Effective Cost Estimating & Risk Quantification Employing Data and Analytics

John K. Hollmann CEP CCP DRMP FAACE
Owner, Validation Estimating LLC
Ashburn VA

John K Hollmann

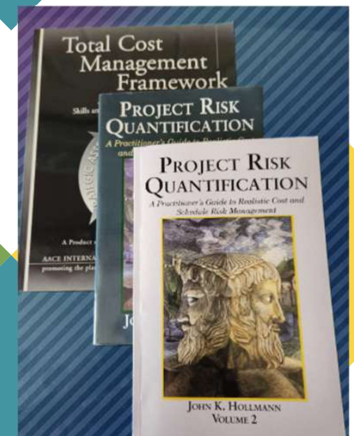
- Owner: Validation Estimating LLC (since 2005)
- Partner: ValidRisk® (since 2022)

Experience

- BS mining engineering (Penn State), MBA (IUP)
- 48 years, engineering, estimating, project controls
- Roles: owner, contractor, consultant, benchmarking
- Industries: process, utilities, infrastructure

Specializations

- Industry technical leader in project estimating, risk quantification, databases and analytics
- Consult with companies to improve their practices + analyses of major (>$1B) project cost, schedule, risk



Achievements

AACE: CEP CCP DRMP

- Fellow, Honorary Life Member, Award of Merit, Lifetime Achievement Award
- Past Technical/Full Board member
- Led DRMP certification development
- Lead author: *AACE TCM Framework* (2006)
- Contributor: > 70 Recommended Practices

Books: *Project Risk Quantification*

- Volume 1 (2016) and Volume 2 (2024)

Agenda

- Importance of Risk Quantification and Risk Analysis in Project Cost Estimating
- Criticality of Data and Analytics in Project Cost Estimating including Common Challenges
- Current and Anticipated Use of AI and ML for Project Cost Estimating.

“A good estimator is one who knows how wrong they are”

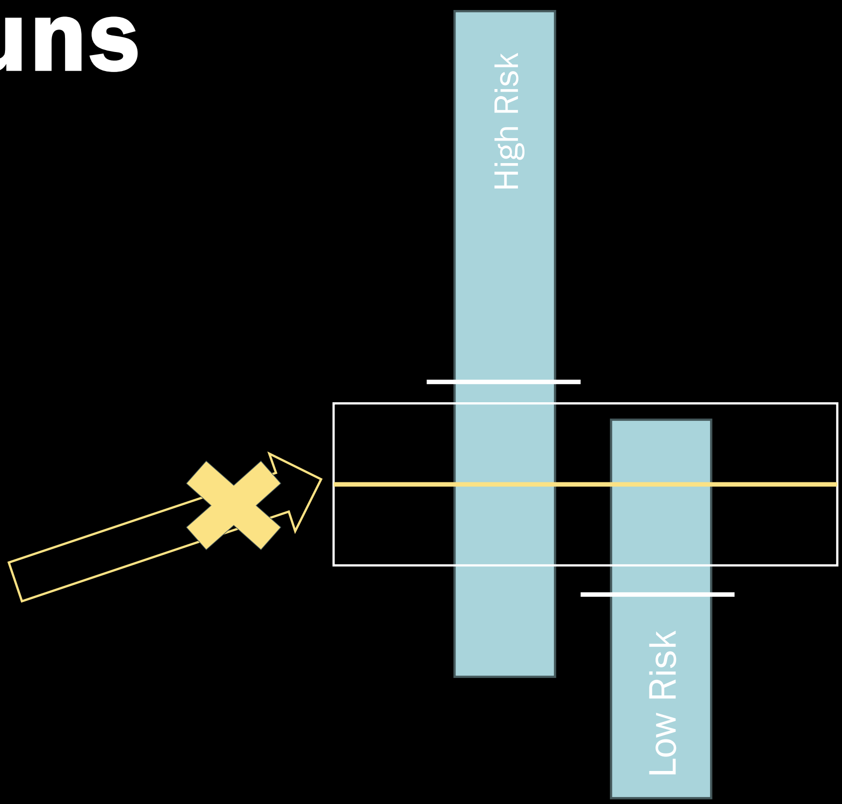
Risk Quantification and "The Number"

- What if the business client just wants "the number"?
- Business cost objectives differ: Research shows that there are two prevalent cultures (and hence two estimating cultures when it comes to risk):
  - ➢ Predictability culture: hit "the number"; minimum interest in risk and ranges
    - Weak risk analysis; generous contingency and other allowances (we hope, but oft not)
    - Common in public, government-funded and utility sectors
  - ➢ Competitive culture: take a chance on least cost (target); focus on risk, ranges and <u>WHY</u>
    - Strong risk analysis; risk-based cost distribution supports contingency
    - More common in for-profit sectors (including contractors (if one's objective is to maximize margin, they must be competitive)

Predictability Culture + High Risk = Overruns

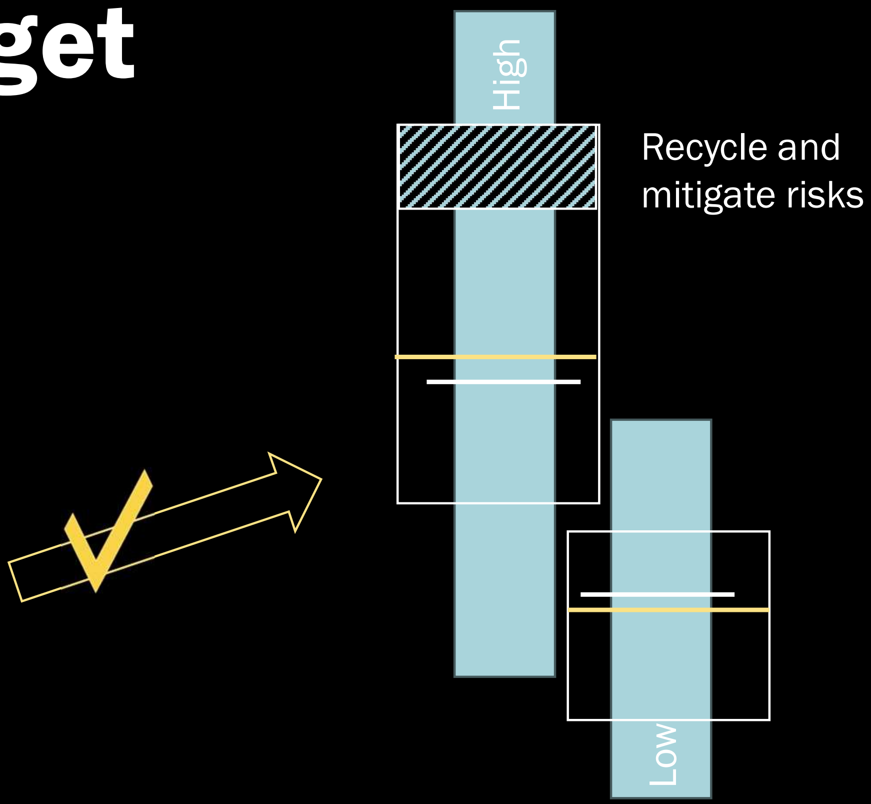
(+ Low Risk = Underruns)

- Research shows that when uncertainty and risks are not well quantified…
  - high risk, large projects tend to overrun
  - low risk, small projects tend to underrun
- Research also shows that the greatest uncertainty/risk drivers are:
  - level of planning and scope definition (phase)
  - physical/execution complexity
  - level of technology (not a big issue for highways)
  - team development (incl stakeholders)
- AACE calls these “systemic risks” (uncertainty)



Competitive Culture + Any Risk = On Target

- Research shows that when uncertainty and risks are well quantified…
  - the team understands WHY costs may vary and can take action to increase VALUE
- Research also shows that the greatest uncertainty/risk drivers are:
  - level of planning and scope definition (phase)
  - physical/execution complexity
  - level of technology (not a big issue for highways)
  - team development (incl stakeholders)
- AACE calls these "systemic risks" (uncertainty)



Estimating Challenges (countering a predictability culture)

Be able to explain WHY the range is what it is and WHY the range matters to VALUE & decisions

- Fundamental estimating/risk analysis practices (based on AACE Recommended Practices):
  - Stage-gate (planning/scope development) system — AACE Class (C): RP 98R-18
    - Scoping (C5), preliminary design (C4), detailed design (C3), final design (C2)
  - Identify risks/uncertainties (start with systemic risks, then project-specific events)
  - Quantity the risks/uncertainties explicitly*
    - Risk-based or risk-driven methods per AACE parlance
    - Probabilistic methods for projects with significant risks
    - Empirically-valid methods (data discussed later)

  QRA* principles in AACE RP 40R-08

** call it quantitative risk analysis (QRA). "Contingency Estimating" reflects a focus on a "number". We are doing more than that.*

QRA Challenges for Estimating

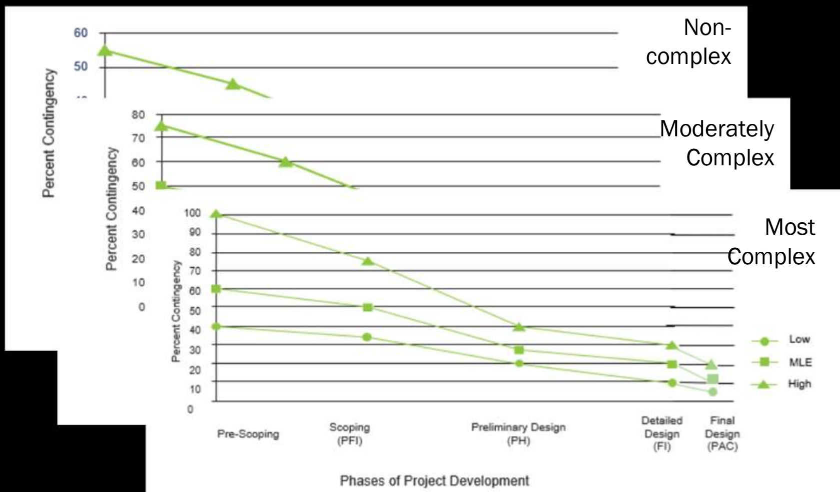
- VDOT Estimating Manual *generally* aligns w/AACE Recommended Practices
- Robust, effective QRA is a challenge for ALL companies and organizations. They make compromises to the extent that the following are roadblocks:
  - Predictability cultures/resources; *just give me the number;* limited buy-in
  - Competencies: estimators focused on detailed base estimating mechanics
  - Data: limited data and analytic skills for reliable probabilistic methods

Risk-Based or Driven QRA (VDOT vs AACE)

- VDOT does it's best to apply best practice considering the challenges (keep it simple)
- Systemic Risks:
  - VDOT employs "**sliding scales**" to address the level of scope definition and complexity
  - AACE recommends a "**parametric model**" that probabilistically addresses the level of scope definition and complexity (along with other systemic risks) (RP 42R-08). There is also a tabular method similar to the sliding scale (RP 119R-21).
- Project-Specific Risks (the risks found in the risk register):
  - VDOT employs deterministic "**expected value**" (prob x impact) for any risk the team identifies as a *comparison to the sliding scale*
  - AACE recommends probabilistic "**expected value**" for critical (red) risks *integrated with the parametric model* for systemic risks (hybrid method; RP 113R-20)

VDOT Estimating Manual

Sliding Scale (definition and complexity)



check

Deterministic Expected Value (if Complex)

Project Cost Element	Estimated Impact	Probability of Occurrence	Cost Contingency
Initial Purchase of RW	$1,200,000	20%	$240,000
Known Hazardous Substance	$125,000	10%	$12,500
Coordination with Railroad Agencies	$50,000	10%	$5,000
Treatment of Water Discharged from Site	$400,000	3%	$12,000
TOTAL			$269,000

Keeping it simple!

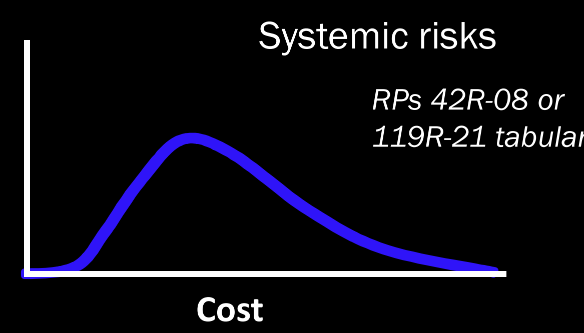
AACE RP 113R-20

Parametric Model (all projects, all phases)

- level of definition
- complexity
- technology
- teams
- bias
- estimate quality
- etc.

data

Calibrated Model

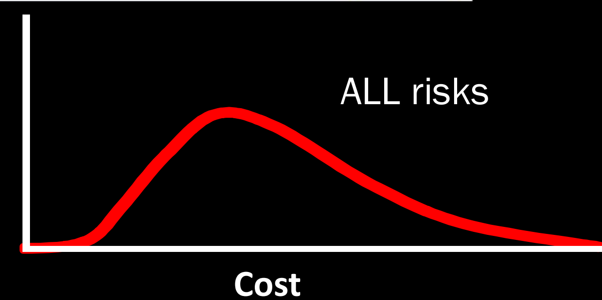


+

Expected Value w/MCS (large, Class 4 or later)

Project Cost Element	Estimated Impact	Probability of Occurrence	Cost Contingency
Initial Purchase of RW	[illegible]	20%	$240,000
Known Hazardous Substance	[illegible]	10%	$12,500
Coordination with Railroad Agencies	[illegible]	10%	$5,000
Treatment of Water Discharged from Site	[illegible]	3%	$12,000
TOTAL			$269,000

MCS

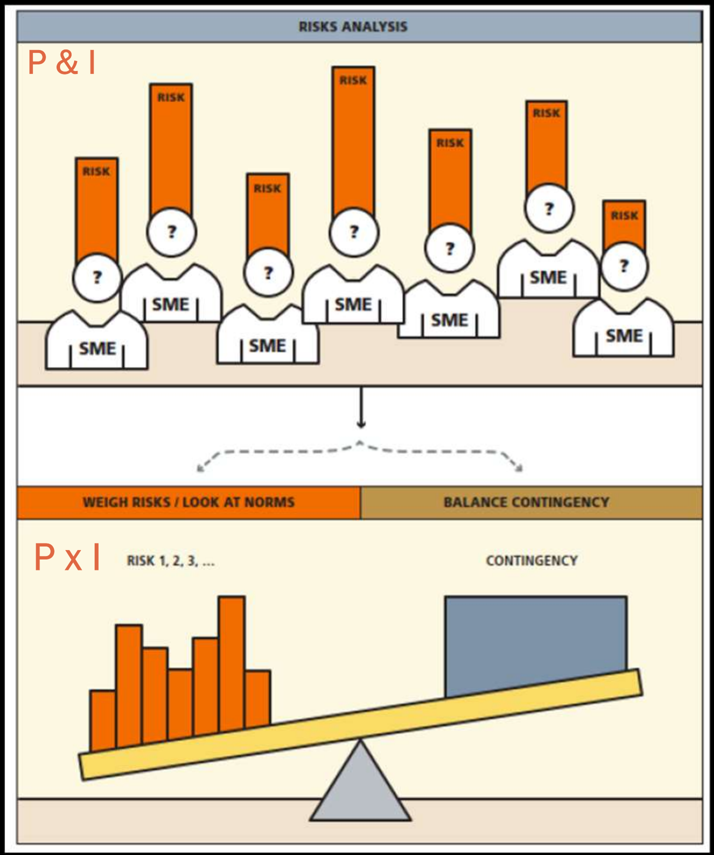


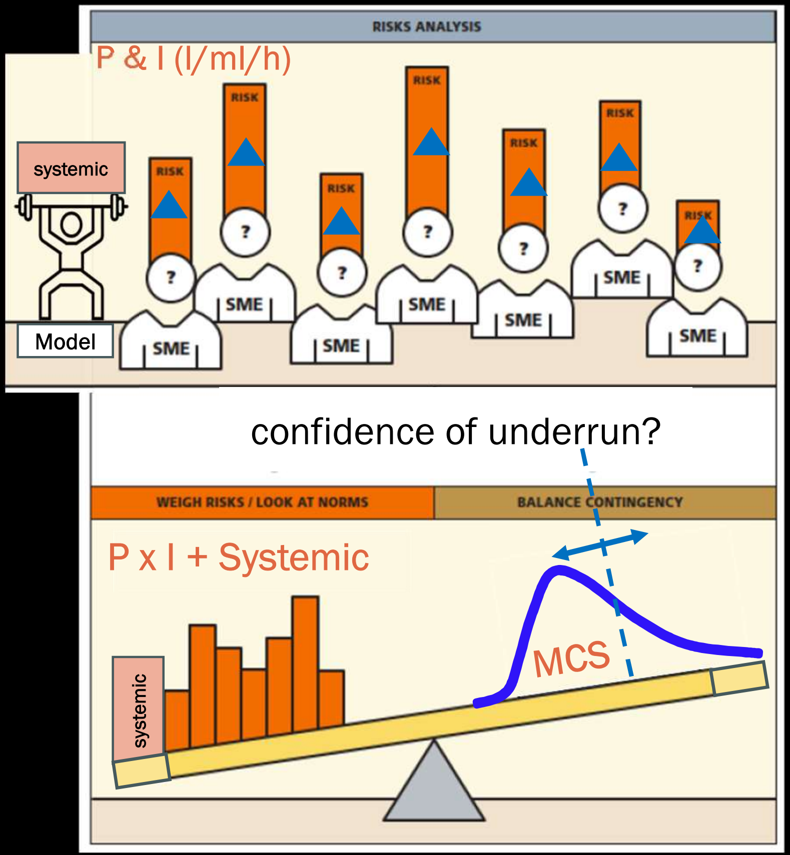
VDOT Estimating Manual

AACE RP 113R-20

Deterministic: Probability x Impact

Probabilistic: Probability x Impact (low, most likely, high)





Expected Value Methods

Importance of Risk Quantification

Given:

- Project costs are uncertain; *the less defined and/or complex, the more uncertain*
- Risks happen; *the more complex, the more susceptible*
- The cost of uncertainty and risks **must be estimated** and included in the budget

Management choice: do we manage RISKS and improve investment VALUE, or not?

  - Predictability cultures accept risks (*just give me a number*)
  - Competitive cultures manage risks (show the range, the WHY, and HOW to improve)

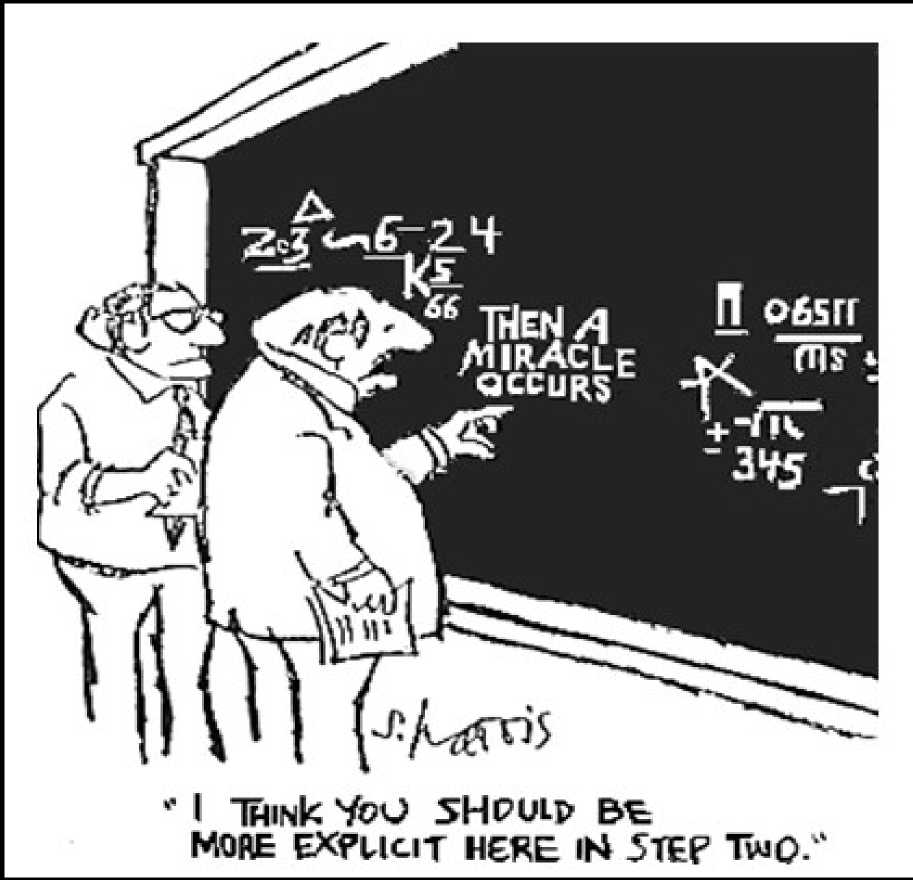
Estimating challenges (compromise/keeping it simple):

- How to identify, price and manage risks with limitations as to buy-in, resources, competencies, and data?

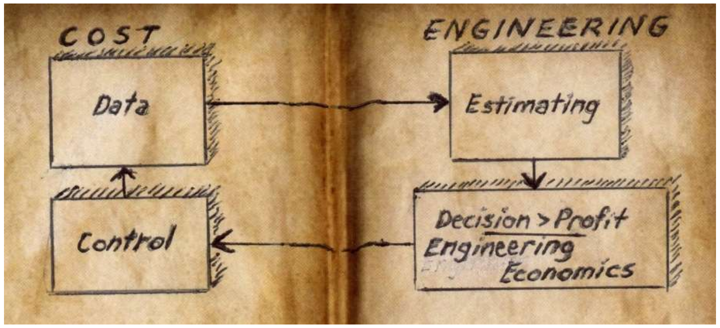
Agenda

- Importance of Risk Quantification and Risk Analysis in Project Cost Estimating
- Criticality of Data and Analytics in Project Cost Estimating including Common Challenges
- Current and Anticipated Use of AI and ML for Project Cost Estimating.

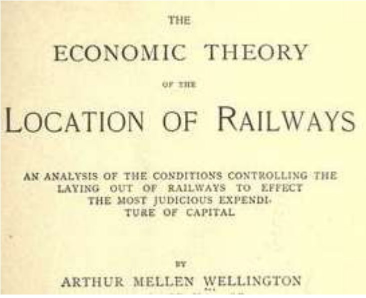
“Without data, estimating is just educated guessing”

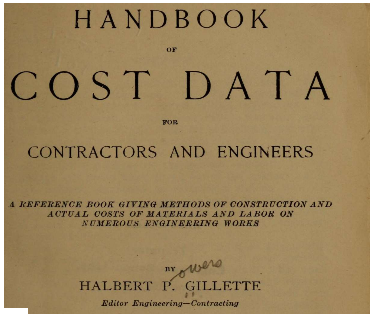


The Foundation of Estimating is Data



- Estimating = algorithms (cost estimating relationships) + data
- The basic algorithm is Quantity x Price. Arthur Wellington's (the founder of Industrial Engineering) 1874 book was about quantity take-off; "*Methods for the Computation from Diagrams of Preliminary and Final Estimates of Railway Earthwork*"*.
- The need for data is "as old as dirt". For example, Halbert Gillette's "*Handbooks of Cost Data*" from 1903 to 1922 covered Rock and Earth Excavation, Concrete, Roads, and other trades.

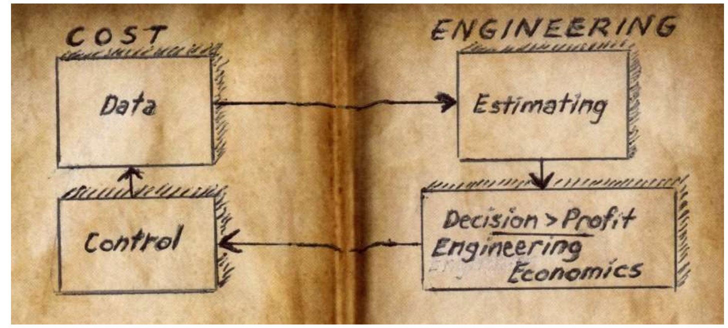

THE
ECONOMIC THEORY
OF THE
LOCATION OF RAILWAYS
AN ANALYSIS OF THE CONDITIONS CONTROLLING THE LAYING OUT OF RAILWAYS TO EFFECT THE MOST JUDICIOUS EXPENDITURE OF CAPITAL
BY
ARTHUR MELLEN WELLINGTON

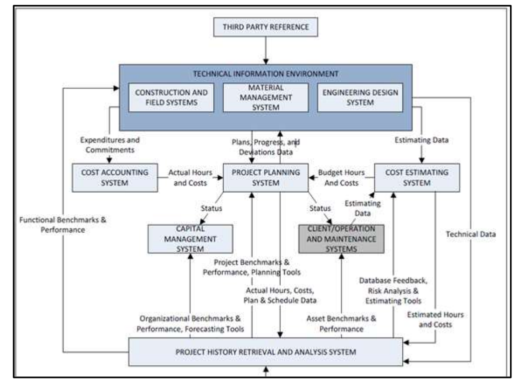

HANDBOOK
OF
COST DATA
FOR
CONTRACTORS AND ENGINEERS
A REFERENCE BOOK GIVING METHODS OF CONSTRUCTION AND ACTUAL COSTS OF MATERIALS AND LABOR ON NUMEROUS ENGINEERING WORKS
BY
HALBERT P. GILLETTE
Editor Engineering—Contracting

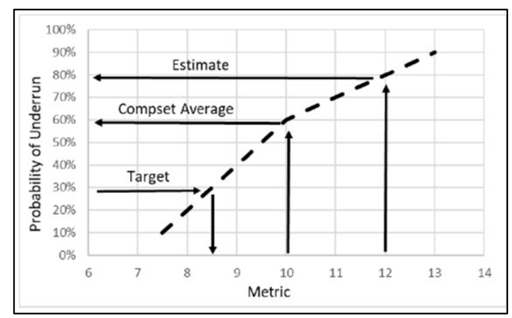
*Quote: "*It would be well if engineering were less generally thought of, and even defined, as the art of constructing. In a certain important sense it is rather the art of not constructing; or, to define it rudely but not inaptly, it is* **the art of doing that well with one dollar, which any bungler can do with two.**"

AACE and **Data**

- AACE was founded in 1956 by engineers seeking to develop better **algorithms and data** to support estimating and engineering economics
- RP 114R-20 Project Historical Database Development
  - Guidelines for evaluating, developing and maintaining project historical data management systems (cost, schedule, risk, etc.)
- RP 110R-20 Cost Estimate Validation
  - Validation is part of estimate review using metrics (ratios such as cost/quantity). **Key quantities** are **essential.**
- Database and application tools are often combined in one system
  - E.g., Statewide Bid Tab Query System, PCES, AASHTOWare,







Database Challenges

- Database development is difficult: few do it or do it well (VDOT's tools are an exception!)
- Accounting data is of limited value. It is often poorly structured and full of noise.
  - *"There is no such thing as actual data"*; useful data is typically "estimated" from records (clean, normalize, backfill, allocate, approximate). Requires the skills of an estimator!
- Capture project attributes, practices and risk drivers; quantity and cost data are just the start
- Databases take time. Projects may take 3 years (including backfill) + dedicated resources
  - Roadblocks: poor records and structure, data allocation, etc. Perfection is enemy of good.
  - Most database projects fail: organizational gyrations and management impatience.

Two Types of Data for Estimating

Base Estimating (e.g., quantity x prices)

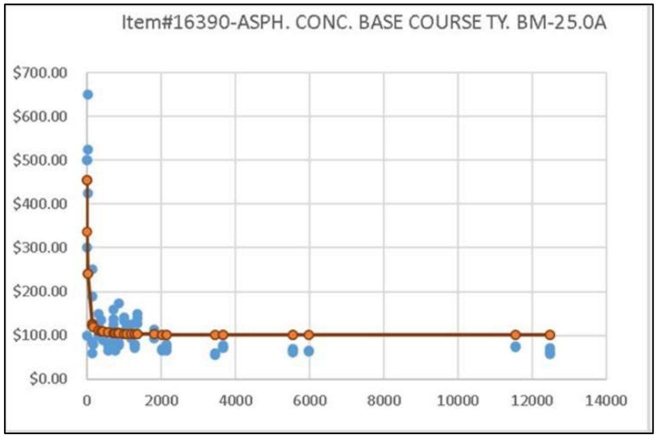
- Data for this is well understood because base estimating methods are easy to understand (quantity x price, factoring, etc.)
- Data = bid unit prices, labor rates, factors, etc.

Risk Quantification (aka contingency estimating)

- Data for this is less understood because risk quantification methods are less mature and more difficult to understand/communicate
- Data for risks = uncertainty and risk-drivers and their impacts

Data and Analytics for Estimating

- Data: unit prices, bids, labor rates, factors, etc., but also quantities and project attributes
- Analytics for Base Estimating; deterministic
  - Use of Descriptive Statistics
    - mean, mode, median, standard deviations, etc.
  - Understand variability; one focus is economy of scale
  - Methods: single variable regression; linear or non-linear
    - E.g., price versus quantity
    - Comparisons (validation, benchmarking)
  - Judgment needed as to WHY there is variance?



Data and Analytics for Risk Quantification

- Data: risk drivers, risk outcomes: cost growth
  - Risk drivers: Systemic vs. Project-Specific (i.e., events)
  - Cost growth: actual costs over/under the base estimate
- Analytics for Risk Quantification: stochastic
  - Use of Inferential Statistics
    - Seek correlations between data (drivers-outputs)
  - Understand causation; to better manage risks/add value
    - Maximize objectivity (less educated guessing)
  - Methods: Regression vs. Monte-Carlo simulation (MCS)
    - Regression=actual data; MCS=assumed data
  - Risk driver aligned with the best QRA method:
    - Systemic risks = Regression (parametric)
    - Risk events = MCS (actual data on risks is sparse)

COST	Calibrate		1.00	1.00
MODEL Risk Inputs	**Input Values**	**Adjusted Values**	**Coefficients Mean**	**Coefficients Std. Deviation**
Maturity	3.20	3.20	0.0663	0.1091
Estimate Inclusiveness	3.0	6.0	(0.0031)	(0.0015)
Estimate Basis	3.0	6.0	(0.0055)	(0.0060)
Estimate Conservativeness	5.0	5.0	(0.0107)	0.0004
PM Effectiveness	3.0	4.0	0.0040	0.0009
New Technology	1	4	0.0010	0.0005
Severity	0	0	0.0055	0.0036
Technical Complexity	0	0	0.0048	0.0013
Execution Complexity	0.9	1.8	0.0054	0.0030
Constant			0.6895	0.5538
Base Parameters			1.21	1.25

Expected Value for Risk Events; Deterministic vs. MCS

Example:
AASHTO Guide
(Deterministic)

Too little?

Project Cost Element	Estimated Impact	Probability of Occurrence	Cost Contingency
Initial Purchase of RW	$1,200,000	20%	$240,000
Known Hazardous Substance	$125,000	10%	$12,500
Coordination with Railroad Agencies	$50,000	10%	$5,000
Treatment of Water Discharged from Site	$400,000	3%	$12,000
TOTAL			$269,000

Example:
AACE RP
113R-20
(w/MCS)

Too much?

1 Trial Risk	Describe impacted work element (for which burn rate applies)			Contracts:	1	N/A	N/A
Assumed risk response:	Describe the assumed risk response if risk occurs (schedule and non-time driven cost impact reflects this response)						
Schedule Impact:	Impact (mos)	Time Driven Cost Burn Rate:	$x1000/Mo	Non-Time Driven Cost Impact:	Impact ($x1000)	
Low	3.0	General & service contracts	$ 626	Low	$1,000	Total
Most likely	5.0	Other contracts (see selected)	$ 256	ValidRisk® Most likely	$5,000	Cost Impact
High	7.0	Burn Rate (can override):	$ 882	High	$20,000	($x1000)
Schedule Months (EV)	4.0	Time Driven $ (EV)	$3,528	Non-Time Driven $ (EV)	$ 6,933	$10,461

3-point estimate of execution schedule completion impact **(MCS function applied)**

Time-driven cost impact (duration x burn-rate for applicable contract(s)

3-point estimate of non-time driven cost impact **(MCS function applied)**

Total cost **(MCS output)**

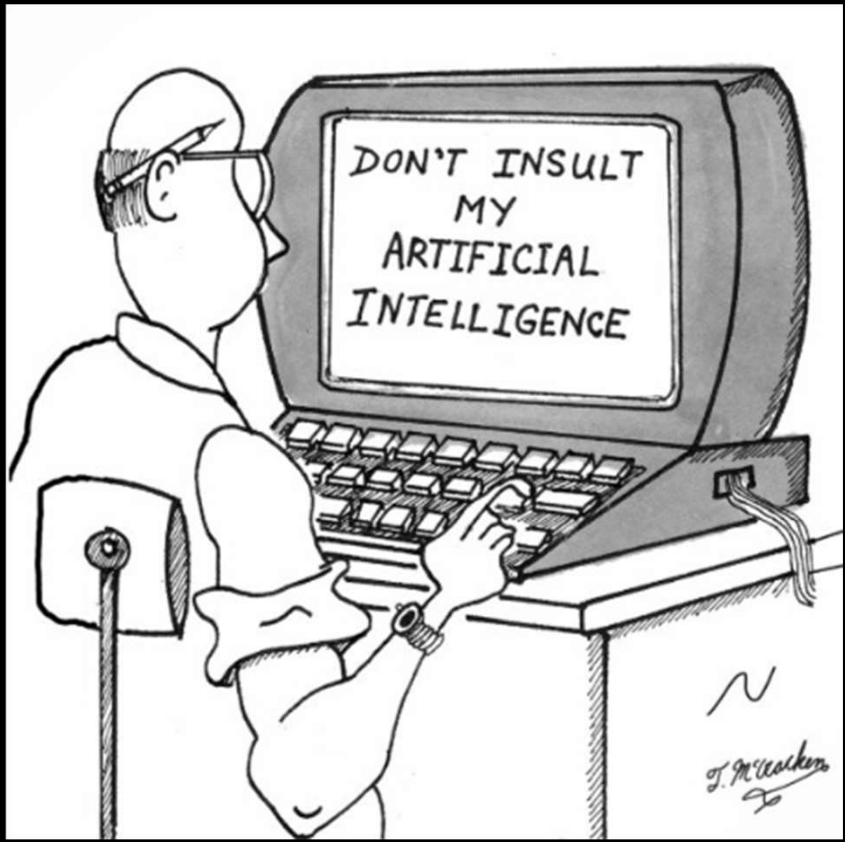
Data to Capture for Risk Quantification Analytics

- Processed Information
  - General: Attributes (scope) & Outcomes (cost, schedule, resources)
    - Project scope (including attributes related to systemic risks)
    - Cost and schedule actual and estimated (by phase) data
    - Resource actual and estimated (by phase) data
  - Risk Focused: Drivers and Responses
    - Systemic risk ratings (definition, complexity, technology, teams, etc.)
    - Critical project-specific risks including actual cost and schedule impacts
    - Lessons Learned; including effective risk responses
- Original Records (searchable electronic format best)
    - Risk analyses reports
    - Risk registers
    - Change control logs

Agenda

- Importance of Risk Quantification and Risk Analysis in Project Cost Estimating
- Criticality of Data and Analytics in Project Cost Estimating including Common Challenges
- Current and Anticipated Use of AI and ML for Project Cost Estimating.

“AI can’t come fast enough”



Estimating Opportunities for Generative AI (I asked Copilot)

It is not hard to see how every one of these meet an estimating need!

- Mathematical Models: Create models that simulate real-world behavior.
- Data Analysis: Analyze large datasets to uncover patterns, trends, and insights.
- Predictive Analytics: Analyze historical data to predict future outcomes.
- Data Cleaning: Automate the process of cleaning and preprocessing data.
- Pattern Recognition: Recognize patterns in datasets for identifying trends
- Natural Language Processing (NLP): Analyze voluminous unstructured text data
- Machine Learning Models: Build models that improve over time with more data.
- Visualization: Create data visualizations to help everyone understand data.
- Time Series Analysis: Forecast trends over time.

Estimating Opportunities for AI

Database Development:

- Structured Data: (e.g., bids) Automate the process of cleaning data
- Unstructured Data: (e.g. bid docs, claims records, change logs, risk registers, etc.). This is a HUGE opportunity that has been largely outside of our capability until now!
- Data analysis (structured and unstructured): patterns, trends, and insights
- Real time: in the past, data capture and processing was slowwww.

Estimating Opportunities for AI

Deterministic steps/functions:

- Base estimating mechanics: AI not so much…it's not where we add our value.
- What price and why? Correlate cost drivers with price & productivity, Now judgmental (the "art" part of estimating). Good opportunity for generative AI.
- Basis of Estimates and reports; Engineers/estimators are poor writers; AI is already here; just needs some teaching.
- Estimate validation/review: Analyze estimates (internal or 3rd party) for quality and competitiveness compared to data.
- Risk registers: Tap unstructured data and project attributes to generate initial drafts: AI is already here; just needs some teaching.

Estimating Opportunities for AI: Looking at the Process Steps

Stochastic steps/higher level functions:

- Conceptual Estimating Models: Scoping (C5) & preliminary design (C4); correlate cost drivers (many extracted from unstructured data) and cost; a natural for AI
- Risk Quantification Models: Systemic Risks: Tap structured and unstructured data and project system attributes and correlate to cost growth and schedule slip data
- Risk Quantification Models: Project-Specific (Event) Risks: Tap structured and unstructured data and project attributes +d estimate probabilities and impacts
- Valuable Insights: Explain the HOWs and WHYs to support risk management and decision making and add VALUE!

In Chapter 8 of Project Risk Quantification Volume 2, I review 5 commercial AI products for QRA (1 for cost, 4 for schedule.) They have a ways to go to address all our expectations, but I have no doubt they (or others) will get there.

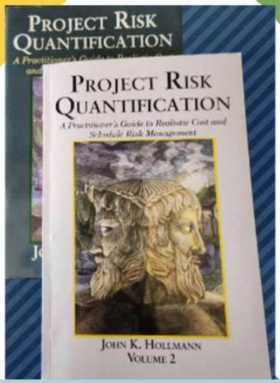
The Future of AI is Here

EVERYTHING on the last three slides is within current AI/ML capabilities!

- Roadblocks: data, resources/budget, and expertise.
  - Data is scattered all over the place, sometimes behind firewalls.
  - Expertise availability is limited but growing, albeit at high charge rates.
- The main benefits will come initially from each company building their own apps (often with consulting support) task by task. Getting the data house in order. Developing expertise. Once a few apps get rolling, it will build until using AI is just “how we do things.”
- Commercial products (e.g., nPlan) will be limited by what data they can tap. But estimating and scheduling software vendors are embedding AI in their confined application spaces.

Conclusions/Takeaways

- Risk quantification and analysis add VALUE to projects when it identifies the HOWs and WHYs
- Estimating does not stop with “a number” (nor is our job to provide a “contingency number”; its QRA!)
- We are Engineers of costs, not Estimators (i.e., be more of a part of optimizing the value of the asset)
- Data and analytics make estimating work; invest in them (regards AI, take a course in data science)
- AI will improve/facilitate just about every estimating and risk quantification task; it is here!
- VDOT is making strides!



Check out "Project Risk Quantification, Volumes 1 and 2 at www.amazon.com

Thank you

John K Hollmann

jhollmann@validest.com

www.validest.com / www.validrisk.com

Connect on LinkedIn