estimate

estimate • analyze • plan • control

ICEAA Keynote March 2014
Estimates Without Bias

Galorath Incorporated

Daniel D. Galorath: Founder & CEO

galorath@galorath.com

SEER by GALORATH

Key Points

Experts are likely providing biased estimates

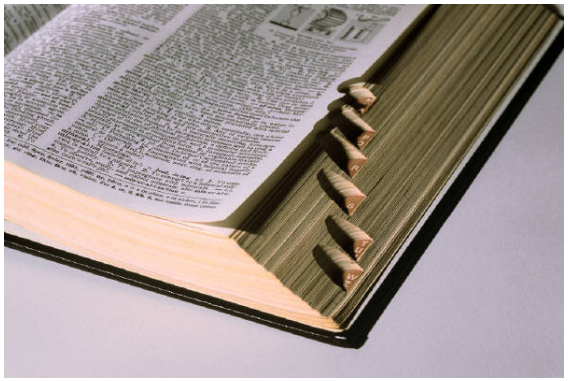
Poor estimates are a root cause of project failure

Estimates can be better, squelching bias & strategic mis-estimation… Parametrics help.

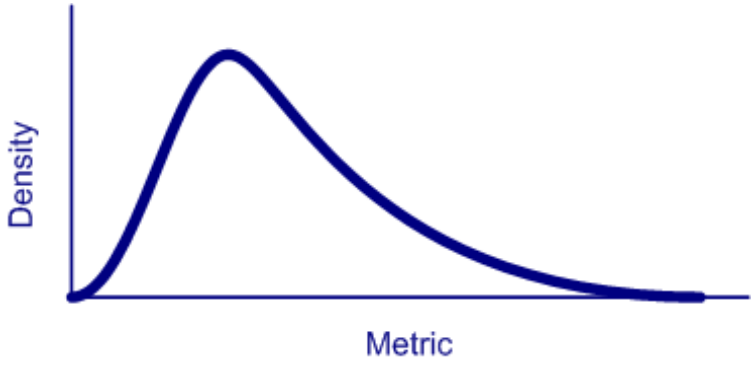
ESTIMATION & PLANNING:

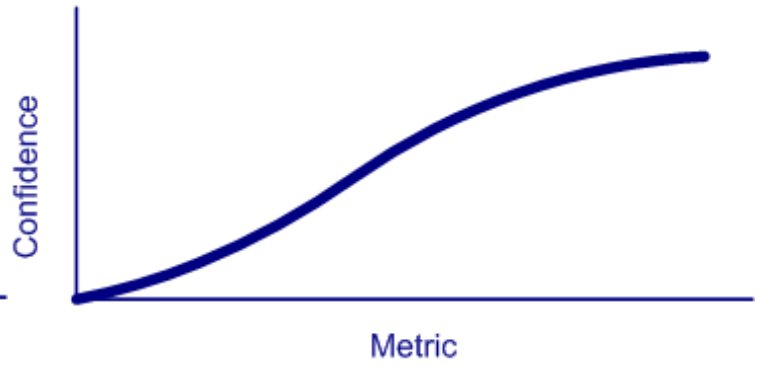
An Estimate Defined

- An ***estimate*** is the most knowledgeable statement you can make ***at a particular point in time*** regarding:
  - Effort / Cost
  - Schedule
  - Staffing
  - Risk
  - Reliability



- Estimates more precise with progress
- ***A WELL FORMED ESTIMATE IS A DISTRIBUTION***





Human Nature: Humans Are Optimists

HBR Article explains this Phenomenon:

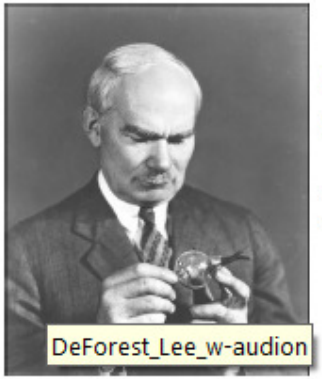
- Humans seem hardwired to be optimists
- We routinely exaggerate benefits and discount costs

Delusions of Success: How Optimism Undermines Executives' Decisions (Source: HBR Articles | Dan Lovallo, Daniel Kahneman | Jul 01, 2003)

Solution - Temper with "outside view":
Past Measurement Results, traditional forecasting, risk analysis and statistical parametrics can help

Don't remove optimism, but balance optimism and realism

While Optimism Needs Tempering, So Does Short Sightedness (Source Northrop)

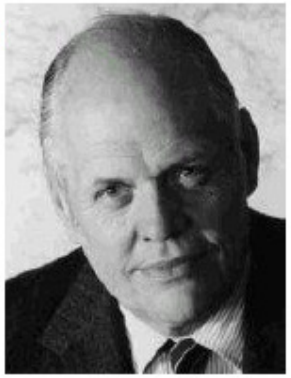


"Man will never reach the moon regardless of all future scientific advances."

- Dr. Lee DeForest, Inventor of Television

"There is no reason anyone would want a computer in their home."

- Ken Olson, president and founder of Digital, 1977



"Airplanes are interesting toys but of no military value."

- Marechal Ferdinand Foch, Professor of Strategy, Ecole Superieure de Guerre

"640K ought to be enough for anybody."

- Bill Gates, 1981

"Any general who's worth his salt knows that war is not a Nintendo game, war is not something that's fought by robots."

- Norman Schwarzkoph, 1991

"To throw bombs from an airplane will do as much damage as throwing bags of flour. It will be my pleasure to stand on the bridge of any ship while it is attacked by airplanes."

- Newton Baker, Sec. of War, 1921

Trouble Starts By Ignoring Project / Program Iron Triangle Realities

- Typical Trouble: Mandated features needed within specific time by given resources

Scope (features, functionality)

Resources **Schedule**

- At least one must vary otherwise quality suffers and system may enter impossible zone!

Pick Two

The Planning Fallacy (Kahneman & Tversky, 1979)

- Judgment errors systematic & predictable, not random
  - Manifesting bias rather than confusion
  - Judgment errors made by experts and laypeople alike
  - Errors continue when estimators aware of their nature
- Optimistic due to overconfidence ignoring uncertainty
  - Underestimate costs, schedule, risks
  - Overestimate benefits of the same actions
- Root cause: Each new venture viewed as unique
  - "inside view" focusing on the components rather than on the outcomes of similar completed actions
  - Typically past more similar assumed
    - even ventures may appear entirely different

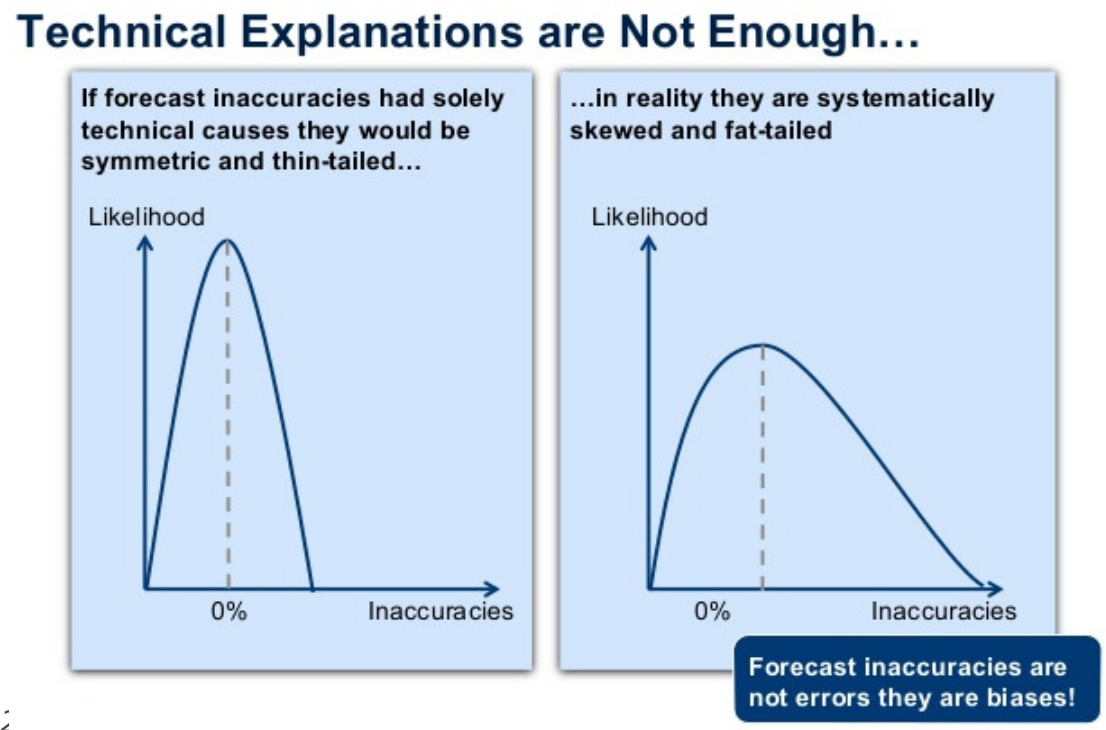
Explanations for Poor Estimating

(Adapted From Source Master Class on Risk, Flybjerg, 2013)

1. Technical: Inadequate data & Models (Vanston)

2. Psychological: Planning Fallacy, Optimism Bias - causes belief that they are less at risks of negative events

3. Political / Economic: Strategic misrepresentation - tendency to underestimate even when experienced with similar tasks overrunning (Flyvberg)



Channel Tunnel Disaster (Source Master Class on Risk, Flybjerg, 2013)

- Actual Costs 200% of Estimates
- Actual Benefits ½ times estimates
- Actaul NPV $-17.8Billion Pounds
- Actual ITT -14.45$

Perform Business Case BUT Eliminate over-optimism in costs and over optimism in benefit

Reference Class Forecasting (adapted from http://www.slideshare.net/assocpm/a-masterclass-in-risk)

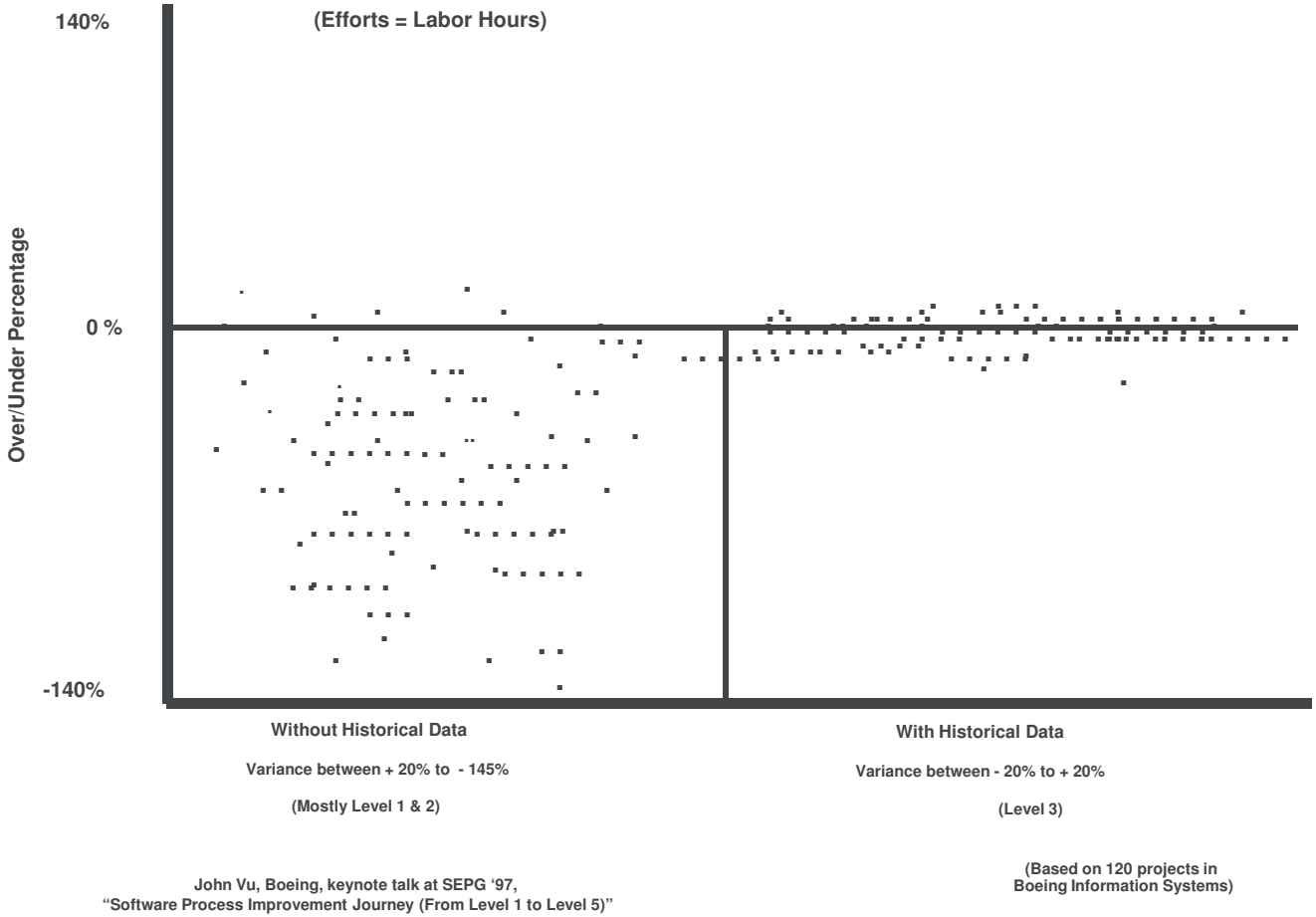
- Best predictor of performance is actual performance of implemented comparable projects (Nobel Prize Economics 2002)
- Provide an "outside view" focus on outcomes of analogous projects
- Reference Class Forecasting attempts to force the outside view and eliminate optimism and misrepresentation
- Choose relevant "reference class" completed analogous projects
- Compute probability distribution
- Compare range of new projects to completed projects

Josiah Stamp Observation On Data & Statistics

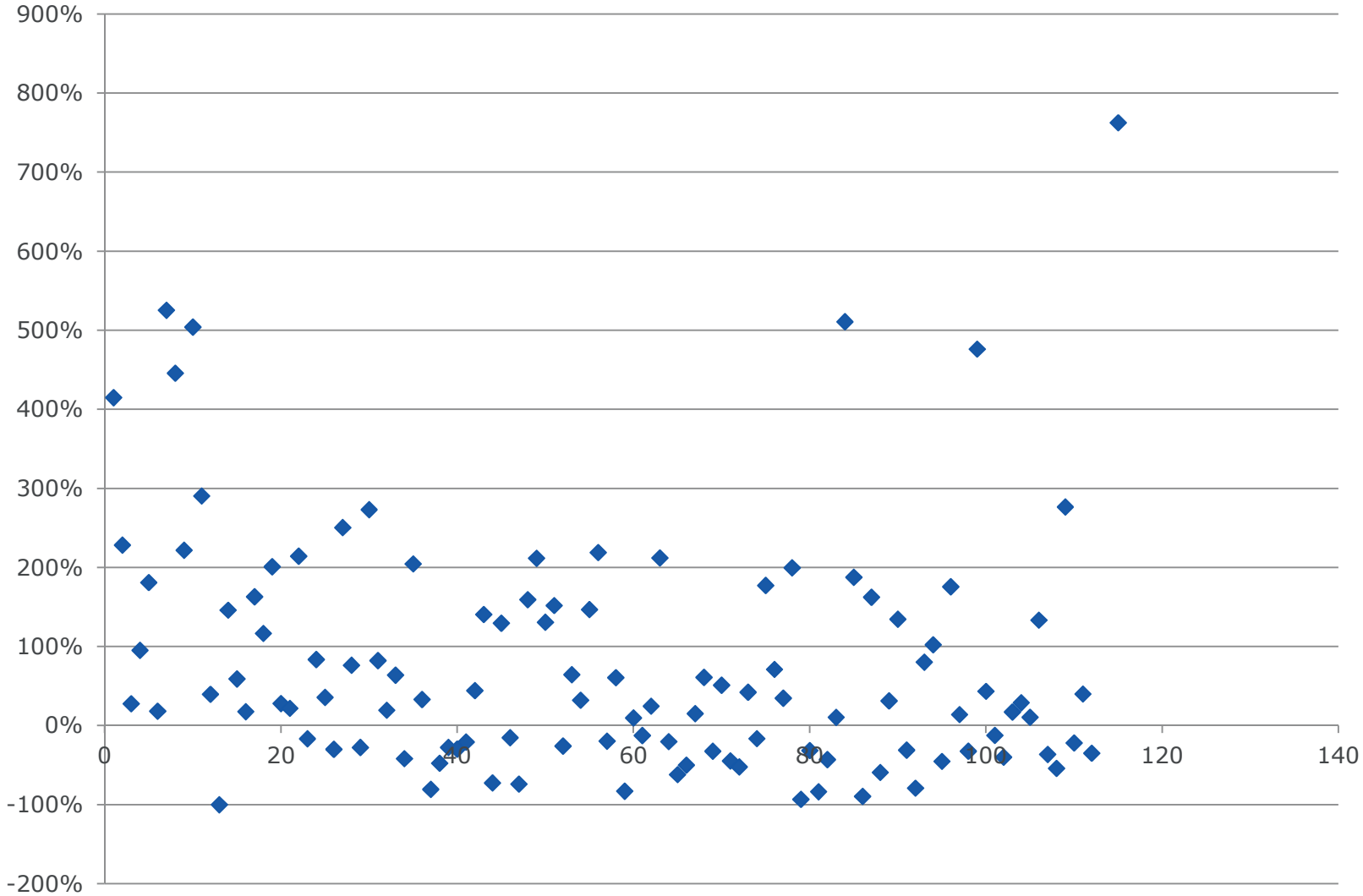
- "The government [is] extremely fond of amassing great quantities of statistics. These are raised to the nth degree, the cube roots are extracted, and the results are arranged into elaborate and impressive displays.
- What must be kept ever in mind, however, is that in every case, the figures are first put down by a village watchman, and he puts down anything he … pleases.
- Attributed to Sir Josiah Stamp, 1840-1941, H.M. collector of inland revenue.

Most Data is imperfect…
And much imperfect data is usable

Data Improves Estimates For New Programs Source: John Vu, Boeing SEPG 1997



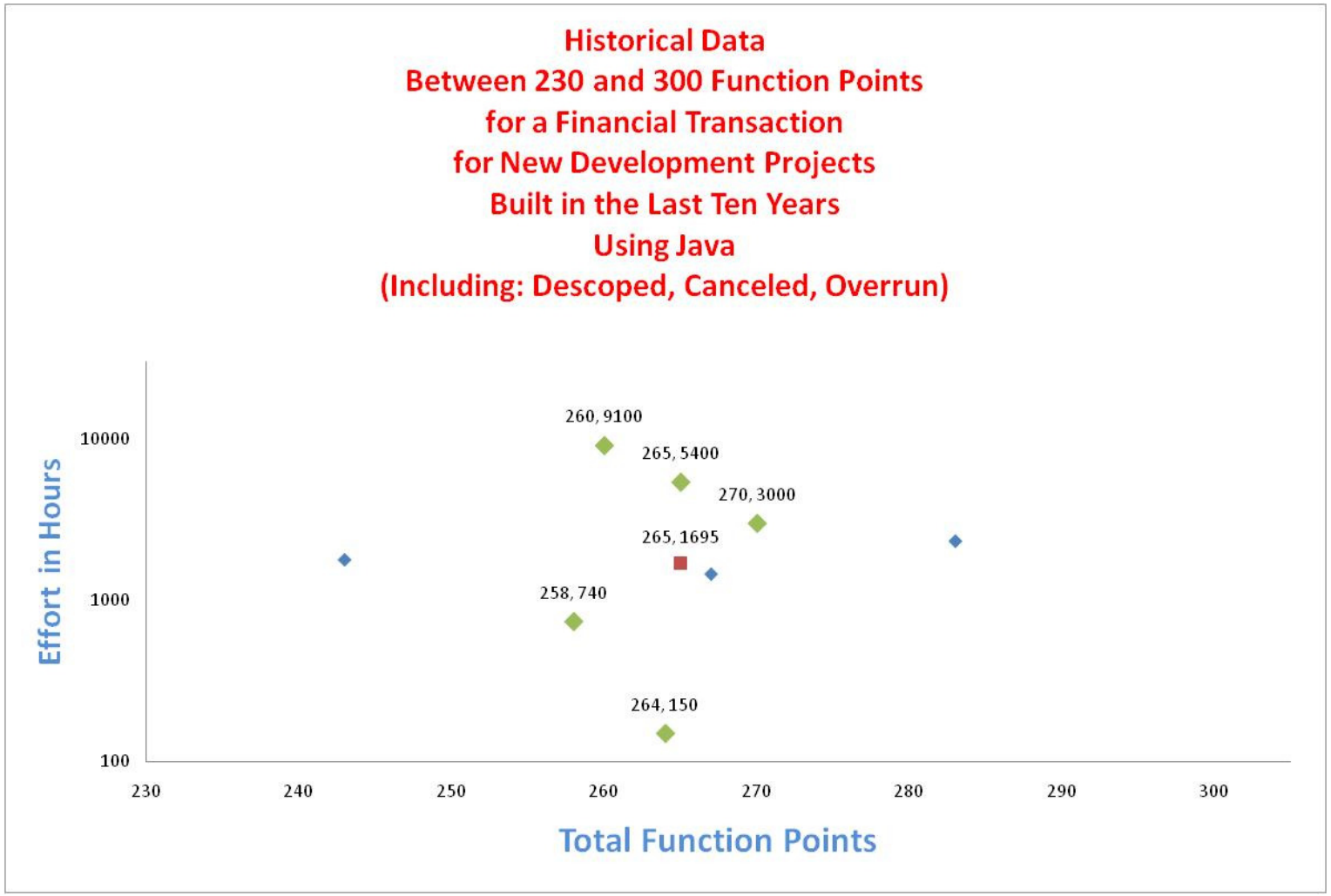
SRDR Estimate New SLOC vs Actual (Note: HUGE outliers removed to make the graph more readable)



Gross underestimation of software size versus actual

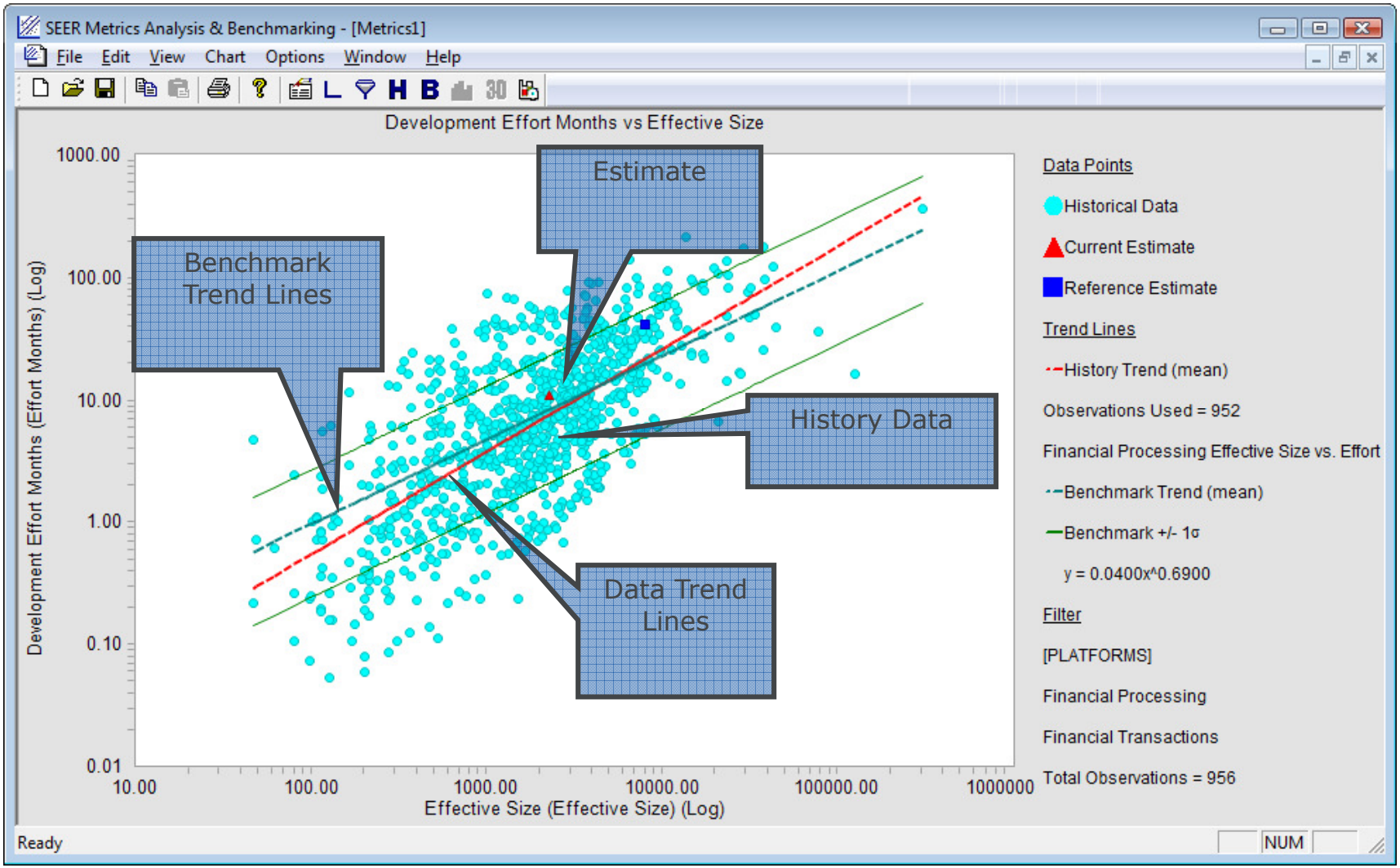
Fallacy of Silent Evidence

What about what we don't know?

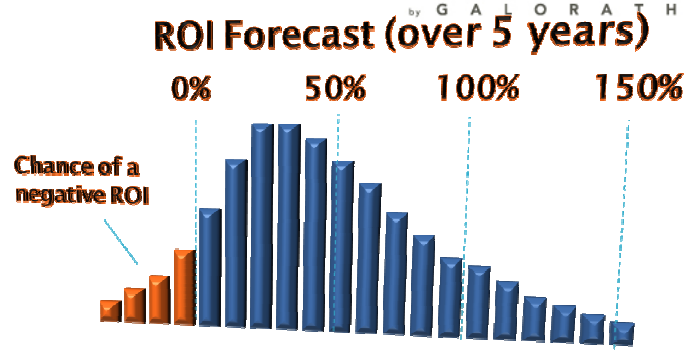


How confident would you feel if the Silent Evidence was visible?

Example: Parametric Estimate Compared With History



ROI Analysis of A New System



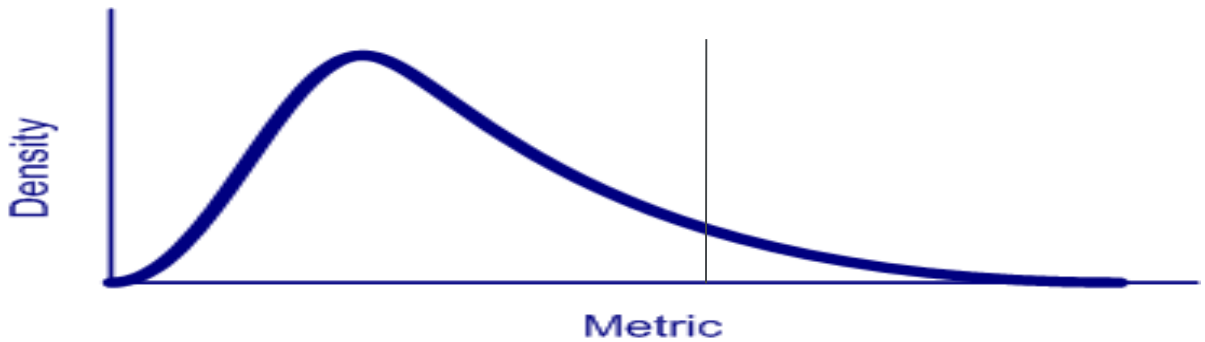
Cost of capital 8.0%

	Initial Investment	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Total Ownership
Investment	$100,000								$100,000
Increase/(dec.) in revenue		($40,000)	$60,000	$110,000	$100,000	$100,000	$150,000	$150,000	$630,000
Increase/(dec.) in op. exp.		$90,000	$70,000	$70,000	$22,000	$24,000	$27,000	$28,000	$331,000
Cash Flow	($100,000)	($130,000)	($10,000)	$40,000	$78,000	$76,000	$123,000	$122,000	$199,000
PV of Cash Flow	($100,000)	($120,370)	($8,573)	$31,753	$57,332	$51,724	$77,511	$71,186	$60,563
NPV	**60,563**								$60,563
IRR	**13.5%**								13.5%
ROI	**121%**								121.1%

A Complete ROI analysis should analysis risk and uncertainty as well as likely

Manual Estimates: Human Reasons For Error (Adapted from Goldratt)

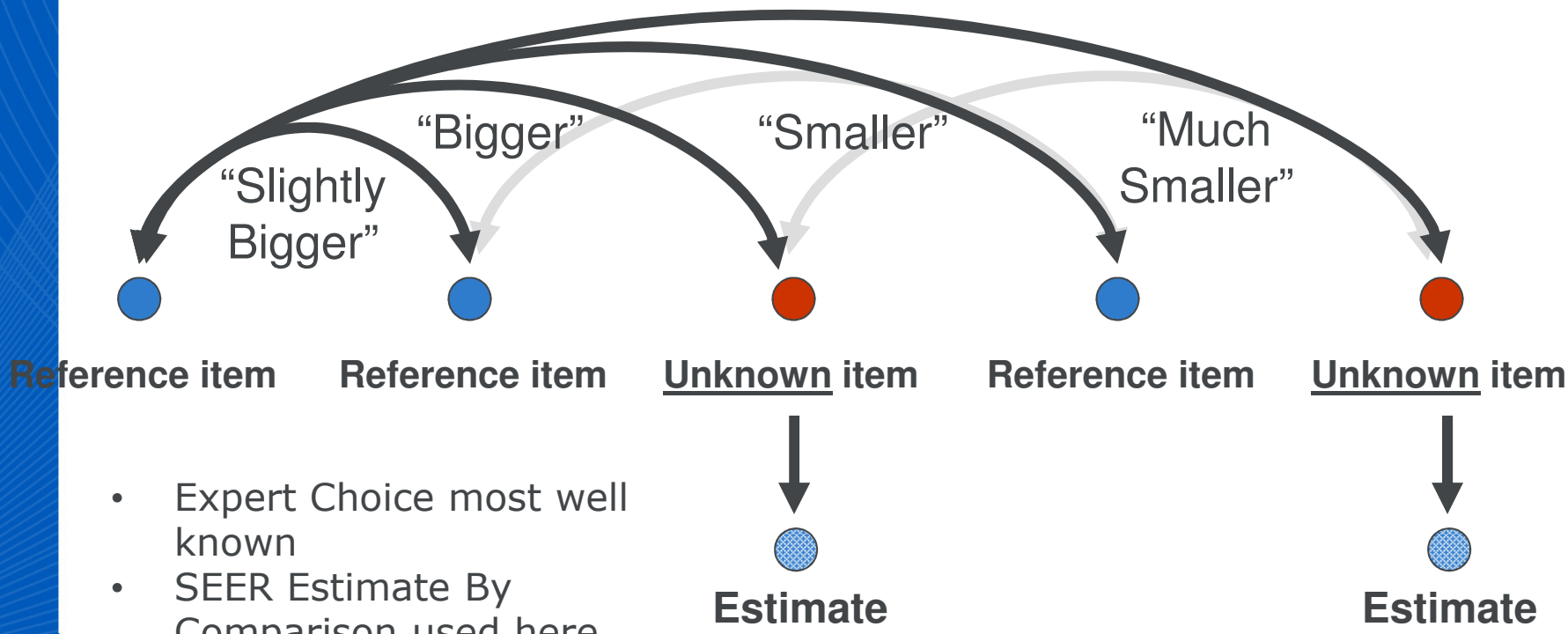
- Desire for "credibility" motivates overestimate behavior (80% probability?)
  - So must spend all the time to be "reliable"
  - Better approach force 50% probability & have "buffer" for overruns
- Technical pride causes underestimates
- Buy-in causes underestimates



Comparison Estimation In A Nutshell

- Reference items have known value
- Unknown items are to be determined

Gross ratios between reference and unknown item pairings are made

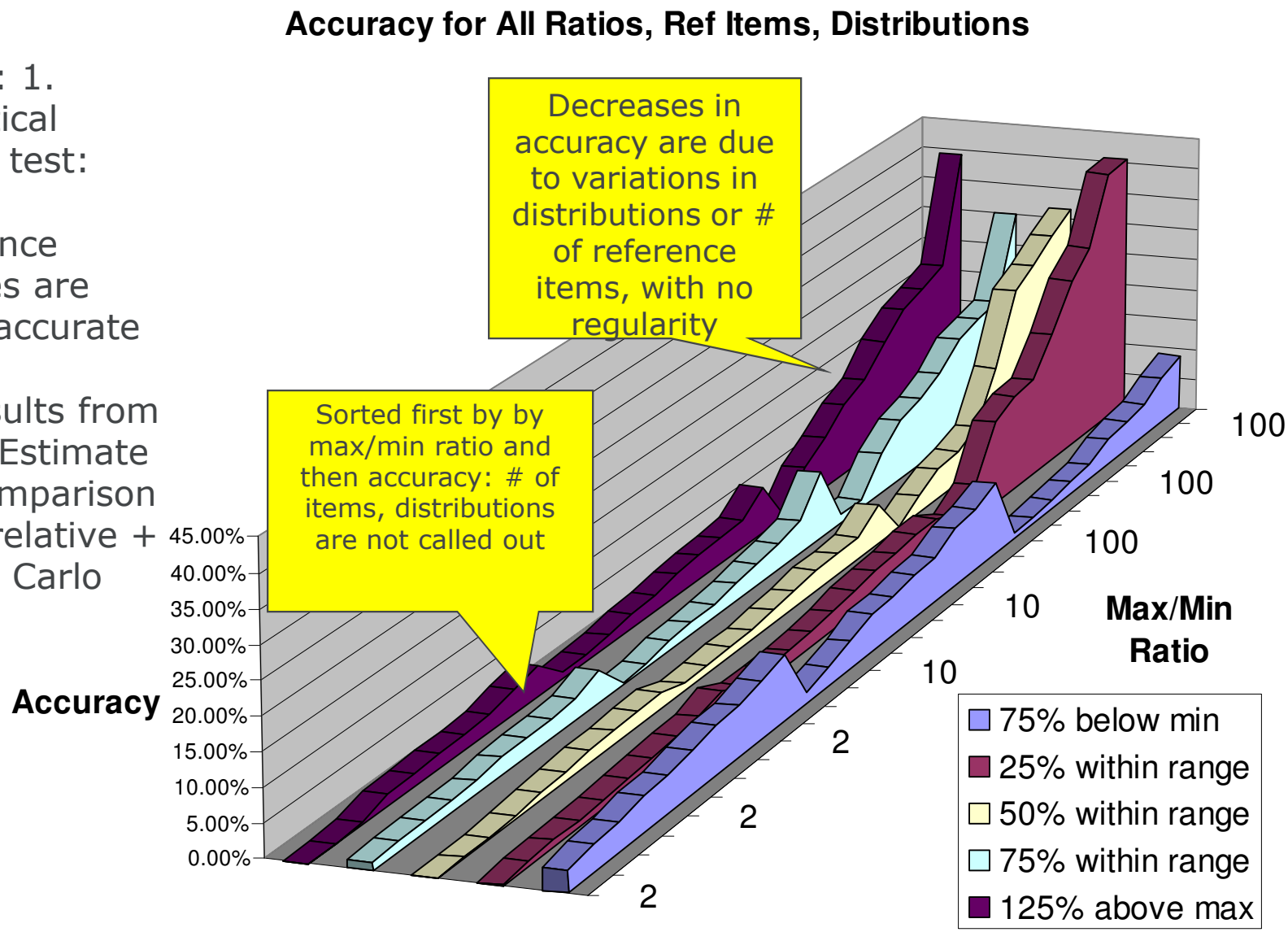


- Expert Choice most well known
- SEER Estimate By Comparison used here

SEER Comparison Estimating Typically Within 10% of Actuals

Notes: 1. statistical stress test: Viable reference choices are most accurate

2. Results from SEER Estimate By Comparison Uses relative + Monte Carlo



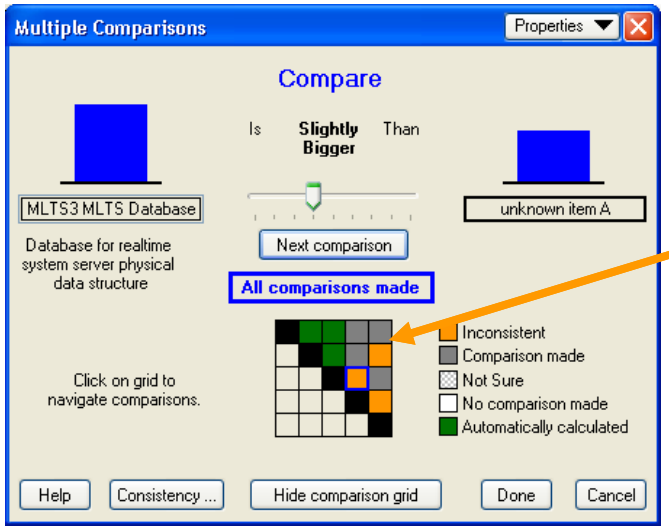
Comparative Estimating Can Flag Inconsistency

What is Consistency?
Not violating transitivity

For A > B > C, saying

"A > B and A > C" is <u>consistent</u>
"A > B and A < C" is <u>inconsistent</u>

Consistency ⇒ Estimate accuracy



SEER-Estimate By Comparison assists the user in identifying inconsistent comparisons

UK Optimism Bias Adjustment Ranges

(http://webarchive.nationalarchives.gov.uk/20130129110402/http://www.hm-treasury.gov.uk/d/5%283%29.pdf)

Project Type	Works Duration (Schedule)		Capital Expenditure (Cost)	
Standard Buildings	4%	1%	25%	2%
Non-Standard Buildings	39%	3%	51%	4%
Standard Civil Engineering	20%	1%	44%	3%
Non-Standard Civil Engineering	25%	3%	66%	6%
Equipment /Development	54%	10%	200%	10%
Outsourcing			41%	

UK Green Book Apply Optimism Bias

ttp://webarchive.nationalarchives.gov.uk/20130129110402/
http://www.hm-treasury.gov.uk/d/5%283%29.pdf

SEER by GALORATH

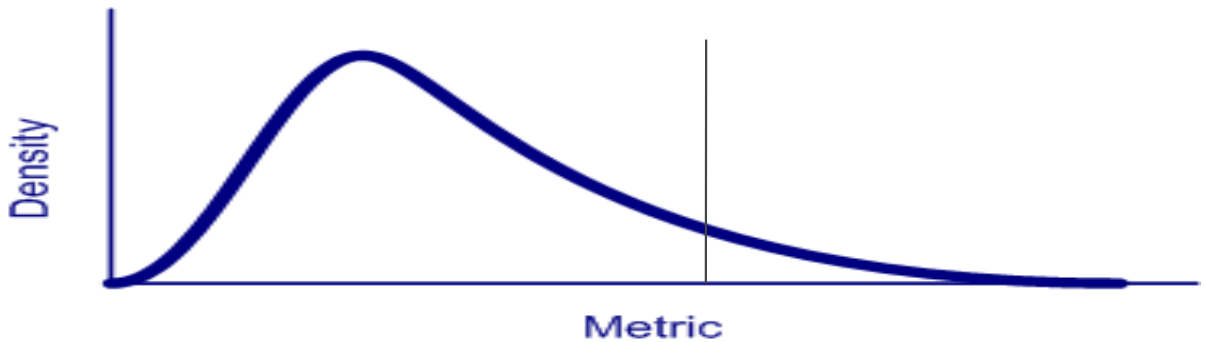
Step 1: Decide which project type(s) to use → Step 2: always start with the upper bound → Step 3: Consider whether optimism bias factor can be reduced → Step 4: Apply the optimism bias factor → Step 5: Review optimism bias adjustment

5 Levels of Risk Management (Adapted from Flyvbierg)

1 Opinions	2 Benchmarking	3 Diligence	4 Rigorous Estimating	5 Risk Analysis
As unbiased as possible	Comparing to viable database	Estimate review	Parametric Relative Reference Class Forecasting	Risk management Black Swan mitigation

Manual Estimates: Human Reasons For Error (Metrics Can Help)

- Manual Task estimates yield SIGNIFICANT error
- Desire for "credibility" motivates overestimate behavior (80% probability?)
  - So must spend all the time to be "reliable"
  - Better approach: force 50% probability & have "buffer" for overruns
- Technical pride sometimes causes underestimates



Myth and Reality (Source Hamaker)

- Myth: The more details, the better the estimate!
  - If a 100 WBS element is good…
    - Then 1000 elements if better still…
    - And 10 elements is to be sneered at…
    - And if it is "parametric" it is worse yet
- But on the contrary, if one *simply wants to know what a project might cost*, details are counterproductive!
- And parametric estimating is the preferred choice

See next charts for substantiating arguments

Comparison of Parametric & Bottoms Up Methods (Source Hamaker)

SEER by GALORATH

Parametric Estimates

- **Top down**
- **Less detail**
- **Based on performance metrics**
- **Less labor intensive**
- **Quicker**
- **Ease of trade-offs analyses**
- **Parametric database**
- **Not always accepted**
- **"Black Magic" aura**
- **Generally more disciplined**
  - Standard methodology
  - Independent
  - Done by trained analysts
  - Captures totality of past programs

Detailed Build-Up Estimates*

- **Bottoms up**
- **More detail**
- **Based on time and material**
- **Labor intensive**
- **Time consuming**
- **Trade offs need details**
- **Performance standards**
- **Accepted method**
- **Generally understood**
- **More susceptible to distortions**
  - Optimism/Pessimism
  - Special interest/buy-in
  - Done by managers/engineers
  - Missing
    - "I forgots"
    - Unknowns

***AKA "labor-material build up", "grass roots", "bottoms up" "engineering estimates"**

Dishwashing Experiment

You have just had a dinner party and the stir-fry, salad, fresh bread, apple pie and coffee were all great. Your guests have gone and it is time to clean up. Your dishwasher is broken and you need to hand-wash the dishes, silver, and pans listed below; and put them in the drying rack next to the sink. The dishes have been sitting randomly stacked in the sink and on the counter for a couple of hours, but no food is burned on.

You need to clean:

- 4 large dinner plates
- 4 desert plates
- 4 sets of silver (2 forks, knife and spoon)
- 4 sets of coffee cups and saucers
- 4 salad bowls
- 2 serving bowls
- Salad tongs
- Bread knife
- Pie serving knife
- 1 wok
- 1 sauce pan
- A pie pan
- A bread pan
- A cream pitcher
- Serving spoon

You have a sponge, scrub brush, dish washing soap and plenty of hot and cold water. After stacking the clean dishes in the drying rack, you need to make sure the 40 in (100cm) square counter top and sink are clean also.

Psychological Effects Tested (Source: JPL http://www.slideshare.net/NASAPMC/arthurchmielewski)

1. Anchoring: Train the managers **not to anchor**
2. Question & Answer Mismatch: Establish **proper Estimation Language** so questions compatible with common interpretation
3. Decomposition: Deep **decompositions may not improve accuracy**
4. Reserve Comfort Calculate the **reserve based on risk**
5. Planning Fallacy: People plan for optimistic case instead of including risk

- 507 volunteers
  - 142 JPLers, 305 college students and 60 other adults. ~2300 data points were collected

Anchoring Causes Flawed Estimates

Objective: Test how easily influenced people may be by wrong answer – "the anchor."

The anchor set asked:

"Estimate how many minutes it will take you to clean the kitchen. One respondent estimated that it will take about 10 minutes to finish cleaning up. He may be wrong of course."

- Nominal 30 min, anchored case 25 min
- Best case estimate was 27 min
- 2 min LONGER than the anchored result
- Conclusion: easy to dramatically skew estimates by asking anchored questions, such as:
  - "We would like you to come in around $6M"
  - "I have a target of $400k for you"
  - "the last robot arm we built cost $7M"…

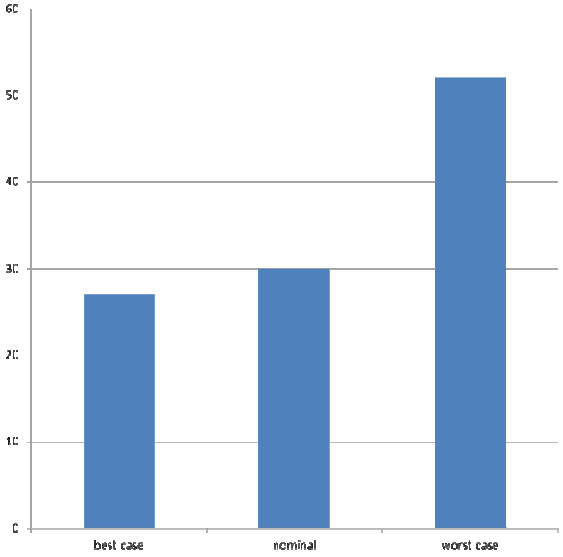
Question & Answer Mismatch (Source: JPL http://www.slideshare.net/NASAPMC/arthurchmielewski)

- Test for mismatch between expected and provided
- Different participants were asked:
- "Estimate how many minutes it will take you to clean the whole kitchen"
  - There is a 50% chance you will finish within ___ min
  - There is a 75% chance you will finish within ___ min
  - There is a 99% chance you will finish within ___ min
- 50% confidence estimate 31 min
- nominal estimate 30 min
- People interpret nominal 50% case (Meaning you will exceed estimate in half the cases)
- But manager probably more reliable result, probably in the 70%-90% confidence range…

This is why we say a complete estimate must include a probability

Planning Fallacy Results (Source: JPL http://www.slideshare.net/NASAPMC/arthurchmielewski)

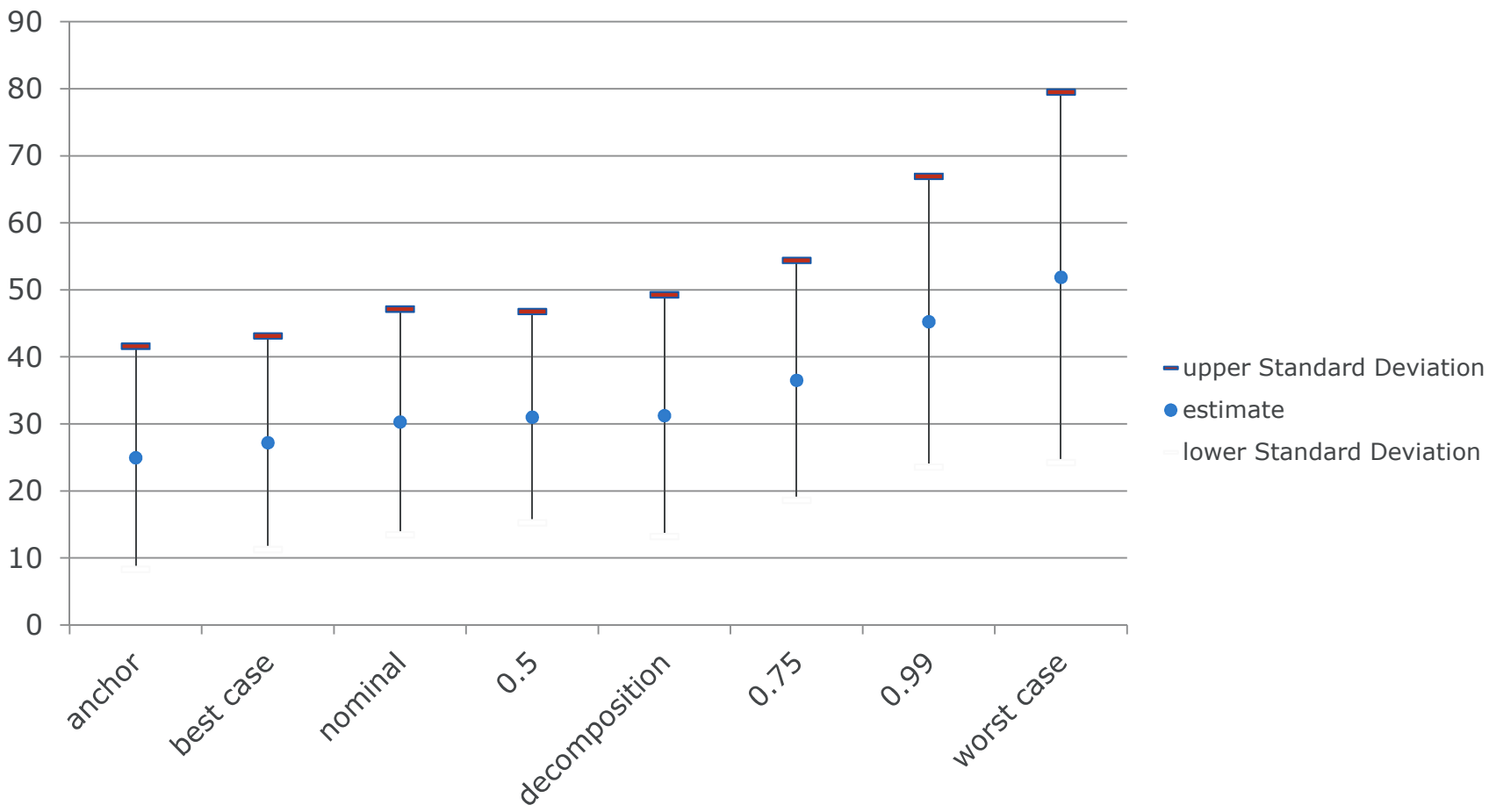
- The following results were obtained:
  - 51 min worst case
  - 45 min 99% confidence
  - 30 min nominal
  - 27 min best case



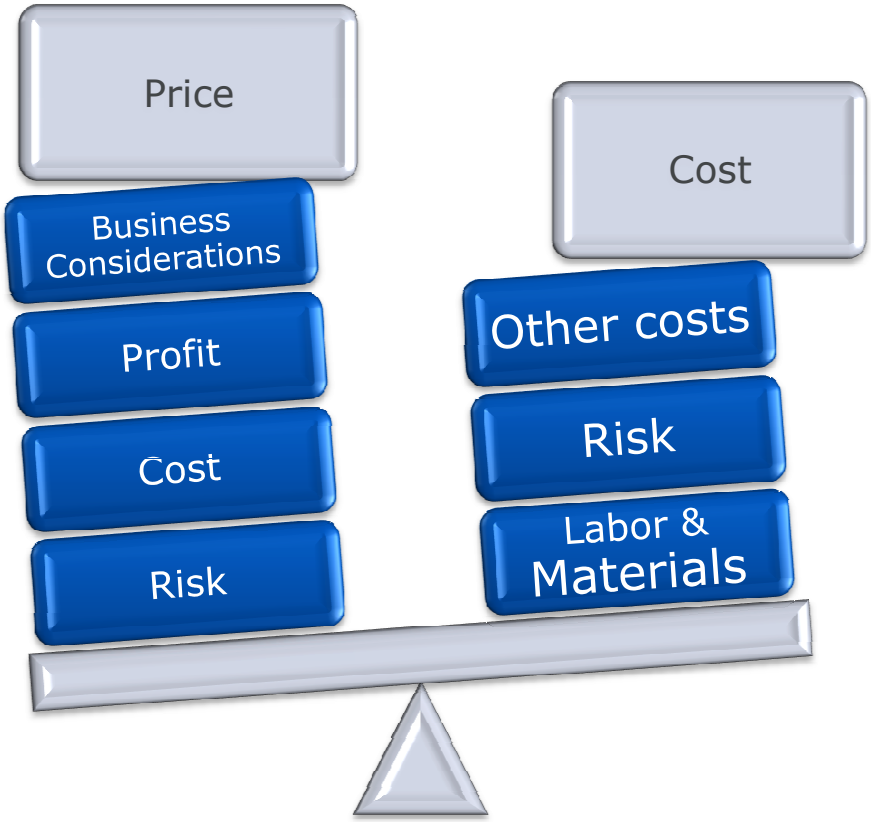
- People skewed people toward optimism
- Nominal estimate 10% longer than best case but 70% shorter than the worst case

People are so optimistic that it was easy to anchor them down but anchoring up failed

Answers Analysis (Source: JPL http://www.slideshare.net/NASAPMC/arthurchmielewski)



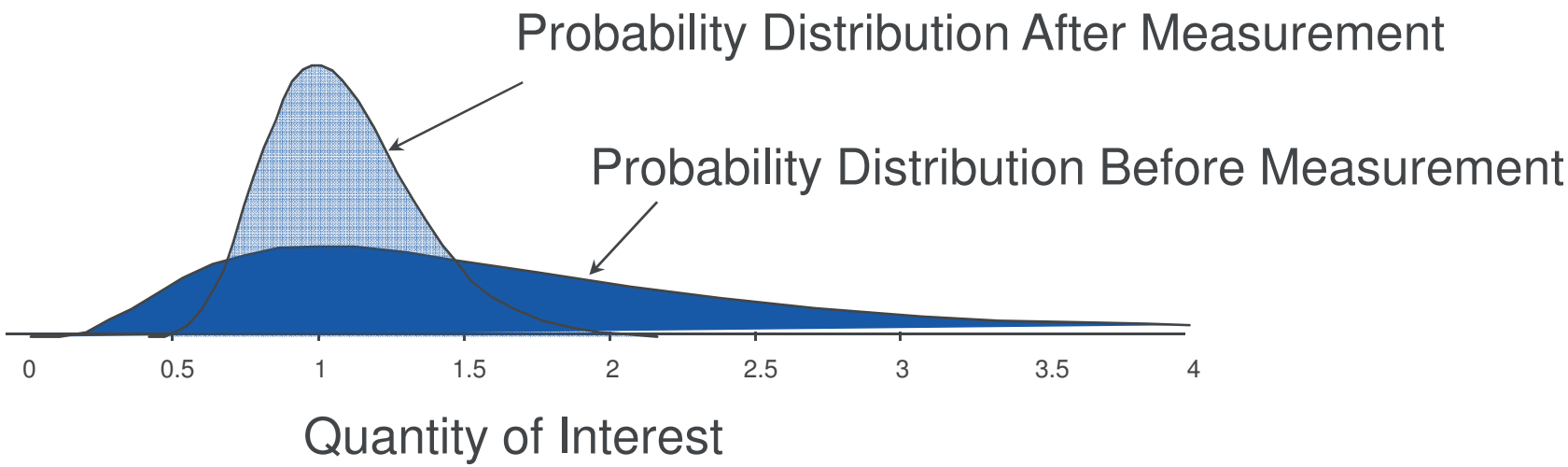
Remember Cost and Price Are Different (Adapted from Morton)



- **Price**: Amount Charged to Customer (considering cost, profit, risk, Price to win, business considerations, etc.)
  - e.g. New Car - Discounts
  - e.g. Machinists - Idle
  - e.g. Golden Gate Bridge - Cables
  - e.g. NASA - Photos

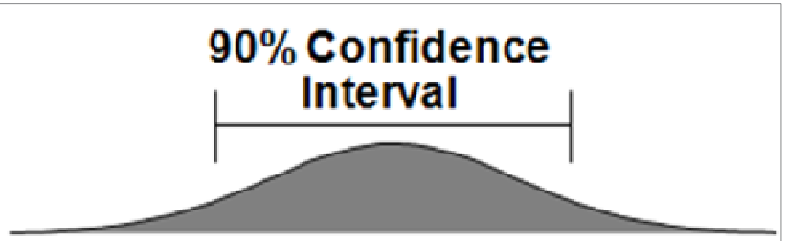
Hubbard: Measure To Reduce Uncertainty

- Perception that measurement is a point value is a key reason why many things are perceived as "immeasurable"
- Measurement: Quantitatively expressed reduction in uncertainty based on observation



Assumptions, Change Drivers & Expert Judgment Need Caution (Source: Hubbard)

- Most people are significantly ***overconfident*** about their estimates ... especially educated professionals



Group	Subject	% Correct (target 90%)
Harvard MBAs	General Trivia	40%
Chemical Co. Employees	General Industry	50%
Chemical Co. Employees	Company-Specific	48%
Computer Co. Managers	General Business	17%
Computer Co. Managers	Company-Specific	36%
AIE Seminar (before training)	General Trivia & IT	35%-50%
AIE Seminar (after training)	General Trivia & IT	~90%

(AIE = Hubbard Generic Calibration Training)

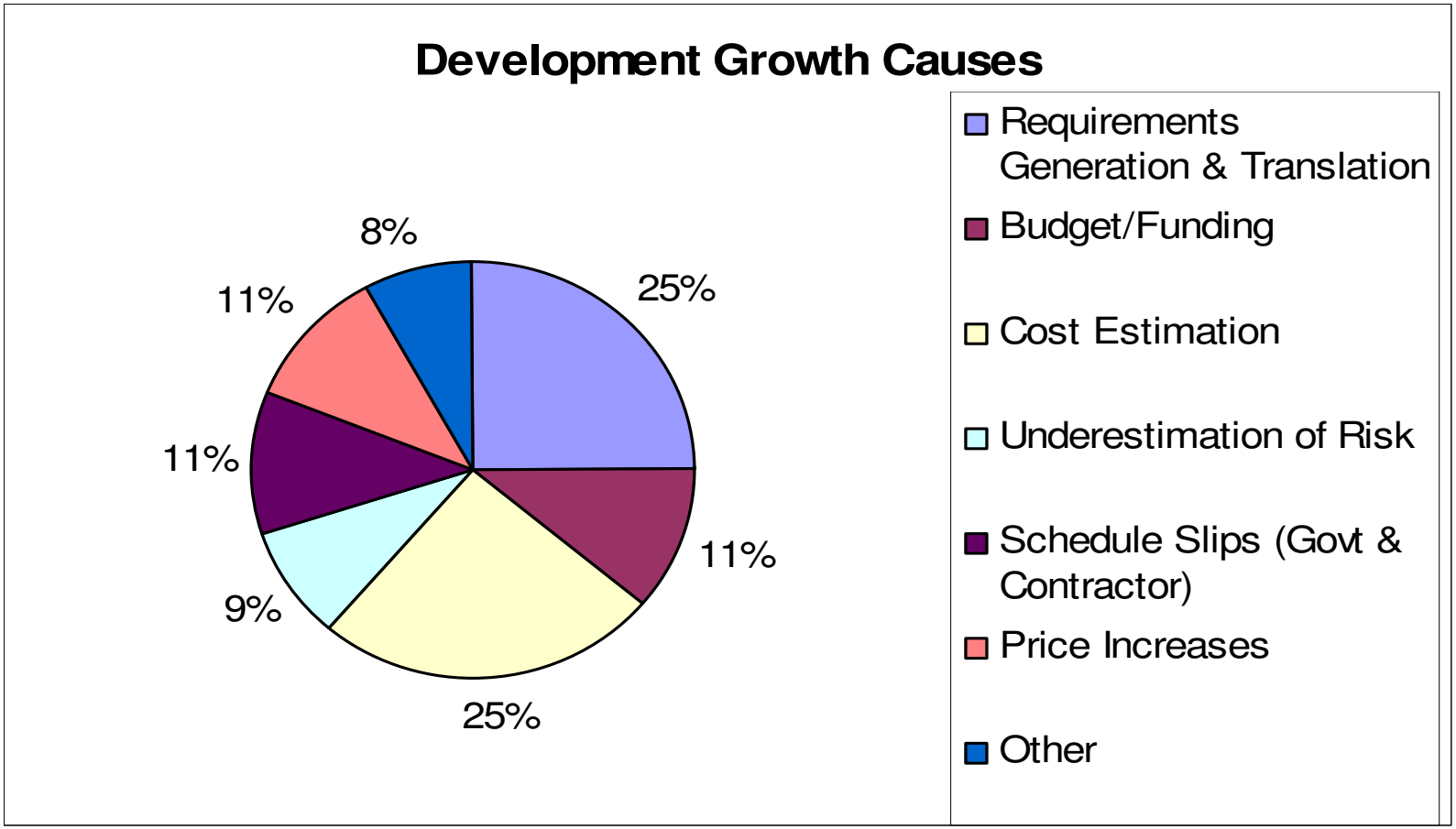
Used with permission from Douglas Hubbard
Copyright HDR 2008 dwhubbard@hubbardresearch.com

Gunning for Models (Adapted from Hubbard)

- Be careful of red herring arguments against models
  - "We cannot model that…it is too complex."
  - "Models will have error and therefore we should not attempt it."
  - "We don't have sufficient data to use for a model."
  - "It works but we cant see all data so we should not use it"
- Build on George E. P. Box: "Essentially, all models are wrong, but some are useful."
  - Some models are more useful than others
  - Everyone uses a model – even if it is intuition or "common sense"
  - So the question is not whether a model is "right" or whether to use a model at all
  - Question is whether one model measurably outperforms another
  - A proposed model (quantitative or otherwise) should be preferred if the error reduction compared to the current model (expert judgment, perhaps) is enough to justify the cost of the new model

Total Cost Growth for Two Space Programs

David Graham, NASA



Quantitative Framework

[5] "The Success Triangle of Cost, Schedule, and Performance: A Blueprint for Development of Large-Scale Systems in an Increasingly Complex Environment" - (Booz|Allen|Hamilton, 2003)

Key Points

Experts are likely providing biased estimates

Poor estimates are a root cause of project failure

Estimates can be better, squelching bias & strategic mis-estimation… Parametrics help.