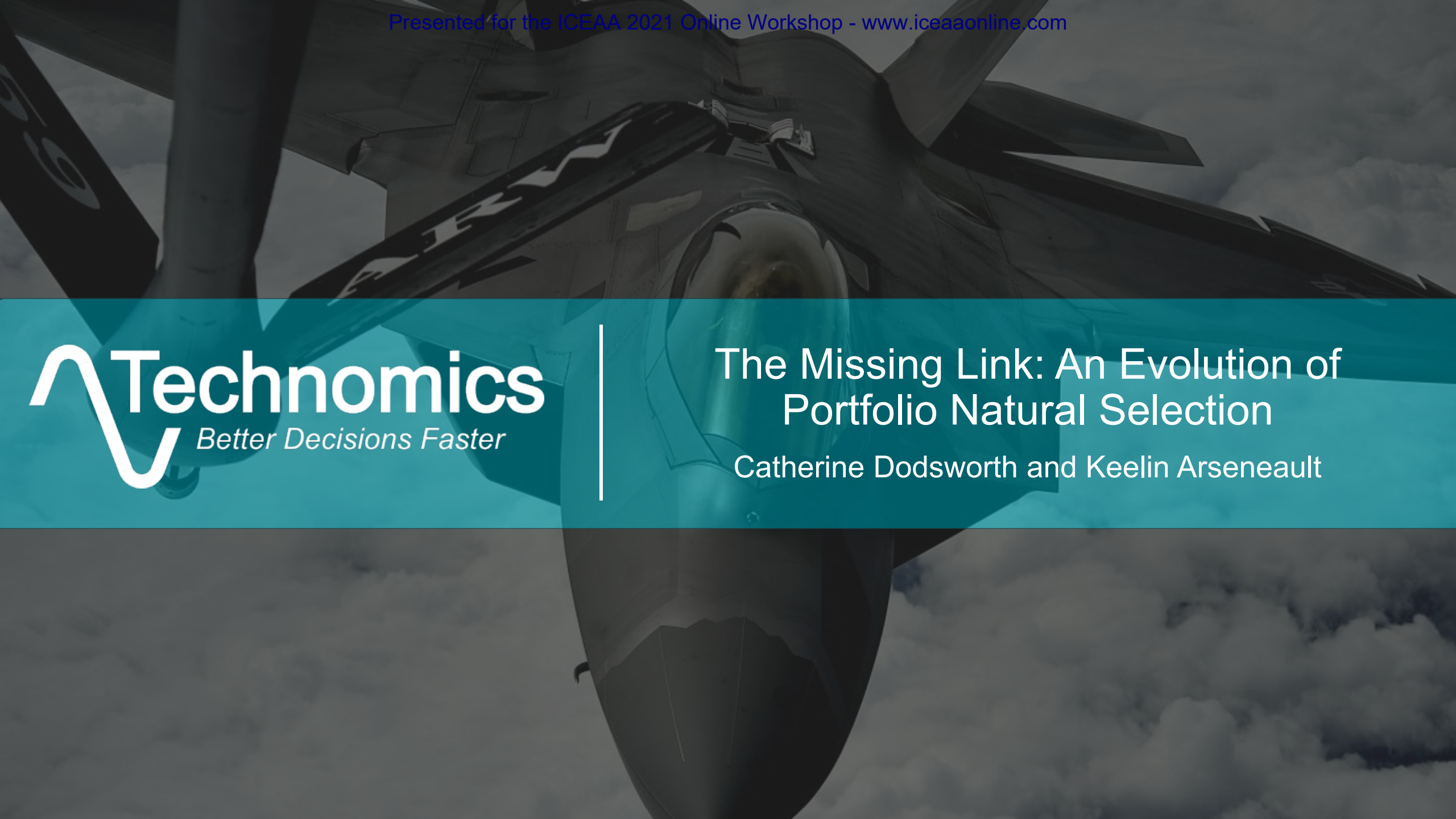
Presented for the ICEAA 2021 Online Workshop - www.iceaaonline.com



Technomics
Better Decisions Faster

The Missing Link: An Evolution of Portfolio Natural Selection

Catherine Dodsworth and Keelin Arseneault

Presenting Today

Catherine Dodsworth

- Experience includes life cycle cost estimating
- Biologist in another life
- B.S. in Biology from University of Mary Washington
- M.S. in Business Analytics from Quinnipiac University

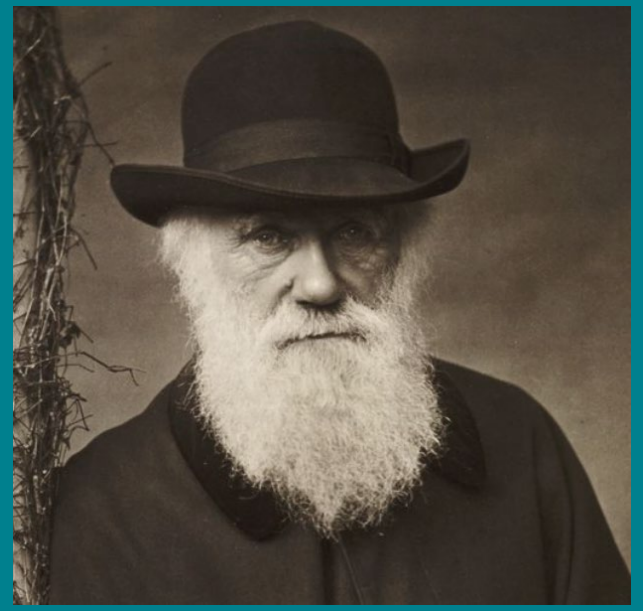
Keelin Arseneault

- Current cost analyst, former biology student researcher
- BA in Biology from College of the Holy Cross
- Pursuing MS in Statistics with concentration in Data Science from George Washington University

Contributors: Remmie Arnold, Brian Flynn, Stella O'Sickey

“It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is the most adaptable to change.”

–Charles Darwin, Origin of Species



https://www.thoughtco.com/charles-darwin-his-origin-of-the-species-1773841

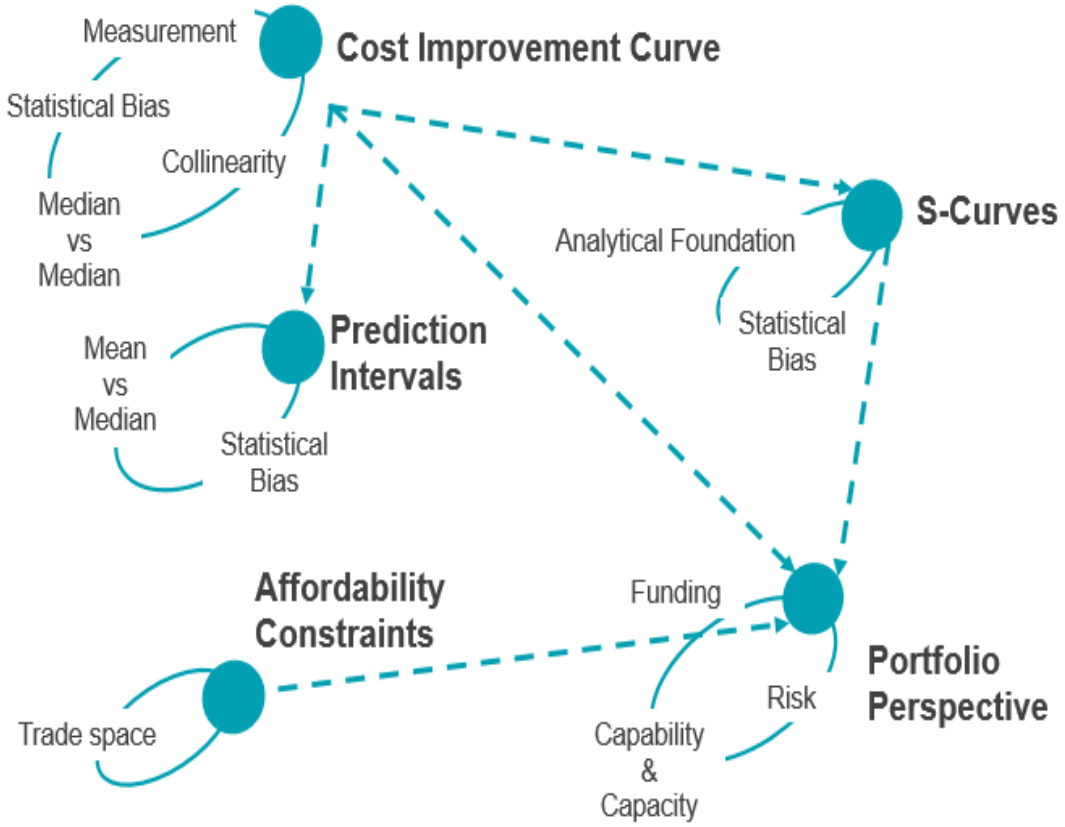
Agenda

- BLUF
- Background and Challenges
- The Missing Links
- Solutions

BLUF

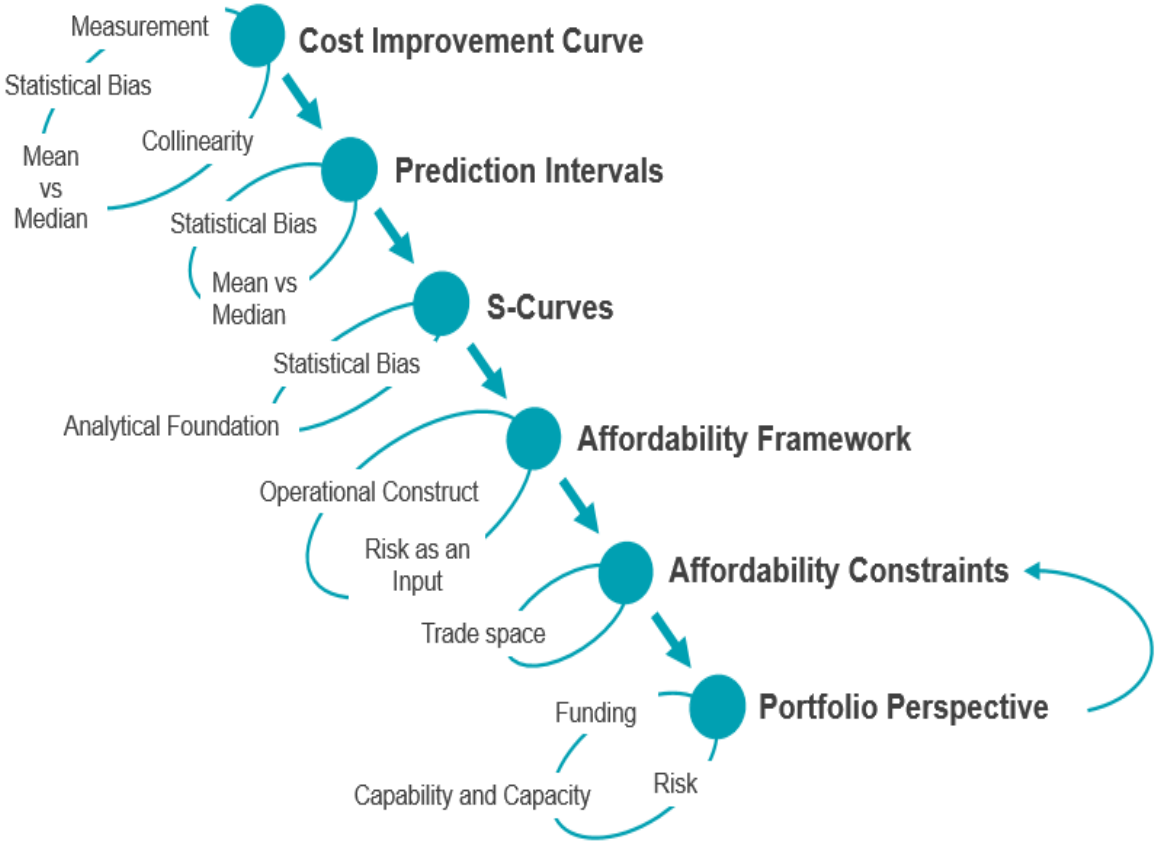
Status Quo

- Biased estimates
- Stove-piped systems
- Myopic view of affordability
- Poor risk analysis

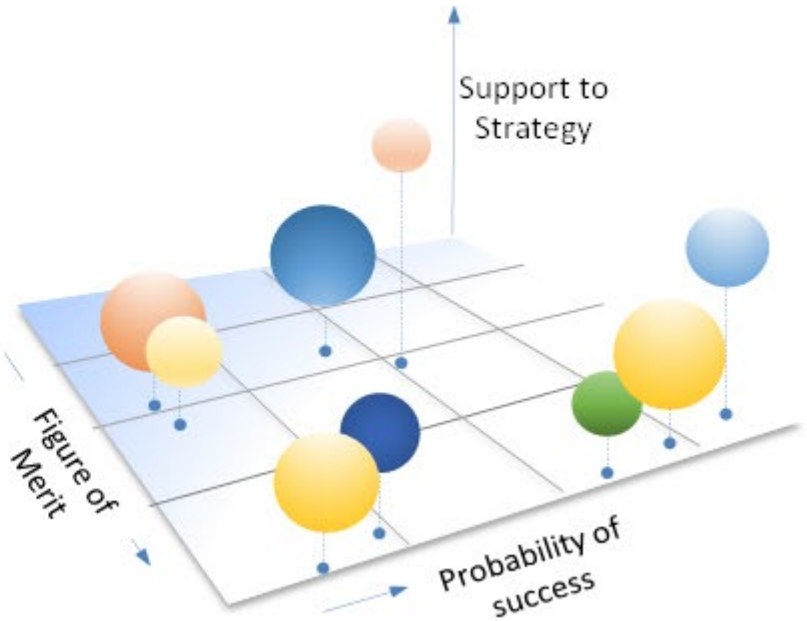


Solution

- Unbiased estimates
- Integrated steps
- Portfolio view of affordability
- Enhanced risk analysis



Strategic Challenges



Imperfect Synchronization

Program & Portfolio Perturbation

Acquisition Tensions

- Costs vs schedule
- User needs
- Risk

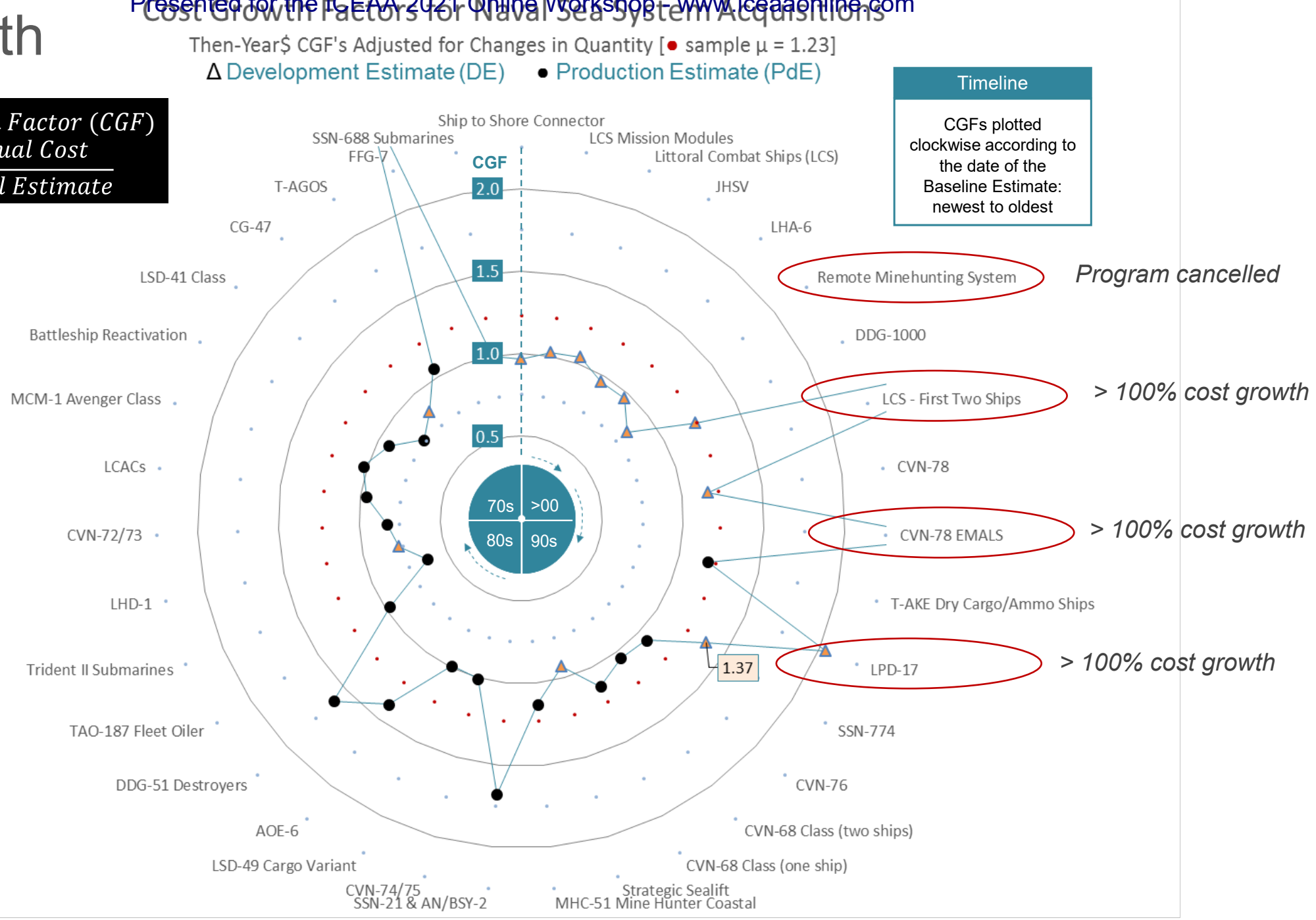
Portfolio Tensions

- Limited Budget
- Competing Demands
- Risk

Inherent tensions lead to poor acquisition outcomes

Cost Growth

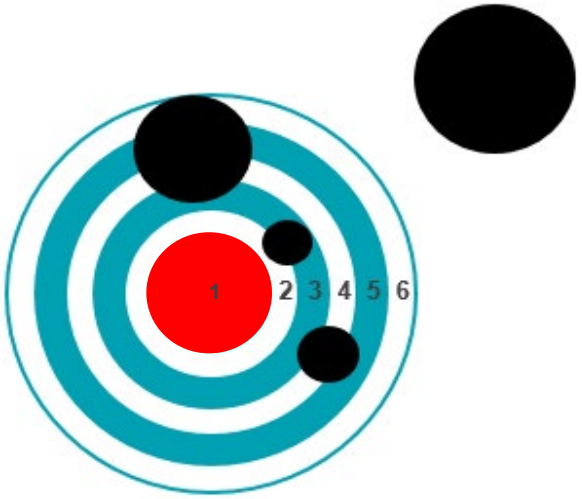
$$CostGrowth\ Factor\ (CGF) = \frac{Actual\ Cost}{Initial\ Estimate}$$



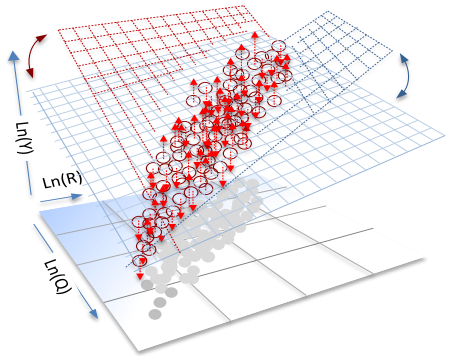
Root-Cause Pain Points

1 Biased Estimates

- Learning curve estimates off target
- Bias (distance from bullseye) increases with uncertainty (bubble size)!



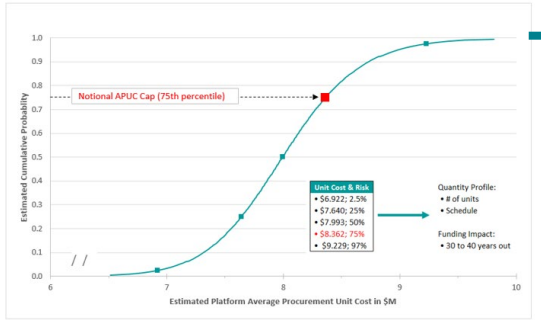
2 Collinearity & Rate



"Bouncing Betas"

3 Poor Risk Analysis

- Biased inputs
- Weak linkage to CERs
- Without affordability options



4 Myopic Perspective

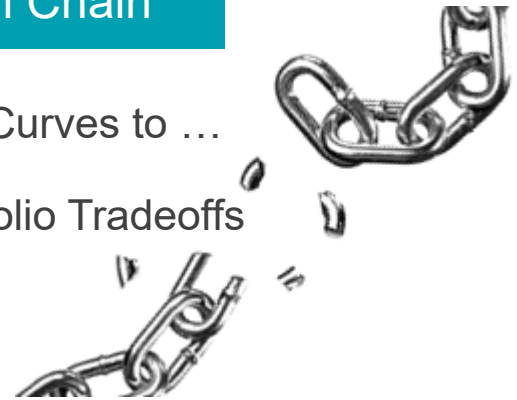
FYDP Yr 1	FYDP Yr 2	FYDP Yr…
Appropriation 1	$	$
Appropriation 2	$	$
Appropriation 3	$	$

Affordability viewed against the budget rather than the portfolio

5 Broken Chain

Learning Curves to …

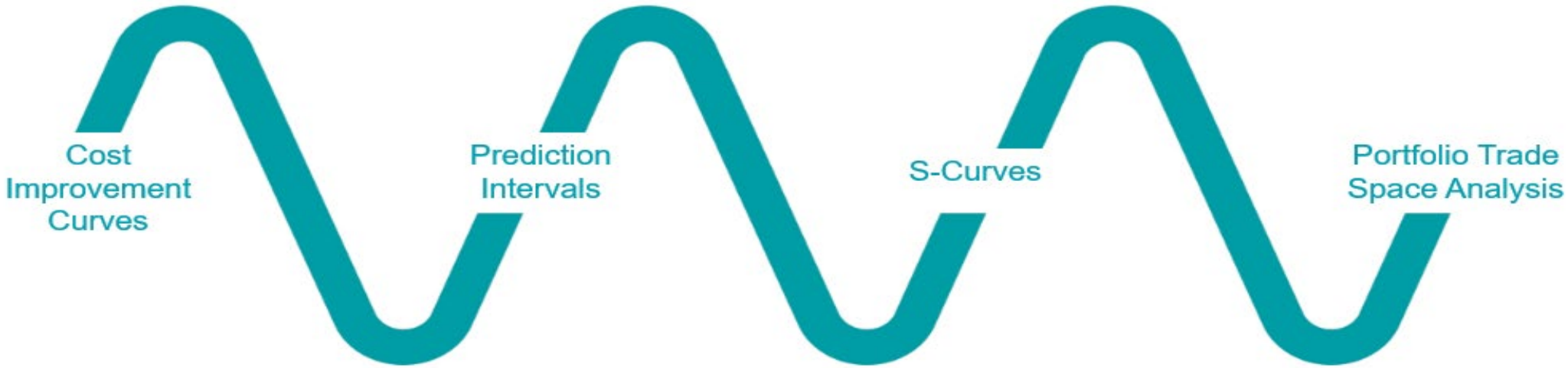
Portfolio Tradeoffs



Technomics
Better Decisions Faster

The Missing Links

Adapt and overcome!



We have the pieces, we just need to forge the linkages!

Cost Improvement Curve – The Estimation Problem

(1) $Y_i = \alpha\, Q_i^b\, R_i^c\, e^{\varepsilon_i}, \varepsilon_i \backsim N(0, \sigma_\varepsilon^2), i = 1\ to\ n,$ where:

$Y_i =$ an observation on the dependent variable (e.g., unit cost or hours),

$Q_i =$ an observation of cumulative quantity,

$R_i =$ an observation of rate of production, and where $\alpha, b,$ and c are population parameters to be estimated.

(2) $\ln(Y_i) = \ln(\alpha) + b\ln(Q_i) + c\ln(R_i) + \varepsilon_i.$

We estimate in Log space

(3) $\hat{\alpha} = e^{\widehat{\ln(\alpha)}},$

Trying to get alpha

(4) $E(\hat{\alpha}) = E\left(e^{\widehat{\ln(\alpha)}}\right) \neq e^{E(\widehat{\ln(\alpha)})}$, but instead

But end up with this!

$$= e^{E(\widehat{\ln(\alpha)}) + \frac{1}{2}\sigma^2_{\widehat{\ln(\alpha)}}} = e^{\ln(\alpha) + \frac{1}{2}\sigma^2_{\widehat{\ln(\alpha)}}} = \alpha e^{\frac{1}{2}\sigma^2_{\widehat{\ln(\alpha)}}}$$

The problem is when we un-log the logs!

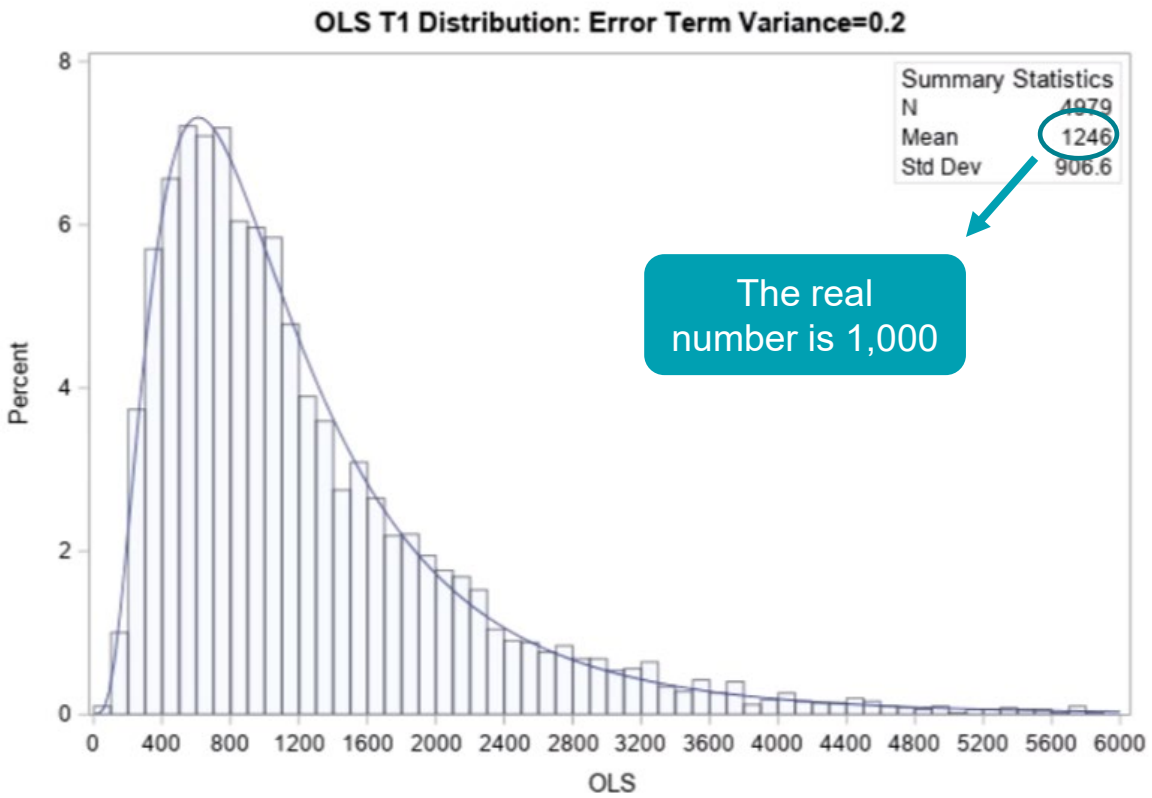
Technomics
Better Decisions Faster

Cost Improvement Curve – OLS Bias

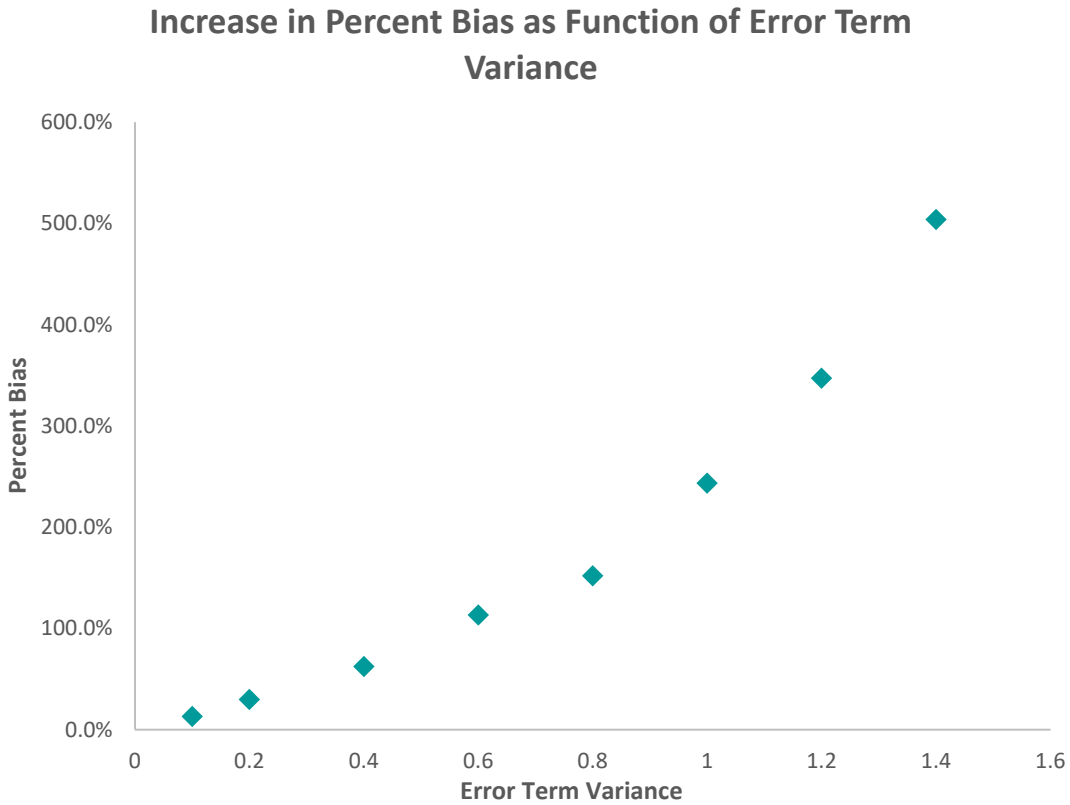
Estimates of T1 & Y: Biased!

Bias increases with σ_ε^2

Leads to inaccurate estimates



The real number is 1,000



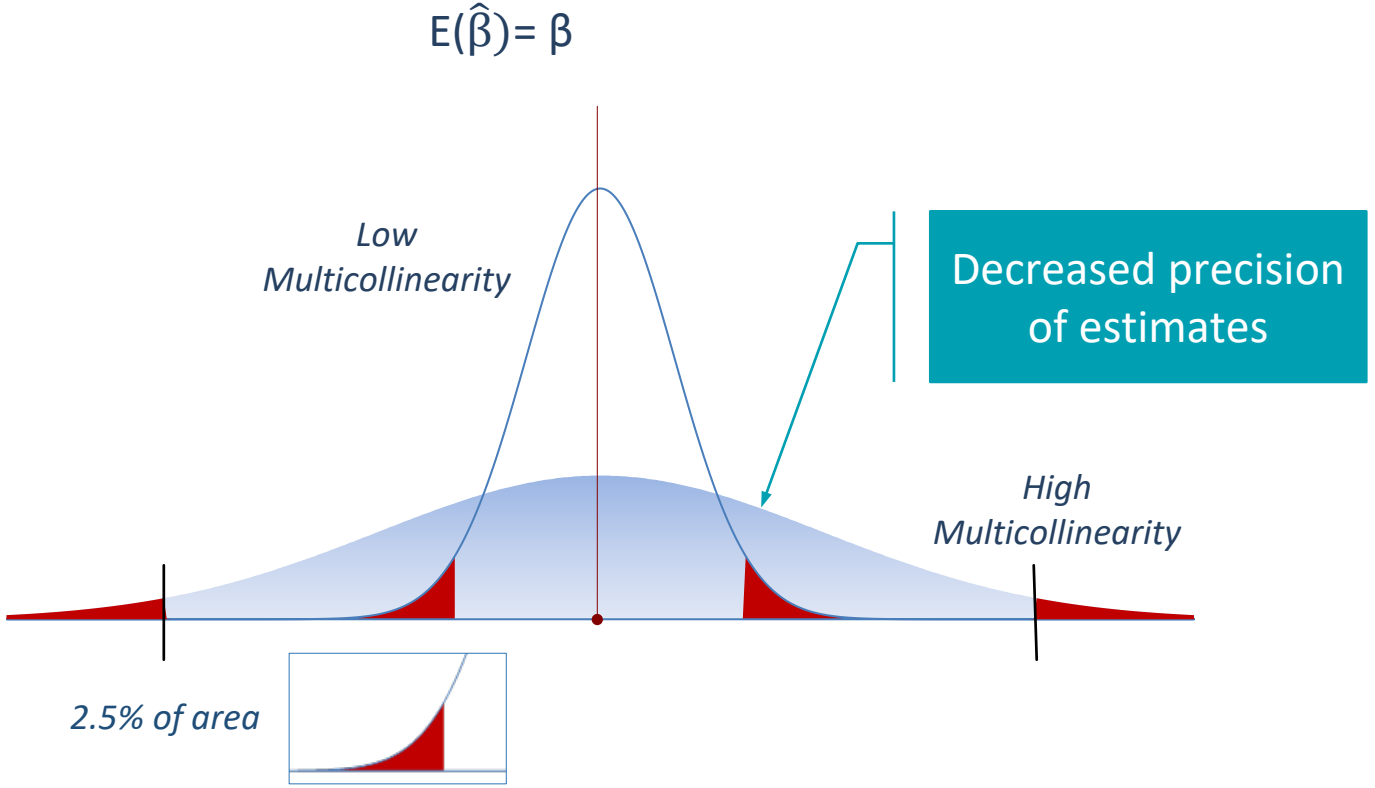
Statistical bias leads to inaccurate cost estimates

Prediction Intervals

Collinearity: Learning & Rate

- Collinearity jeopardizes accuracy of predictions, which requires perpetuation of:
  - Stable relationship between Y and the X's, and
  - Stable interdependency amongst the X's [often not the case]

- Widens confidence and prediction intervals for Y_o set of x values



OLS: Biased Estimates of Y

$$E(Y) = \alpha\, Q_i^b\, R_i^c\, e^{0.5\sigma_\varepsilon^2}$$

Biased

$$\hat{Y} - tstatistic * \sigma_F \;\leq\; Y \;\leq\; \hat{Y} + tstatistic * \sigma_F, where\ \sigma_F\ is\ the\ square\ root\ of\ \sigma_F^2$$

Technomics
Better Decisions Faster

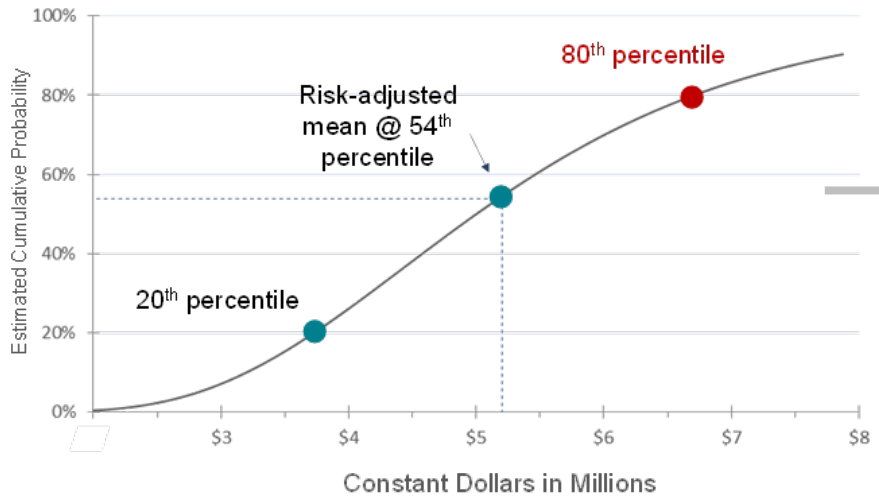
S-Curves and Affordability Constraints

S-Curves Constructed from Cost Improvement Curves

Y-coordinate: Vary λ (level of significance) from 1% to 100%
X-coordinate: Compute **Predicted** Cost or Hours

But OLS is biased!

Cumulative Distribution Function



National Defense Authorization Act of 2017 Requires Affordability Constraints

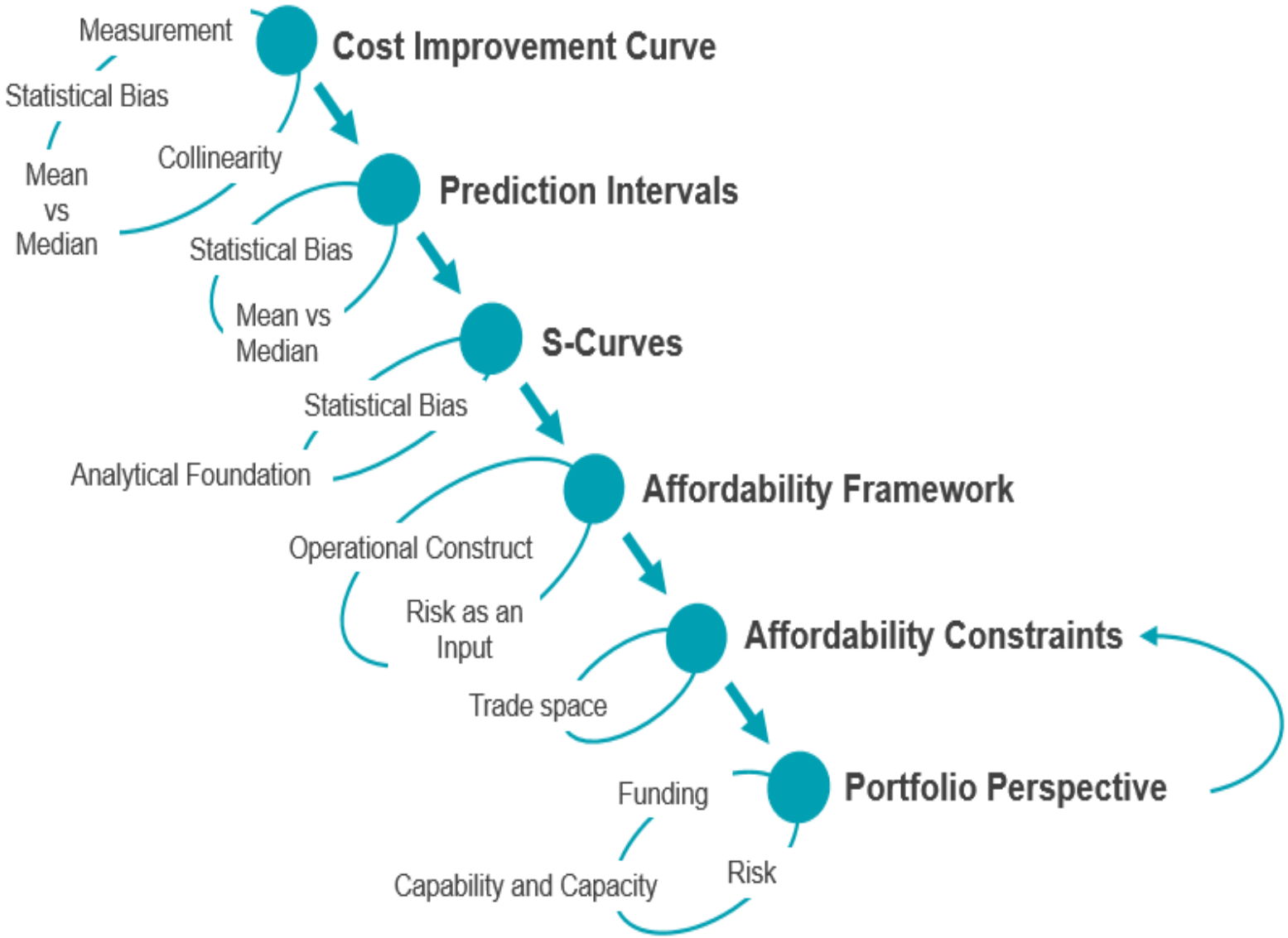
But, current practice:

- Focus on one acquisition program viewed in isolation
- Reliance on program-office point estimates
- No framework for analysis or operational construct for implementation
- No consideration of alternatives on the S-curve
- No view of portfolio tradeoffs

Navy Resource Sponsor: "It's the Wild, Wild, West"

Technomics
Better Decisions Faster

Solution


Measurement
Cost Improvement Curve
Statistical Bias
Collinearity
Mean vs Median
Prediction Intervals
Statistical Bias
Mean vs Median
S-Curves
Statistical Bias
Analytical Foundation
Affordability Framework
Operational Construct
Risk as an Input
Affordability Constraints
Trade space
Portfolio Perspective
Funding
Capability and Capacity
Risk

Unbiased T1 Estimates

The math ...

$Let\ \tilde{\alpha} = e^{\hat{\gamma}} F$, where

$\hat{\gamma} = \widehat{\ln(\alpha)}$ or the OLS estimate of $\ln(\alpha)$, and $F = \sum_{j=0}^{\infty} \frac{f_j (cw)^j}{j!}$, where

$c = -\frac{1}{2} m^{00}$, with m^{00} denoting the estimated, unscaled variance of $\widehat{\ln(\alpha)}$, i.e., the upper left corner element of the inverse of $\boldsymbol{X'X}$, with $\boldsymbol{X}$ the data matrix.

$w = s^2$, or the residual variance from OLS estimation

$$f_j = (0.5v)^j \frac{\Gamma(0.5v)}{\Gamma(0.5v + j)}$$

$v = n - k - 1$ degrees of freedom, where

n = number of observation and k = number of explanatory variables.

The infinite series usually converges to 4 decimal places after only six iterations

Unbiased T1 Estimates

$$f_j = (0.5v)^j \frac{\Gamma(0.5v)}{\Gamma(0.5v+j)}$$

The calculations …

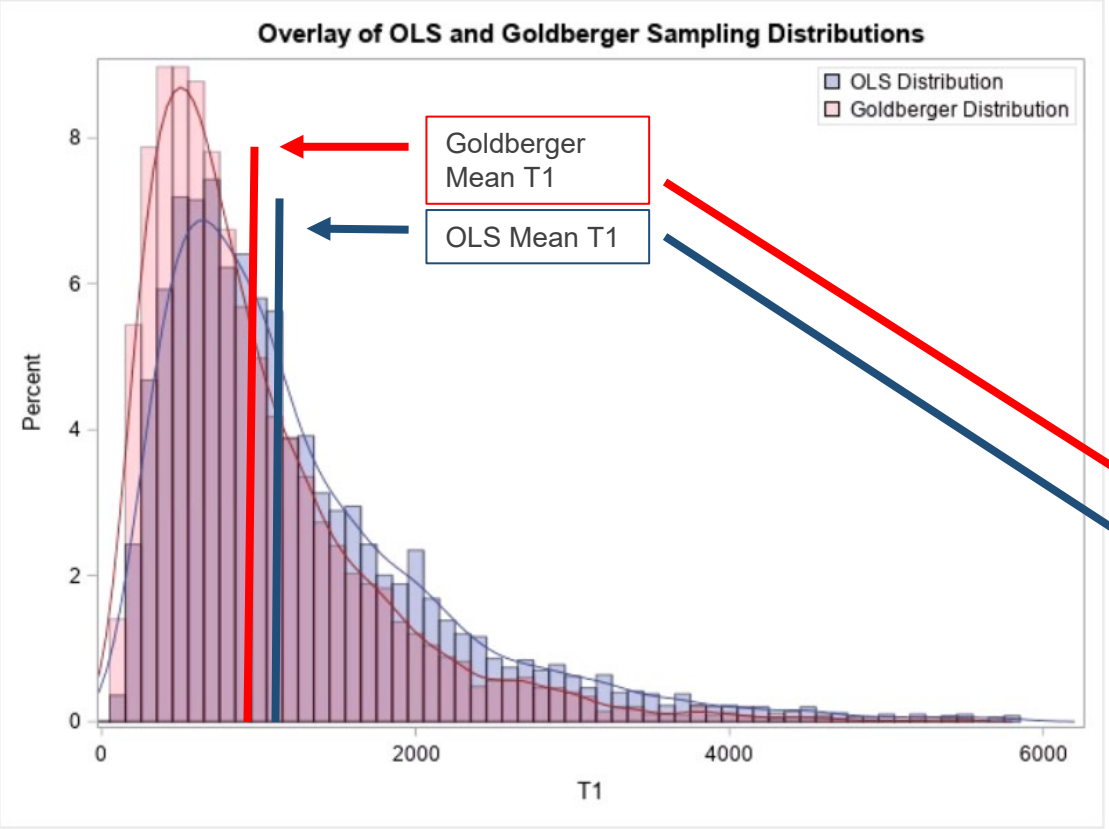
$$\frac{f_j (cw)^j}{j!}$$

$$F = \sum_{j=0}^{\infty} \frac{f_j (cw)^j}{j!}$$

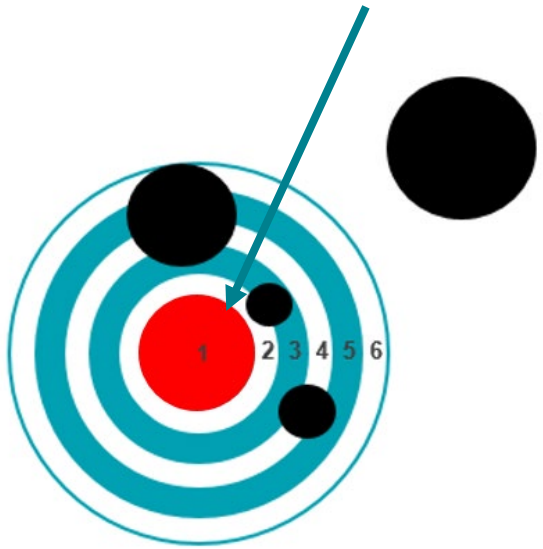
j	f_j	j!	Ratio	F
0	1.0000	1	1.00000000000	1.00000000000
1	1.0000	1	-0.65474971848	0.34525028152
2	0.8333	2	0.17862383077	0.52387411229
3	0.5952	6	-0.02784616736	0.49602794493
4	0.3720	24	0.00284879222	0.49887673716
5	0.2067	120	-0.00020724955	0.49866948761
6	0.1033	720	0.00001130805	0.49868079566
7	0.0470	5040	-0.00000048078	0.49868031488
8	0.0196	40320	0.00000001640	0.49868033128
9	0.0075	362880	-0.00000000046	0.49868033082
10	0.0027	3628800	0.00000000001	0.49868033083
11	0.0009	39916800	0.00000000000	0.49868033083
12	0.0003	479001600	0.00000000000	0.49868033083
13	0.0001	6227020800	0.00000000000	0.49868033083
14	0.0000	87178291200	0.00000000000	0.49868033083
15	0.0000	1307674368000	0.00000000000	0.49868033083
16	0.0000	20922789888000	0.00000000000	0.49868033083
17	0.0000	355687428096000	0.00000000000	0.49868033083
18	0.0000	6402373705728000	0.00000000000	0.49868033083
19	0.0000	121645100408832000	0.00000000000	0.49868033083
20	0.0000	2432902008176640000	0.00000000000	0.49868033083

Unbiased T1 Estimates

The results … Bullseye!



Goldberger hits the bullseye!



Estimator	Sample Values: T1		Sample Values: beta-hat	
	Mean	Variance	Mean	Variance
Goldberger	996	616247	-0.4997	0.0132
OLS	1246	1007827	-0.4997	0.0132
True Value	1000		-0.5000	
Var(ε) = 0.2				

The paper presents expressions for the unbiased conditional mean and median of Y

Real World Cost Model T1

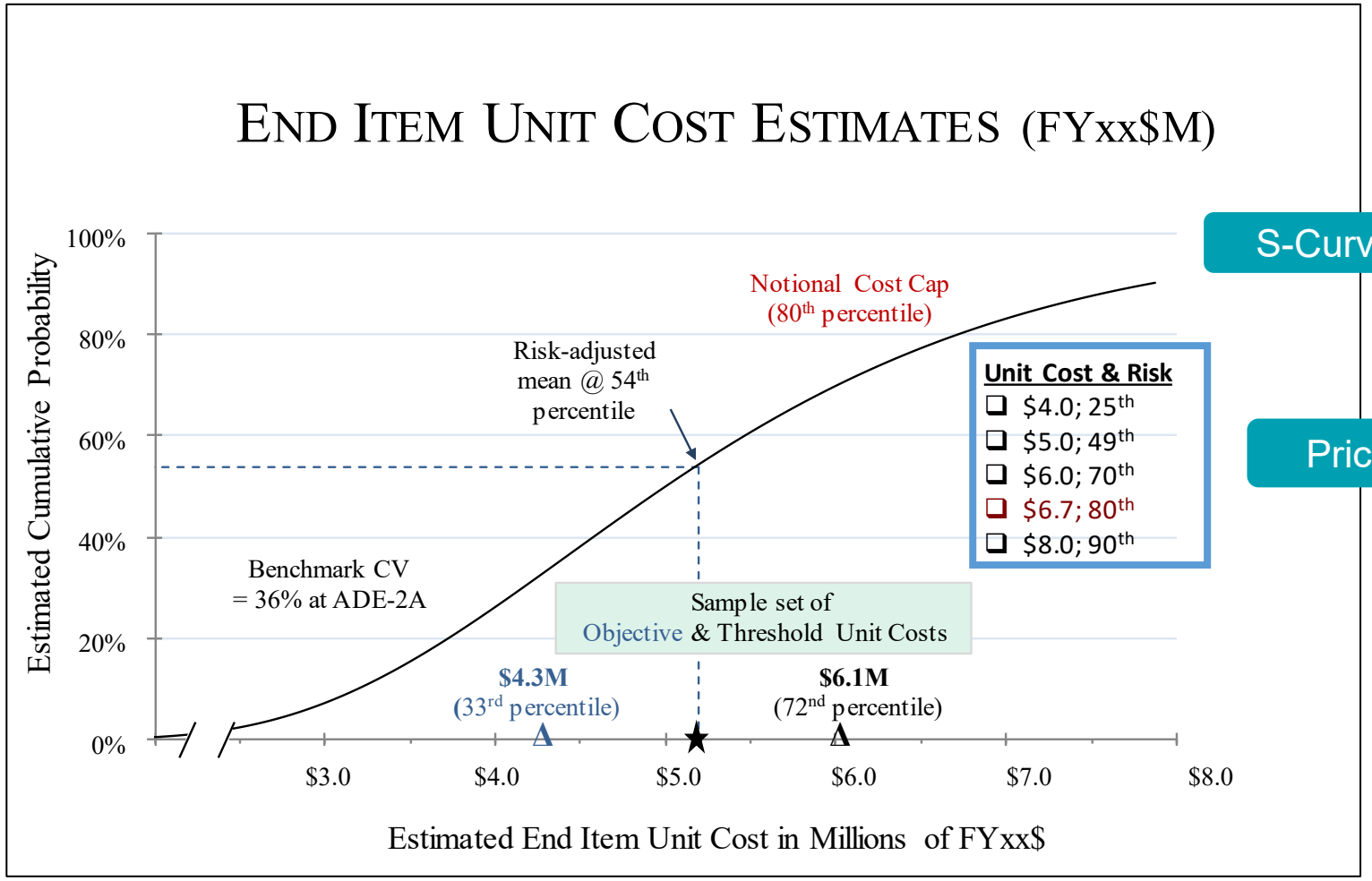
CER	Value of X	Predicted Costs			% Delta
		Goldberger	OLS	Delta	(OLS from Goldberger)
CER 1					
Lowest	48.0	424.3	439.9	15.6	3.7%
Mean	1065.0	3314.1	3329.7	15.6	0.5%
4*Mean	4260.0	8089.7	8234.7	145.0	1.8%
CER 2					
Lowest	40.2	1374.0	1447.8	73.9	5.4%
Mean	509.1	13082.0	13240.7	158.7	1.2%
4*Mean	2036.1	42464.2	44339.2	1875.0	4.4%
CER 3					
Lowest	81.1	2674.6	2773.2	98.5	3.7%
Mean	663.4	18014.2	18167.3	153.1	0.8%
4*Mean	2654.4	60567.5	62762.1	2194.5	3.6%
CER 4					
Lowest	9.4	246.0	256.3	10.3	4.2%
Mean	327.4	7339.6	7380.8	41.2	0.6%
4*Mean	1309.7	26910.1	27374.1	463.9	1.7%
CER 5					
Lowest	712.1	989.2	1022.6	33.4	3.4%
Mean	508020.2	7012.6	7056.8	44.3	0.6%
4*Mean	2032081.0	10480.0	10607.6	127.5	1.2%

OLS constantly over estimates the Mean value!

Affordability Constraints

Cost Constraints:
- PAUC, APUC (Threshold and Objective values)
- Unit sustainment
- Caps

PAUC = Program Acquisition Unit Cost
APUC = Average Procurement Unit Cost



S-Curve from previous steps

Pricing Options

Threshold: Maximum allowable value for cost

Objective: What the user desires and expects (a should-cost value)

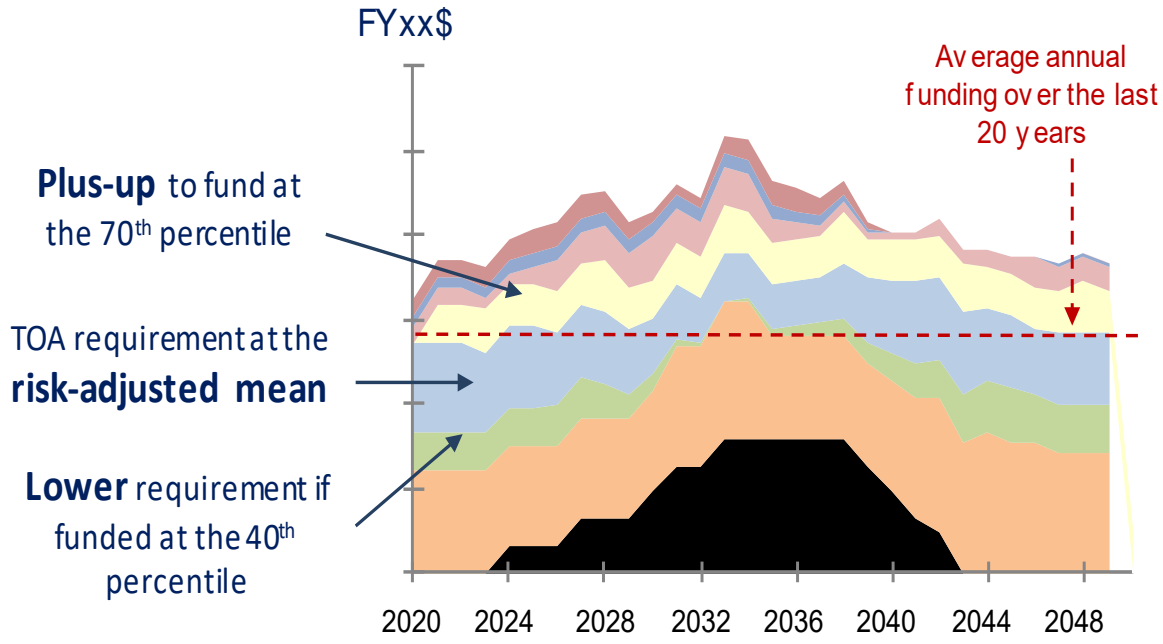
Constraints support the establishment of cost controls

Linkage to Resources

Pricing Options

Unit Cost & Risk

- ❑ $4.0; 25th
- ❑ $5.0; 49th
- ❑ $6.0; 70th
- ❑ $6.7; 80th
- ❑ $8.0; 90th



Portfolio Tradeoffs

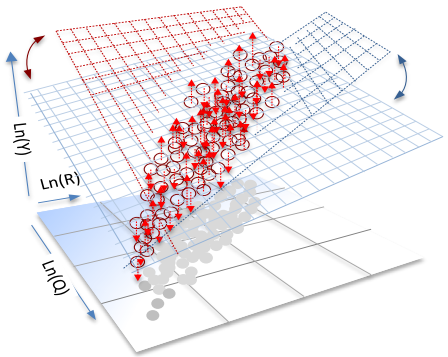
Example COA's	Impacts
• Funding at 40th percentile	Add a program
• Funding at risk-adjusted mean	No change
• Fund at 80th percentile	Subtract a program

Recap

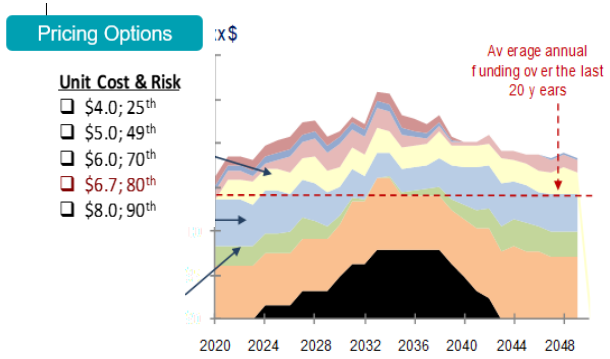
1 Un-Biased Estimates

- Learning curve estimates ON target
- Bias corrected

2 Collinearity & Rate Corrected and Incorporated

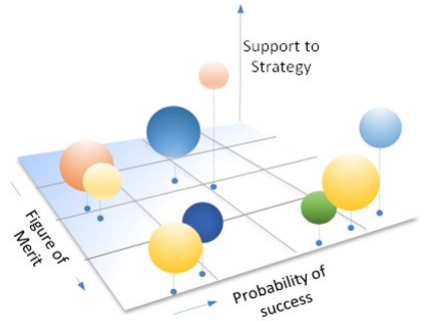


3 Accurate Risk Analysis



- Un-Biased inputs
- Linkage to CERs
- Menu of affordability options

4 Portfolio Perspective

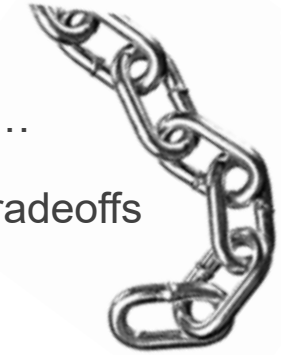


- Affordability constraints analytically set
- Affordability viewed against the portfolio

5 Forged Chain

Learning Curves linked to S-curves…

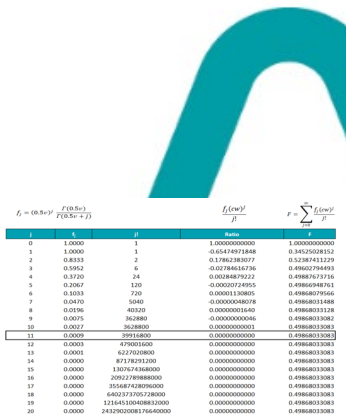
S-curves linked to Portfolio Tradeoffs



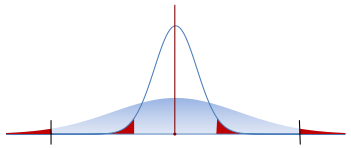
Technomics
Better Decisions Faster

Added Value

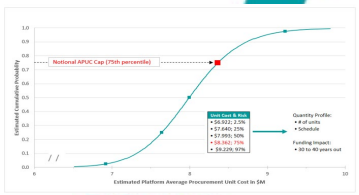
Forged the Missing Links!



Cost Improvement Curves



Prediction Intervals



S-Curves

Portfolio Trade Space Analysis

More accurate estimates with menu of options for analytically informed decisions!

“Great is the power of steady misrepresentation”

–Charles Darwin

Questions?

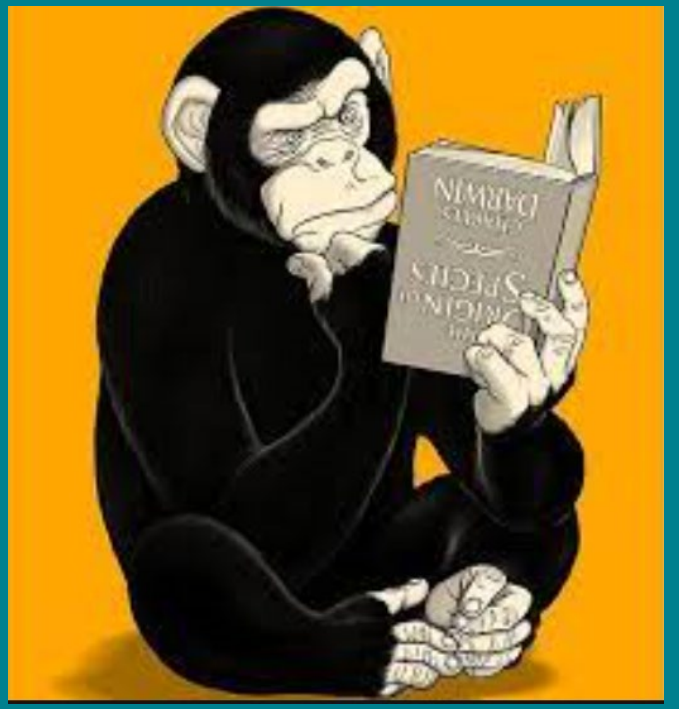


Image from Google, link below in speaker notes