[ simula . research laboratory ]

# Software Cost Estimation:
## When to Use Estimation Models and When to Use Expert Judgment[1]

Magne Jørgensen

---

# Present situation (1)

- Effort estimation models have been around for decades, supported by several tools and are recommended by authorities. In spite of this, they are not much in use:

  – 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 fulfill the basic requirements for the development, calibration and use of these tools."*

  – Our 2007-study (not published) suggests a decrease in use of estimation models in Norway compared to 5 years ago.

[ simula . research laboratory ]

## Present situation (2)

- Our ongoing collection and analysis of estimation work of 40 outsourcing companies shows that:
  - Very few of them use a formalized estimation approach, e.g., based on user stories, function points, use case points, and, COCOMO II.
  - All of the model users (5-6?) were examples of model misuse or "expert judgment in disguise".
    - Example of model misuse: Use of "industry standard" productivity factor when using Function points or Use Case points.
    - Example of expert judgment in disguise: Gut feeling based lines of code input to the COCOMO model.

## Present situation (3)

- In most fields the researchers find that the models are more accurate than the experts:
  - 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.*"

**So, why don't software professionals use models?**

- Are they ignoring empirical evidence and best-practices?

- Or, is it possible that effort estimation models are not the answer to the estimation inaccuracy problems we have?

- And, by the way, what is actually the difference between model and judgment-based estimates?

---

**Main difference between judgment and model-based effort estimates**

- **Quantification step:** The step from the understanding of the problem to the number (the numerical estimate).

- **Model**: Mechanical, explicit quantification step (analysis-based)

- **Judgment**: Quantification step based on unconscious, non-explicit processes (based on "gut feeling", intuition, "what feels right")

## Main difference between judgment and model-based effort estimates

- Model and expert judgment may be based on exactly the same information about the project, the same historical data, and the same list of identified activities. It is the quantification step that separates them.

- In real-world software effort estimation, both model estimates and expert estimates are subjective:
  - Baruch Fischhoff: "*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*".

## We don't know much about judgment-based estimation. Why not?

- There is an essential difference between ability to estimate and ability to know how the estimates are derived.

  - Lagnado et al. 2006: "*Studies suggest that quite different regions of the brain are involved in learning and insight about learning.*"

- The quantification step is unconscious/intuition-based. We don't have easy access to such processes.

## We don't know much about judgment-based estimation. Why not?

- Ask a software professional about his judgment-based estimation process or use a think-aloud protocol to collect this information, and you will NOT get much valuable information.
  - They typically respond with "don't know", "it felt right" or vague statements about their use of experience.
  - The may also feel that they should know how they did the estimation work, and start to rationalize, e.g., by describing how they believe they should have done this as rational beings.

- It is consequently not gain much insight into these processes by asking people. (We have tried ...)

## The feeling that an estimate is "right" seems to involve brain regions different from those involved in conscious, analytic processes ...

- "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.
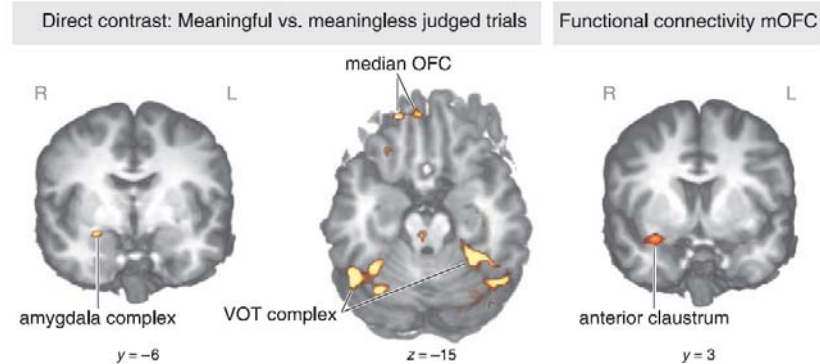
# The dual theory of cognition ...

- "*Both theory and a substantial body of evidence, some of it derived from neuro-imagining 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)

- The "gut feeling" (intuitive) based system is probably the oldest and the one that feels most natural to follow.

- When our "gut feeling" (e.g., judgment-based estimation) says one thing, while your "head" (e.g., an analytic quantification step) says something else, we have a conflict between the two thinking systems.
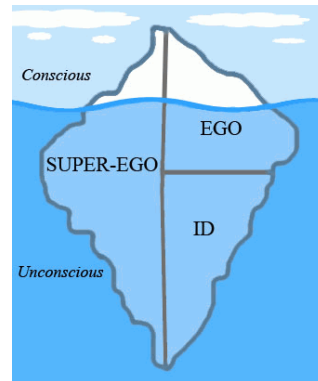


[ simula . research laboratory ]

---

# More on differences between these two systems (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 ]

12

**A minor distraction: Do women base their judgments more on intuition than men?**

- NO. Only small differences in use of intuition (unconscious processes) in judgment and decision processes.

- Men, however, seem to have a larger need to explain judgments analytically!
  - 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

- All of us, independent of gender and profession, are strongly dependent on intuition!

---

# Example of conflict: Are the lines parallel?
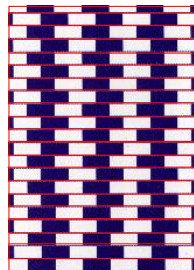
## Experiment: (Denesraj, V, Epstein, S: Conflict between intuitive and rational processing – when people behave against their better judgment)

- From the paper abstract:
  - "When offered an opportunity to win $1 on every "win" trial in which they drew a red jelly bean, subjects frequently elected to draw from a bowl that contained a greater absolute number, but a smaller proportion, of red beans (e.g., 7 in 100) than from a bowl with fewer red beans but better odds (e.g., 1 in 10). **Subjects reported that although they knew [analytically] the probabilities were against them, they felt [intuitively] they had a better chance when there were more red beans**."

- Even some of those selecting the "right" bowl described that they had to fight against the desire of selecting the non-optimal bowl.

[ simula . research laboratory ]

15

---

## The same conflict (analysis vs. intuition) is present when estimating effort

- Suppose that we have a simple model, e.g., the rule that a medium complex "user story" takes 8 work-hours.

- Use of that model implies that a task with five medium complex user stories should take about 40 work-hours.

- The estimator, however, feels that 40 work-hours is too high, and, that 30 work-hours should be sufficient. We now have a conflict between analysis and intuition.

- As reported earlier, we tend to have more confidence in the analytical **process**, but at the same time more confidence in the intuition-based **output** (our expert judgment). How is this conflict solved?
  - A strongly analytical person: Trust the model
  - A strongly intuitive person: Trust the intuition
  - Conflict-averse person: Adjust the model input so that it gives the desired output. In the example, this may be achieved through categorization of some of the medium complex user stories as "simple". This conflict-avoiding adjustment may happen both consciously and unconsciously.

[ simula . research laboratory ]

16

8

## Ongoing analysis on expert judgment strategies (1)

- **Participants:** 28 software developers estimating the same 10 maintenance tasks.

- **Research question**: How do software professionals select among closest analogy and aggregation-based estimation strategies?

- **Relevance**: We suspect that closest analogy-based estimation strategies are not replaced with aggregation-based, even when the latter is clearly more accurate. It is possible that many estimators are poor at changing estimation strategy, because an analogy-based strategy intuitively feels more right.

[ simula . research laboratory ]

17

## Study: Estimation strategies

- The only possible, rational strategy in the beginning was closest analogy.

- When estimating the last tasks, however, it should be clear that aggregation-based estimation strategies (e.g., mean productivity of tasks of the same type) are more accurate.

- The figure shows that those with more aggregation-based strategies had the most accurate estimates
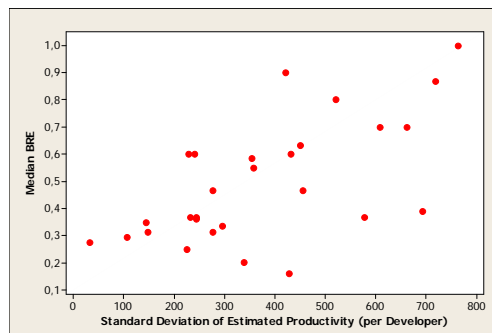


[ simula . research laboratory ]

18

# Estimation strategy

- What is the relevance for the model vs expert discussion?
  - Estimation accuracy, of both model and expert, is relative to fit between estimation strategy and data set properties.
  - There is a large variation in the estimation strategy chosen by software professionals. [Supported by two more recent studies.]
  - There is a large variation in ability to change estimation strategy when needed.
  - The perhaps most important expert estimation ability is to have a "toolbox" of strategies and learn to select between them?

# We conducted a review

- **Inclusion criteria**:
  - Empirical comparison of model and judgment-based software development effort estimates.

- **Search process**:
  - Manual scan of all journals potentially including relevant studies.
  - Library search applying the terms ("effort estimation" OR "cost estimation") AND "software development".
  - Last search: February 2006.

- **Identified papers**:
  - Seventeen papers.
  - One paper excluded due to incomplete information about how the estimates were derived.

## The Review

- The following properties of each study was described:
  - Study design
  - Estimation model selection process
  - Estimation models
  - Calibration level
  - Model use expertise and whether the use was mechanical or not
  - Expert judgment process
  - Expert judgment expertise
  - Possible motivational biases
  - Estimation input
  - Contextual information
  - Estimation complexity
  - Fairness limitations
  - Other design issues
- The authors of the reviewed papers were contacted and asked to quality assured the review of their own study.

## Review Limitations

- Expert judgment is not one process.

- Model use is not one process, either.

- Field use of models may be "expert judgment in disguise".

- Different estimation methods on different types of estimation tasks.
  - Models not used when the required model input is not there. Expert judgment more frequent on complex estimation tasks?
  - Higher willingness to use models, when more important to have accurate estimates. Expert judgment less frequent on complex estimation tasks?

- Imprecise use of "estimate".

# The results

**Table 1: Experts vs Models**

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

[ simula . research laboratory ]

23

---

# Which factors decides when expert judgment will be better or worse?

- The selected estimation model
  - Does it fit the estimation situation/historical data?

- The use of the estimation model
  - Meaningful use seem to be difficult due to lack of company-specific data

- The experience and estimation ability of the software professionals in charge of the estimation work
  - Personal experience from very similar projects seems to be difficult to beat

- The expert estimation process
  - Varies from pure gut feeling to use of structured processes with much use of historical data.

- The estimation context
  - Experts may for example be more misled by irrelevant information

[ simula . research laboratory ]

24

## Model advantages:

- Better weighting of variables
  - Experts tend to put too much weight on variables with low importance

- More consistent estimates
  - In one of our studies we let the same software professionals estimate the same tasks two times with more than one month in-between. The median difference in estimate for same person and same task was approx. 50%!

- Less "wishful thinking"

- Better opportunities to learn
  - An explicit quantification step can be improved more easily than unconscious processes

## Expert advantages:

- Ability to use highly specific information not part of the model
  - Integration of highly context-specific information in the models would typically lead to overfitting, i.e., to poor models.

- More flexible estimation processes
  - The expert may know better when the history is less relevant.
  - Missing information does not easily stop the experts' estimation process (not necessarily an advantage)

## Indicators of estimation expertise

- Length of experience?
  - Not a good indicator.

- Experience from similar projects?
  - Definitively yes, but remember that expertise is "narrower" than typically assumed.

- The best developer?
  - Not always. The best developer may not be suited for the estimation of work effort for novices.
  - "Outside view" sometimes better.

## Indicators of estimation expertise

- The one with highest confidence in his/her estimate?
  - No. We observed the opposite. The most confident are typically the most over-optimistic.

- Those historically most accurate?
  - Yes, but not a very good indicator. We observed that the software professional (out of two) most over-optimistic on previous estimate had a 70% probability of being the most over-optimistic on the next estimate.

- Personality?
  - Probably not of much help.

- Slightly depressive people?
  - Yes ☺. They are on average most realistic regarding own abilities.

## How to select process for expert estimation?

- Impose structure and give support:
  - Use checklists as experience databases, standardized WBS, historical data and require justification.

- Combine estimates from different sources (group-based estimates)
  - Wideband Delphi
  - Planning Poker

## Final words ...

- Use expert estimation when:
  - The models are not based on company-specific data (or data from a similar software development context),
  - The experts has essential information (contextual information) not part of the model, or
  - The expert has experience with very similar projects.

- Use models (analytic quantification step) together with expert estimation when:
  - There is a lack of experts with relevant experience
  - The risk of impact from irrelevant and misleading information is high.