunne sammenligne

med andre undersøkelser som har brukt denne typen målestørrelser, samt at konsekvensene trolig ikke er så store i situasjoner som i denne studien, der vi har relativt små estimeringsavvik. Vi vil i tillegg bruke målestørrelser for estimeringsfeil som unngår problemet beskrevet over.

I Tabell 4-2 har vi oppsummert målestørrelsene vi kommer til å bruke i evalueringen av punktestimatene, med tilhørende tapsfunksjon og type estimat som minimerer tapsfunksjonen. For å oppnå samsvar mellom type estimat og hvordan vi evaluerer disse, er dette viktige sammenhenger.

Vi ønsker i tillegg å vite om det er skjevheter i punktestimatene, det vil si om P50-estimatene har en tendens til å over- eller underestimere faktiske kostnader. Noen typiske målestørrelser man kan bruke for å evaluere skjevheten i punktestimatene er vist i Tabell 4-3. I stedet for tapsfunksjon har vi brukt ordet evalueringsfunksjon, siden målet her er å komme nærmest mulig 0, ikke å minimere funksjonen. For alle målestørrelsene er positive verdier forbundet med mer overskridelse og negative verdier forbundet med mer underskridelser.

Tabell 4-2: Målestørrelser for estimatfeil med samsvarende tapsfunksjon og punktestimat

Målestørrelse for estimeringsfeil	Tapsfunksjon	Samsvarende type punktestimat
Absoluttverdi av estimeringsavvik (i kroner)	$ kostnad - estimat $	P50-estimat (median til F)
Andel “unøyaktige” estimater	$\begin{cases} 1, & kostnad - estimat \geq c \\ 0, & \text{ellers} \end{cases}$, der c er en konstant større enn 0	Mest sannsynlig estimat (modalverdi til F)
Prosentvis absolutt estimatavvik	$\frac{ kostnad - estimat }{estimat}$	Medianen til en stokastisk variabel med en tetthetsfunksjon symmetrisk til $kostnad \cdot f(kostnad)$, der f er tetthetsfunksjon til F (Gneiting 2011, s 754).
Symmetrisk prosentvis absolutt estimeringsavvik (lik vektning av over og underskridelser)	$\left \ln\left(\frac{est}{act}\right) \right = \ln(est) - \ln(act) = \left \ln\left(\frac{act}{est}\right) \right $ (For å finne medianen til symmetrisk prosentvis feil, bruker vi $e^{Md \cdot \left \ln\left(\frac{act}{est}\right) \right } - 1$, der Md er medianen til n estimater.)	Estimater der målet er å minimere medianen til forventet relativ estimeringsfeil, med symmetrisk vektning av over og underskridelser (Mer om målestørrelsen i (Morely m.fl., 2018))

Tabell 4-3: Målestørrelser for estimatskjevhet (Md=Median)

Målestørrelse for estimatskjevhet	Evalueringsfunksjon	Samsvarende type punkttestimat
Estimatskjevhet	$Md(kostnad - estimat)$	P50-estimat (median)
Estimatskjevhet	$M(kostnad - estimat)$	Forventningsverdi
Asymmetrisk relativ estimatskjevhet	$Md\left(\frac{kostnad - estimat}{estimat}\right)$	P50-estimat (median)
Asymmetrisk relativ estimatskjevhet	$M\left(\frac{kostnad - estimat}{estimat}\right)$	Forventningsverdi
Symmetrisk relativ estimatskjevhet	$sg * e^{Md(\ln(\frac{kostnad}{estimat}))} - 1$, Hvor sg er fortegnet til $Md(\ln(\frac{kostnad}{estimat}))$	P50-estimat (median)
Summert estimatskjevhet	$\Sigma kostnad - \Sigma estimat$	Forventningsverdi

Den viktigste konsekvensen av sammenhengene vist i tabellen ovenfor er at dersom vi evaluerer estimatskjevhet for P50-estimer, så bør vi bruke medianen av alle målingene. En måling av, for eksempel, gjennomsnittet til prosentvis avvik for P50-estimer vil være vanskelig å tolke fordi man vil oppnå mindre skjevhet ved å oppgi estimer som er høyere enn P50, gitt at fordeling ikke er symmetrisk.

Evaluering av prediksjonsintervaller og estimerte utfallsfordelinger

Som beskrevet ovenfor kan man evaluere P50-estimat som punkttestimat, hvor fokuset er på avvik mellom estimert og faktisk kostnad. Man kan også evaluere P50-estimat, samt alle andre PX-estimat, som ensidige prediksjonsintervaller fra P0 til PX. I tillegg kan man evaluere prediksjonsintervaller basert på to PX-estimat (for eksempel tilsvarer verdiene mellom P10 og P90 et 80 prosent prediksjonsintervall), og man kan evaluere i hvilken grad estimatfordelingen F (for eksempel utledet fra PX estimer) samsvarer med faktiske enkeltutfall. I teksten under viser vi hvordan man kan evaluere prediksjonsintervaller og estimatfordelinger.

Tradisjonelt, også i tidligere Concept-rapporter, har det vært et sterkt fokus på hvor godt kalibrerte prediksjonsintervallene er. For eksempel at P50-estimer for kostnader (altså prediksjonsintervallet fra P0 til P50) skal inneholde faktisk kostnad i ca. 50 prosent av tilfellene. Dette er en viktig egenskap ved et

probabilistisk estimat av denne typen, men er ikke tilstrekkelig for å vurdere hvor gode estimatene er, noe følgende eksempel (typologi) illustrerer:

Anta at vi har estimert en rekke prosjekter av omtrent samme størrelse og angitt helt korrekte P50-estimer for alle prosjektene og at prosjektene enten er av lav-risiko eller av høy-risiko typen. Anta videre at forholdet mellom faktisk kostnad og P50-estimatet (kostnad/P50-estimatet) for lav-risiko typen er normalfordelt med gjennomsnitt 0,0 og standardavvik på 0,2 mens det for høy-risiko typen er normalfordelt med gjennomsnitt 0,3 prosent og standardavvik på 0,2. Dette tilsvarer at vi trenger å gange P50-estimatet med 1,21 for lavrisiko og med 1,51 for høyrisiko-prosjekter for å finne P85-estimatet. Noe overraskende er det flere strategier som *alle* vil gi perfekt kalibrerte P85-estimer (gitt et stort utvalg prosjekter):

- Estimator har evne til å gjenkjenne et lav-risiko prosjekt og multipliserer P50-estimatet med en faktor på 1,21 for å finne P85-estimatet. Estimator gjenkjenner også høy-risiko prosjekt og multipliserer P50-estimatet med en faktor på 1,51 for å finne P85-estimatet. I tilfellet over gir dette perfekte estimer og dermed også perfekt kalibrering.
- Estimator gjenkjenner ikke lav og høy-risiko, vet hva gjennomsnittlig påslag bør være, og behandler alle prosjekter likt. Dette kan gjøres ved å ha det samme påslaget på P50 for alle prosjekter, der påslaget er det som historisk sett har vært nødvendig for å omfatte 85 prosent av prosjektkostnadene (85 prosent-prosentilen av påslagene). Dette tilsvarer i dette eksemplet multiplikasjon av P50 med en faktor på 1,42 for alle prosjekter. En slik estimerer kalles i estimeringslitteraturen, se for eksempel (Gneiting m.fl., 2007) for en klimatologisk estimerer. Det gir ikke perfekte estimer, men vi oppnår likevel perfekt kalibrering på tvers av prosjekter.
- Estimator gjenkjenner ikke lav og høy-risiko, vet hva gjennomsnittlig påslag bør være, og varierer påslagene tilfeldig rundt dette gjennomsnittet. Dette kan, for eksempel, være en som i 75 prosent av tilfellene legger til 0,1 på den gjennomsnittlige (klimatologiske) faktoren, det vil si multipliserer med 1,52, og i 25 prosent av tilfellene trekker fra 0,1, det vil si multipliserer med 1,32. Om det blir lagt til eller trukket fra er ikke korrelert med om det er lav eller høy risiko i prosjektet. En slik estimerer kan (vårt forslag) kalles ufokusert, klimatologisk. Det gir ikke perfekte estimer, men vi oppnår likevel perfekt kalibrering på tvers av prosjekter.

Alle disse strategiene gir at ca. 85 prosent av kostnadene vil være lavere enn P85-estimatet, det vil si at alle har perfekt kalibrerte P85-estimerer. Men de er ikke like nyttige. Vi kan for eksempel utlede at den klimatologiske estimerer i gjennomsnitt har 17 prosent høyere og den ufokuserte, klimatologiske estimerer i gjennomsnitt 31 prosent høyere P85-estimerer enn den perfekte estimerer. De vil altså kreve en høyere kostnadsramme, og mer midler bundet opp. Det vi ønsker oss er altså estimerer som varierer usikkerhetspåslag i samsvar med underliggende usikkerhet.

Tabell 4-4: Målestørrelser for evaluering av prediksjonsintervall og estimatfordelinger

Intervall-estimerer	
Treffrate (kalibrering)	$\frac{1}{n} \sum_i h_i, \quad h_i = \begin{cases} 1, & \text{minimum}_i \leq \text{kostnad}_i \leq \text{maksimum}_i \\ 0, & \text{ellers} \end{cases}$ der minimum og maksimum er endepunktene i prediksjonsintervallet. Måler andel prosjekter med faktisk kostnad innenfor prediksjonsintervallet.
Relativ bredde («sharpness»)	$\frac{(\text{maksimum} - \text{minimum})}{P50}$ der minimum og maksimum er endepunktene i prediksjonsintervallet. Måler informasjonsverdi. For likt kalibrerte intervaller vil smalere gjennomsnittlig relativ bredde indikere bedre estimatkvalitet.
Korrelasjon	τ (Relativ bredde, prosentvis absolutt estimatavvik), Spearman's rang-korrelasjonskoeffisient. Måler informasjonsverdi. Indikator på i hvilken grad variasjon i bredden på P10-P90 prediksjonsintervallene reflekterer variasjon i usikkerhet i faktisk kostnadsbruk.
Estimatfordeling	
Kontinuerlig rangert sannsynlighets-skåre (CRPS)	$CRPS(F_i, y_i) = \int_{\mathbb{R}} (F(z) - \mathbb{I}(\text{kostnad} \leq z))^2 dz,$ der $\mathbb{I}(\text{kostnad} \leq z)$ er «identifikator-funksjon» som 1 dersom $\text{kostnad} \leq z$, ellers 0. Verdien (integralet) minimeres når estimatfordeling F er lik utfallsfordelingen G. CRPS integrerer kalibrering og «sharpness» i samme målestørrelse. En lavere gjennomsnittlig CRPS indikerer bedre estimater.
PIT (probability integral transform)	$F_i(\text{kostnad}_i)$, der F_i er estimatfordelingen for prosjekt i. PIT er verdien (mellom 0 og 1) man får når faktisk utfall (kostnad) brukes som inndata til estimatfordelingen F.
PIT-histogram	Histogram over PIT-verdier for en mengde med faktiske kostnadsutfall. Dersom perfekt kalibrert estimatfordeling og tilstrekkelig mange observasjoner, så vil PIT histogrammet vise likt antall observasjoner i hver sannsynlighetskategori, det vil si at PIT-verdiene er uniformt fordelt. Et PIT-histogram med U-form tyder på en for smal estimatfordeling, mens et PIT-histogram med \cap -form tyder på en for bred estimatfordeling.

En konsekvens av at flere ulike strategier, som ikke alle er like gode, gir godt kalibrerte prediksjonsintervaller og PX-estimerer gjør at vi trenger å evaluere disse både med hensyn til hvor godt de samsvarer med faktisk utfallsfordeling (kalibrerthet) og hvor informative de er. Tabell 4-4 gir en oversikt over målestørrelser for dette. Flere av disse målestørrelsene er noe tekniske og noen av de har kun mening når flere verdier blir sammenlignet mot hverandre, så når vi presenterer resultatene forutsetter vi ikke at leseren forstår den underliggende matematikken bak målestørrelsene.

Negativ, eller svært lav korrelasjon mellom bredde på prediksjonsintervallene og estimatavvik vil tyde på at usikkerhetsanalysene ikke klarer å fange opp variasjon i reell usikkerhet. Med andre ord, at de som estimerer mangler evnen til skille mellom prosjekter med relativt sett lav og høy kostnadsusikkerhet.

Antatt at faktisk utfallsfordeling er den samme som den estimerte (perfekt kalibrering), så vil korrelasjonen τ i KS2-datasettet være på ca. 0,3. Vi må derfor forvente at selv om de som estimerer er gode til å skille høy og lav kostnadsusikkerhet, så vil korrelasjonen være lavere enn dette.

Vi vil i analysene også bruke korrelasjon mellom bredden på prediksjonsintervallet P10 og P90 og avvik for å se om det er korrelasjon mellom estimatskjevhet (bias) og bredde på prediksjonsintervallene. Dette vil kunne avdekke om større eller mindre usikkerhet i estimatene (bredere prediksjonsintervaller) skiller mellom høy og lav risiko for overskridelser i stedet for (som det optimalt sett burde) høy eller lav risiko for estimatavvik begge veger.

5 Styrings- og kostnadsrammer

I dette kapittelet ser vi på i hvilken grad prosjekter som har vært gjennom KS2 har blitt gjennomført innenfor de vedtatte styrings- og kostnadsrammene. Analysene tilsvarende om lag de som er gjort i tidligere studier (jf. Kap. 2), men med et noe større utvalg.

5.1 Overholdelse av styrings- og kostnadsrammer

Tabell 5-1 viser avvik fra styrings- og kostnadsrammene for prosjektene i utvalget.

Tabell 5-1: Overholdelse av styrings- og kostnadsrammer

	Styrings- ramme	Kostnads- ramme
Antall prosjekter	83	85
Gjennomsnittlig over-/underskridelse (skjevhet)	3,4 %	-5,6 %
Median overskridelse (skjevhet)	1,5 %	-4,9 %
Standardavvik overskridelse	16,9 %	15,5 %
Standardfeil overskridelse	1,9 %	1,7 %
Minimum overskridelse (underskridelse)	-41,9 %	-43,4 %
Maksimum overskridelse	48,6 %	39,6 %
Gjennomsnittlig avvik	12,5 %	12,5 %
Median avvik	9,8 %	10,7 %
Andel innenfor rammen (dersom kalibrert, bør andel innen styrings- og kostnadsramme være hhv. ca. 50 % og 85 %)	40 %	73 %
Andel med mer enn 10 % overskridelse av ramme	31 %	15 %
Andel innenfor +/- 10 % av ramme	49 %	48 %
Gjennomsnittlig overskridelse blant prosjekter med overskridelse av ramme	13,4 %	12,9 %

Med et noe større utvalg enn i Welde (2017) finner vi en gjennomsnittlig overskridelse av styringsrammen på 3,4 prosent og median overskridelse på 1,5 prosent. Spredningen rundt styringsrammen er relativt stor med et standardavvik på 17 prosentpoeng. Om lag 50 prosent av prosjektene i utvalget har avvik fra styringsrammen større enn +/- 10 prosent.

Dersom det er omtrent like mange og store overskridelser som det er underskridelser (målt mot styringsrammen) vil gjennomsnittlig overskridelse være i nærheten av null. Gjennomsnittlig estimatoverskridelse måler derfor ikke nøyaktigheten til estimatene/rammene, kun skjevheten. De fleste internasjonale studier har imidlertid vist en klar tendens til overskridelser og hevdet at problemet er skjevhet og ikke så mye unøyaktighet. En median overskridelse av styringsrammen på 1,5 prosent er i så måte bedre enn det man har funnet i andre studier. Dersom vi tester verdien opp mot hypotesen at medianen skal være 0 prosent (selv om dette ikke er en helt riktig antagelse, se Kapittel 4), får vi en p-verdi på 0.12 (Wilcoxon signed-rank test). Dette betyr at i datasettet vårt vil median avvik på 1,5 prosent eller mer kunne forventes med 12 prosent sannsynlighet dersom man antar at median avvik i virkeligheten er 0 prosent. Dette kan tolkes som at det kan være en tendens til overskridelser, men det er heller ikke urimelig å anta at avviket fra 0 prosent skyldes tilfeldigheter.

For å måle estimeringsevnen er det hensiktsmessig også å se på avvik fra kostnader uavhengig av retning på avviket. Estimataavvik eller «mean absolute percentage error» er sjelden inkludert i internasjonale studier, men i en analyse av 1045 vegprosjekter gjennomført i perioden 1993-2007 fant Odeck (2014) et avvik på 17,6 prosent. Blant KS-prosjektene finner vi et gjennomsnittlig avvik på 12,5 prosent, og median på 9,8 prosent. Dette bør anses å være et lavt avvik gitt usikkerheten i bruk av kostnader i store prosjekter. Som argumentert i Kapittel 4 så vil det forventede minimumsnivået på estimataavvik, gitt usikkerheten i bruk av kostnader for relevant type prosjekter, være på 8-10 prosent, antatt at man ikke tilpasser leveransene og arbeidet til styringsrammen.

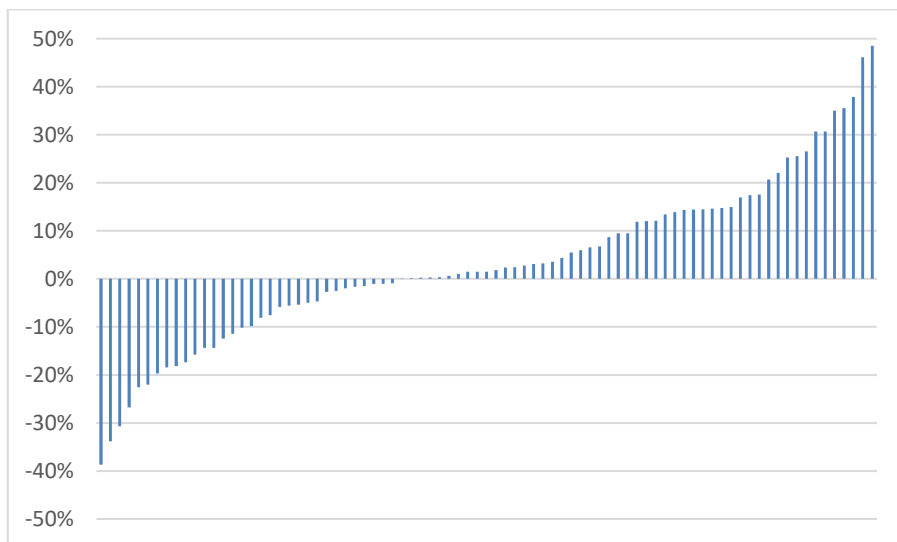
12 prosent av prosjektene (10 prosjekter) har overskridelser over 25 prosent målt mot styringsrammen. Gjennomsnittet blant disse «verste» prosjektene er en overskridelse på 34 prosent. Dette er betryggende for en risikoavers beslutningstaker som vil unngå store kostnadsoverskridelser. Selv blant prosjektene som går over rammene så er ikke overskridelsene urovekkende høye sammenlignet med skrekkeksemplene fra den internasjonale litteraturen.

Gjennomsnittlig overskridelse blant prosjektene med overskridelse av kostnadsrammen er 12,9 prosent. Noen prosjekter har store og problematiske overskridelser, opptil 48 prosent over styringsrammen og 40 prosent over kostnadsrammen. Selv dette er ikke mye sett i lys av de store overskridelsene

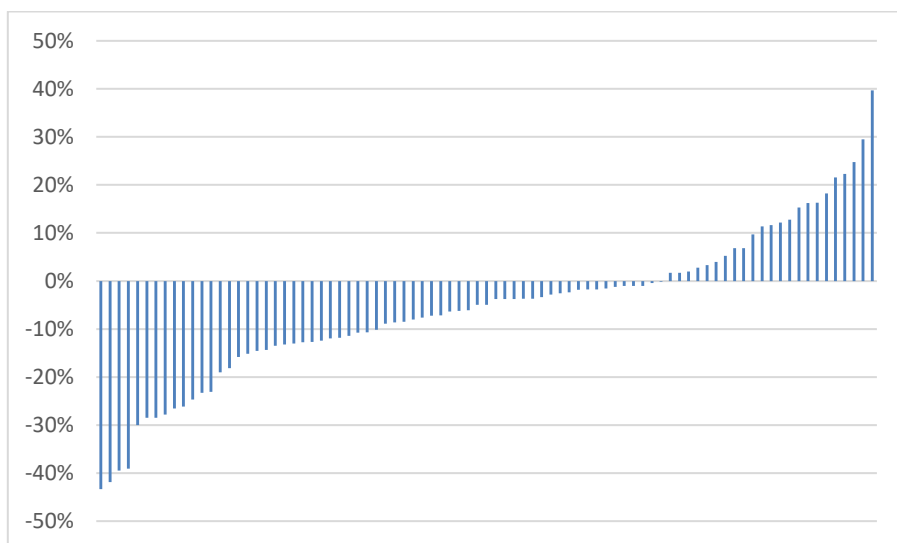
som har blitt rapportert i studier fra andre land. På den annen side, og som diskutert kort i Kapittel 3, så kan enkelte av prosjektene ha reelle overskridelser ut over det som vises i dette datasettet ettersom vi her måler mot stortingsvedtak etter gjennomført KS2. Enkelte prosjekter kan dessuten ha hatt en lang og problematisk periode preget av kostnadsøkninger *før* de er kommet over terskelverdien for KS2.

Hvis styrings- og kostnadsrammen settes til henholdsvis P50 og P85 burde andelen prosjekter under rammene være om lag 50 og 85 prosent. Som Tabell 5-1 viser så er ikke det målet helt oppnådd, med 40 prosent av prosjektene under styringsrammen og 73 prosent under kostnadsrammen. En test av andelen prosjekter under styringsrammen oppnår en p-verdi på 0.10 når nullhypotesen er satt til at 50 prosent skal være innenfor rammen (eksakt binomial test). Dette betyr at det er 10 prosent sannsynlig å se et avvik likt eller større enn det vi observerer i data. Det er altså ikke urimelig å anta at «feilen» skyldes tilfeldig variasjon, men man kan også se på det som en indikasjon på en underliggende systematisk feil. Samme test for kostnadsrammen, der nullhypotesen er satt til at 85 prosent av prosjektene bør være innenfor rammen, viser en p-verdi på 0.008. Dette avviket fra den forventede andelen er såpass usannsynlig dersom nullhypotesen er sann (et lignende avvik eller større vil skje med bare 0.8 prosent sannsynlighet) at det er naturlig å tenke at prosjektporteføljen systematisk har bommet på P85 (statistisk signifikant avvik).

Andelene over rammene er høyere enn i Welde (2017). Figur 5-1 og Figur 5-2 viser hvordan prosjektenes kostnader fordeler seg i forhold til styrings- og kostnadsrammene.



Figur 5-1: Prosjekter med kostnad under/over styringsrammen



Figur 5-2: Prosjekter med kostnad under/over kostnadsrammen

Tabell 5-2 viser overskridelser av styringsramme og kostnadsramme i de ulike etatene.

Tabell 5-2: Kostnadskontroll i de ulike etatene

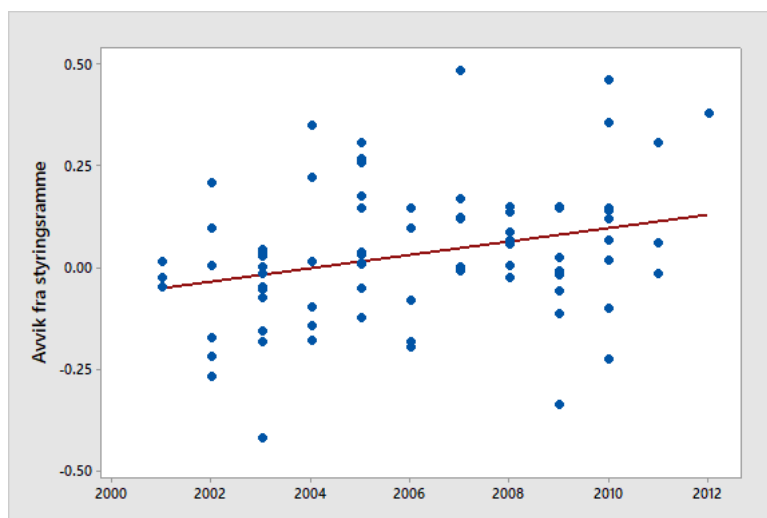
Ansvarlig etat	Utvalg	Gjennomsnittlig overskridelse styringsramme	Median overskridelse styringsramme	Andel under styringsramme	Andel under kostnadsramme
Statens vegvesen	43	3,8 %	2,8 %	37 %	65 %
Jernbaneverket / Bane NOR	8	4,4 %	0,7 %	38 %	87 %
Statsbygg	14	3,6 %	2,4 %	36 %	79 %
Forsvaret	13	-6,1 %	-4,7 %	70 %	92 %
Andre	7	13,6 %	13,2 %	29 %	57 %

Vi ser at de store etatene som opererer innenfor bygg- og anleggsbransjen - Statens vegvesen, Jernbaneverket/Bane NOR og Statsbygg, har nokså like resultater. Overskridelsene er om lag som for utvalget samlet. En analyse av andel prosjekter under styringsrammen per etat viser p-verdier fra 0.12 og oppover. Vi klarer altså ikke å avdekke om noen systematisk har en større eller mindre andel av prosjekter over styringsrammen (antatt P50) enn de burde. Når det gjelder kostnadsrammen, utmerker Statens vegvesen seg med en relativt lav andel prosjekter med en sluttkostnad under kostnadsrammen. Det er lite sannsynlig å oppnå en slik andel dersom man antar at den underliggende sannsynligheten for prosjekter under styringsrammen er 85 prosent. En binomial test ga en p-verdi på 0.001 for Statens vegvesen. Men om vi ser på antall prosjekter er det her mulig at Statens vegvesen blir «straffet» fordi de har mange prosjekter. Med det mener vi at det er godt mulig at de andre etatene har en underliggende kalibrering som er like dårlig, men at vi ikke har nok data til å avdekke dette. En test av forskjeller mellom etatene viser at det er ganske sannsynlig å observere forskjeller som de i datasettet bare ved tilfeldigheter ($p = 0.4$).

5.2 Utvikling av estimatskjevhet og avvik over tid

Welde (2017) så på en rekke ulike mulige forklaringsfaktorer for kostnadsoverskridelser, men fant at verken prosjektstørrelse, gjennomføringstid, antall år mellom beslutning og faktisk oppstart eller antall kontrakter hadde noen signifikant påvirkning på sannsynligheten for overskridelser. Man fant imidlertid at prosjekter vedtatt i perioden 2004-2008, det vil si i årene før finanskrisen, hadde signifikant flere overskridelser enn

prosjekter vedtatt før og etter denne perioden. Figur 5-3 viser utviklingen i avvik fra styringsrammen med prosjekter helt frem til oppstartsåret 2012.



Figur 5-3: Avvik fra styringsramme etter tidspunkt for investeringsbeslutning

Figuren viser at omfanget av kostnadsoverskridelser var lavere i den første perioden etter innføring av ekstern kvalitetssikring. En enkel lineær regresjon gir følgende resultat:

$$\text{Overskridelse} = -30.2 + 0.015 \text{ ÅR} + \varepsilon$$

Koeffisienten for tidspunkt (ÅR) er positiv og statistisk signifikant ($t = 2.496$, $p = 0.015$) noe som bekrefter inntrykket fra Figur 5-3 om at det er en økende tendens til kostnadsoverskridelser i perioden.

Tabell 5-3 viser utvikling i overskridelser og kalibrering for prosjekter vedtatt i periodene 2001-2003, 2004-2006, 2007-2009 og 2010-2012.

Tabellen bekrefter at det har vært en synkende andel prosjekter med en sluttkostnad under styrings- og kostnadsrammen. En test av forskjeller mellom periodene gir p-verdien 0.09 for andelen prosjekter med kostnader under styringsramme og 0.0009 for andelen prosjekter med kostnader under kostnadsramme (Fisher's exact test), noe som kan tyde på at det er forskjeller mellom periodene som ikke skyldes tilfeldigheter. For prosjektene fra den siste delen av perioden (2010-2012) var andelen prosjekter med kostnader lavere enn styringsrammen 21 prosent og lavere enn kostnadsrammen 43 prosent.

Dette er langt under målsetningen dersom man antar at styrings- og kostnadsrammene omtrent skal tilsvare P50 og P85.

Tabell 5-3: under/overskridelser og kalibrering for ulike perioder

Periode	Antall	Under/overskridelser (gjennomsnitt)	Under/overskridelser (median)	Andel under styringsramme	Andel under kostnadsramme
2001-2003	22	-6 %	-2,7 %	62 %	100 %
2004-2006	24	5,1 %	3,1 %	35 %	70 %
2007-2009	23	5,0 %	2,4 %	35 %	75 %
2010-2012	14	12,3 %	9,3 %	21 %	43 %

Hva dette skyldes er usikkert. I Welde (2017) ble det, som vist til ovenfor, antydnet at finanskrisen var årsaken til en høyere andel overskridelser i perioden 2004-2007 ettersom andelen overskridelser var signifikant høyere i den perioden sammenliknet med prosjekter vedtatt før og etter, men som Figur 5-3 viser så har tendensen til en høyere andel overskridelser fortsatt også i årene etter.

Finanskrisen markerte slutten på en internasjonal høykonjunktur, og Norge opplevde en særnorsk nedgang da oljeprisen falt fra 2014-2016, men siden da har veksten i norsk økonomi tiltatt og kapasitetsutnyttelsen er høy. Vi vet at den systemiske markedsusikkerheten er en viktig usikkerhetsfaktor i mange prosjekter og som kan ha stor påvirkning på prosjekters evne til å nå sine resultatmål. Selv om man til en viss grad kan ta høyde for dette i forbindelse med usikkerhetsanalysen der markedsusikkerheten påvirker viktige elementer i kalkylen, er det vanskelig å fullt ut ta høyde for økende råvarepriser og manglende kapasitet i entreprenør- og leverandørmarkedet. Et eksempel på det er prosjektet IFI2 som ble oppført i Gaustadbekkdalen i Oslo i perioden 2006-2010. På forhånd ble markedsusikkerheten identifisert som den viktigste usikkerhetsfaktoren og prosjektet var forberedt på svak respons på enkelte av utlysningene. Likevel ble konkurransen mer krevende enn man hadde sett for seg. På enkelte entrepriser var det få eller ingen tilbud. Man forsøkte også å lyse ut oppdrag i Danmark, men responsen var den samme. Tilbudssummene lå vesentlig over Statsbygg sine egne kalkyler. I tillegg til at et presset marked vil gi høyere priser på innsatsfaktorer og derigjennom høyere priser i markedet

for entreprenørtjenester, kan det også føre til at man er nødt til å inngå avtale med en entreprenør man kanskje ikke ville ha valgt hvis konkurransen hadde vært bedre. Sammen med en rekke andre utfordringer førte dette til at prosjektet opplevde en relativt betydelig kostnadsoverskridelse (Welde m.fl., 2018).

Variasjon i kostnadskontroll over tid kan være knyttet til variasjon i markedsprisene. Eksempelvis er byggekostnadsindeksene som benyttes i bygg- og anleggsnæringen «input-indeks» som måler utvikling i innsatsfaktorene, som timepriser og materialpriser, og ikke i markedsprisene direkte. I høykonjunktur opplever man gjerne dårligere produktivitet og høyere marginer, som medfører at veksten i markedsprisene kan være vesentlig høyere enn det byggekostnadsindeksene tilsier. I lavkonjunktur har man i perioder opplevd at markedsprisene har gått betydelig ned samtidig som byggekostnadsindeksene går opp. Usikkerhetsanalysene skal ta høyde for alle forhold som kan medføre prisutvikling forskjellig fra avtalt priskompensasjonen, men to forhold kan føre til at usikkerhetsanalysene likevel ikke får tatt tilstrekkelig høyde for dette. For det første er kostnadsestimatene vanligvis basert på enten kontraktverdier eller sluttkostnader fra lignende, tidligere utførte prosjekter, hvor kostnadstallene prisjusteres med en byggekostnadsindeks. Dersom det har vært høykonjunktur etter at referanseprosjektene var utført, vil denne tilnærmingen gi en undervurdering av reelle markedspriser. Videre vil man i vurderingen av markedsusikkerhet normalt ta høyde for at markedet kan gå begge veier. Markedsusikkerheten blir dermed reflektert i usikkerhetsspenning, men ikke i forventet kostnad. Selv om det innebærer en viss (kvalifisert) gjetning om markedet vil utvikle seg opp eller ned fremover, burde man kanskje bli flinkere til å anslå dette ut fra dagens situasjon og prognoser fremover. Concept rapport nr. 1 (Berntsen og Sunde, 2003) drøftet håndtering av markedsusikkerhet og kom med noen anbefalinger. Resultatene med hensyn på avvik over tid kan tyde på at vi ikke fullt ut har vært i stand til å omsette anbefalinger eller erfaringer fra tidligere prosjekter i usikkerhetsanalyser for fremtidige prosjekter.

En annen mulig årsak til økende utfordringer knyttet til kostnadskontroll er utslag av valutausikkerhet. De siste 7-8 årene har kronen svekket seg mot blant annet euro og dollar. Det kan påvirke kostnadene i store prosjekter, men i rammeavtalene frem til nå har det eksplisitt blitt sagt at valutausikkerhet ikke skal medtas i usikkerhetsanalysene ved KS2. Valutausikkerhet vil gjøre seg

gjeldende direkte der kontraktene er priset i utenlandsk valuta (spesielt ved ulike typer materiellanskaffelser), eller indirekte, som risikopåslag, der kontraktene er priset i norske kroner. Til en viss grad vil valutasingningene kunne sies å være inkludert i byggekostnadsindeksene, men siden valutausikkerhet ikke er et tema i KS2 er det begrenset kunnskap og erfaring om hvordan dette slår ut i praksis.

Vi skal med andre ord være varsomme med å fastslå hva denne utviklingen skyldes, men det kan synes som om man ikke i tilstrekkelig grad har evnet å lære fra gjennomførte prosjekter. Statens prosjektmodell er en standardisert og faseinndelt modell for hvordan prosjekter planlegges og besluttes. Det gir et godt grunnlag for sammenlikning. De ansvarlige etatene burde ideelt sett etter hvert ha opparbeidet seg solid erfaring for estimering og for hvordan store prosjekter gjennomføres med sikte på å unngå overskridelser. Selv om mange av utfordringene som prosjekter opplever er relativt like, kan det synes som om mange av de feilene som begås gjentas over tid. Etterevalueringer av enkeltprosjekter har pekt på at det kan være behov for bedre systemer for tilgjengeliggjøring av data, analyse, og ikke minst erfaringsoverføring mellom prosjekter internt i og mellom sektorer. Det er selvsagt en mulighet for at prosjektene over tid har blitt større, mer komplekse og mer krevende, og med større bruk av ikke-standardiserte løsninger slik at risikoen for overskridelser øker.

Til slutt vil vi peke på at Tabell 5-3 også viser et annet potensielt interessant forhold. I årene før finanskrisen og med krevende markedsforhold¹ så vedtok Stortinget spesielt mange store prosjekter. Staten er en betydelig oppdragsgiver i fastlands-Norge og hva Statens vegvesen, Statsbygg, Bane NOR og de øvrige statlige etatene samlet foretar seg kan ha stor betydning for kapasiteten i markedet. Å «gi gass» når temperaturen i økonomien ellers er høy kan være ugunstig hvis man ønsker å unngå kostnadsoverskridelser.

¹ Byggekostnadsindeksen for veganlegg benyttes ofte som en indikator for temperaturen i fastlands-økonomien. Den måler prisutviklingen på innsatsfaktorene til bygging av veganlegg og benyttes for justering av rammer og kontrakter i ulike bransjer som jernbane, vannkraft, m.fl. I perioden 2001-2003 var den gjennomsnittlige årlige økningen i indeksen 1,9 prosent, i 2004-2008 6,2 prosent, og i 2009-2012 var den 4,1 prosent

6 Estimaterne i KS2-rapportene

Dette kapitlet undersøker selve estimatene som er gitt som grunnlag for å utarbeide styrings- og kostnadsrammene. Det vil være et naturlig overlapp med analysene gitt i det foregående kapitlet, men estimatene er ikke alltid identiske med de vedtatte rammene. I tillegg vil kapitlet gi en analyse av andre egenskaper ved estimatene, i særlig grad i hvilken grad estimeringsarbeidet klarer å fange opp variasjonen i usikkerhet mellom ulike prosjekter. Til dette har vi tatt i bruk estimatene i KS2-rapportene², herunder grunnkalkyle/basisestimat (som begge antas å ha blitt brukt om omtrent det samme estimatgrunnlaget); P10, P50, P85, P90; og forventningsverdi. Analysene er basert på datainnhentingene beskrevet i Kapittel 3 og bruker metoderammeverket beskrevet i Kapittel 4.3. For 14 av prosjektene var det ikke mulig å få tak i estimeringsdata, slik at estimeringsanalysen er basert på færre (N = 70) prosjekter enn i analysen av styrings- og kostnadsrammene. Alle kostnadstall er justerte for kostnadsøkning i perioden, slik at vi opererer med kroneverdier som gjenspeiler samme kostnadsår for samme prosjekt. Når absolutte kostnadsverdier sammenlignes på tvers av prosjekter, vil kronebeløp med ulik verdi likevel sammenlignes og vi vil derfor ha hovedvekten på relative, f. eks. prosentvise, avvik i analysen.

Kapitlet består av følgende underkapitler:

- Kapittel 6.1 analyserer avvik og skjevheter til P50-estimatene, samt diskuterer i hvilken grad justeringene fra P50 til styringsrammen har økt eller minnet kostnadsavvik og/eller skjevheter.
- Kapittel 6.2 analyserer hvor godt kalibrerte og hvor informative P50- og P85-estimatene og de totale estimatfordelingene er.
- Kapittel 6.3 analyserer forskjeller mellom etater, konsultentselskap ansvarlige for KS-estimatene og utvikling over tid.

² KS2-rapportene er tilgjengelig fra Concepts hjemmeside:

<https://www.ntnu.no/concept/ks-rapporter>

- Kapittel 6.4 analyserer hvor gode usikkerhetsanalysene er målt mot enkle påslagsmodeller, og innsparingspotensialet ved å bli bedre på å skille mellom prosjekter med høy og med lav kostnadsusikkerhet.

6.1 P50 som punktestimat: Avvik og skjevhet

I dette kapitlet vurderer vi ulike målestørrelser for avvik og skjevhet for P50-estimatene. Det er her også naturlig å samtidig se på styringsrammene og forholdet mellom P50 og styringsrammene. Sammenligner vi med styringsrammene er det små forskjeller i relativ estimeringsskjevhet. Median skjevhet i kroner er noe høyere for P50 (25 millioner) enn styringsrammen (18 millioner), men det mest iøynefallende resultatet er at prosjektporteføljen totalt sett har en overskridelse på 5422 millioner når vi sammenligner med styringsrammene vedtatt av Stortinget, mot 1692 når vi sammenligner med P50-estimatene fra KS2-rapportene

Tabell 6-1 viser resultater fra analysen av estimeringsavvik og skjevhet for P50-estimatene og styringsramme for de samme prosjektene. Det er kun tatt med prosjekter der både estimer, rammer og faktisk kostnad finnes (N=67)

Det gjennomsnittlig absolutte estimeringsavviket var på 216 millioner, det vil si at estimatene i gjennomsnitt ligger 216 millioner over eller under de faktiske kostnadene. For prosentvise absolutte avvik var gjennomsnittet på 14 prosent og medianavviket på 12 prosent. Litt over 50 prosent av prosjektene hadde avvik på 10 prosent eller mer, omtrent en fjerdedel hadde avvik på 20 prosent eller mer og 15 prosent hadde avvik på 30 prosent eller mer.

Estimeringsavvikene for P50 var relativt lik avvikene fra styringsrammene.

Ser vi på estimeringsskjevhet er median estimatskjevhet 25 millioner, altså er det slik at det midterste prosjektet i fordelingen av estimeringsfeil hadde faktiske kostnader 25 millioner over P50. For prosentvise avvik tilsvarte medianen 3 prosent høyere kostnader enn estimert P50. Dersom man vektet relative over- og underskridelser likt var medianavviket negativt, på -1 prosent. Midten av fordelingen av denne type estimeringsfeil viser altså underskridelse heller enn overskridelse. Totaloverskridelsen for hele prosjektporteføljen var på 1692 millioner, men merk at P50 ikke er et estimat som minimerer total overskridelse. For å minimere forskjellen mellom summen av kostnader og summen av estimer, burde forventningsverdien blitt brukt som estimat (se Kapittel 4.3).

Sammenligner vi med styringsrammene er det små forskjeller i relativ estimeringsskjevhet. Median skjevhet i kroner er noe høyere for P50 (25 millioner) enn styringsrammen (18 millioner), men det mest iøynefallende resultatet er at prosjektporteføljen totalt sett har en overskridelse på 5422 millioner når vi sammenligner med styringsrammene vedtatt av Stortinget, mot 1692 når vi sammenligner med P50 estimatene fra KS2-rapportene.

Tabell 6-1: Analyse av estimeringsavvik og skjevhet

Evaluerings	Målestørrelse (Md = median)	Resultat
Estimeringsavvik		
Gjennomsnittlig estimeringsavvik	$\frac{1}{N} \sum kostnad - P50_Estimat $	216 mill. kr (P50-estimat) 194 mill. kr (styringsramme)
Gjennomsnittlig relativt avvik (asymmetrisk)	$\frac{1}{N} \sum \frac{ kostnad - P50_Estimat }{P50_Estimat}$	14 % (P50-estimat) 14 % (styringsramme)
Median relativt avvik (asymmetrisk)	$Md\left(\frac{ kostnad - P50_Estimat }{P50_Estimat}\right)$	12 % (P50-estimat) 12 % (styringsramme)
Median relativt avvik (symmetrisk)	$e^{Md(\ln(\frac{act}{est}))} - 1$	12 % (P50-estimat) 12 % (styringsramme)
Andel unøyaktige estimater (>10%, >20%, >50% relativt avvik)	$\frac{1}{N} \sum \left\{ \begin{array}{l} 1, \text{ hvis } \frac{ kostnad - P50_Estimat }{P50_Estimat} \geq c \\ 0, \text{ ellers} \end{array} \right.$, der c=0.1, 0.2 and 0.3	c=0.1 52 % (P50-est.) 55 % (styrings.)
		c=0.2 27 % (P50-est.) 24 % (styrings.)
		c=0.3 15 % (P50-est.) 13 % (styrings.)
Estimeringsskjevhet		
Median estimatskjevhet	$Md(kostnad - P50_Estimat)$	25 mill. kr (P50-estimat) 18 mill. kr (styringsramme)
Median asymmetrisk relativ estimatskjevhet	$Md\left(\frac{kostnad - P50_estimat}{P50_estimat}\right)$	3 % (P50-estimat) 2 % (styringsramme)
Symmetrisk relativt estimatskjevhet	$sg * e^{\left Md\left(\ln\left(\frac{kostnad}{P50_estimat}\right)\right)\right } - 1$, Hvor sg er fortegnet til $Md\left(\ln\left(\frac{kostnad}{P50_estimat}\right)\right)$	-1 % (P50-estimat) 3 % (styringsramme)
Summert estimatskjevhet (totaloverskridelse for hele porteføljen)	$\Sigma kostnad - \Sigma estimat$	1692 mill. kr (P50-estimat) 5422 mill. kr (styringsramme)

Justeringene fra estimat til ramme utgjør samlet en forskjell på 5422 – 1692 = 3730 millioner kroner. Dette betyr at overskridelsene i totalkostnader ville ha vært ca. 3,7 milliarder kroner lavere dersom de opprinnelige P50-estimatene hadde vært beholdt.

P50-estimatene var i gjennomsnitt på 1605 millioner kroner, det vil si noe høyere enn gjennomsnittet til styringsrammene, som for de samme prosjektene var på 1502 millioner. Dette tilsvarer at det i sum skjer en reduksjon på 7 prosent fra P50-estimat til styringsramme. 72 prosent av prosjektene fikk en reduksjon fra P50-estimat til styringsramme og 28 prosent en økning.

Tilsvarende gjaldt forøvrig også justeringer fra P85 til kostnadsramme. For P85-estimatene var gjennomsnittet på 1800 millioner kroner, noe som er 7 prosent høyere enn gjennomsnittlig kostnadsramme på 1682 millioner. 54 prosent fikk en reduksjon, 7 prosent fikk samme kostnadsramme som P85, og 39 prosent fikk en økning i kostnadsrammen.

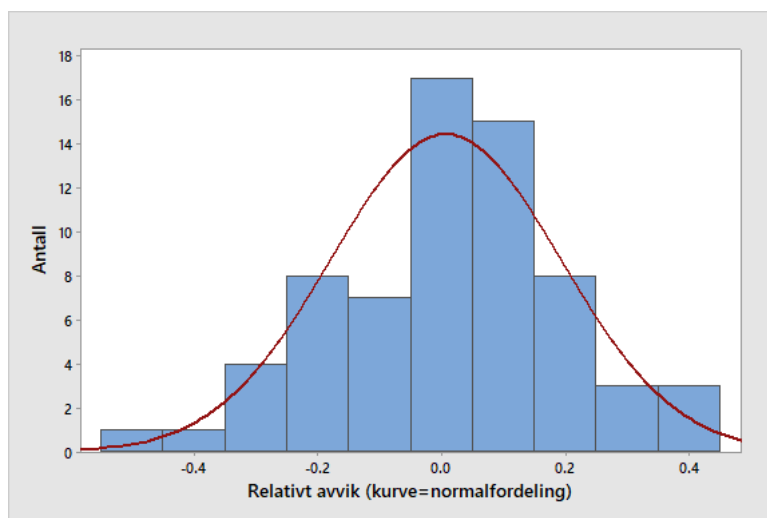
Det skjer altså typisk en reduksjon fra estimatene både for P50 og P85 til henholdsvis styrings- og kostnadsramme, og det er da rimelig å tenke at reduksjoner fra P50-estimat til ramme medfører hyppigere overskridelser. Dette er det imidlertid ikke grunnlag for å hevde. En gjennomgang av de største justeringene finner at like mange av disse har økt som redusert realismen, ved at justeringene ikke har redusert sannsynligheten for over- eller underskridelser.

For eksempel var P50-estimatene for prosjektet «Panserbekjempelse – middels rekkevidde (PBM)» på 1488 millioner kroner, mens styringsrammen ble på 991 millioner. Faktisk kostnad var på 726 millioner, som er mer enn 50 prosent lavere enn P50-estimatet. Årsaken til det var at P50 var basert på en vesentlig høyere dollarkurs enn det utbetalingene ble gjennomført til. Der var det åpenbart gode grunner til å redusere fra P50-estimatet til styringsrammen. Det finnes også eksempler på det motsatte. «Nytt operahus», hadde for eksempel et P50-estimatet som var på 4140 millioner, styringsrammen ble mye lavere, 3544 millioner, mens faktisk kostnad ble på 4279 millioner.

Totalt sett økte realismen i estimatene etter justeringer i 31 av prosjektene, var den samme i 5, og ble redusert i 31. Annen forskning på justeringer av estimater tyder på at det å etter-justere på estimater stort sett gjør dem mindre nøyaktige, og at det er kun når det er særlige forhold som nødvendiggjør det at

det bør gjøres. Selv om dette ikke var tilfelle her, så peker dette likevel på viktighet av å kreve en god argumentasjon og dokumentasjon av grunnene for å justere på estimatene (Goodwin, 2000).

En grafisk visning (histogram) av fordelingen av relativt estimeringsavvik for P50-estimatene, se Figur 6-1, tyder på en relativt symmetrisk fordeling (svakt høyreskjev dersom vi vektlegger at median er noe lavere enn forventningsverdi, men svakt venstreskjev dersom vi vektlegger den lengre halen til venstre). Dette avviker fra hva som synes å være vanlig i prosjektsammenheng, der høyreskjeve avviksfordelinger er typiske (Halkjelsvik og Jørgensen, 2018). Det at fordelingen av relativt estimeringsavvik i våre analyser er nær symmetrisk, sammen med at avvikene er relativt små, betyr at det ikke gjør så stor forskjell om vi bruker asymmetrisk eller symmetrisk målestørrelser på estimeringsfeil og skjevhet. Dette forsvarer en bruk av de tradisjonelle prosentbaserte (asymmetriske) målestørrelsene for estimeringsfeil og skjevhet i analysene.



Figur 6-1: Histogram over relativt estimeringsfeil

6.2 Prediksjonsintervall og estimatfordelinger: Kalibrering og informasjonsverdi

Tabell 6-2 og figurene i dette kapittelet rapporterer hvor godt kalibrerte og informative prediksjonsintervallene og estimatfordelingene er. Estimatomfordelingene er avledet fra P10-estimatene, forventningsverdiene og

P90-estimatene angitt i KS2-analysene, samt antagelsen om at estimatfordelingen er tilnærmet log-normal (til dette brukte vi programtillegget @Risk og funksjonen «RiskLogNormAlb»).

Tabell 6-2: Kalibrering og informasjonsverdi (N=69)

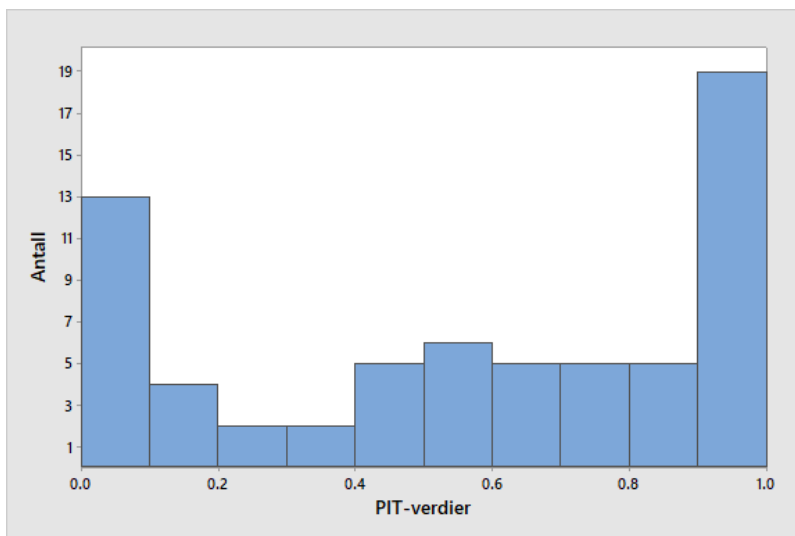
Evalueringskriterium	Målestørrelse	Resultat
Kalibrering		
Treffrate for PX-estimer	$\frac{1}{n} \sum_i h_i, \quad h_i = \begin{cases} 1, & \text{kostnad}_i \leq PX \\ 0, & \text{ellers} \end{cases}$ <p>p-verdi angir sannsynlighet for å finne forskjellen vi har funnet (eller større forskjeller), gitt at faktisk andel er den normative (binomial test).</p>	P10: 19% (p=0.02)
		P50: 42% (p=0.22)
		P85: 75% (p=0.04)
		P90: 80% (p<0.01)
Kalibrering av estimatfordeling som helhet (PIT-verdier)	$F_i(\text{kostnad}_i)$, der F_i er estimatfordelingen for prosjekt i.	Se Figur 6-2 for histogram av PIT-verdier
Informasjonsverdi		
Gjennomsnittlig relativ bredde (80% PI)	$\frac{1}{N} \sum \frac{(P90 - P10)}{P50}$	0.29 (standardavvik: 0.10) Se Figur 6-3 for histogram av relativ intervallbredde.
Korrelasjon mellom intervall-bredde og estimeringsavvik	τ (Relativ bredde for 80% PI, relativt avvik), Spearmans rang-korrelasjonskoeffisient τ .	-0.01 (p=0.94) Se Figur 6-4 for diagram
Korrelasjon mellom intervall-bredde og estimeringsskjevhet	τ (Relativ bredde for 80% PI, relativ estimatskjevhet), Spearmans rang-korrelasjonskoeffisient τ .	-0.11 (p=0.364) Se Figur 6-5 for diagram
Andre /kombinerte egenskaper		
Gjennomsnittlig skjevhet/symmetri i intervall	$\frac{1}{N} \sum \frac{(P90 - P50)}{(P50 - P10)} - 1$	0.04 (standardavvik 0.13) Se Figur 6-6 for histogram
Gjennomsnittlig CRPS-verdi (kalibrering + konsentrasjon) til estimatfordeling som helhet	$CRPS(F_i, y_i) = \int_{\mathbb{R}} (F(z) - \mathbb{I}(\text{kostnad} \leq z))^2 dz,$ <p>der $\mathbb{I}(\text{kostnad} \leq z)$ er «identifikator-funksjon» som 1 dersom $\text{kostnad} \leq z$, ellers 0. (Ingen triviell interpretasjon, vil kun brukes til senere sammenligninger)</p>	167 (standardavvik 214)

Kalibrering av PX-estimatene og estimatfordelingene

Et PX-estimat er, som tidligere beskrevet, et prediksjonsintervall fra 0 til PX-estimatet der vi anslår det å være X-prosent sannsynlig at virkelig verdi skal være innenfor intervallet. Dersom disse er godt kalibrerte vil vi altså finne 10 prosent, 50 prosent, 85 prosent og 90 prosent av prosjektene innenfor henholdsvis P10, P50, P85 og P90-estimatene.

Som vi ser i tabellen så er P10-estimatene generelt satt for høyt i utvalget sett under ett, siden nesten 20 prosent av prosjektene ender opp under P10-estimatet ($p=0.02$). P50-estimatet er satt noe for lavt og inneholder 42 i stedet for 50 prosent, men feilen her kan godt skyldes tilfeldig variasjoner ($p=0.22$). Det er derimot mindre sannsynlig at tilfeldigheter gjorde det slik at kun 75 prosent av P85-estimatene inneholdt faktisk kostnad ($p=0.04$), og det er svært usannsynlig at bare 80 prosent av prosjektene er under P90-estimatet dersom disse var godt kalibrerte ($p<0.01$).

Dette bildet blir enda tydeligere i histogrammet over PIT-verdier (se Figur 6-2). Perfekt kalibrerte estimatfordelinger vil være uniformt fordelte (samme verdier på alle søylene), mens vi ser at det er altfor mange i den første og siste søylen. Dette viser med stor tydelighet at usikkerheten i kostnader er undervurdert i estimatene. Dette gjelder både forholdet at prosjekter for ofte koster mye mindre enn P10-estimatene, men enda mer at prosjekter for ofte koster mer enn P85 og P90-estimatene. Denne undervurderingen av usikkerhet samsvarer godt med resultater fra litteraturen på usikkerhetsvurderinger, hvor man typisk observerer for smale prediksjonsintervaller sammenlignet med faktiske utfall. Se diskusjon av dette i Kapittel 2.3.



Figur 6-2: Kalibrering til estimatfordelingen (uniform fordeling = perfekt kalibrering)

Dersom vi bruker P10 og P90-estimatene for å lage 80 prosent prediksjonsintervaller (andel innenfor P90-estimatene minus andel innenfor P10-estimatene) ser vi at bare 61 prosent av prosjektene havner innenfor dette intervallet, og ikke 80 prosent. Dette er en statistisk signifikant forskjell ($p < 0.001$), og er igjen et uttrykk for at fordelingene er for smale i forhold til den faktiske usikkerheten i prosjektene.

Hva denne undervurderingen av usikkerhet er drøftet i ulike studier om usikkerhetsvurderinger (Halkjelsvik og Jørgensen, 2018). Mye tyder på at «ignoranse» har betydning. Det vil si at den som estimerer kostnader på aktivitetsnivå ikke har noe forhold til hva det vil si å være for eksempel 90 prosent sikker. Det er ikke enkelt å gå fra erfaring med tidligere kostnader til å vite hvor man bør legge et P90-estimat. Det vil da kunne være andre hensyn som teller mer, som at man vil forsøke å være mest mulig informativ. Svært vide, men mer realistiske, prediksjonsintervaller er lite informative, og vil dermed unngås.

Et tiltak som kan forbedre realismen er å gå den tradisjonelle måten å be om minimum (for eksempel P10-estimer) og maksimum (for eksempel P90-estimer) estimer, er bruk av den alternative metoden beskrevet i Jørgensen (2004) og Halkjelsvik og Jørgensen (2018). Metoden bygger på følgende steg:

- Estimer mest sannsynlig kostnad.

- Avgjør hvor ofte, for lignende aktiviteter, faktisk kostnad er blitt overskredet med for eksempel 50 prosent.
- Avgjør hvor ofte, for lignende aktiviteter, faktisk kostnad er blitt underskredet med for eksempel 30 prosent.

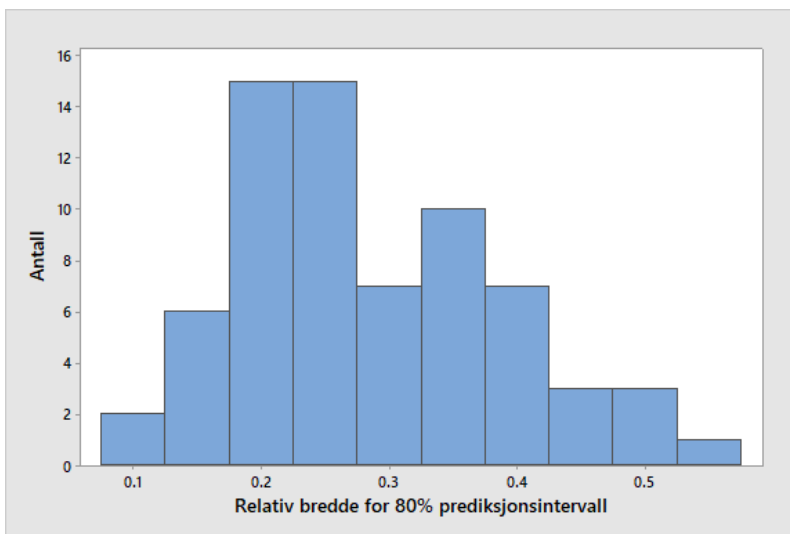
Dersom en aktivitet er estimert til mest sannsynlig å koste 100.000 kroner, og man tror at faktiske kostnader i 20 prosent av tidligere tilfeller, for lignende aktiviteter, har blitt overskredet med mer enn 50 prosent, så har man at P80-estimatet er på 150 prosent av 100.000, det vil si at P80-estimatet bør være på 150.000 kroner. Tilsvarende, om man tror at man har underskredet med mer enn 30 prosent i 5 prosent av tilfellene, så har man at P5-estimatet er på 70.000 kroner.

Styrken til denne metoden er at man bedre vil kunne overføre tidligere erfaringer til realistiske prediksjonsintervaller, og slipper å stole på en ofte lite pålitelig, ofte intuisjonsbasert, direkte ekspertvurdering av P10 og P90-estimer. Se Halkjelsvik og Jørgensen (2018, s. 74-78) for mer om metoden.

Andre mulige årsaker til undervurdering av usikkerhet ligger i hvordan prediksjonsintervaller aggregeres (summeres). Om ikke avhengigheter, for eksempel i form av korrelasjon, tas tilstrekkelig hensyn til, så vil dette også føre til at blant annet P10 og P90-estimer på aggregert nivå (sum-nivå) vil undervurderes. Vi har ikke hatt tilgang til faktisk aggregeringsprosess. En gjennomgang av hvordan aggregeringen gjøres vil kunne bidra til med kunnskap om årsak til at prediksjonsintervallene blir for smale. En egnet metode for å ta finne frem til realistiske korrelasjoner mellom aktiviteter er beskrevet i Cho (2006).

Relativ bredde for 80 prosent prediksjonsintervall

Figur 6-3 viser fordelingen av prosjektenes relative bredde for 80 prosent prediksjonsintervall, det vil si differansen mellom P90 og P10, delt på P50. Figuren viser at estimatfordelingene til prosjektene varierer en del med hensyn til hvor usikre prosjektene antas å være. Dette viser blant annet at de som estimerer ikke opererer klimatologisk, altså er det ikke slik at de opererer med omtrent samme prosentvise risikopåslag på grunnkalkyle for å finne P50 og P85 uavhengig av prosjekttype og kontekst. De neste to analysene undersøker om denne variasjonen i estimert usikkerhet, vist ved bredere estimatfordeling, korrelerer med faktisk usikkerhet (vist ved økt gjennomsnittlig estimatavvik).

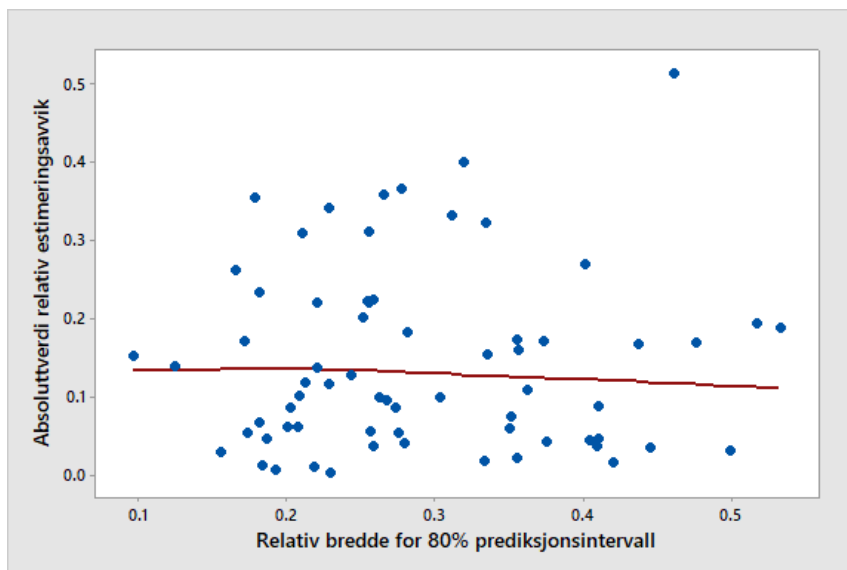


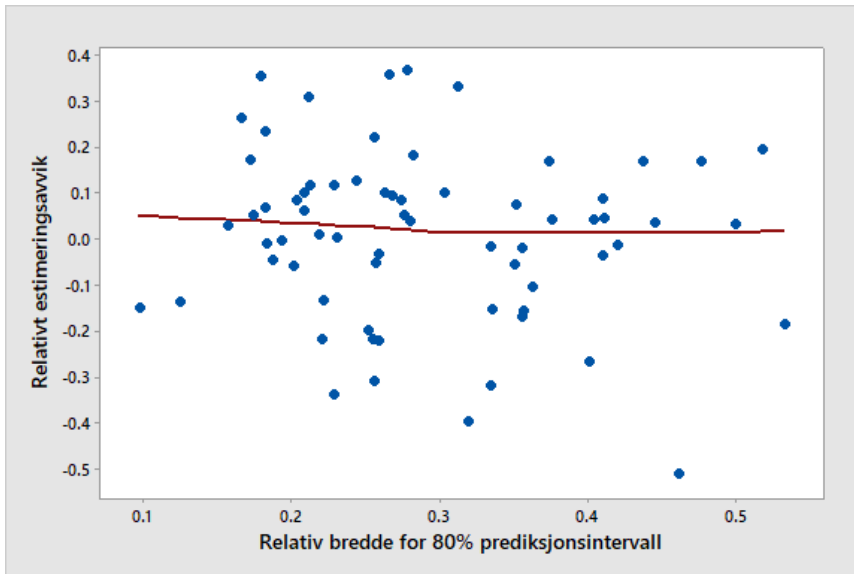
Figur 6-3: Fordeling bredde til 80 prosent prediksjonsintervallene

Informasjonsverdi

Som argumentert tidligere er det fullt mulig å oppnå god kalibrering uten at man klarer å skille mellom prosjekter med høy og lav usikkerhet. For KS2-prosjektene vil man for eksempel kunne oppnå perfekt kalibrerte P50 og P85-estimerer (treffrate på 50 prosent og 85 prosent) ved *alltid* å legge til henholdsvis ca. 13 prosent og 38 prosent på kostnadsestimatet i grunnkalkylen. Dette vil imidlertid gi for lave påslag for prosjekter med høy og for høye påslag for prosjekter med lav usikkerhet, med andre ord gi liten informasjon om faktisk behov for usikkerhetspåslag for ulike prosjekter.

I Tabell 6-2 og i Figur 6-4 har vi sett på i hvilken grad det er korrelasjon mellom relativ bredde for 80 prosent prediksjonsintervallene og estimeringsavvik (relativ forskjell mellom P50 og faktisk kostnad). Dersom estimatene er i stand til å skille mellom høy og lav-risikoprojekter, så vil vi forvente en positiv korrelasjon her, det vil si at bredere 80 prosent prediksjonsintervaller var korrelert med i gjennomsnitt høyere estimeringsavvik. Gitt variasjon i utfall (faktisk kostnad er en stokastisk variabel) så vil vi ikke, selv ved perfekt estimering, kunne forvente 100 prosent korrelasjon, men trolig maksimalt ca. 30 prosent (se Kapittel 4.3). Det er likevel klart at en korrelasjon rundt 0 eller med negativ verdi tyder på dårlig evne til å skille mellom høy og lav-risikoprojekter. Dette er, som vist i tabellen ovenfor og i Figur 6-4 tilfelle, med en korrelasjon på kun -0.01.



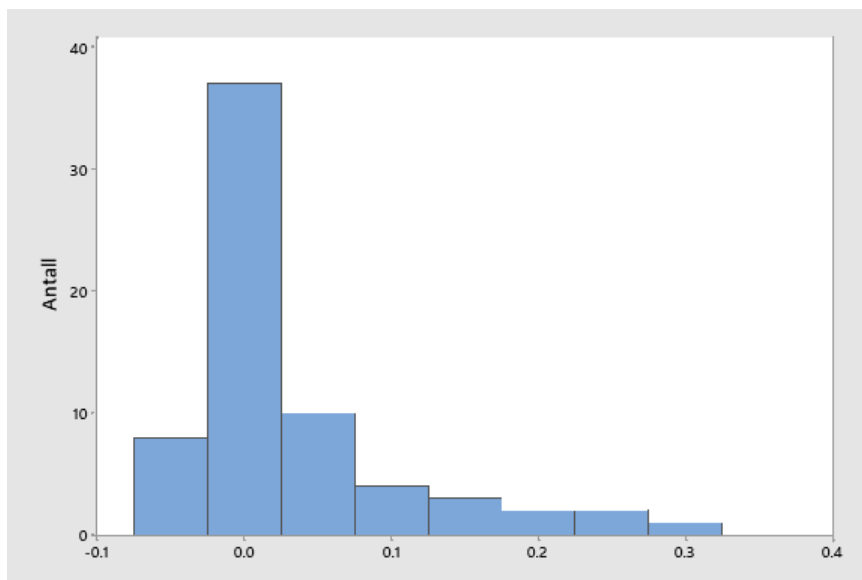


Figur 6-5: Estimert usikkerhet (intervallbredde) vs. prosentvis over/underskridelse

Denne observasjonen antyder også at selve estimatene vil kunne erstattes med en enkel påslagsmodell med utgangspunkt i grunnkalkyle eller lignende, og gjennom dette få bedre kalibrering med samme informasjonsverdi. Det kan imidlertid være andre egenskaper, for eksempel ved estimeringsprosessen, som gjør at man heller bør beholde dagens modell. Vi utarbeider en slik modell og tester denne opp mot KS-estimatene i Kapittel 6.4.

Symmetri

Tabellen ovenfor, og Figur 6-6 (to ekstremverdier - en høy og en lav, er utelatt), viser at de aller fleste 80 prosent prediksjonsintervallene er symmetriske rundt P50. Få av prosjektene (16 prosent) har for eksempel P90-estimatet som er mer enn 10 prosent lenger fra P50-estimatet enn P10-estimatet er fra det samme P50-estimatet. I lys av at faktisk estimatskjevhet (se Figur 6-1) er nokså symmetrisk så er dette trolig akseptabelt. Vi har for eksempel at andelen prosjekter med mer enn 25 prosent kostnadsunderskridelse, målt mot P50-estimatet, er 9 prosent, som er identisk med andel prosjekter med mer enn 25 prosent kostnads-overskridelse. Hovedproblemet er i dette tilfelle altså ikke at estimatfordelingene er for symmetriske, men at de (som viste tidligere) er for smale, og ikke klarer å skille prosjekter med lav og høy kostnadsrisiko.



Figur 6-6: Symmetri (rundt P50-estimatet) til 80% prediksjonsintervall

Den siste målestørrelsen vist i tabellen ovenfor, CRPS, integrerer informasjon om hvor godt estimatfordelingene er kalibrert og hvor konsentrert estimatfordeling er. Den er vanskelig å tolke som enkeltstående verdi, men vil bli brukt i senere analyser, der vi sammenligner etater, konsultentselskap, ser utvikling over tid og sammenligner ulike strategier for å anslå usikkerhet.

6.3 Forskjeller mellom etater, KS-rådgivere, sammenheng med prosjektstørrelse, og utvikling over tid

Analysene i dette kapitlet inkluderer kun målestørrelser der det er tilstrekkelig antall observasjoner per etat og KS-rådgivere til å gjøre meningsfulle analyser.

Forskjeller mellom etater

Tabell 6-3 viser treffraten for KS2-estimatene til de etatene som har mer enn fem prosjekter. Tabellen viser at det er Forsvaret som har størst gjennomsnittlig relativt estimatavvik. Dette skyldes imidlertid ikke kostnadsoverskridelser, men kostnadsunderskridelser, noe som fremgår av målingene av estimatskjevhet. Andel med kostnad lavere enn P10 er på hele 42 prosent. Hva dette skyldes, for eksempel om det er levert mindre enn

opprinnelig planlagt eller om estimatene generelt sett er for pessimistiske, krever mer detaljerte analyser. De andre etatene har KS-estimer med gjennomsnittlige avvik på 8 til 15 prosent, noe som må sies å være rimelig lavt, sett i lys av andre undersøkelser (se Kapittel 2.2).

Tabell 6-3: Forskjell mellom etater

	Forsvaret	Jernbane- verket/Bane NOR	Statens vegvesen	Statsbygg
N	7	8	35	12
Gjennomsnittlig estimeringsfeil (P50-estimat)	0.25	0.08	0.15	0.11
Median estimeringsskjevhet (P50-estimat)	-0.19	-0.01	0.03	0.05
Andel med kostnad lavere enn P10	42 %	0 %	17 %	17 %
Andel med kostnad lavere enn P90	71 %	100 %	74 %	83 %
Treffrate for 80 % prediksjonsintervall	29 %	100 %	57 %	67 %
Test av forskjell til 80 % (p-verdier)	0.005	0.37	0.002	0.27
Gjennomsnittlig relativ intervallbredde (P90-P10)/P50	0.29	0.33	0.27	0.28
Andel med kostnad lavere enn P50	57 %	38 %	43 %	33 %
Andel med kostnad lavere enn P85	71 %	100 %	69 %	75 %

Treffraten for 80-prosent prediksjonsintervallene (som optimalt sett bør ligge rundt 80 prosent) er kun 29 prosent for Forsvaret, 100 prosent for Jernbaneverket, 57 prosent for Statens vegvesen og 67 prosent for Statsbygg. Forskjellene mellom etatene kan være tilfeldige, ($p = 0.03$, chi-square test), men sannsynligheten for dette er lav.

Forsvaret og Statens vegvesen har altså vesentlig lavere treffrate for 80 prosent prediksjonsintervaller enn 80 prosent. Dersom vi tester den faktiske treffraten opp mot den normativt riktige 0,8 i binomiale tester, ser man at det er mindre enn 1 prosent sannsynlig å få slike eller mer ekstreme resultater av rene tilfeldigheter ($p < 0.01$). Dette er ikke tilfelle for de to andre etatene ($p > 0.25$). Man kunne kanskje tenke seg at for eksempel prediksjonsintervallene i

prosjektene til Forsvaret er smalere (og mer informative) og derfor kom dårligere ut, men det var ingen tydelige forskjeller mellom etatene med tanke på bredden av prediksjonsintervallene (se siste kolonne i tabellen ovenfor). Gjennomsnittlig lå relativ bredde for 80 prosent prediksjonsintervallene for de fire etatene mellom 0.27 og 0.33. Dette kan tyde på at prosjektene til Forsvaret har en mer feil-kalibrert estimatfordeling, for eksempel ved at prosjektene har større usikkerhet enn det er tatt høyde for i estimatfordelingen. Ser vi på andelen innenfor P50 og P85-estimatene ser vi at prosjektene til Forsvaret har rimelig godt kalibrert P50-estimer, men typisk for lave P85-estimer. Prosjektene til Statsbygg på den annen side har hatt både for lave P50 og P85-estimer.

Forskjeller mellom KS-rådgivere

Det er som regel konsulentselskapet, i rollen som kvalitetssikrer/KS-rådgiver, som har ansvaret for å utarbeide prediksjonsintervallene og estimatfordelingene. Tabell 6-4 viser om disse har ulik treffrate for estimatene de er ansvarlige for. Ser vi på 80-prosent prediksjonsintervallet som er basert på P10 og P90-estimatene var den beste treffraten blant fem firmaer 79 prosent og den dårligste 36 prosent. Selv om dette kan virke som store forskjeller er antall prosjekter små og en test på tvers av de fem firmaene viser at forskjellene kanskje skyldes tilfeldig variasjon ($p = 0,08$, chi-squared test).

Tabellen viser kun små forskjeller i gjennomsnittlig estimeringsfeil og heller ikke så store forskjeller i estimatskjevhet (ikke statistisk signifikant forskjell på gjennomsnittsverdier) mellom konsulentselskapene.

Vi kan for hvert konsulentselskap teste hvor sannsynlig det er at avvikene fra den normative treffraten på 80 prosent kan skyldes tilfeldigheter. Merk at det her kan være en fordel å ha få prosjekter, siden det er vanskelig å skille systematisk fra tilfeldig variasjon når utvalget er lite, det vil si at p-verdien i testen blir høy fordi det er få observasjoner og ikke fordi forskjellen til normativ andel på 80 prosent er lav. For de fleste selskapene er avvikene innenfor det som kan forventes av rene tilfeldigheter, men ett av selskapene (Metier) ser ut til å ha en statistisk signifikant for lav treffrate i forhold til hva som normativt kan forventes. Dette ser ut til å skyldes både for høye P10-estimat og for lave P90-estimat. Merk at vi ikke har tatt høyde for at det er forskjeller mellom firma med tanke på type prosjekter og andre faktorer, så man bør være varsom med tolkingen.

Tabell 6-4: Forskjell mellom KS-rådgivere

	DNV/ Advansia	Dovre	Holte	Metier	Terramar
N	9	19	15	19	7
Gjennomsnittlig estimeringsfeil (P50-estimat)	0.12	0.16	0.12	0.17	0.13
Median estimeringsskjevhet (P50-estimat)	0.10	-0.02	-0.01	0.05	0.09
Andel med kostnad lavere enn P10	11 %	16 %	20 %	32 %	0 %
Andel med kostnad lavere enn P90	67 %	95 %	93 %	68 %	57 %
80% prediksjonsintervall	56 %	79 %	63 %	36 %	57 %
Test av forskjell til 80 % (p- verdier)	0.22	1.00	0.75	<0.001	0.36
Gjennomsnittlig relativ bredde (P90-P10)/P50	0.22	0.40	0.26	0.23	0.28
Andel med kostnad lavere enn P50	11 %	57 %	53 %	42 %	14 %
Andel med kostnad lavere enn P85	55 %	95 %	87 %	68 %	43 %

Merknad: Dovre inkluderer også Dovre/Transportøkonomisk Institutt, Terramar inkluderer også Terramar/Asplan Viak.

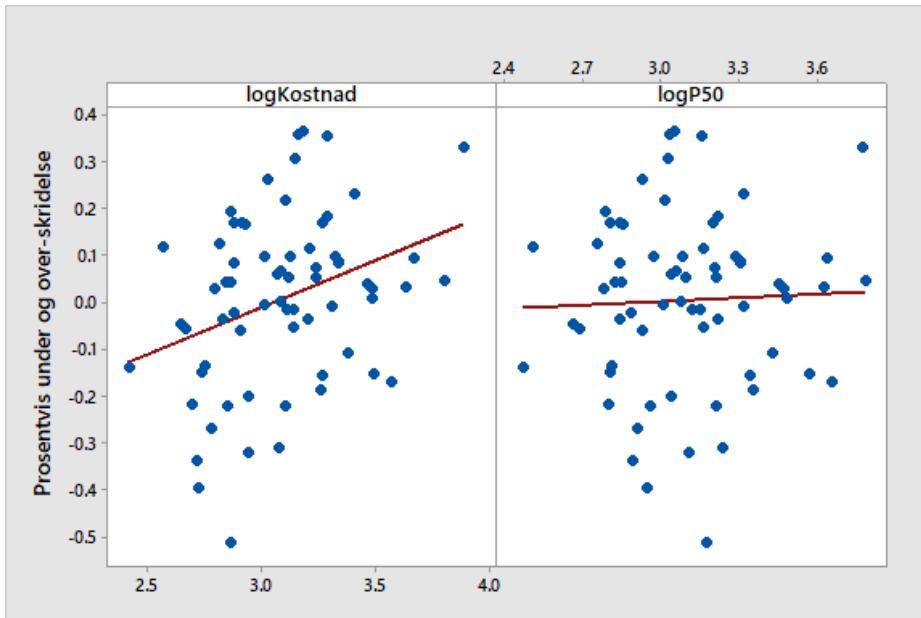
Gjennomsnittlig relativ bredde av 80 prosent prediksjonsintervallene var ganske lik for konsulentselskapene, med unntak av Dovre som hadde vesentlig bredere intervaller. Det er mulig at noen selskap har ansvar for mer usikre typer prosjekter, men det er også mulig at noen selskaper har estimeringsmetoder som gir bredere fordeling av estimater.

Prosjektstørrelse

Størrelsen på prosjektene er en annen faktor som kanskje kan påvirke grad av overskridelser. Spørsmålet om store prosjekter er mer utsatt for kostnadsoverskridelser enn små prosjekter høres i utgangspunktet ut som et enkelt spørsmål, men svaret avhenger av hva man mer konkret legger i dette. Ett spørsmål er om prosjekter som har store kostnader ofte er underestimert. Et annet spørsmål er om prosjekter med høyere estimat ofte er underestimert. Og et siste spørsmål er om den iboende størrelsen til et prosjekt kan knyttes til underestimert og overskridelser. En måte å se for seg et prosjekts iboende størrelse er om man utførte samme prosjekt en rekke ganger og fant

gjennomsnittet av disse (altså størrelsen til prosjektet med gjennomsnittlig komplikasjoner, gjennomsnittlig flaks osv.). Det er kanskje sistnevnte problemstilling som er interessant, fordi den kan legge føringer for strukturering av et prosjekt (gitt at sammenhengen mellom størrelse på prosjektet og sannsynligheten for overskridelser er kausal). For eksempel kan man tenke seg at det i noen tilfeller er mulig å dele opp prosjekter i mindre uavhengige prosjekter eller samle relaterte prosjekter under felles administrasjon med en felles arbeidsstyrke som kan allokeres mer dynamisk enn i et enkeltstående prosjekt. Dessverre er det vanskelig å studere sammenhengen mellom et prosjekts iboende størrelse og kostnadsoverskridelser, fordi dette ville kreve eksperimentelle forskningsmetoder (mer om dette under). Forskningslitteraturen har derimot kun omfattet de to første problemstillingene, uten å skille mellom dem, og har gjerne formidlet resultatene som om det var prosjektenes iboende størrelse som ble undersøkt.

Jørgensen m.fl. (2012) samlet originaldata fra en rekke studier som hadde studert overskridelser og prosjektstørrelse og viste hvordan forholdet mellom kostnadsoverskridelser og prosjektstørrelse endrer seg etter hvordan prosjektstørrelse blir definert (operasjonalisert). Dersom man måler prosjektstørrelse som faktiske kostnader vil man typisk finne høyere kostnadsoverskridelser for større prosjekter. Dersom man måler prosjektstørrelse som estimerte kostnader finner man ofte ingen eller en motsatt effekt. Vi kan se denne tendensen også i KS2-dataene. Figur 6-7 viser prosentvise avvik for alle prosjektene mot enten faktiske kostnader (graf til venstre) eller estimerte (P50) kostnader (graf til høyre), med regresjonslinjer. Her har vi log-transformert kostnader og P50-estimatene fordi antall prosjekter varierer med størrelsen (det er få veldig store prosjekter i utvalget).



Figur 6-7: Sammenheng mellom prosjektstørrelse og estimatskjevhet avhengig av om størrelse er definer ved hjelp av faktisk kostnad (graf til venstre) eller estimert kostnad (graf til høyre).

Svaret på spørsmålet om store KS2-prosjekter har større overskridelser enn mindre KS2-prosjekter er dermed *ja* dersom man definerer et stort prosjekt som et prosjekt med store kostnader, men *nei* dersom man definerer et stort prosjekt som et prosjekt med høye estimerte kostnader. Men hva er mest meningsfylt, og hvordan henger dette sammen med vår tredje måte å formulere spørsmålet om størrelse og overskridelser på, som handlet noe mystisk om den iboende størrelsen til et prosjekt?

Det at prosjekter som viser seg å få store kostnader ender opp med å få overskridelser er på mange måter ganske opplagt, siden større overskridelser fører til mer kostnader og dermed skaper en sammenheng mellom overskridelser og kostnader (se Jørgensen og Halkjelsvik, 2018). Altså vil stokastisk variasjon i kostnader direkte være forbundet med variasjon i kostnadsoverskridelser (med andre ord er det en direkte matematisk kobling mellom kostnader og kostnader minus styringsramme). Det vil derfor virke som om størrelse operasjonalisert som faktiske kostnader ikke er en god måte å se på forholdet mellom størrelse og overskridelser på. Det som derimot kanskje ikke er så opplagt, er at det også kan være problematisk å bruke estimerte kostnader for å studere forholdet mellom størrelse og overskridelser.

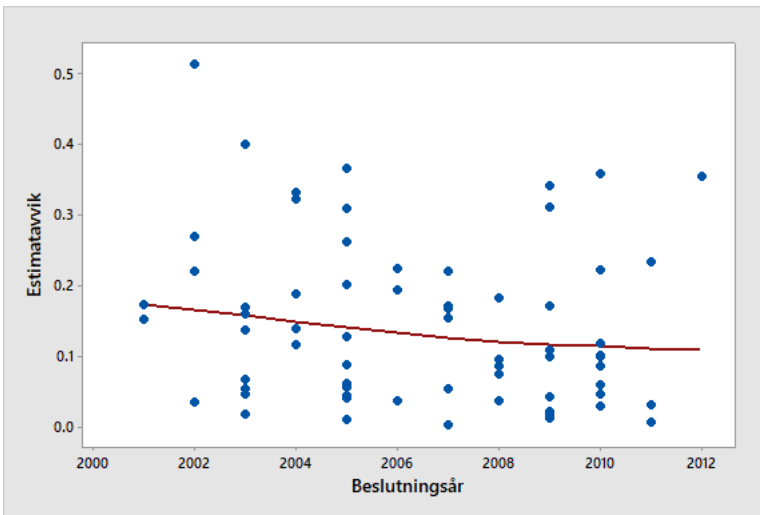
På samme som faktiske kostnader, er også estimerte kostnader en stokastisk variabel som varierer med for eksempel hvem som gjør estimeringsarbeidet, hvilke konsulteselskap som kvalitetssikrer og hvor godt prosjektet er planlagt. Siden det også er en matematisk kobling mellom prosjektstørrelse definert som P50 og overskridelser (siden sistnevnte er kostnader minus P50) vil man her også få en kunstig sammenheng der prosjekter som av en eller annen grunn er for høyt estimert både har mindre overskridelser og blir definert som et større prosjekt enn det burde være. Altså kan vi statistisk sett vente en negativ sammenheng mellom størrelse og overskridelser dersom estimatet eller styringsrammen definerer hva vi mener med et større prosjekt. I tilfellet med KS2-data som vist i figuren ovenfor, kan det tenkes at en slik negativ effekt utligner en underliggende reell effekt av prosjektstørrelse.

Selv om vi ikke har fått svar på hva som er det egentlige forholdet mellom den iboende prosjektstørrelsen og overskridelser i analysene over, kan vi anta at forholdet er et sted mellom de to regresjonslinjene. Det kan da se ut som at det er en svak sammenheng med noe høyere overskridelser for store prosjekter. Det som er problemet med denne slutningen er at prosjektene er selektert inn i KS-ordningen på bakgrunn av estimerte kostnader og ikke faktiske kostnader. Eventuelle prosjekter med estimerte kostnader som er lavere enn grensen for å bli med i KS-ordningen, men med reelle kostnader som er over KS-grensen vil ikke bli med i datasettet. Denne seleksjonen vil teoretisk sett føre til at sammenhengene i figuren er svakere enn de ville vært om alle prosjektstørrelser hadde vært inkludert. Det er også velkjent at tilfeldig variasjon i avhengige variable reduserer effekten (lavere b-verdi i regresjonslinjen). Se Jørgensen m.fl. (2012) for mer om dette.

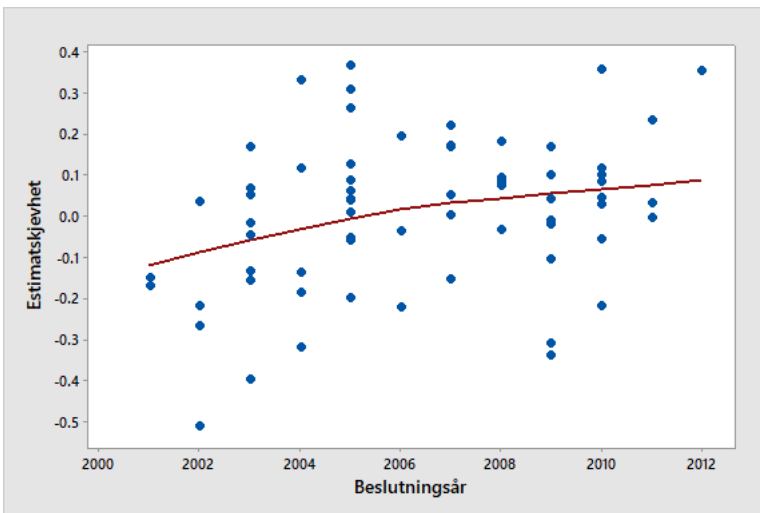
Det er med andre ord vanskelig å nærme seg spørsmålet om den iboende størrelse på et prosjekt har noe å si for overskridelser, men det kan se ut som en eventuell sammenheng er positiv, altså at det faktisk er større kostnadsoverskridelser for større prosjekter uavhengig om vi måler størrelse som estimert eller faktisk kostnad. Merk at det da ikke nødvendigvis er snakk om en kausal sammenheng der for eksempel størrelsen fører til dårligere kostnadsstyring. Det kan være slik at de største prosjektene er fundamentalt forskjellige fra de mindre prosjektene med tanke på andre faktorer som for eksempel kompleksiteten i arbeidet.

Utvikling over tid

Figurene nedenfor viser estimatavvik og skjevhet målt mot P50-estimatene for prosjekter med besluttet oppstart i årene 2001 til 2012. For alle grafene så er linjene generert ved en såkalt loess-funksjon (som i stedet for å finne en rett linje gjennom hele datasettet heller setter sammen vektete ikke-lineære regresjoner for å kunne illustrere mer kortsiktige trender).



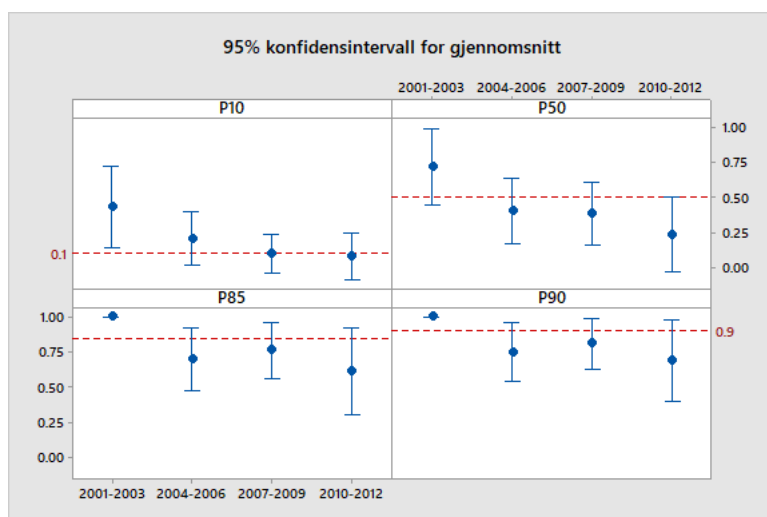
Figur 6-8: Utvikling av estimatavvik over tid



Figur 6-9: Utvikling av estimatskjevhet (under og overskridelser) over tid

Figurene tyder på at mens det har vært en svak nedgang i estimatavvik (en lineær regresjonsanalyse gir en statistisk ikke-signifikant nedgang, $p = 0.18$), så har det vært en mer tydelig tendens fra typiske underskridelser til overskridelser (en lineær regresjonsanalyse gir en statistisk signifikant økning, $p = 0.008$). Det siste resultatet samsvarer med resultatene rapportert i Kapittel 5.2 for styringsramme.

Figur 6-10 viser andelen av prosjekter med faktiske kostnader under de ulike ensidige prediksjonsintervallene (PX-estimatene), og hvordan disse andelen utvikler seg over tid. De stiplede linjene tilsvarer andelen man ville forvente med perfekt kalibrering av prediksjonsintervallene. Det er henholdsvis 14, 20, 21 og 13 prosjekter for tidsintervallene 2001-2003, 2004-2006, 2007-2009 og 2010-2012 (beslutningsår).



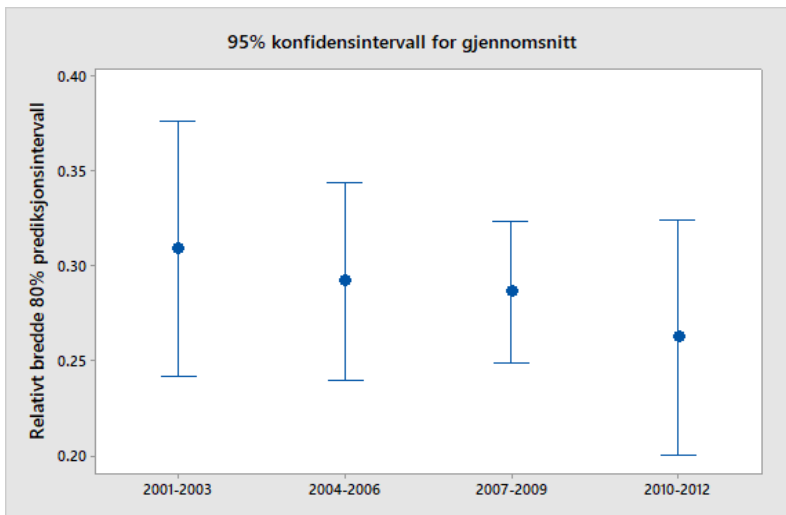
Figur 6-10: Utvikling av andel prosjekter under P10, P50, P85 og P90-estimatene over tid

Figuren tyder på at P10-estimatene har gått fra å ha vært for høy til å bli bedre kalibrert over tid. P50-estimatene har gått fra å være for høye til å være for lave (som tidligere vist i

Figur 6-9) og kun 23 prosent av faktisk kostnad var innenfor i siste tidsperiode. P85 og P90-estimatene var noe for høye i første periode, mens de de senere tidsperiodene har vært for lave. For eksempel var kun 62 prosent av faktisk kostnad innenfor P85-estimatene i siste tidsperiode. Totalt sett tyder figuren på at det i perioden 2001-2012 har vært en dreining fra for høye til

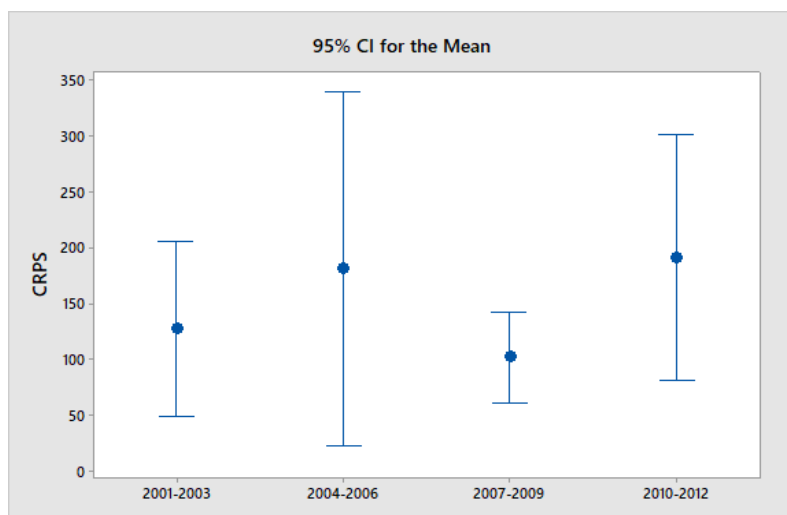
lavere verdier (relativt til faktisk kostnad) for alle PX-estimatene. Det kan med andre ord se ut som om KS2-estimatene har gått fra å overestimere usikkerheten i prosjektene med besluttet oppstart i 2001-2003 til å underestimere usikkerheten i prosjektene med besluttet oppstart i 2010-2012. Dette kan tyde på en uheldig utvikling for realisme i styrings- og kostnadsrammer, og dermed også for styringen av prosjektene.

For lave P50-, P85- og P90-estimer i den seneste tidsperioden samsvarer med en reduksjon i relativ bredde på 80 prosent prediksjonsintervallet, det vil si intervallet [P10, P90], over tid, som vist i figuren nedenfor. Forskjellene er imidlertid, som vist ved 95 prosent konfidensintervallene, relativt små og ikke statistisk signifikante.



Figur 6-11: Utvikling av relative intervallbredde over tid

Ser vi på CRPS-verdiene i figuren nedenfor så ser vi at de «beste» estimatene (når vi kombinerer kalibrering og informasjonsverdi var de gitt for prosjekter med oppstart i 2001-2003 og 2007-2009. Heller ikke her er forskjellene statistisk signifikante.



Figur 6-12: Utvikling av CRPS over tid

6.4 Hvor gode er estimatene mot en enklere referansemodell?

Det brukes mye tid på usikkerhetsanalysene etter at grunnkalkylen er utarbeidet. Dersom dette arbeidet ikke gir vesentlig bedre bilde av usikkerheten enn det vi kan få fra enkle, lite arbeidskrevende modeller, så bør man vurdere om metoden er bra nok, og eventuelt vurdere om enklere modeller kan erstatte de tidkrevende usikkerhetsanalysene. I Kapittel 6.2 fant vi at korrelasjonen mellom intervallbredde, som skal indikere grad av usikkerhet, og estimeringsfeil var nær null. Vi tolket dette som at KS2-estimatene ikke var gode til å skille mellom prosjekter med lav og høy kostnadsrisiko. Dette skulle tilsi at en enkel modell kan gi minst like gode resultater.

For å undersøke dette gjennomførte vi en sammenligning av KS2-estimatene og estimerer basert på en enkel modell-basert usikkerhetsanalyse etter prinsipper formulert i (Jørgensen og Sjøberg, 2003). Den enkle modellen baserer seg på å bruke tidligere estimeringsavvik som utgangspunkt for påslag på grunnkalkylen for å finne P50 og P85-estimatene.

For å få en tilstrekkelig mengde data for å lage den enkle modellen utarbeider vi modellbaserte P50 og P85-estimerer kun for prosjekter med beslutningsår 2007 eller tidligere. Vi utarbeider (trener) altså modellen på prosjekter med

beslutningsår i 2007 eller tidligere og evaluerer modellen basert på prosjekter med beslutningsår i 2008 eller senere.

Modellen har følgende to steg:

- 1) Finn forholdet mellom faktisk kostnad og grunnkalkyle for prosjektene med beslutningsår 2007 eller tidligere (Q =fordelingen av faktisk kostnad/grunnkalkyle).
- 2) Bruk grunnkalylen for prosjektene med beslutningsår etter 2007 og gang denne med 50-prosentilen av Q for å regne ut P50-estimatet og med 85-prosentilen av Q for finne P85-estimatet for prosjektene etter 2007 (klimatologisk estimering).

Vi ser i Tabell 6-5 at KS2-estimatene gjør det bedre enn estimatene til den enkle (klimatologiske) modellen både på relativt estimatavvik og estimatskjevhet.

Tabell 6-5: Sammenligning av modell-basert (klimatologisk) og KS-baserte estimater for prosjekter med beslutningsår 2008 eller senere

Målestørrelse	Resultat	
	Modell	KS2-estimat
Median relativt avvik	13 %	10 %
	p<0.001 ⁱ	
Median relativ estimatskjevhet	-7 %	4 %
	p<0.001 ⁱ	
Treffrate P50-estimater	41 %	35 %
	p=0.81 ⁱⁱ	
Treffrate P85-estimater	76 %	73 %
	p=1.00 ⁱⁱ	
Gjennomsnittlig intervallbredde: $\frac{P85 - P50}{\frac{1}{2}(P85 + P50)}$	0.18	0.11
	p<0.001 ⁱ	
Gjennomsnittlig CRPS	179	146
	p=0.25 ⁱ	

i Wilcoxon signed rank test (pairwise differences)

ii Fisher's exact test for forskjell i proporsjoner

Treffraten for P50 og P85-estimatene er i favør av modellen, men forskjellen er liten og ikke statistisk signifikant. Bredden mellom P50 og P85 er vesentlig lavere (og dermed mer informative og krever mindre kostnader i reserve for kostnadsramme) for KS2-estimatene sammenlignet med modell-estimatene. Totalt sett tyder dette på at det gir merverdi med de arbeidskrevende KS2-estimatene, sammenlignet med det å bruke en enkel påslagsmodell med utgangspunkt i grunnkalkylen. Denne analysen gir altså et annet, og mer positivt, bilde av kvaliteten på estimatene i KS2-analysen enn den manglende korrelasjonen mellom bredde i estimatfordelingen og estimeringsfeil rapportert i Kapittel 6.2.

Det skal altså mer til enn en enkel påslagsmodell for å oppnå like bra estimeringsnøyaktighet, tilsvarende treffrate med samme informasjonsverdi (like liten bredde i intervallene), som den oppnådd ved estimatene fremkommet ved hjelp av KS2-gjennomgangen. På den andre siden er det ikke utenkelig at en litt mer kompleks modell, der man for eksempel skiller mellom type prosjekter (for eksempel jernbane versus bygg), kunne ha konkurrert bedre mot den omfattende KS-prosessen. Per dags dato er det litt for lite data til å bryte ned på prosjekttype eller risikoklasse dersom vi både skal lage modellene og teste modellene.

7 Avsluttende oppsummering

I diskusjonen nedenfor oppsummer vi det vi oppfatter som hovedfunnene i studien.

Generelt sett er estimeringsavvik og estimatskjevhet lave

Resultatene tyder på at estimatavvik og estimatskjevhet er lave, målt mot hva som rapporteres i internasjonale undersøkelser (se kapittel 2.2), med gjennomsnittlig 14 prosent avvik fra P50-estimat, og en median på 2 prosent overskridelse. Styringsrammene som er basert på P50-estimatene viste omtrent tilsvarende tall for avvik og overskridelser. I kapittel 4.2 illustrerte vi at avvik kan forventes selv med perfekt estimering. Det gjennomsnittlige avviket på 14 prosent ligger relativt nært det teoretisk forventede avviket ved perfekt estimering på mellom 8 og 10 prosent. Betyr dette at estimatene gitt i KS2-sammenheng er nesten så gode som det lar seg gjøre, og at forbedringspotensialet er lite? Trolig ikke. En effekt som også må tas hensyn til er at arbeidet i noen grad tilpasses budsjettene (Parkinson's prinsipp), og dermed blir estimatavviket mindre enn det ellers ville ha vært. Hvor stor denne tilpasningen er vet vi lite om. Et viktig poeng er likevel at det forventede avviket for perfekte estimater ikke er 0 prosent, og at man for et enkeltprosjekt ikke kan si at estimatet var feil dersom sluttkostnaden ble høyere eller lavere enn for eksempel P50-estimatet.

Treffrater, for alle prosjektene over ett, på 42 og 75 prosent for henholdsvis P50 og P85-estimatene, er også akseptable, særlig i lys av den typiske tendensen til å undervurdere kostnadsusikkerhet i prosjekter (se Kapittel 2.3). Dette tyder på at estimeringsarbeidet i KS-ordningen generelt sett blir godt gjennomført. På den andre siden viste det seg at estimatene som tilsvarende ytterpunkter i estimeringsdistribusjonen, nemlig P10 og P90, er for nær hverandre. Dette betyr at usikkerheten, særlig for store under og overskridelser, er betydelig undervurdert. Vi har kort beskrevet en mulig metode for å øke realismen til disse i kapittel 6.2.

Nedenfor følger også en del andre forhold som fortjener oppmerksomhet og kanskje kan forbedres.

Negativ tendens over tid – mot større kostnadsoverskridelser

Mens prosjektene med beslutningstidspunkt i perioden 2001-2003 viste en underskridelse på 2,7 prosent (median av prosentvis skjevhet), så viste prosjektene i perioden 2010-2012 en median overskridelse på 9,3 prosent (gjennomsnittene var henholdsvis -6 og 12,3 prosent). Mens vi også i fremtiden må regne med estimatavvik/overskridelser (selv med perfekte P50-estimer vil gjennomsnittlig estimatavvik for prosjekter med tilsvarende grad av kostnadsusikkerhet trolig være på 8-10 prosent), så er den systematiske, og økende, tendensen til underestimering/overskridelser av kostnader verdt å merke seg. Departementene og etatene som mer ansvarlige for disse prosjektene bør vurdere tiltak mot dette.

En annen markant endring over tid, som samsvarer med økt tendens til underestimering, er at treffratene for P50 og P85-estimatene har gått ned fra et for høyt til et for lavt nivå. Mens treffratene var 72 prosent for P50-estimatene (som skulle være ca. 50 prosent) og 100 prosent for P85-estimatene (som skulle være ca. 85 prosent) i perioden 2001-2003, så var de tilsvarende treffratene kun 23 og 62 prosent i perioden 2010-2012. Den relativt gode treffraten på P50 og P85-estimer alle prosjektene over ett skjuler altså at det i den senere tid har vært mange flere prosjekter enn forventet ut fra intensjonen med de sannsynlighetsbaserte estimatene som har overskredet P50-estimatet (og styringsrammen) og P85-estimatet (og kostnadsrammen).

Tilsynelatende liten læring over tid og mellom prosjekter

Med en standardisert faseinndelt prosjektmodell, og med et høyt antall prosjekter som ikke er vesentlig forskjellige fra hverandre så burde omfanget av kostnadsoverskridelser over tid gå ned. Hvis det ikke skjer, kan det tyde på at mange av de feilene som begås gjentas over tid, og at erfaringsoverføring internt i og mellom sektorer ikke har vært tilstrekkelig. Ettorevalueringer kan bidra til å lære mer om årsaker til overskridelser, og hva som er årsakene til disse sammenliknet med de faktorene som ble identifisert i forkant.

Noen etater og KS-rådgivere har dårligere realisme i estimatene enn andre

Det er forskjeller i estimatenes avvik, overskridelser, kalibrering og informasjonsverdi avhengig av hvilken etat som har ansvar for prosjektet og hvilken kvalitetssikrer som har vært ansvarlig for utarbeidelse av estimatene. Det har for eksempel vært en tendens til overestimering i Forsvarets prosjekter og underestimering i Statens vegvesens prosjekter. Begge typer prosjekter har

hatt for smale prediksjonsintervaller, det vil si at en har undervurdert kostnadsusikkerheten. Blant KS-rådgiverne så er det ingen store forskjeller i gjennomsnittlig estimatavvik målt mot P50-estimatet, mens Metier har hatt den klart laveste treffraten på 80 prosent prediksjonsintervallet. Det er imidlertid mange faktorer som spiller inn her, som forskjeller i prosjektenes kompleksitet, slik at det er vanskelig å bruke disse resultatene til å forklare forskjellene i prestasjoner.

Større prosjekter har noe større sannsynlighet for kostnadsoverskridelser

Sammenhengen mellom relativ estimatskjevhet og prosjektenes størrelse er komplisert å analysere på en meningsfull måte. Det kan imidlertid se ut som om det er slik at større prosjekter har noe større sannsynlighet for å overskride P50-estimatet enn de mindre.

KS-aktørene bør bli bedre på å identifisere og skille mellom prosjekter med høy og lav kostnadsrisiko

Flere av analysene tyder på at det er svakheter med usikkerhetsanalysenes evne til å skille mellom prosjekter med lav og høy kostnadsrisiko. Dersom usikkerhetsanalysene hadde vært gode på å avdekke grad av kostnadsusikkerhet så skulle vi forvente at et bredere prediksjonsintervall, for eksempel større avstand mellom P50 og P85-estimer, var korrelert med et høyere estimatavvik. I stedet er korrelasjonen så å si lik null. Dette tyder på at påslaget fra grunnkalkyle til P50 og P85-estimer i stor grad bestemmes av andre forhold enn en realistisk vurdering av kostnadsusikkerhet. Likevel, og heldigvis, er det ikke slik at estimeringsarbeidet i KS-ordningen ikke gir merverdi målt mot en modell med et mekanisk påslag ut fra hva som har vært nødvendig i tidligere prosjekter. En sammenligning med en slik enkel påslagsmodell gir at KS2-estimatene gjør det bedre på de aller fleste områdene (estimeringsavvik, estimeringsskjevhet og informasjonsverdi/bredde på prediksjonsintervallene).

Avslutningsvis vil vi peke på et annet relevant, men ofte oversett forhold. Kostnadskontroll dreier seg om mer enn estimering og usikkerhetsanalyse. Prosjektene må også styres på en slik måte at ressursene omsettes i resultater på en produktiv måte. Graden av usikkerhet i kostnadsestimaterne må derfor påvirke styringen, gjennom blant annet prosjekteierstyring, gjennomførings- og kontraktsstrategi og prosjektorganisering. Her kan fleksibilitet i prosjektgjennomføringen, bruk av kuttlistor, og andre forhold knyttet til prosjektstyringen være relevante stikkord for senere studier. Kostnadsavvik er

med andre ord ikke bare et uttrykk for kvaliteten på estimatene, men også et uttrykk for kvaliteten i styringen; blant annet i hvor stor grad anbefalingene fra kvalitetssikrer følges opp og i hvilken grad kuttlisten realiseres.

Referanser

AACE, 2019. *Cost engineering terminology* [online]. Tilgjengelig fra:
<http://library.aacei.org/terminology/>

Ahiaga-Dagbui, D.D. og Smith, S.D., 2014. Rethinking construction cost overruns: cognition, learning and estimation. *Journal of Financial Management of Property and Construction*, 19 (1), 38-54.

Aspenberg Jordal, H., 2019. *Kostnads- og nyttentvikling i tidligfasen*. Concept arbeidsrapport 2019-4. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Austeng, K., Midtbø, J.T., Jordanger, I., Magnussen, O.M. og Torp, O., 2005a. *Usikkerhetsanalyse – kontekst og grunnlag*. Concept rapport nr. 10. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Austeng, K., Torp, O., Midtbø, J.T., Helland, V. og Jordanger, I., 2005b. *Usikkerhetsanalyse – metoder*. Concept rapport nr. 12. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Austeng, K., Binz, V. og Drevland, F., 2005c. *Usikkerhetsanalyse- feilkilder i metode og beregning*. Concept rapport nr. 13. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Berntsen, S. og Sunde, T., 2003. *Styring av statlige porteføljer i staten*. *Usikkerhetsansetning på porteføljenivå*. Concept rapport nr. 1. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Carvalho, A., 2016. An Overview of Applications of Proper Scoring Rules. *Decision Analysis*, 13 (4), 223-242.

Cho, S., 2006. An exploratory project expert system for eliciting correlation coefficient and sequential updating of duration estimation. *Expert Systems with Applications*, 30 (4), 553-560.

Connolly, T. og Dean, D., 1997. Decomposed versus holistic estimates of effort required for software writing tasks. *Management Science*, 43 (7), 1029-1045.

Concept rapport nr. 59

-
- Drevland, F., Austeng, K. og Torp, O., 2005. *Usikkerhetsanalyse - Modellering, estimering og beregning*. Concept rapport nr. 11. Trondheim: Norges teknisk-naturvitenskapelige universitet.
- Drevland, F., 2013. *Kostnadsestimering under usikkerhet*. Concept temahefte nr. 4. Trondheim: Norges teknisk-naturvitenskapelige universitet.
- Finansdepartementet, 2008. *Felles begrepsapparat KS2*. Veileder nr. 2.
- Flyvbjerg, B., Ansar, A., Budzier, A., Buhl, S., Cantarelli, C., Garbuio, M., Glenting, C., Skamris Holm, M., Lovallo, D., Lunn, D., Molin, E., Rønne, A., Steward, A. og van Wee, B., 2018. Five things you should know about cost overrun, *Transportation Research Part A. Policy and Practice*, 118, 174-190.
- Forskningsprogrammet Concept, 2019. *KS-ordningen* [online]. Tilgjengelig fra: <https://www.ntnu.no/concept/ks-ordningen1>
- Foss, T., Stensrud, E., Kitchenham, B. og Myrtveit, I., 2003. A simulation study of the model evaluation criterion MMRE. *IEEE Transactions on Software Engineering*, 29 (11), 985-995.
- Gneiting, T., 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106 (494), 746-762.
- Gneiting, T., Balabdaoui, F. og Raftery, A. E., 2007. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69 (2), 243-268.
- Goodwin, P., 2000. Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, 16 (1), 85-99.
- Granger, C.W.J. og Pesaran, H.M., 2000. Economic and Statistical Measures of Forecast Accuracy. *Journal of Forecasting*, 19, 537-560.
- Grimstad, S., Jørgensen, M., og Moløkken-Østvold, K., 2006. Software effort estimation terminology: The tower of Babel. *Information and Software Technology*, 48 (4), 302-310.
- Halkjelsvik, T. og Jørgensen, M., 2018. *Time predictions: Understanding and avoiding unrealism in project planning and everyday life*. Simula SpringerBriefs on Computing, Concept rapport nr. 59

5, 110p. Tilgjengelig fra: <https://link.springer.com/book/10.1007%2F978-3-319-74953-2>

Halkjelsvik, T. og Jørgensen, M. 2019. *When 2 + 2 should be 5: The work effort summation fallacy in judgment-based estimation*. Upublisert.

Jergeas G.F. og Ruwanpura, J., 2010. Why Cost and Schedule Overruns on Mega Oil Sands Projects? *Practice Periodical on Structural Design and Construction*, 15 (1), 40-43.

Jørgensen, M. og Sjøberg, D.I.K., 2003. An effort prediction interval approach based on the empirical distribution of previous estimation accuracy. *Information and Software Technology*, 45 (3), 123-136.

Jørgensen, M., Teigen, K.H. og Moløkken-Østvold, K., 2004. Better sure than safe? Over-confidence in judgment based software development effort prediction intervals. *Journal of Systems and Software*, 70 (1), 79–93.

Jørgensen, M., 2014. Communication of software cost estimates. I: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (p. 28). ACM.

Jørgensen, M., Halkjelsvik, T. og Kitchenham, B., 2012. How does project size affect cost estimation error? Statistical artifacts and methodological challenges. *International Journal of Project Management*, 30 (7), 839-849.

Jørgensen, M., 2004. Realism in assessment of effort estimation uncertainty: It matters how you ask. *IEEE Transactions on Software Engineering*, 30 (4), 209-217.

Kaming, P.F., Holt, O. og Harris F.C., 1997. Factors influencing construction time and cost overruns on high-rise projects in Indonesia. *Construction Management and Economics*, 15, 83–94.

Kommunal og Moderniseringsdepartementet, 2018. *Tildelingsbrev 2019 – Statsbygg*. Brev til Statsbygg, 20. desember 2018.

Love, P.E.D., Smith, J., Simpson, I., Regan, M. og Olatunji, O., 2015. Understanding the landscape of overruns in transportation infrastructure projects. *Environment and Planning B: Planning and Design*, 42 (3), 490–509.

-
- Love, P.E.D. og Ahiaga-Dagbui, D., 2018. De-bunking 'Fake News' in a Post-Truth Era: The Plausible Untruths of Cost Underestimation in Transport Infrastructure Projects. *Transportation Research Part A Policy and Practice*, 113, 357–368.
- Morley, S. K., Brito, T. V. og Welling D. T., 2018. Measures of model performance based on the log accuracy ratio. *Space Weather*, 16(1), 69-88.
- Nan N. og Harter, D.E., 2009. Impact of budget and schedule pressure on software development cycle time and effort. *IEEE Transactions on Software Engineering*, 35 (5), 624–637.
- Teigen, K.H. og Jørgensen, M., 2005. When 90% confidence intervals are 50% certain: on the credibility of credible intervals. *Applied Cognitive Psychology*, 19 (4), 455-475.
- Teigen, K. H., Andersen, B., og Alnes, S. L., 2018. *Hvordan oppfattes og omtales usikre kostnadsestimater?* Concept arbeidsrapport 2018-3. Trondheim: Norges teknisk-naturvitenskapelige universitet.
- Torp, O., Drevland, F. og Austeng, K., 2015. *Prosess for kostnadsestimering under usikkerhet*. Concept temahefte nr. 6. Trondheim: Norges teknisk-naturvitenskapelige universitet.
- Volden, G.H. og Samset, K., 2017. *Statlige investeringsiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene*. Concept rapport nr. 52. Trondheim: Ex ante akademisk forlag.
- Welde, M., 2014a. *Kostnadsutvikling i vegprosjekter underlagt KS2 – fra første offisielle omtale til ferdigstillelse*. Concept arbeidsrapport. Trondheim: Norges teknisk-naturvitenskapelige universitet.
- Welde, M., 2014b. *Oppdaterte sluttkostnader - prosjekter som har vært underlagt KS2 per mai 2014*. Concept arbeidsrapport. Trondheim: Norges teknisk-naturvitenskapelige universitet.
- Welde, M., 2014c. *Prisomregning på tvers av sektorer. Praksis, konsekvenser, harmonisering*. Concept arbeidsrapport. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Welde, M., 2015. *Oppdaterte sluttkostnader - prosjekter som har vært underlagt KS2 per september 2015*. Concept arbeidsrapport. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Welde, M., 2016. *Kostnadsutvikling i store statlige investeringsprosjekter fra KS1 til KS2*. Concept arbeidsrapport. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Welde, M., 2017. *Kostnadskontroll i store statlige investeringer underlagt ordningen med ekstern kvalitets sikring*. Concept rapport nr. 51. Trondheim: Ex ante akademisk forlag.

Welde, M. og Torp, O., 2016. *Kostnads estimeringsmetodikk i etatene omfattet av KS-ordningen. En kartlegging*. Concept arbeidsrapport. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Welde, M., Whist, E. og Hjelm brekke, H., 2018. *Etterevaluering av informatikkbygget ved Universitetet i Oslo (IFI2)*. Concept evalueringsrapport. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Wifstad, K., Haugland, L.M., Aalen, P., Gulbrandsen, M., Ulstein, H., Nilsen, Ø. L., Melbø, O. og Welde, M., 2018. *Evaluering av Rv. 13 Hardangerbrua*. Oslo: Menon Economics.

Ulstein, H., Wifstad, K., Syrstad, R.S., Seeberg, A.R., Gulbrandsen, M. og Welde, M., 2017. *Evaluering av E6 Østfold*. Oslo: Menon-publikasjon nr. 4/2017.

Zellner, A., 1986. A tale of forecasting 1001 series: The Bayesian knight strikes again. *International Journal of Forecasting*, 2 (4), 491-494.

Aass, T., 2013. *Kostnadskontroll i prosjekter som har vært underlagt ekstern kvalitets sikring KS2*. Concept arbeidsrapport. Trondheim: Norges teknisk-naturvitenskapelige universitet.

Vedlegg: Liste over prosjekter benyttet i studien

Prosjektnavn	Faktisk sluttkostnad	Felles prisnivå	Opprinnelig kostnadsramme	Avvik	Relativt avvik	Opprinnelig styringsramme	Avvik	Relativt avvik	Investeringsbeslutning år	Ferdig år
E6 Skjerdingsstad - Jaktøyen	595	2006	824	-229	-27,8 %	720	-125	-17,4 %	2003	2005
E6 Ny Svinesund-forbindelse	1 167	2014	1 324	-157	-11,8 %	1 164	3	0,2 %	2003	2005
E6 Assurtjern Klemetsrud	489	2006	662	-173	-26,1 %	627	-138	-22,0 %	2003	2004
Eiksundsambandet	1 062	2009	1 145	-83	-7,2 %	1 078	-16	-1,5 %	2004	2008
E18 Høvik Frydenhaug	1 162	2009	1 293	-131	-10,1 %	1 159	3	0,3 %	2003	2007
Ringveg nord i Tønsbergpakken fase 1	1 210	2009	1 239	-29	-2,3 %	1 172	38	3,2 %	2003	2008
Lofoten fastlandsforbindelse	1 380	2009	1 500	-120	-8,0 %	1 343	37	2,8 %	2004	2008
E18 Langåker Bommestad	1 316	2011	1 711	-395	-23,1 %	1 639	-323	-19,7 %	2007	2010
E18 Kopstad Gulli	1 827	2011	2 382	-555	-23,3 %	2 169	-342	-15,8 %	2004	2008
Rv 2 Kløfta Nybakk	747	2008	759	-12	-1,6 %	736	11	1,5 %	2004	2008
E6 Åsgård Halmstad	526	2015	905	-379	-41,9 %	858	-332	-38,7 %	2004	2006
Rv. 150 Ulven-Sinsen	4 167	2013	3 836	331	8,6 %	3 521	646	18,3 %	2005	2016
E18 Bjørvikaprojektet	7 770	2015	6 571	1 199	18,2 %	5 755	2 014	35,0 %	2006	2010
E16 Wøyen Bjørum	1 520	2011	1 308	212	16,2 %	1 163	357	30,7 %	2005	2011
E6 Nordre avlastningsveg	1 600	2010	1 376	224	16,3 %	1 311	289	22,0 %	2006	2009
E6 Svingenskogen Åsgård	3 034	2015	3 089	-55	-1,8 %	2 944	90	3,1 %	2006	2010

Prosjektnavn	Faktisk sluttkostnad	Felles prisnivå	Opprinnelig kostnadsramme	Avvik	Relativt avvik	Opprinnelig styringsramme	Avvik	Relativt avvik	Investeringsbeslutning år	Ferdig år
Rv 519 Fastlandsforb. Finnøy (Finnfast)	669	2014	769	-101	-13,1 %	728	-59	-8,2 %	2006	2010
Rv 64 Atlanterhavstunnelen	1 058	2019	943	115	12,2 %	900	158	17,6 %	2006	2014
Rv 7 Hardangerbrua	2 914	2018	3 264	-350	-10,7 %	2 896	18	0,6 %	2006	2009
E18 Frydenhaug Eik	793	2009	824	-31	-3,7 %	766	27	3,6 %	2007	2010
E6 Assurtjern Vinterbro	1 253	2011	1 753	-500	-28,5 %	1 537	-284	-18,5 %	2006	2008
E18 Momarken Sekkelsten	643	2010	611	32	5,2 %	561	82	14,6 %	2010	2013
Fv 107 Jondalstunnelen	889	2012	1 048	-159	-15,2 %	944	-55	-5,9 %	2007	2012
Fv 108 ny Kråkerøyforbindelse	833	2011	597	237	39,6 %	561	272	48,6 %	2008	2011
E18 Krosby Knapstad	1 194	2011	1 256	-62	-4,9 %	1 192	2	0,2 %	2006	2011
Ringveg vest byggetrinn 1	1 363	2011	1 576	-213	-13,5 %	1 440	-77	-5,3 %	2008	2015
E6 Gardemoen-Kolomoen	1 830	2012	1 643	187	11,4 %	1 565	265	16,9 %	2009	2014
Fv. 47 T-forbindelsen	1 924	2013	1 944	-20	-1,0 %	1 696	228	13,4 %	2010	2013
E6 Øyer-Tretten	1 447	2015	1 160	287	24,8 %	1 067	379	35,6 %	2010	2012
E6 Dal-Minnesund	2 375	2013	2 934	-559	-19,1 %	2 682	-307	-11,4 %	2010	2015
Rv456 Kolsdalen-Lumberkrysset	1 284	2015	1 480	-195	-13,2 %	1 310	-26	-2,0 %	2009	2015
E6 øst Trondheim-Stjørdal	4 655	2015	4 529	127	2,8 %	4 284	371	8,7 %	2009	2013
E39 Kvivsvegen	1 245	2012	1 258	-13	-1,0 %	1 166	79	6,8 %	2009	2013
E18 Sky-Langangen	2 038	2013	2 170	-132	-6,1 %	2 059	-21	-1,0 %	2010	2014
Bremangersambandet 2	515	2015	851	-336	-39,5 %	778	-263	-33,8 %	2011	2015
E16 Smedalsosen - Borlaug	705	2015	1 007	-302	-30,0 %	910	-205	-22,5 %	2011	2015
Rv70 Oppdølsstranda	1 013	2016	1 107	-94	-8,5 %	1 029	-17	-1,6 %	2011	2015

Prosjektnavn	Faktisk sluttkostnad	Felles prisnivå	Opprinnelig kostnadsramme	Avvik	Relativt avvik	Opprinnelig styringsramme	Avvik	Relativt avvik	Investeringsbeslutning år	Ferdig år
E18 Gulli-Langåker	3 025	2015	2 976	50	1,7 %	2 840	186	6,5 %	2011	2015
Rv7 Sokna-Ørgenvika	2 094	2017	1 960	134	6,8 %	1 839	255	13,8 %	2011	2014
E6 Alta vest, Møllnes-Hjemmeluft	849	2015	835	14	1,7 %	742	107	14,5 %	2010	2014
Fv609/Fv57 Dalsfjordsambandet	1 327	2015	1 276	51	4,0 %	1 160	167	14,4 %	2011	2014
Rv80 Løding-Vikan	619	2015	698	-80	-11,4 %	584	35	5,9 %	2011	2015
E16 Slomarka-Kongsvinger	2 532	2017	2 070	462	22,3 %	1 938	594	30,7 %	2012	2018
E136 Tresfjordbrua og Vågastrandtunnelen	1 915	2017	1 479	436	29,5 %	1 389	526	37,9 %	2002	2006
Dobbeltspor Sandvika Asker	3 687	2007	4 209	-522	-12,4 %	3 790	-103	-2,7 %	2004	2007
GSM-R	1 709	2008	1 915	-206	-10,8 %	1 810	-101	-5,6 %	2006	2010
Lysaker stasjon	1 155	2009	1 250	-95	-7,6 %	1 143	12	1,0 %	2006	2010
Dobbeltspor Sandnes Stavanger	2 158	2011	1 933	225	11,6 %	1 705	453	26,6 %	2008	2012
Dobbeltspor Lysaker Sandvika	3 105	2013	3 635	-530	-14,6 %	3 138	-33	-1,1 %	2010	2012
Gevingåsen tunnel	730	2012	731	-1	-0,1 %	635	95	15,0 %	2009	2012
Barkåker-Tønsberg	1 575	2011	1 679	-104	-6,2 %	1 569	6	0,4 %	2011	2017
Nytt operahus	4 278	2008	4 356	-78	-1,8 %	3 544	734	20,7 %	2002	2008
Høgskolen i Østfold Remmen (HiØ)	560	2007	665	-105	-15,8 %	606	-46	-7,6 %	2004	2007
Svalbard forskningspark	342	2006	454	-112	-24,7 %	418	-76	-18,2 %	2004	2006

Prosjektnavn	Faktisk sluttkostnad	Felles prisnivå	Opprinnelig kostnadsramme	Avvik	Relativt avvik	Opprinnelig styringsramme	Avvik	Relativt avvik	Investeringsbeslutning år	Ferdig år
Statlig kontrollområde Svinesund	257	2006	301	-44	-14,6 %	285	-28	-9,8 %	2005	2005
Høgskolen i Vestfold (HiV)	694	2011	718	-24	-3,3 %	591	103	17,4 %	2006	2011
Patologibygg HiOA (Sykepleiehøgskolen)	746	2015	775	-29	-3,7 %	729	18	2,4 %	2010	2014
Nytt informatikkbygg UiO (IFI2)	1 382	2011	1 199	183	15,3 %	1 101	281	25,5 %	2006	2011
Odontologibygg Bergen	746	2012	775	-29	-3,8 %	707	39	5,5 %	2009	2013
Halden fengsel	1 300	2009	1 217	83	6,8 %	1 160	140	12,1 %	2008	2010
Høgskolen i Bergen Samlokalisering	2 156	2015	2 469	-313	-12,7 %	2 400	-244	-10,2 %	2011	2015
R6 Regjeringsbygg	1 254	2015	1 214	40	3,3 %	1 121	133	11,9 %	2008	2015
Domus Medica (Gaustad UiO)	455	2013	517	-62	-12,0 %	448	7	1,5 %	2011	2013
Domus Media (Sentrum)	819	2015	862	-43	-5,0 %	840	-21	-2,5 %	2009	2015
HiO Sykepleierutdanningen	746	2016	775	-29	-3,7 %	729	17	2,3 %	2010	2014
Hærens taktiske treningssenter (HTTS)	544	2008	586	-42	-7,2 %	571	-27	-4,7 %	2002	2009
FISBasis	1 520	2003	1 560	-40	-2,6 %	1 498	22	1,5 %	2001	2005
Panserbekjempelse – middels rekkevidde	726	2016	1 192	-466	-39,1 %	991	-265	-26,7 %	2003	2012
Regionfelt Østlandet	1 809	2018	2 210	-401	-18,1 %	2 210	2 113	105,4 %	2005	2012
Golf leveranseprosjekt 1 (LP1)	439	2004	482	-43	-8,9 %	462	-23	-5,0 %	2004	2006
Luft-til-luft kortholdsmisil	1 024	2017	1 808	-784	-43,4 %	1 477	-453	-30,7 %		

Prosjektnavn	Faktisk sluttkostnad	Felles prisnivå	Opprinnelig kostnadsramme	Avvik	Relativt avvik	Opprinnelig styringsramme	Avvik	Relativt avvik	Investeringsbeslutning år	Ferdig år
NASAMS II	1 239	2016	1 287	-48	-3,8 %	1 132	107	9,5 %	2003	2014
Nye sjømålsmissiler (NSM)	4 104	2017	4 180	-76	-1,8 %	N/A			2007	
Pansrede brukte spesialkjøretøy	1 186	2016	1 220	-34	-2,8 %	1 197	-11	-0,9 %	2010	2013
Taktisk Data Link for fregatt og F16	863	2016	1 207	-344	-28,5 %	1 008	-145	-14,4 %	2004	2015
Integrert strategisk ledelsesbygg (ISL-bygget)	574	2006	580	-6	-1,0 %	550	24	4,4 %	2004	2007
P3 Orion strukturoppdatering	1 127	2018	927	200	21,6 %	N/A				
LOS Økonomiprojektet (ØP)	724	2009	773	-49	-6,3 %	633	91	14,4 %	2006	2009
Pensjonsprogrammet fase 2	963	2009	975	-12	-1,3 %	880	83	9,5 %	2007	2009
Pensjonsprogrammet fase 3	1 713	2011	1 874	-161	-8,6 %	1 493	220	14,7 %	2009	2011
NAV IKT Basis	867	2009	1 180	-313	-26,5 %	990	-123	-12,4 %	2005	2009
LØFT Modulis	812	2015	815	-3	-0,4 %	812	0	0,0 %	2004	2015
Perform	1 023	2012	933	90	9,7 %	700	323	46,2 %	2011	2012
EFFEKT-programmet	363	2014	356	7	2,0 %	324	39	12,0 %	2011	2014
Tetra nødnettprosjektet	5873	2017	5208	665	12,8 %	4687	1186	25,3 %	2004	2018

Concept rapportserie

Papirtrykk: ISSN 0803-9763

Elektronisk utgave på internett: ISSN 0804-5585

Lastes ned fra: <https://www.ntnu.no/concept/concept-rapportserie>

Rapport	Tittel	Forfatter
Nr. 1	Styring av prosjektporteføljer i staten. Usikkerhetsavsetning på porteføljenivå <i>Project Portfolio Management. Estimating Provisions for Uncertainty at Portfolio Level.</i>	Stein Berntsen og Thorleif Sunde
Nr. 2	Statlig styring av prosjektleidelse. Empiri og økonomiske prinsipper. <i>Economic Incentives in Public Project Management</i>	Dag Morten Dalen, Ola Lædre og Christian Riis
Nr. 3	Beslutningsunderlag og beslutninger i store statlige investeringsprosjekt <i>Decisions and the Basis for Decisions in Major Public Investment Projects</i>	Stein V. Larsen, Eilif Holte og Sverre Haanæs
Nr. 4	Konseptutvikling og evaluering i store statlige investeringsprosjekt <i>Concept Development and Evaluation in Major Public Investment Projects</i>	Hege Gry Solheim, Erik Dammen, Håvard O. Skaldebø, Eystein Myking, Elisabeth K. Svendsen og Paul Torgersen
Nr. 5	Bedre behovsanalyser. Erfaringer og anbefalinger om behovsanalyser i store offentlige investeringsprosjekt <i>Needs Analysis in Major Public Investment Projects. Lessons and Recommendations</i>	Petter Næss
Nr. 6	Målformulering i store statlige investeringsprosjekt <i>Alignment of Objectives in Major Public Investment Projects</i>	Ole Jonny Klakegg
Nr. 7	Hvordan trur vi at det blir? Effektvurderinger av store offentlige prosjekt <i>Up-front Conjecture of Anticipated Effects of Major Public Investment Projects</i>	Nils Olsson

Nr. 8	<p>Realopsjoner og fleksibilitet i store offentlige investeringsprosjekt</p> <p><i>Real Options and Flexibility in Major Public Investment Projects</i></p>	Kjell Arne Brekke
Nr. 9	<p>Bedre utforming av store offentlige investeringsprosjekter. Vurdering av behov, mål og effekt i tidligfasen</p> <p><i>Improved Design of Public Investment Projects. Up-front Appraisal of Needs, Objectives and Effects</i></p>	<p>Petter Næss med bidrag fra Kjell Arne Brekke, Nils Olsson og Ole Jonny Klakegg</p>
Nr. 10	<p>Usikkerhetsanalyse – Kontekst og grunnlag</p> <p><i>Uncertainty Analysis – Context and Foundations</i></p>	<p>Kjell Austeng, Olav Torp, Jon Terje Midtbø, Ingemund Jordanger, og Ole M Magnussen</p>
Nr. 11	<p>Usikkerhetsanalyse – Modellering, estimering og beregning</p> <p><i>Uncertainty Analysis – Modeling, Estimation and Calculation</i></p>	<p>Frode Drevland, Kjell Austeng og Olav Torp</p>
Nr. 12	<p>Metoder for usikkerhetsanalyse</p> <p><i>Uncertainty Analysis – Methodology</i></p>	<p>Kjell Austeng, Jon Terje Midtbø, Vidar Helland, Olav Torp og Ingemund Jordanger</p>
Nr. 13	<p>Usikkerhetsanalyse – Feilkilder i metode og beregning</p> <p><i>Uncertainty Analysis – Methodological Errors in Data and Analysis</i></p>	<p>Kjell Austeng, Vibeke Binz og Frode Drevland</p>
Nr. 14	<p>Positiv usikkerhet og økt verdiskaping</p> <p><i>Positive Uncertainty and Increasing Return on Investments</i></p>	Ingemund Jordanger
Nr. 15	<p>Kostnadsusikkerhet i store statlige investeringsprosjekter; Empiriske studier basert på KS2</p> <p><i>Cost Uncertainty in Large Public Investment Projects. Empirical Studies</i></p>	<p>Olav Torp (red.), Ole M Magnussen, Nils Olsson og Ole Jonny Klakegg</p>
Nr. 16	<p>Kontrahering i prosjektets tidligfase. Forsvarets anskaffelser.</p> <p><i>Procurement in a Project's Early Phases. Defense Aquisitions</i></p>	Erik N. Warberg

Nr. 17	Beslutninger på svakt informasjonsgrunnlag. Tilnærminger og utfordringer i prosjekters tidlige fase <i>Decisions Based on Scant Information. Challenges and Tools During the Front-end Phases of Projects</i>	Kjell Sunnevåg (red.)
Nr. 18	Flermålsanalyser i store statlige investeringsprosjekt <i>Multi-Criteria Decision Analysis In Major Public Investment Projects</i>	Ingemund Jordanger, Stein Malerud, Harald Minken, Arvid Strand
Nr. 19	Effektvurdering av store statlige investeringsprosjekter <i>Impact Assessment of Major Public Investment Projects</i>	Bjørn Andersen, Svein Bråthen, Tom Fagerhaug, Ola Nafstad, Petter Næss og Nils Olsson
Nr. 20	Investorers vurdering av prosjekters godhet <i>Investors' Appraisal of Project Feasibility</i>	Nils Olsson, Stein Frydenberg, Erik W. Jakobsen, Svein Arne Jessen, Roger Sørheim og Lillian Waagø
Nr. 21	Logisk minimalisme, rasjonalitet - og de avgjørende valg <i>Major Projects: Logical Minimalism, Rationality and Grand Choices</i>	Knut Samset, Arvid Strand og Vincent F. Hendricks
Nr. 22	Miljøøkonomi og samfunnsøkonomisk lønnsomhet <i>Environmental Economics and Economic Viability</i>	Kåre P. Hagen
Nr. 23	The Norwegian Front-End Governance Regime of Major Public Projects – A <i>Theoretically Based Analysis and Evaluation</i>	Tom Christensen
Nr. 24	Markedsorienterte styringsmetoder i miljøpolitikken <i>Market oriented approaches to environmental policy</i>	Kåre P. Hagen
Nr. 25	Regime for planlegging og beslutning i sykehusprosjekter <i>Planning and Decision Making in Hospital Projects. Lessons with the Norwegian Governance Scheme.</i>	Asmund Myrbostad, Tarald Rohde, Pål Martinussen og Marte Lauvsnes

- | | | |
|--------|---|--|
| Nr. 26 | <p>Politisk styring, lokal rasjonalitet og komplekse koalisjoner. Tidligfaseprosessen i store offentlige investeringsprosjekter</p> <p><i>Political Control, Local Rationality and Complex Coalitions. Focus on the Front-End of Large Public Investment Projects</i></p> | Erik Whist, Tom Christensen |
| Nr. 27 | <p>Verdsetting av fremtiden. Tidshorisont og diskonteringsrenter</p> <p><i>Valuing the future. Time Horizon and Discount Rates</i></p> | Kåre P. Hagen |
| Nr. 28 | <p>Fjorden, byen og operaen. En evaluering av Bjørvikautbyggingen i et beslutningsteoretisk perspektiv <i>The Fjord, the City and the Opera. An Evaluation of Bjørvika Urban Development</i></p> | Erik Whist, Tom Christensen |
| Nr. 29 | <p>Levedyktighet og investeringstiltak. Erfaringer fra kvalitetssikring av statlige investeringsprosjekter</p> <p><i>Sustainability and Public Investments. Lessons from Major Public Investment Projects</i></p> | Ola Lædre, Gro Holst Volden, Tore Haavaldsen |
| Nr. 30 | <p>Etterevaluering av statlige investeringsprosjekter. Konklusjoner, erfaringer og råd basert på pilotevaluering av fire prosjekter</p> <p><i>Evaluating Public Investment Projects. Lessons and Advice from a Meta-Evaluation of Four Projects</i></p> | Gro Holst Volden og Knut Samset |
| Nr. 31 | <p>Store statlige investeringers betydning for konkurranse- og markedsutviklingen. Håndtering av konkurransemessige problemstillinger i utredningsfasen</p> <p><i>Major Public Investments' Impact on Competition. How to Deal with Competition Issues as Part of the Project Appraisal</i></p> | Asbjørn Englund, Harald Bergh, Aleksander Møll og Ove Skaug Halsos |
| Nr. 32 | <p>Analyse av systematisk usikkerhet i norsk økonomi.</p> <p><i>Analysis of Systematic Uncertainty in the Norwegian Economy.</i></p> | Haakon Vennemo, Michael Hoel og Henning Wahlquist |

- | | | |
|--------|--|---|
| Nr. 33 | <p>Planprosesser, beregningsverktøy og bruk av nytte-kostnadsanalyser i vegsektoren. En sammenlikning av praksis i Norge og Sverige.</p> <p><i>Planning, Analytic Tools and the Use of Cost-Benefit Analysis in the Transport Sector in Norway and Sweden.</i></p> | <p>Morten Welde, Jonas Eliasson, James Odeck, Maria Börjesson</p> |
| Nr. 34 | <p>Mulighetsrommet. En studie om konseptutredninger og konseptvalg</p> <p><i>The Opportunity Space. A Study of Conceptual Appraisals and the Choice of Conceptual Solutions.</i></p> | <p>Knut Samset, Bjørn Andersen og Kjell Austeng</p> |
| Nr. 35 | <p>Statens prosjektmodell. Bedre kostnadsstyring. Erfaringer med de første investeringsiltakene som har vært gjennom ekstern kvalitetssikring</p> | <p>Knut Samset og Gro Holst Volden</p> |
| Nr. 36 | <p>Investing for Impact. Lessons with the Norwegian State Project Model and the First Investment Projects that Have Been Subjected to External Quality Assurance</p> | <p>Knut Samset og Gro Holst Volden</p> |
| Nr. 37 | <p>Bruk av karbonpriser i praktiske samfunnsøkonomiske analyser. En oversikt over praksis fra analyser av statlige investeringsprosjekter under KVU-/KS1-ordningen.</p> <p><i>Use of Carbon Prices in Cost-Benefit Analysis. Practices in Project Appraisals of Major Public Investment Projects under the Norwegian State Project Model</i></p> | <p>Gro Holst Volden</p> |
| Nr. 38 | <p>Ikke-prissatte virkninger i samfunnsøkonomisk analyse. Praksis og erfaringer i statlige investeringsprosjekter</p> <p><i>Non-Monetized Impacts in Economic Analysis. Practice and Lessons from Public Investment Projects</i></p> | <p>Heidi Bull-Berg, Gro Holst Volden og Inger Lise Tyholt Grindvoll</p> |
| Nr. 39 | <p>Lav prising – store valg. En studie av underestimering av kostnader i prosjekters tidligfase</p> <p><i>Low estimates – high stakes. A study of underestimation of costs in projects' earliest phase</i></p> | <p>Morten Welde, Knut Samset, Bjørn Andersen, Kjell Austeng</p> |

Nr. 40	<p>Mot sin hensikt. Perverse insentiver – om offentlige investerings-prosjekter som ikke forplikter</p> <p><i>Perverse incentives and counterproductive investments. Public funding without liabilities for the recipients</i></p>	Knut Samset, Gro Holst Volden, Morten Welde og Heidi Bull-Berg
Nr. 41	<p>Transportmodeller på randen. En utforskning av NTM5-modellens anvendelsesområde</p> <p><i>Transport models and extreme scenarios. A test of the NTM5 model</i></p>	Christian Steinsland og Lasse Fridstrøm
Nr. 42	<p>Brukeravgifter i veisektoren</p> <p><i>User fees in the road sector</i></p>	Kåre Petter Hagen og Karl Rolf Pedersen
Nr. 43	<p>Norsk vegplanlegging: Hvilke hensyn styrer anbefalingene</p> <p><i>Road Planning in Norway: What governs the selection of projects?</i></p>	Arvid Strand, Silvia Olsen, Merethe Dotterud Leiren og Askill Harkjerr Halse
Nr. 44	<p>Ressursbruk i transportsektoren – noen mulige forbedringer</p> <p><i>Resource allocation in the transport sector – some potential improvements</i></p>	James Odeck (red.) og Morten Welde (red.)
Nr. 45	<p>Kommunale investeringsprosjekter. Prosjektmodeller og krav til beslutningsunderlag.</p> <p><i>Municipal investment practices in Norway</i></p>	Morten Welde, Jostein Aksdal og Inger Lise Tyholt Grindvoll
Nr. 46	<p>Styringsregimer for store offentlige prosjekter. En sammenliknende studie av prinsipper og praksis i seks land.</p> <p><i>Governance schemes for major public investment projects: A comparative study of principles and practices in six countries</i></p>	Knut F. Samset, Gro Holst Volden, Nils Olsson og Eirik Vårdal Kvalheim
Nr. 47	<p>Governance Schemes for Major Public Investment Projects. A comparative study of principles and practices in six countries.</p>	Knut F. Samset, Gro Holst Volden, Nils Olsson og Eirik Vårdal Kvalheim
Nr. 48	<p>Investeringsprosjekter og miljøkonsekvenser. En antologi med bidrag fra 16 forskere.</p> <p><i>Environmental Impact of Large Investment Projects. An Anthology by 16 Norwegian Experts.</i></p>	Kåre P. Hagen og Gro Holst Volden

Nr. 49	<p>Finansiering av vegprosjekter med bompenger. Behandling av og konsekvenser av bompenger i samfunnsøkonomiske analyser.</p> <p><i>Financing road projects with tolls. The treatment of and consequences of tolls in cost benefit analyses.</i></p>	Morten Welde, Svein Bråthen, Jens Rekdal og Wei Zhang
Nr. 50	<p>Prosjektmodeller og prosjekteierstyring i statlige virksomheter.</p> <p><i>Project governance and the use of project models in public agencies and line ministries in Norway.</i></p>	Bjørn Andersen, Eirik Vårdal Kvalheim og Gro Holst Volden
Nr. 51	<p>Kostnadskontroll i store statlige investeringer underlagt ordningen med ekstern kvalitetssikring.</p> <p><i>Cost performance in government investment projects that have been subjected to external quality assurance.</i></p>	Morten Welde
Nr. 52	<p>Statlige investeringer under lupen. Erfaring med evaluering av de 20 første KS-prosjektene.</p> <p><i>A Close-up on Public Investment Cases. Lessons from Ex-post Evaluations of 20 Major Norwegian Projects</i></p>	Gro Holst Volden og Knut Samset
Nr. 53	<p>Fremsynsmetoder</p> <p><i>Foresight methods</i></p>	Tore Sager
Nr. 54	<p>Neglected and underestimated impacts of transport investments</p>	Petter Næss, Gro Holst Volden, James Odeck og Tim Richardson
Nr. 55	<p>Kostnadsstyring i entreprisekontrakter</p> <p><i>Cost performance of construction contracts</i></p>	Morten Welde, Roy Endre Dahl, Olav Torp og Torbjørn Aass
Nr. 56	<p>Styring og gjennomføring av store statlige IKT-prosjekter</p> <p><i>Governance of Major Public ICT-projects</i></p>	Håkon Finne
Nr. 57	<p>Effektivitet og produktivitet i norsk veibyging 2007-2016</p> <p><i>Efficiency and productivity in Norwegian road construction 2007-2016</i></p>	Kenneth Løvold Rødseth, Rasmus Bøgh Holmen, Finn R. Førstund og Sverre A.C. Kittelsen

- Nr. 58 Mandater for konseptvalgutredninger. En gjennomgang av praksis. Knut Samset og Morten Welde
The Terms of Reference Document for Conceptual Appraisal. A Review of Current Practice.
- Nr. 59 Estimering av kostnader i store statlige prosjekter: Hvor gode er estimatene og usikkerhetsanalysene i KS2-rapportene? Morten Welde, Magne Jørgensen, Per Fridtjof Larsen og Torleif Halkjelsvik
Estimating costs in large government investment projects. How good are the estimates and uncertainty analyses in the QA2-reports?

Forskningsprogrammet Concept skal utvikle kunnskap som sikrer bedre ressursutnytting og effekt av store, statlige investeringer. Programmet driver følgeforskning knyttet til de største statlige investeringsprosjektene over en rekke år. En skal trekke erfaringer fra disse som kan bedre utformingen og kvalitetssikringen av nye investeringsprosjekter før de settes i gang.

Concept er lokalisert ved Norges teknisk- naturvitenskapelige universitet i Trondheim (NTNU), ved Fakultet for ingeniørvitenskap og teknologi. Programmet samarbeider med ledende norske og internasjonale fagmiljøer og universiteter, og er finansiert av Finansdepartementet.

The Concept research program aims to develop know-how to help make more efficient use of resources and improve the effect of major public investments. The Program is designed to follow up on the largest public projects over a period of several years, and help improve design and quality assurance of future public projects before they are formally approved.

The program is based at The Norwegian University of Science and Technology (NTNU), Faculty of Engineering Science and Technology. It cooperates with key Norwegian and international professional institutions and universities, and is financed by the Norwegian Ministry of Finance.

Address:

The Concept Research Program
Høgskoleringen 7A
N-7491 NTNU
Trondheim
NORWAY

ISSN: 0803-9763 (papirversjon)
ISSN: 0804-5588 (nettversjon)
ISBN: 978-82-93253-81-5 (papirversjon)
ISBN: 978-82-93253-82-2 (nettversjon)

