

Hvordan tilrettelegge for realisme i estimer av arbeidsmengde i IT-prosjekter

Magne Jørgensen,
magnej@simula.no
Linpro, 8. June, 2009

Kilder

- "Practical Guidelines for Expert-Judgment-Based Software Effort Estimation". Magne Jørgensen, IEEE Software, May/June 2005.
- "Avoiding Irrelevant and Misleading Information When Estimating Development Effort", Magne Jørgensen & Stein Grimstad, May/June 2008.
- "How to Avoid Selecting Bids Based on Overoptimistic Cost Estimates", Magne Jørgensen, May/June 2009.
- Disse (og mye mer estimeringsforskning) kan lastes ned fra:
 - simula.no/research/engineering/projects/best/bibliography

EN OPPVARMING OM VÅRE VURDERINGSEVNER ...

[**simula** . research laboratory]

3

Clouds Make Nerds Look Good?



- Studie av søkere til et universitet (USA)
 - Reelle ekspertvurderinger (n=682)
 - Gjennomsnittlig 12% høyere sjanse for å bli tatt opp på dager med solskinn, i forhold til dager med mest overskyet.
 - “Nerder” hadde større sjanse, sammenlignet med “ikke-nerder” på dager med overskyet vær.
 - “Nerd-faktor” målt som “akademisk score” delt på “sosial score”
 - “*Clouds Make Nerds Look Good*”, Journal of Behavioral Decision Making, Simonsohn, 2007 vol:20 side:143

[**simula** . research laboratory]

4

The Cocktail Party Effect



[simula . research laboratory]

5

Mennesker har to "hovedsystemer" for tenkning

- "The dual-process theory of cognition":
 - *"Both theory and a substantial body of evidence, some of it derived from neuro-imaging studies of the brain employing fMRI technology, support the view that humans employ at least two distinct systems to process information, a rational system and an intuitively-oriented experiential system"* (Goel & Dolan, 2003)
- Når magefølelsen (hjertet, intuisjonen, din indre stemme, ...) sier en ting, mens "hodet" (analysen) sier noe annet har vi en konflikt mellom disse to systemene.
- De intuitive systemet er høyst sannsynlig det evolusjonsmessige eldste, og det som oftest føles(!) mest "naturlig" å følge.

[simula . research laboratory]

6

Noen forskjeller mellom analyse og intuisjon

(Hammond et al, 1987)

Analysis:

- High insight into judgment process, and, hence publicly retraceable
- Low confidence in outcome, high confidence in method
- Slow rate of processing
- High cognitive consistency

Intuition:

- Low insight into judgment process, and, hence difficult to retrace and defend
- High confidence in outcome, low confidence in method
- Fast rate of processing
- Low cognitive consistency

[**simula** . research laboratory]

7

En liten avsporing: Er det forskjell mellom kjønnene?

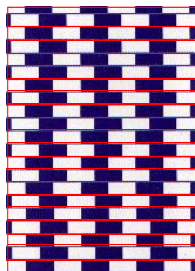
- Ser ut til å være små forskjeller mhp faktisk bruk av intuisjon.
- Menn synes imidlertid å ha et merkbart større behov for å rapportere at prosesser har vært analytiske!
 - Intuisjon som en typisk kvinneegenskap er derfor trolig mer en kulturell stereotype, enn en reell forskjell i tenkemåte.
 - Individual Differences in Intuitive-Experiential and Analytical-Rational Thinking Styles, Seymour Epstein and Rosemary Pacini, Journal of Personality and Social Psychology, 1996, Vol. 71, No. 2, 390-405
- Vi er alle, uavhengig av kjønn og profesjon, sterkt avhengige av intuisjon!



[**simula** . research laboratory]

8

Ligger mursteinene parallelt?



Konflikt

(Denesraj, V, Epstein, S: Conflict between intuitive and rational processing – when people behave against their better judgment)

- To urner med røde og hvite kuler:
 - Urne 1: 7 røde og 100 hvite
 - Urne 2: 1 rød og 10 hvite
- Du vinner et pengebeløp ved å trekke rød kule og får opplyst antall røde og hvite per urne. Hvilken vil du velge?
 - Det er åpenbart at det er ulønnsomt å velg Urne 1, selv om den har flere røde kuler.
 - En del av deltakerne valgte likevel Urne 1, **til tross for** at de på sett og vis skjønnte at det var større sannsynlighet å finne en rød kule i Urne 2. De valgte å følge magesfølelsen, til tross for en korrekt analyse.
 - Flere av de som valgte Urne 2 (analytisk valg), beskrev at de måtte kjempe mot fristelsen å velge Urne 1 fordi den var mer "tiltrekkende".
 - De mest analytiske var imidlertid ikke i tvil og hadde ingen følelsesmessige problemer med å velge Urne 2.

Vi har de samme konfliktene innen estimering av IT-prosjekter

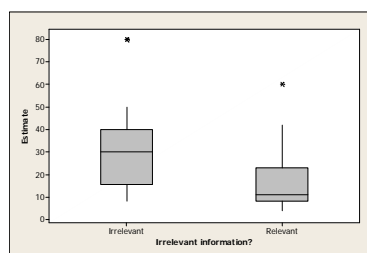
- Anta at vi basert på historiske data har laget en enkel estimeringsregel (analytisk modell) basert på antall timeverk per enkle, middels og komplekse "moduler".
- Estimeringsregelen (eller andre historiske data) gir at prosjektet vil ta ca. 1500 timeverk.
- Magefølelsen ("intuitiv erfaring") sier oss imidlertid at det umulig kan kreve så mye arbeid og at 1000 timeverk må være mer enn nok. Vi har nå en konflikt mellom de to tenkesystemene.
- Vi vil ofte ha mest tillit til den analytiske prosessen som prosess, men likevel stole mest på estimatet basert på magesfølelsen. Hvordan løse denne konflikten?
 - Sterkt analytisk person: Stoler på modellen (gitt at modellen er bra)
 - Sterkt intuitiv person: Stoler på magesfølelsen (nesten uansett)
 - Konfliktunngående person: Justering av de komplekse "use case-ene" til å bli middels e.l.. Justeringen (rasjonaliseringen av magesfølelsen) kan skje både bevisst og ubevisst.

HVA PÅVIRKER EKSPERTVURDERINGER?

Bakgrunnsinformasjon med lav/ingen relevans for arbeidsmengden?

- Gruppe A fikk den originale spesifikasjonen.
- Group B fikk den samme spesifikasjonen, men vi la til estimerings-irrelevant informasjon om sluttbrukernes kontorstøtte-programmer, webdesign, brukeres passord, osv..

Resultater



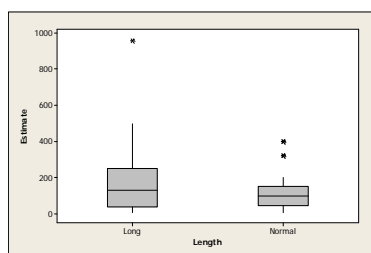
	Irrelevant	Relevant	Difference
Mean	34	17	100%
StDev	20	14	43%

Lengden på spesifikasjonen?

- IFI-studenter estimerte arbeidsmengde til den samme programmeringsoppgaven
 - Gruppe A: Fikk den originale spesifikasjonen, som var en side lang
 - Group B: Fikk en versjon av spesifikasjonen som hadde identisk tekst, men var på syv sider. Økningen i lengde skyldes dobbel linjeavstand, vide marger, større font-størrelse og mer avstand mellom avsnittene



Resultater



	Long	Normal	Difference
Mean	170	117	45%
StDev	173	98	77%

Kravenes rekkefølge

- Gruppe EASY-HARD fikk en spesifikasjon som inneholdt fire krav. Kravene var sortert etter stigende vanskelighetsgrad (lette først, vanskelige sist)
- Gruppe HARD-EASY fikk den samme spesifikasjonen, men kravene var sortert i motsatt rekkefølge (vanskelige først, lette sist)

$(A+B) = (B+A) ?$

Resultater

- Group EASY-HARD : 46,5 work-hours
- Group HARD-EASY : 65 work-hours
- Forskjell: 40%

Fremtidige muligheter

- Deltakerne skulle estimere hvor lang tid de trodde de ville bruke på å implementere den samme oppgaven.
- Gruppe WISHFUL ble fortalt at kunden hadde bedt ti leverandører implementere denne oppgaven, og at de ville velge fremtidige leverandører basert på leverandørenes effektivitet.
- Gruppe CONTROL fikk ikke denne informasjonen
- I begge gruppene fikk deltakerne beskjed om at estimatet kun var til eget bruk og at det ikke skulle presenteres for kunden.



Resultater

- Group-WISHFUL : 40 work-hours
- Group-CONTROL : 100 work-hours

Kundens forventninger

- HIGH (LOW) group: *“The customer has indicated that he believes that **1000 (50)** work-hours is a reasonable effort estimate for the specified system. However, the customer knows very little about the implications of his specification on the development effort and you shall not let the customer’s expectations impact your estimate. Your task is to provide a realistic effort estimate of a system that meets the requirements specification and has a sufficient quality.”*
- Erfarne systemutviklere estimerte.

Resultater

- *HIGH group gjennomsnitt: 555 timeverk*
- *CONTROL group (uten forventinger) gjennomsnitt: 456 timeverk*
- *LOW group gjennomsnitt: 99 timeverk!!!*
- *Ingen av utviklerne opplevde at de hadde blitt mye påvirket. De fleste mente at de ikke var påvirket i det hele tatt av kundens forventninger.*

Felt-studie (replikering med 100% realisme)

“The preliminary budget of the new system is \$10 000 [corresponding to about 100 work-hours with typical pricing in the country in which it will be built]. The preliminary budget is not built on any knowledge about the actual cost of developing the new system, and will, if needed, be extended to cover the expenses necessary to build a quality system with the desired functionality.”

100 timeverk var en svært lav verdi for dette prosjektet og alle firmaene ble instruert om å IKKE bruke dette som input til estimatet.

Felt-studie: Resultater

Numerical Anchor

Group	Median estimate
Manipulated (client's expectation, 100 work-hours)	724 work-hours (n=23)
Ordinary	956 work-hours (n=23)

Oppsummering av resultatene

- **De som estimerer har store problemer med å ignorere estimeringsirrelevant og villedende informasjon**
 - Dette til tross for at de har profesjonell erfaring med lignende oppgaver og de er i stand til å skille mellom relevant og irrelevant informasjon.
- **De som estimerer er i liten grad klar over at de blir påvirket.**
 - Dette indikerer at deltakerne ikke bruker den irrelevante informasjonen bevisst, dvs. at effekten hovedsakelig er ubevisst.
- **Det er vanskelig å forutsi størrelsen av effekten på forhånd**
 - Det virker å være en kompleks sammenheng mellom størrelsen på effekten og faktorer slik som mengden irrelevant informasjon, ekstremiteten, plasseringen, tidspresset, osv.

HVA BØR VI GJØRE?

Uthev relevant/fjern irrelevant informasjon?

- Studert i et eksperiment hvor vi la til irrelevant informasjon i kravspesikasjonen:
 - Group IRRELEVANT: Irrelevant informasjon
 - Group HIGHLIGHT RELEVANT: Deltakerne ble bedt om å highlighte relevant informasjon før de estimerte (gul tusj)
 - Group REMOVE IRRELEVANT: Deltakerne ble bedt om å fjerne irrelevant informasjon med en svart penn før de estimerte
 - Group CONTROL : Ingen irrelevant informasjon

Resultater

Group	Estimate
Group IRRELEVANT	40
Group HIGHLIGHT RELEVANT	40
Group REMOVE IRRELEVANT	35
Group CONTROL	25

Kompetente personer?

- Vi undersøkte om de mest kompetente utviklerne var mindre påvirket av irrelevant informasjon i et studie hvor vi varierte formuleringene:
 - Group LOW: Noen ord som vanligvis assosieres med små oppgaver
 - Group HIGH: Noen ord som vanligvis assosieres med store oppgaver
 - Group CONTROL: Nøytrale formuleringer
- Etter estimeringsarbeidet, så anslo utviklerne selv kompetansenivået sitt

Resultater

	Low Skill	Medium Skill	High Skill
Group LOW	65	40	27
Group CONTROL	180	55	24
Group HIGH	250	80	50

Anbefaling for å unngå å bli påvirket av irrelevant og misvisende informasjon

- Vi har ingen god metode for å nøytralisere effekten av irrelevant/misvisende informasjon, etter at man er har vært eksponert for denne.
- Vi anbefaler derfor at følgende elementer inkluderes i estimeringsprosessen
 - Trinn 1: Preparering av estimeringsinformasjonen
Irrelevant/villedende informasjon fjernes/nøytraliseres
 - Trinn 2: Estimeringsarbeid
Av personer som ikke er involvert i trinn 1
 - Trinn 3: Justeringer
Evaluering av personer som var med i trinn 1, re-estimering av personer som var med i trinn 2

USIKKERHETSANSLAG

Antall innbyggere i Norge i 2001?

Minimum



Maksimum

Vær 99% sikker på å inkludere virkelig verdi!

Radius til (dverg)planeten Pluto

Minimum



Maksimum

Vær 99% sikker!

OL i Montreal i 1976

- *“It can no more lose money than a man can have a baby.” - Montreal Mayor Jean Drapeau.*



[simula . research laboratory]

35

“Sikker” er ikke det samme som “sikkert”

- “Så å si helt sikker” = “60-70% sikkert”
 - Dette gjelder normalprosjekter.
 - Mye mer overoptimisme ved mega-prosjekter!
- “50% sikker” = “75% sikker” = “90% sikker” = “99% sikker”
 - Gjelder min-maks intervaller i situasjoner der usikkerheten er høy

[simula . research laboratory]

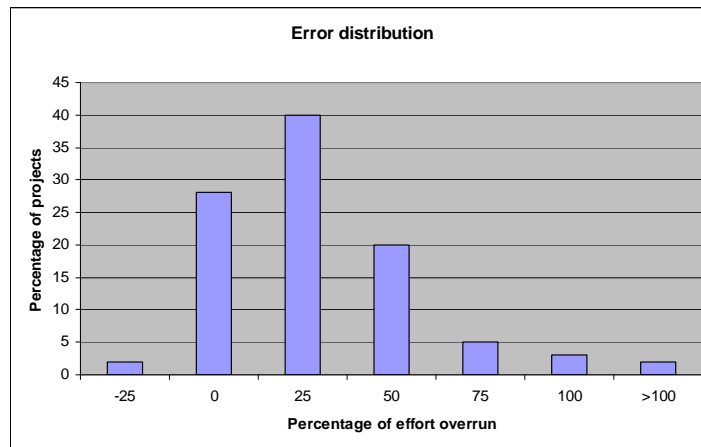
36

Hvorfor undervurderer vi systematisk usikkerheten?

- **Biologi:** Dårlig intuisjon og læring mhp sannsynlighet.
- **Prosess:** Mangel på egnede metoder for å komme fram til TOTALUSIKKERHET.
- **Psykologi:** Realisme i konflikt med:
 - Ønske om å få oppdrag
 - Ønske om at prosjektet skal startes
 - Ønske om å vise dyktighet

Det kommer an på hvordan man spør...

- En prosjektleder har estimert utviklingsarbeid til å ta ca. 1250 timeverk. Du skal utarbeide et budsjett og vil vite hvor usikkert dette estimatet er.
- **Ikke spør slik:** - Hvor mye arbeid vil utviklingen *maksimalt* kreve? Vær 98% sikker.
- **Spør heller slik:** - Av tidligere prosjekter som ligner på dette, hvor stor andel har hatt mer enn 50% overskridelse?



Viser at: 30% av lignende prosjekter har overskredet estimatene med 50% eller mer.

Tre råd jeg gjerne vil dere skal huske:

- Gå ut fra at usikkerheten i store og/eller uvanlige prosjekter er kraftig undervurdert.
- Spør slik at den som svarer kan fokusere på realisme, bruke historiske data (se tilbake) og ikke blir mest opptatt av å demonstrere egen effektivitet og kompetanse.
- Krev at realismen i usikkerhetsangivelsen sannsynliggjøres gjennom henvisning til faktisk overskridelser i lignende prosjekter (historiske data).

VINNERENS FORBANNELSE ...

KUNDENS ROLLE

[**simula** . research laboratory]

Spørsmål

- Du deltar i en budrunde på en hytte der det er mange budgivere. Hvilket av utsagnene er riktig?
 - Ekstra aggressiv budgivning er nødvendig for å vinne anbudsrunderen.
 - Ekstra forsiktig budgivning er nødvendig for å unngå å betale for mye for hytta.
 - Antall budgivere bør ikke påvirke budgivingen din.

[**simula** . research laboratory]

42

Vinnerens forbannelse innen IT-utvikling

Utsagn fra leverandører og kunder:

- "Vi vinner bare budrunder der vi har vært overoptimistiske."
- "Vi kan ikke stole på estimatene til (under)leverandørene våre. De er alltid for optimistiske. Budsjettene våre sprekker og vi mister kunder fordi leverandørene våre får problemer."
- Vinnerens forbannelse →Kundens forbannelse

Eksempel på vinnerens forbannelse

Outsourcing 1992-2000:

- Data med erfaringer finnes på www.umsl.edu/~lacity/cases.htm.
- Leverandørene tapte penger i 20% av tilfellene.
- Av disse tapsavtalene ble:
 - 7 avsluttet tidlig (med tap for begge parter)
 - 1 terminert og deretter restrukturert (med store kostnader for begge)
 - 1 ikke fornyet (også med tap for begge parter)
 - 3 fortsatt, men ansett som svært problematiske av begge parter
 - 3 vellykkede pga langsiktige leverandørgevinster (f eks stordriftsfordeler)

”Forbannelsens” omgang avhenger av ...

- Din egen estimeringskompetanse
- De andre budgivernes estimeringskomptanse
 - Særlig viktig er antall budgivere med dårlig estimeringskomptanse (f eks som blander strategien ”price-to-win” med ”most likely cost”)
- Antall budgivere
- Usikkerhet mhp avtalens verdi/kostnad
- Systematisk over-optimisme i estimatene
- Strategi for valg av leverandør, formatet til auksjonen, utforming av avtalen:
 - Prisfokus, Pre-seleksjon, Risikodeling (PS 2000), Fleksibilitet i leveranse
- Kunnskap om ”vinnerens forbannelse”
 - F eks: strategier hos de som legger inn bud for å unngå vinnerens forbannelse.
 - Der kunnskapen om vinnerens forbannelse er stor (særlig finans og repeterende auksjoner som f eks ”risauksjoner”), fører mange budgivere til lavere priser.

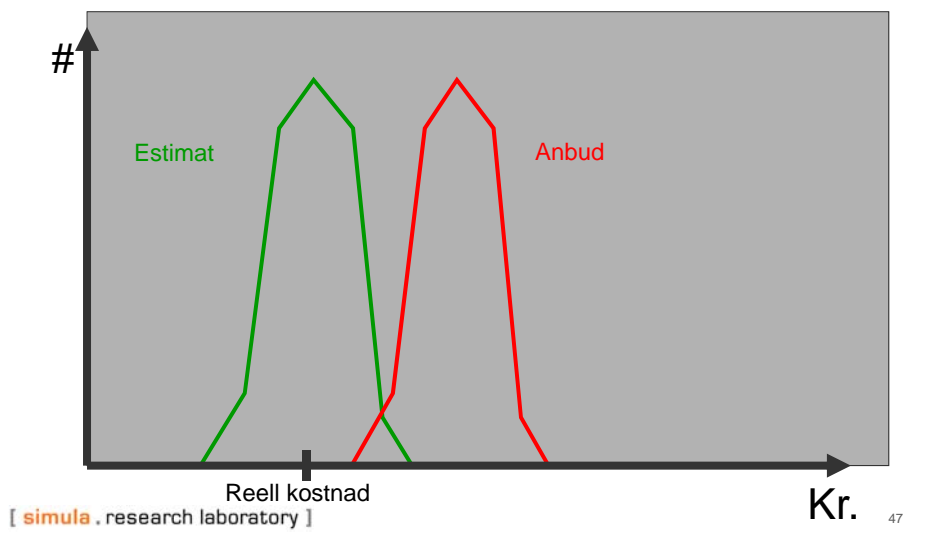
Noen illustrerende scenarier

Antagelser for scenarioene:

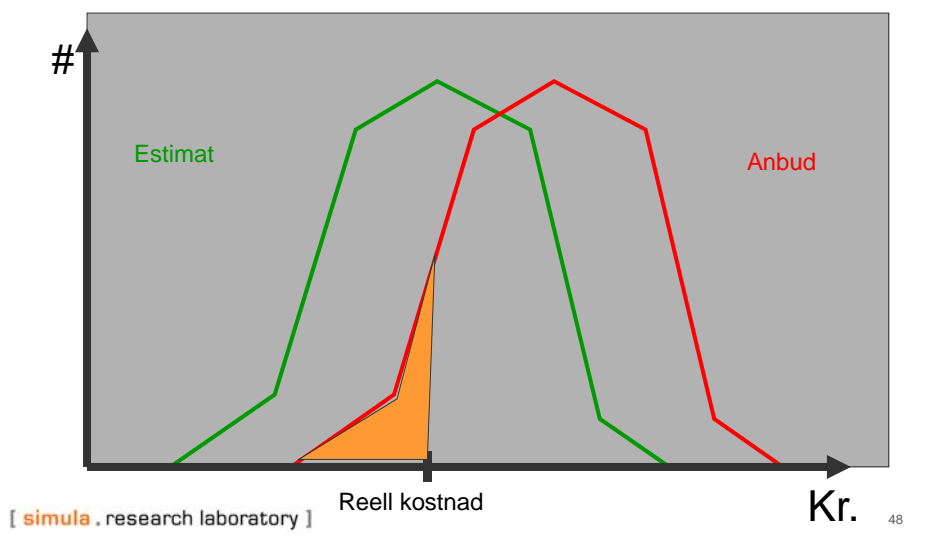
- Mange budgivere
- Samme reelle kostnad og marginer (=bud-estimert kostnad) for alle leverandører
 - Noe urealistisk antagelse, men endrer ikke hovedpoenget i vesentlig grad.
 - Antagelsen om samme reelle kostnad for alle medfører heller en reduksjon av ”vinnerens forbannelse”, siden:
 - de med lavest kompetanse oftest har de mest optimistiske estimatene og de minste marginene.

Tilsvarende scenarier (med samme argumentasjon) kan gjøres for samme leverandør med mange anbud.

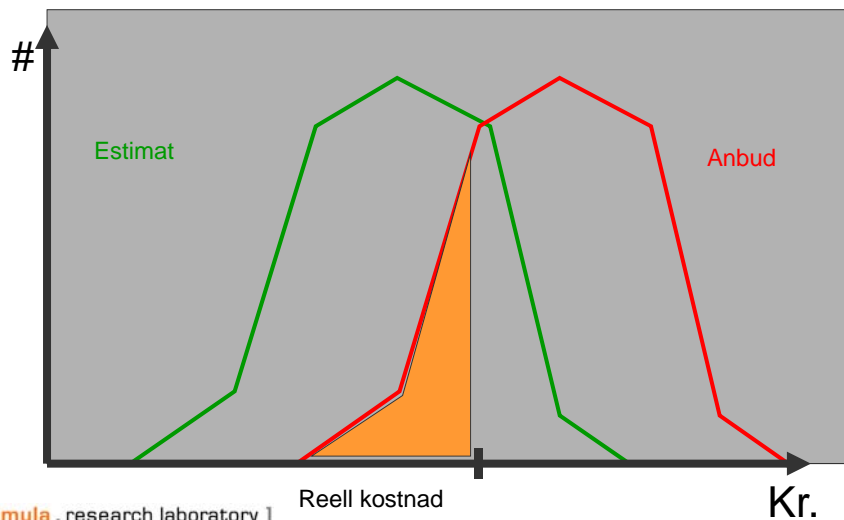
SCENARIO 1: Liten spredning, gode marginer, nøytrale estimater



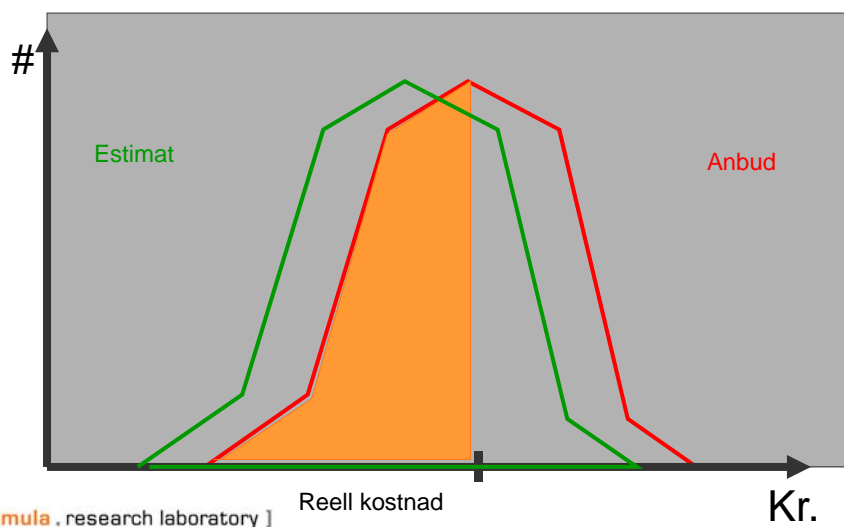
SCENARIO 2: Stor spredning, nøytrale estimater, gode marginer



SCENARIO 3: Stor spredning, tendens til overoptimisme, gode marginer



SCENARIO 4: Stor spredning, tendens til overoptimisme, lave marginer



Oppdragsgiver sitter med mye av nøkkelen til realisme!

- Det er oppdragsgiver som sitter med nøkkelen til å unngå uheldige effekter av "vinnerens forbannelse" gjennom å bestemme:
 - Antall tilbydere som inviteres
 - Kvalitet på spesifikasjon (usikkerhet)
 - Fleksibilitet i leveranse
 - Risikodeling/insitament
 - Fokus på pris vs kompetanse ved valg av leverandør
 - "Ankerinformasjon", f eks ved å lekke informasjon om tilgjengelig budsjett
 - Tro på at de vinner på å forhandle seg til lavere pris

Alternativ valgstrategi for kunde (Eksempel 1)

- Få en uavhengig part til å anslå kostnader for å få bygget et kvalitetssystem ihht spesifikasjon. Denne bør **ikke** selv være kandidat for å utvikle systemet, men bør ha kompetanse i å kunne sammenligne med tilsvarende systemer.
 - F eks: Uavhengig part anslår at kostnadene vil ligge mellom 5 og 10 millioner. Forutsetningene for dette bør angis.
 - Veibygging i noen av delstatene i USA gjøres på denne måten (disse prosjektene viser seg å være de som er minst utsatt for "the winner's curse" og har de mest realistiske budsjettene).
 - Amerikanske flyvåpenet gjør også dette.

Alternativ valgstrategi for kunde (Eksempel 1)

- Gjør en siling av mulige leverandører slik at få, men kompetente leverandører gjenstår, f eks 4-5 leverandører.
- Inviter disse til en anbudsrunde der det uavhengige kostnadsestimatet er oppgitt.
- Be om tilbud. Krev at større avvik fra det uavhengige anslag (både høyere og lavere) blir grundig argumentert for.
 - Slik argumentasjon har f eks forsvaret i USA begynt å kreve for å unngå forsinkelser og dårlig kvalitet.

Alternativ valgstrategi (Eksempel 2)

- Tilbudsrunde med angremulighet for leverandør
 - Mulighet for å trekke tilbudet ("bidding with default").
 - Kunde velger leverandør.
 - Tilbyder som er foretrukket får vite de andre tilbudene (anonymisert) og mulighet til å trekke seg.
 - Dette vil typisk skje dersom de andre pristilbudene er mye høyere.
 - Dersom tilbud trekkes utløses "gebyr".
 - Dersom førstevalg trekker seg, velges nest beste tilbud, osv.
- Hensikt:
 - Leverandører har en sikkerhet mot høy grad av winner's curse (pga misforståelser, høy overoptimisme etc.).
 - Kunden unngår "client's curse".

Hva kan du som leverandør gjøre?

- Leverandøren er som oftest prisgitt rammene til oppdragsgiver og kan gjøre relativt lite annet enn å la være å delta i budrunder med mange tilbydere, sterk pridfokus og ingen særfordeler!
- Vær mer opptatt av kunnskap om de andre budgiverne. Det at andre har mindre erfaring, fører ofte til at de har lavere tilbud. Se etter muligheter for å kompensere for dette, f.eks. ved å dokumentere at lavere anbud er lite troverdige.
- Legge inn **høyere** risikotillegg når det er mange tilbydere med og sterk pridfokus og kunde (med tilsvarende reduksjon i mulighet for å vinne budrunder).
 - Legg inn ekstra risikotillegg basert på analysen: "Gitt at jeg vinner med risikotillegg X, hvor sannsynlig er det at jeg har vært overoptimistisk?"
- Legge inn større "fleksibilitet" i leveransetilbudet når du kun kan vinne med svært lav pris.
- Forsøk å endre rammevilkårene ved å gjøre kunde oppmerksom på risiko ved måten anbudsrunden er lagt opp på og fordeler ved en langsiktig relasjon + fokus på kompetanse - som begge har gevinster av.

NOEN RÅD TIL SLUTT ...

Elleve råd for å redusere overoptimisme

1. Utdann en "cost engineer" som evalueres mhp realisme i estimater og ikke selv er del av prosjektene han/hun estimerer.
2. Skill tydelig mellom prosesser for estimering, planlegging og prising.
3. Unngå irrelevant informasjon som leder til for lave "ankere" (f eks kundeforventninger)
4. Bruk historiske data.
5. Ikke stol på at prosjektforbruk huskes riktig. Dokumenter reelt tidsforbruk.
6. Krev at argumentasjon for estimater er etterprøvbart, dvs viser til lignende erfaringer. Krev en svært god argumentasjon dersom ytelsen estimeres å være bedre enn historiske data skulle tilsi.
7. Ikke regn med at dere har lært noe særlig fra forrige prosjekt.
8. Dersom historiske data ikke finnes, forsøk å finne eksperter med relevant bakgrunn (også utenfor organisasjonen) og historiske data.
9. Ikke la ekspertutviklere estimere arbeidsmengde til uerfarne (bruk middels dyktige til å estimere middels og under middels dyktige).
10. Dersom en person har motiver for lave estimater (svært engasjert i å få prosjektet startet, indirekte belønnes for å angi lave estimater for eget arbeid, etc), finn en annen person til å estimere.
11. Kombiner estimater fra uavhengige kilder. Sørg for at disse har ulik bakgrunn og roller. Bruk en Delphi-prosess for å kombinere vurderingsbaserte estimater.

Momenter for å få kunder til å forstå at "winner's curse" er "client's curse"

- Fullstendige spesifikasjoner finnes ikke! Ingen spesifikasjoner er f eks gode på vedlikeholdbarhet. Fokus på laveste pris, gitt svært ufullstendig spesifikasjon gir høyere risiko for at de ikke-spesifiserte delene blir underprioritert. Dette gir høyere total kostnader (utvikling + vedlikehold) for kunde, og større inntekter for leverandør. ("Moral hazard").
- Få fram at kompetanseevaluering er viktigere enn prisevaluering. Dette gjelder særlig i situasjoner der leverandøren sitter med stor grad av mer kunnskap om systemet. ("Information asymmetry").
- Vis til resultater (f eks fra Simula) som tyder på at lav pris er en indikator på lav kompetanse, mao at det er høyrisiko å ikke kunne vurdere kompetanse til leverandør og kun se på tilbudet/CV-er.
- La de lese "*How to Avoid Selecting Bids Based on Overoptimistic Cost Estimates*".