

Estimering av IT-prosjekter: Ekspertestimat eller modell?

Magne Jørgensen



Nå-situasjon

- Bruk av estimeringsmodeller anbefales av autoriteter, men brukes lite.
 - Mores and Edwards: *“Even though estimation is regarded as a problem by almost all the respondents [91%], the use of commercial tools or the development of in-house estimating models remains low. This result holds even though more than three-quarters of companies fulfil the basic requirements for the development, calibration and use of these tools.”*
 - Vår studier (foreløpig analyse) antyder en nedgang heller enn en økning i bruk av estimeringsmodeller.
- I de fleste andre felt finner forskerne at modeller er mer treffsikre enn ekspertene.
 - Paul Meehl: *“When you are pushing 90 investigations, predicting everything from the outcomes of football games to the diagnosis of liver disease and when you can hardly come up with a half dozen studies showing even a weak tendency in favor of the clinician [the expert judgment], it is time to draw a practical conclusion.”*

Hvorfor brukes ikke modeller mer hos oss?

- Er systemutviklere ignoranter som ikke vet sitt eget beste?
- Eller kan det tenkes at modellene ikke gir de gevinstene man tror og håper?
- Og hva er egentlig forskjellen mellom modell og ekspert-baserte estimerer?

Hovedforskjellen mellom modell og ekspertestimering

- **Modell:** Mekanisk, eksplisitt kvantifiseringssteg (analytisk)
- **Ekspertvurdering:** Kvantifiseringssteg basert på ubevisste, ikke-eksplisitte prosesser (magefølelse)
- Modell og ekspertvurdering kan være basert på nøyaktig samme informasjon om prosjektet, samme historiske data, samme sjekklister og samme nedbryting i aktiviteter.
- Det er feil å si at modeller er objektive og ekspertestimerer subjektive. I IT-estimering er begge metoder basert på subjektive data.
 - Baruch Fischhoff (“forecasting”-guru): *“Serious forecasts are seldom entirely subjective, never entirely objective. At one extreme, even the most automatic (or ‘objective’) forecasting requires some exercise of judgment: Someone must choose the model, set its initial parameters, and decide whether to use its results”.*

Vi vet lite om ekspertestimering. Hvorfor?

- Kvantifiseringen skjer uten at vi legger merke til det, dvs er basert på ubevisste prosesser.
- Stor forskjell på å kunne noe og å vite hvordan man gjør det:
 - Lagnado et al. 2006: “*Studies suggest that quite different regions of the brain are involved in learning and insight about learning.*”
- Spør en systemutvikler (som ikke har brukt en modell) om hvordan ”kvantifiseringssteget” ble gjort, og du vil få rasjonaliseringer og/eller vage beskrivelser om følelser, ekspertise og erfaringer.
 - Ekspertene vet sjelden hvordan de estimerer, men føler at du burde vite det.

Magefølelse, intuisjon, ekspertvurdering

Noen forsøk på definisjoner:

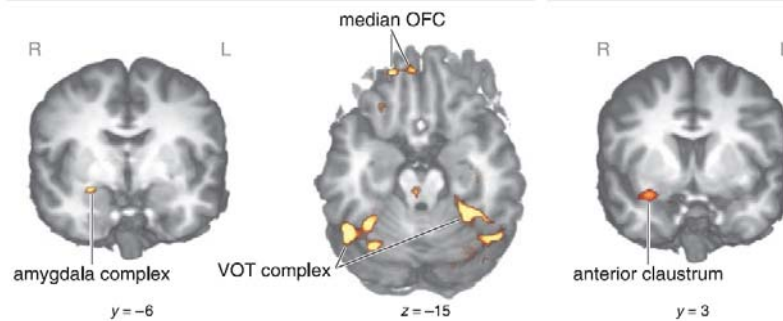
- “*knowing something without knowing how you know it*”
- “*the ability to understand or know something immediately, without conscious reasoning*”
- “*preliminary perception of coherence (pattern, meaning, structure) that is at first not consciously represented*”

Magefølelsen (som sitter i hodet!) synes å involvere ...

- "the median OFC, the lateral portion of the amygdala, anterior insula, and ventral occipito-temporal regions ..."
 - *What Neuroscience Can Tell about Intuitive Processes in the Context of Perceptual Discovery*, by Kirsten G. Volz and D. Yves von Cramon, 2006.

Direct contrast: Meaningful vs. meaningless judged trials

Functional connectivity mOFC



[simula . research laboratory]

7

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 magesfølelsen (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]

8

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]

9

En liten avsporing: Er det forskjell mellom kjønnene?

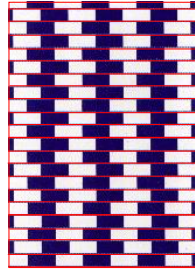
- Ser ut til å være små forskjeller mhp faktisk bruk av intuisjon.
- Menn synes imidlertid å ha et merkbart større behov for å rapportere at prosesser har vært analytiske!
 - Intuisjon som en typisk kvinneegenskap er derfor trolig mer en kulturell stereotype, enn en reell forskjell i tenkemåte.
 - Individual 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]

10

Eksempel 1: Ligger mursteinene parallelt?



[**simula** . research laboratory]

11

Eksempel 2: Hvilken av bananene har noe gult i seg?



A



B

[**simula** . research laboratory]

12

Eksperiment: 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 magefølelsen, 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.

[**simula** . research laboratory]

13

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 "use cases".
- Estimeringsregelen (eller andre historiske data) gir at prosjektet vil ta ca. 1500 timeverk.
- Magefølelsen 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 tenke-systemene.
- Vi vil ofte ha mest tillit til den analytiske prosessen som prosess, men likevel stole mest på estimatet basert på magefølelsen. Hvordan løse denne konflikten?
 - Sterkt analytisk person: Stoler på modellen (gitt at modellen er bra)
 - Sterkt intuitiv person: Stoler på magefølelsen (nesten uansett)
 - Konfliktunngående person: Justering av de komplekse "use case-ene" til å bli middels e.l.. Justeringen (rasjonaliseringen av magefølelsen) kan skje både bevisst og ubevisst.

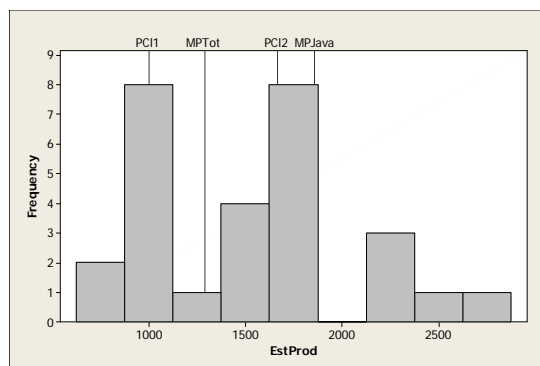
[**simula** . research laboratory]

14

Eget studium: Estimeringsstrategier (1)

- **Deltakere:** 28 systemutviklere som estimerte de samme 10 vedlikeholdsoppgavene.
- **Forskningsspørsmål:** Når bruker de en "analogi-basert" og når en "aggregerings-basert" estimeringsstrategi?
- **Relevans:** Mistanke om at:
 - "Nærmeste analogi"-basert estimeringsstrategi brukes av noen selv når det er klart bedre å bruke gjennomsnittet fra flere oppgaver av samme type. (Kanskje fordi den intuitivt føles mer riktig.)
 - Mange er for dårlige til å endre strategi etter behov.
- **Antagelse:**
 - Det er rasjonelt å bytte strategi, dersom en annen enkel strategi ville ha vært klart bedre på foregående estimeringsoppgaver.
- **NB:** Dette studiet er under arbeid ...

Estimeringsstrategier (2)

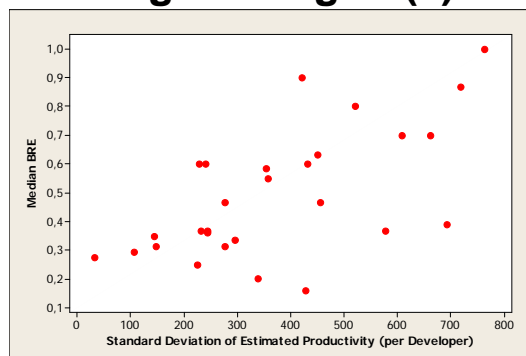


- Figuren antyder at de fleste (for Oppgave 7) valgte en av strategiene "nærmeste analogi" (PCI1), "nærmeste to" (PCI2) eller gjennomsnitt på alle lignende Java-oppgaver (MPJava).
- Noen få velger annerledes (>2200), kanskje basert på forventninger om læring?

Estimeringsstrategier (3)

- Ca. halvparten innså underveis at "nærmeste analogi" strategien burde endres til mer gjennomsnittsbetraktninger.
 - Alle startet med "nærmeste analogi" som eneste valgmulighet pga få historiske data.
- En del ble imidlertid værendes i en uhensiktsmessig "nærmeste analogi" strategi.
 - Indikerer valg av strategi basert på magefølelse mer enn analyse?
 - Eller at men mentale verktøykassen av estimeringsstrategier ikke inneholder strategier basert på gjennomsnittsbetraktninger?

Estimeringsstrategier (4)



- Lav variasjon (standard deviation) antyder større bruk av aggregert informasjon. Korrelerer med lav BRE (= mer nøyaktige estimater).
- Viser at de som valgte å anta en produktivitet nært opp til gjennomsnittet for samme type oppgave (her: samme utviklingsplattform) gjorde det bedre enn de som varierte mer, dvs var mer påvirket av den mer intuitive "nærmeste analogi"- strategien.

Estimeringsstrategier (5)

- Hva sier dette oss om modell vs ekspert?
 - Egnethet av estimeringsstrategi (både for modell og ekspert) er relativ til egenskaper ved historiske data/nytt prosjekt som skal estimeres.
 - Det er stor variasjon mellom "eksperter" i bruk av strategi og, ikke minst, i evne til å endre strategi etter forholdene.
 - Det viktigste er å ha en verktøykasse av estimeringsstrategier og lære å velge mellom dem?

Hvilke faktorer avgjør hva som lønner seg av ekspertestimat og modell:

- Estimeringsmodellen
 - Passer den til situasjonen?
- Brukerne av estimeringsmodellen
 - Bruker de modellen riktig eller som "magefølelse i forkledning"?
- Ekspertene som estimerer
 - Hvor relevant erfaring har de?
- Prosessen for ekspertestimerting
 - Stor forskjell på en strukturert gruppeprosess (f eks Planning Poker) med mye historiske data og en kjapp "magefølelse" basert på dårlig forståelse av prosjektet.
- Estimeringssituasjonen
 - Er det mye som kan påvirke ekspertene irrasjonelt?

Modellfordeler:

- Bedre vekting av variable
 - Ekspertene overvektede ofte informasjon av lav viktighet.
- Større grad av konsistens
 - Ekspertene kan gi svært forskjellige estimater med samme input
 - I en av våre studier lot vi samme person estimerte samme oppgaver flere ganger med minst en måned mellom. Gjennomsnittlig forskjell (median) mellom estimatene for samme person på samme oppgave var ca. 50%.
- Mindre "biases" (f eks mindre ønsketenkning)
 - Ekspertene har ofte andre mål enn nøyaktighet, f eks å ubevisst forsøke å bevare et selv bilde av seg selv som en effektiv person med god kontroll.
- Bedre læringsmuligheter
 - Ekspertene vet typisk ikke hvilken informasjon de har brukt og hvor mye ulik informasjon er blitt vektlagt (selv om de gir uttrykk og tror at de vet det ...).

Ekspertfordeler

- Kan bruke viktig informasjon som ikke er del av modellen
 - Svært konkret kunnskap om kunden Kåre eller utvikleren Gro er lite hensiktsmessig å integrere i modellen ("overfitting" gir dårlige modeller).
- Har mer fleksibilitet i estimeringsprosess
 - Eksperten kan ha kunnskap om når man står overfor uvanlige tilfelle der modell-antagelser ikke er gyldige.
 - Manglende informasjon stopper ikke eksperten så lett, på godt og ondt.

Empiri (Jørgensen, 2007): Seksten studier

Table 1: Experts vs Models

	Most Accurate Model	Average Accuracy of Models	Least Accurate Model
Most Accurate Expert	Experts vs Models: 2 - 0 Expert more accurate: Studies 2 and 12	Experts vs Models: 2 - 0 Expert more accurate: Studies 2 and 12	Experts vs Models: 2 - 0 Expert more accurate: Studies 2 and 12
Average Accuracy of Experts	Expert vs Models: 1 - 7 Expert more accurate: Study 6 Model more accurate: Studies 1, 2, 7, 9, 11, 12, and, 14	Expert vs Models: 10 - 6 Expert more accurate: Studies 1, 2, 3, 5, 6, 7, 9, 10, 11, and, 13 Model more accurate: Studies 4, 8, 12, 14, 15, and, 16	Experts vs Models: 6 - 2 Expert more accurate: Studies 1, 2, 6, 7, 9, 11 Model more accurate: Studies 12, and, 14
Least Accurate Expert	Experts vs Models: 0 - 2 Model more accurate: Studies 2 and 12	Experts vs Models: 0 - 2 Model more accurate: Studies 2 and 12	Experts vs Models: 0 - 2 Model more accurate: Study 2 and 12

[simula . research laboratory]

23

Hovedanbefalinger:

- Bruk ekspertestimering når:
 - Modellene ikke er tilpasset estimeringssituasjonen (dvs ikke basert på organisasjonens egne data)
 - Ekspertene har vesentlig informasjon (kontekstuell informasjon) som ikke gis som input til modellene.
- Bruk modeller (analytisk kvantifiseringssteg) sammen med ekspertestimering når:
 - Organisasjonen mangler gode eksperter.
 - Det er fare for mye irrasjonell påvirkning i estimeringssituasjonen.

[simula . research laboratory]

24

Hvordan velge estimeringsekspert?

- Lengde erfaring?
 - Nei. Ingen god indikator.
- Erfaring fra lignende prosjekter?
 - Ja! Gitt at prosjekt er "svært likt" med tidligere prosjekt. "Litt likt", er en dårlig indikator.
- Den flinkeste utvikleren?
 - Tja. Den flinkeste er ofte uegnet for å estimere arbeidsmengde for de uerfarne. Kanskje bruke de flinkeste til å identifisere aktiviteter og en med middels utviklingsekspertise til å estimere disse?
- Den de andre tror er en god estimerer?
 - Tja. I mangel av noe bedre måte å velge på, og dersom disse har grunnlag for hva de tror. Det har de ikke alltid.

Hvordan velge estimeringsekspert?

- Den som er mest sikker på estimatet sitt?
 - Nei. Dette er oftere en indikator på det motsatte. De mest sikre er som oftest de mest overoptimistiske!
- Den som har vært mest nøyaktige tidligere?
 - Ja, men det er ingen svært god indikator. En studie indikerer at den (av to) som var mest over-optimistisk i foregående prosjekt var med ca. 70% sannsynlighet den som var mest over-optimistisk i det neste prosjektet.
- Basert på tester?
 - Tvilsoomt. Liten hjelp i personlighetstester m.m..
- Lett depressive mennesker?
 - Ja ☺. De viser seg å være de mest realistiske mhp egne evner.

Hvordan velge prosess for ekspertestimering?

- Struktur prosessen med hjelpemidler:
 - Bruk sjekklister (en enkel form for erfaringsdatabase), standardiserte oppdelinger i aktiviteter (WBS), historiske data og krav om begrunnelse.
- Anvend gruppeestimering og kombinasjon av estimater
 - Wideband Delphi
 - Planning Poker